Loop Motion in Triosephosphate Isomerase Is Not a Simple Open and Shut Case

Qinghua Liao,[†] Yashraj Kulkarni,[†] Ushnish Sengupta,[‡,§] Dušan Petrović,[†,‡] Adrian J. Mulholland,[∥] Marc W. van der Kamp,[∥,⊥] Birgit Strodel,[‡,#] and Shina Caroline Lynn Kamerlin*,[†]

[†]Department of Chemistry − BMC, Uppsala University, BMC Box 576, 751 23 Uppsala, Sweden

[‡]Institute of Complex Systems: Structural Biochemistry (ICS-6), Forschungszentrum Jülich, 52425 Jülich, Germany

[§]German Research School for Simulation Sciences, RWTH Aachen University, 52062 Aachen, Germany

[∥]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, BS8 1TS Bristol, United Kingdom

[⊥]School of Biochemistry, University of Bristol, University Walk, BS8 1TD Bristol, United Kingdom

[#]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

S *Supporting Information*

ABSTRACT: Conformational changes are crucial for the catalytic action of many enzymes. A prototypical and well-studied example is loop opening and closure in triosephosphate isomerase (TIM), which is thought to determine the rate of catalytic turnover in many circumstances. Specifically, TIM loop 6 "grips" the phosphodianion of the substrate and, together with a change in loop 7, sets up the TIM active site for efficient catalysis. Crystal structures of TIM typically show an open or a closed conformation of loop 6, with the tip of the loop moving ∼7 Å between conformations. Many studies have interpreted this motion as a two-state, rigid-body transition. Here, we use extensive molecular dynamics simulations, with both conventional and enhanced sampling techniques, to analyze loop motion in apo and substrate-bound TIM in detail, using five crystal structures of the dimeric TIM from *Saccharomyces cerevisiae*. We find that loop 6 is highly flexible and samples multiple conformational states. Empirical valence bond simulations of the first reaction step show that slight displacements away from the fully closed-loop conformation can be sufficient to abolish most of the catalytic activity; full closure is required for efficient reaction. The conformational change of the loops in TIM is thus not a simple "open and shut" case and is crucial for its catalytic action. Our detailed analysis of loop motion in a highly efficient enzyme highlights the complexity of loop conformational changes and their role in biological catalysis.
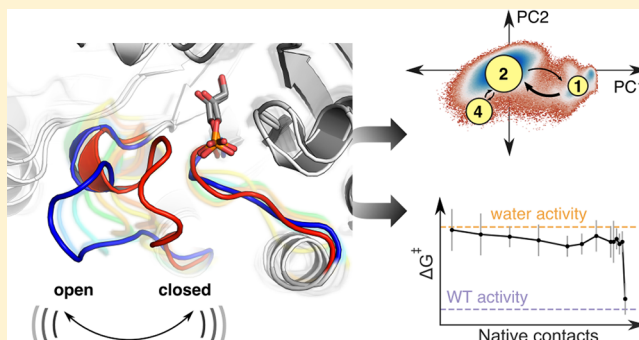
INTRODUCTION

Triosephosphate isomerase (TIM) is an enzyme that catalyzes a simple reversible isomerization reaction, namely the isomerization of dihydroxyacetone phosphate (DHAP) and (*R*)-glyceraldehyde-3-phosphate (GAP, Figure 1A).[1−5] It does so with tremendous catalytic proficiency, such that the turnover of this enzyme has been argued to be limited by diffusion.[2,3] As a result, TIM has often been described as a "catalytically perfect" enzyme[3] and has been the subject of extensive experimental and computational studies, as a model system for understanding the factors that drive enzyme catalysis (see refs 2 and 4−27 as just a few examples).

Structurally, TIM is usually a homodimer and gives its name to the archetypal TIM barrel fold, which consists of eight α-helices and eight parallel β-sheets alternating along the protein backbone (Figure 1B). We note that exceptions exist: In some organisms, TIM is a tetramer, and the change in oligomerization state can be functionally important.[28−30] TIM barrels, which are by far one of the most commonly occurring protein folds,[31] provide a versatile[11] and highly evolvable[11,15,24,32] scaffold and have been argued to have facilitated the early evolution of protein-mediated metabolism.[24] In addition, TIM barrel enzymes are excellent model systems for understanding enzymatic thermal adaptation.[33,34] A defining feature of reaction in TIM is the large motion (up to 7 Å) of a phosphate gripper loop, loop 6 (residues Pro166-Ala176 for yeast TIM from *Saccharomyces cerevisiae*, *y*TIM, PDB ID: 1NEY),[16,35] which interacts first with loop 5 in the unliganded form of TIM, and subsequently with loop 7 of TIM and with the phosphodianion of the bound substrate (Figure 1C). This dianion is, in turn, anchored to the enzyme through (up to) four hydrogen bonds with the protein backbone (Figure 1D).[16]
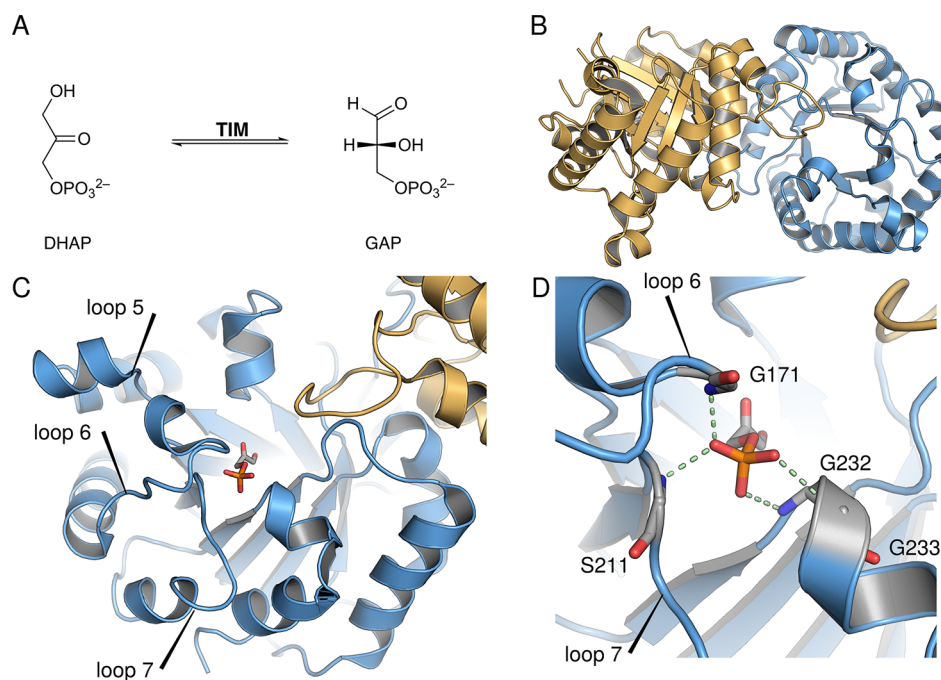
Figure 1. (A) Scheme showing the interconversion of DHAP and GAP. (B) Overall structure of yTIM (PDB ID: 1NEY),[16,35] highlighting also (C) loops 5, 6, and 7 and (D) the stabilization of the phosphodianion of the substrate DHAP by four discrete H-bonds formed with the protein backbone.

There has been extensive elegant experimental work on activation of TIM by phosphite dianions.[19,21,23,25]

The movement of the TIM loop was among the first functional enzyme motions to be investigated by molecular dynamics (MD) simulations. Seminal work by Kollman,[7] Karplus,[8] and co-workers reported MD simulations *in vacuo* using simple reaction coordinates and argued from the results that the loop closure is essentially a rigid-body type displacement, in which the loop moves like a "lid" attached to the protein by two hinges (Figure 2), the sequences of which are not conserved.[36] This simple picture has been highly influential and has become a widely accepted model for how this loop moves.[4,7,8] Specifically, previous NMR data were interpreted based on a two-state system in which the loop is only open and closed.[4,7−9,13,14,37] However, more recent studies have indicated that there is some degree of independence in the motions of the N- and C-terminal hinges of the loop.[14] Desamero et al.[17] have suggested that substrate binding to TIM is very fast, followed by slow unimolecular loop closure, with a population ratio between two states that favors the closed-loop conformation, but that is substantially less temperature dependent than the subsequent product release step. This is in contrast to previous NMR studies that suggested that ligand-free TIM is mostly open, but with small amplitude motions on the micro- to nanosecond time scale,[14] as well as work by Williams and McDermott[9] who argued that irrespective of whether TIM is substrate free, or bound to substrate GAP, or transition state (2-phosphoglycolate, PGA) analogues, the loop movement occurs at a rate similar to the empty enzyme, with a similar population ratio for the two conformers and with a measured rate for loop movement that is approximately matched to the turnover time. McDermott and colleagues[9,13,14,17] have similarly found that the rate of loop opening and closing is on the 10^4 s$^{−1}$ time scale for TIM, with both open and closed states being substantially populated.

Tawfik and co-workers have argued, based on theoretical considerations, that k_{cat} is unlikely to be higher than $10^6−10^7$ s$^{−1}$ for any enzyme, and, additionally, the apparent second-order rate constant (k_{cat}/K_M) for a diffusion-limited enzyme-catalyzed reaction with a single low-molecular mass substrate cannot exceed ∼$10^8−10^9$ s$^{−1}$ M$^{−1}$ (see ref 38 and references cited therein). In the case of TbbTIM, k_{cat} values of 300 s$^{−1}$ and 2100 s$^{−1}$, and corresponding k_{cat}/K_M values of 4.3 × 10^5 M$^{−1}$ s$^{−1}$ and 8.4 × 10^6 M$^{−1}$ s$^{−1}$, have been measured for the deprotonation of substrates DHAP and GAP, respectively.[21] Therefore, TIM is only partially diffusion-limited, and, while full loop closure is essential for the catalytic activity of the enzyme, the loop motion itself is only partially rate-limiting. Finally, more (comparatively) recent MD simulations of TIM loop motion surprisingly suggested multiple loop opening and closing events on a 100 ns time scale,[39,40] although this is on a much faster time scale than that derived from experiments, and could plausibly be either a force field artifact or an artifact of how loop opening and closure was measured and/or defined.

Clearly, therefore, despite being a prototype system for understanding functionally (and potentially catalytically) important loop motions,[4] the nature of the motion of these loops in TIM is far from fully understood. Addressing this issue in a satisfactory way would expand fundamental understanding of these regulatory loop motions. It might superficially be thought that loop motions provide a relatively easy case for molecular simulations, but in fact, as we show here, even simple loop motions provide a significant challenge for biomolecular simulation, because modeling loop motions is far from trivial computationally.[45,46] Loops that are regulated by ligand-gated conformational changes are a common feature of not just TIM barrel proteins,[47,48] but also unrelated enzymes such as orotidine 5′-monophosphate decarboxylase[49] (OD-Case) and glycerol 3-phosphate dehydrogenase (GPDH).[25,50] Analyzing how the motion of these loops is regulated and
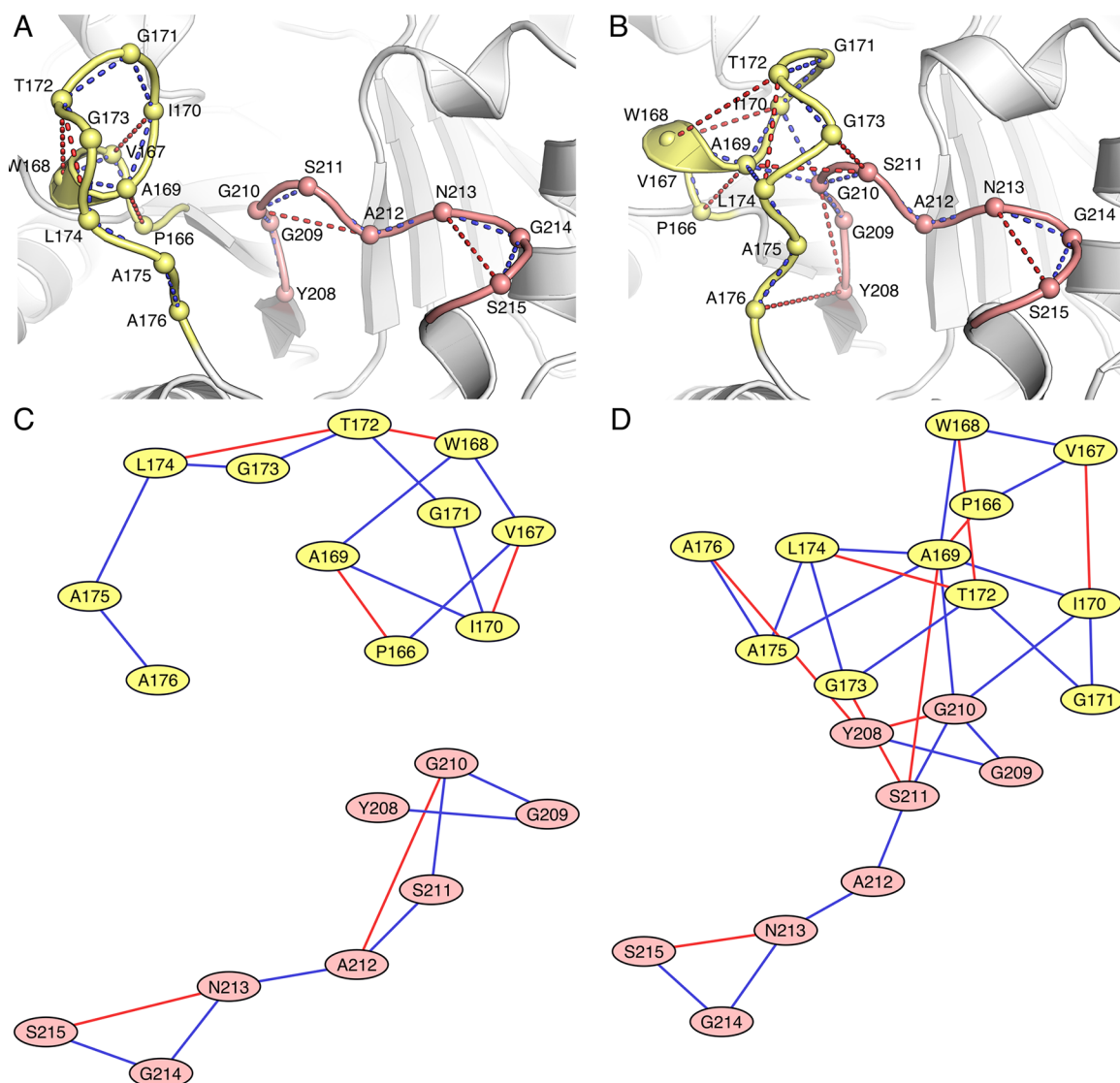
Figure 2. Structural differences between the (A) open- and (B) closed-state conformations of loops 6 (yellow) and 7 (pink) in the crystal structures of the dimeric TIM from *S. cerevisiae* (*y*TIM) with loop 6 in the open (PDB ID: 1YPI)[35,41] and closed (PDB ID: 1NEY)[16,35] conformations, respectively, with the spheres in panels (A) and (B) representing the C_α atoms of the annotated residues. All $C_\alpha-C_\alpha$ contacts within a 5 Å cutoff were mapped with blue lines, while the H-bonds are represented with red lines. The same interactions were mapped in (C, D), using Cytoscape 3.6.1,[42] RINalyzer 2.0,[43] and UCSF Chimera 1.13,[44] in order to show the interaction network.

potentially linked to their catalytic activity is central to understanding the catalytic proficiencies of these enzymes as well as, more generally, the high evolvability of the TIM barrel scaffold.[11,15,24,32−34]

The extensive experimental data on TIM loop dynamics (e.g., refs 9, 13, 14, and 17 among others) make TIM an ideal model system. We perform here detailed simulations of the conformational dynamics of TIM loop 6, using a range of different techniques, including long-time scale conventional MD simulations (which are used to construct Markov state models),[51−53] principal component analysis (PCA),[54−56] Hamiltonian replica exchange molecular dynamics (HREX-MD),[57] and bias-exchange metadynamics (BE-METAD)[58,59] simulations. These are complemented by empirical valence bond (EVB) simulations[60,61] of the TIM-catalyzed deprotonation of the substrate DHAP, which test the effects of varying loop conformations on the activation free energy for this process, using the conformational coordinates defined by the motion of loop 6. We illustrate that accurately modeling the

loop dynamics of even the apparently simple case of triosephosphate isomerase is challenging. Extensive conformational sampling is required to obtain a reliable model. Further, we demonstrate that loop 6 visits multiple stable conformational substates rather than being an example of a two-state rigid-body motion as widely thought. Also, this conformational dynamics has a clear impact on the catalytic activity of TIM. Therefore, in contrast to the simple two-state model often put forward in the literature,[4,7−9,13,14,37] the dynamics of the TIM active site loops is in fact far from a simple open and shut case.

■ METHODOLOGY

System Setup. In order to study the motion of loops 6 and 7 of both substrate-free TIM (apo-TIM) and TIM in complex with DHAP (DHAP-TIM), five crystal structures of the dimeric TIM from *S. cerevisiae* (*y*TIM) were chosen as starting structures (PDB IDs: 1NEY,[16,35] 1YPI,[35,41] 1I45,[13,35] 1NF0,[16,35] and 7TIM.[35,62] The three tryptophan mutations (Trp90Tyr, Trp157Phe, and Trp168Ftr, where Ftr = 5′fluorotryptophan) introduced into the structure of 1NEY were

mutated back so that all the five structures have an identical sequence, as in our previous work[26,27,63] (note also that these mutations have been shown not to affect the catalytic activity of the enzyme[16]). Two (1YPI and 1I45) of the five crystal structures are ligand-free, with loop 6 in its open state, while the other three were crystallized in complex with either a substrate or a transition state analog. Of these three structures, 1NEY and 7TIM are complexed with DHAP and phosphoglycolohydroxamate (PGH), respectively, with loop 6 in its closed state in both chains of the two structures. In 1NF0, DHAP is bound in the active sites of both chains, but, curiously, loop 6 is in its closed state in one chain, but in its open state in the other chain. All water molecules present in the crystal structures were removed, and in the case of the simulations of substrate-free yTIM, the ligands in 1NEY, 7TIM, and 1NF0 were also removed, considering only the protein coordinates as starting conformations for the simulations. In the case of the simulations of yTIM in complex with DHAP, the substrate was placed in the active site using Chimera[44] to perform structural alignment of the different structures onto PDB ID 1NEY, which is a structure of yTIM in complex with DHAP at 1.2 Å resolution.[16,35] We maintained the same histidine protonation patterns (protonating at the δ- vs ε-positions) as in our previous work,[26,27] as outlined in the Supporting Information of ref 26, and all other ionizable residues were kept in their default protonation states at physiological pH. Each system was placed into an octahedral box filled with TIP3P water molecules[64] with a distance of at least 10 Å from the solutes to the surface of the box. To neutralize the system, 10 Na$^+$ counterions were added to the TIM-DHAP complex, while 6 were added to the apo-TIM system. The protein was described using the Amber ff99SB-ILDN force field,[65] while the General AMBER Force Field 2 (GAFF2)[66] was used to obtain parameters to describe the substrate DHAP. The partial charges for DHAP were calculated using the standard restrained electrostatic potential (RESP) approach,[67] using Antechamber[68] as implemented into AMBER16,[69] and based on the electrostatic potential calculated in vacuum at the HF/6-31G(d) level of theory, using Gaussian 09 rev. D.01.[70] The GAFF2 parameters and partial charges of DHAP are listed in Table S1 in the Supporting Information. To keep the substrate DHAP stable in the TIM binding sites, several distance restraints were applied for all the DHAP-TIM simulations, as listed in Figure S1 and Table S2. Monitoring the P−O1−C1−C2 dihedral of DHAP indicates that the distance restraints do not distort the conformation of DHAP (Figure S2). These distance restraints were applied to all DHAP-TIM simulations *except* the empirical valence bond simulations, as described below.

Conventional Molecular Dynamics Simulations. The LEaP module of AMBER16[69] was used to generate the topology and coordinate files for the conventional MD simulations, which were carried out using the CUDA version of the PMEMD module[71−73] of the AMBER16 simulation package,[69] with system preparation as described above. The solvated system was first subjected to a 2000 step steepest descent minimization, followed by a 3000 step conjugate gradient minimization with positional restraints on all heavy atoms of the solute, using a 5 kcal mol^{-1} Å^{-2} harmonic potential. The minimized system was then heated up to 300 K using the Berendsen thermostat,[74] with a time constant of 1 ps for the coupling, and 5 kcal mol^{-1} Å^{-2} positional restraints (again a harmonic potential) applied during the heating process. The positional restraints were then gradually decreased to 1 kcal mol^{-1} Å^{-2} over five 500 ps steps of NPT equilibration, using the Berendsen thermostat[74] and barostat[74] to keep the system at 300 K and 1 atm. For the production runs, each system was subjected to either 200 or 400 ns of sampling in an NPT ensemble at constant temperature (300 K) and constant pressure (1 atm), controlled by the Langevin thermostat,[75,76] with a collision frequency of 2.0 ps^{-1}, and the Berendsen barostat with a coupling constant of 1.0 ps. A 2 fs time step was used for all simulations, and snapshots were saved from the simulation every 10 ps. The SHAKE algorithm[77] was applied to constrain all bonds involving hydrogen atoms. A 10 Å cutoff was applied to all nonbonded interactions, with the electrostatic interactions being treated with the particle mesh Ewald (PME) approach.[78] Between 10 and 22 independent

simulations were performed for each of the 5 starting structures, resulting in a cumulative sampling time of 22.0 and 26.4 μs for all simulations of apo- and DHAP-TIM, respectively. A summary of all the conventional MD simulations (time scale and number of replicas) performed in this work is shown in Table S3.

Hamiltonian Replica Exchange Molecular Dynamics Simulations. HREX-MD[57] simulations, an advanced sampling technique, were performed in order to further explore the conformational space of loops 6 and 7 of TIM. These simulations were based on the crystal structure of yTIM in complex with DHAP (PDB ID: 1NEY),[16,35] and in the case of the substrate-free simulations, the coordinates of the substrates were simply deleted from the PDB structure. The system setup was performed as described above for the conventional MD simulations. As with the conventional MD simulations, the resulting systems were then subjected to 5000 steps of each of steepest descent and conjugate gradient minimizations, after which they were heated up from 0 to 300 K over 500 ps of simulation time in an NVT ensemble, using the velocity scaling algorithm,[79] with a time constant of 1 ps. They were then equilibrated for a further 500 ps in an NPT ensemble at 300 K and 1 atm, controlled by the same thermostat and a Parrinello−Rahman barostat,[80] with a time constant of 2.0 ps. During the equilibration steps, 2.39 kcal mol^{-1} Å^{-2} positional restraints were applied to all heavy atoms, in each of the xyz directions. The final configuration of the system after equilibrations was then used as the starting structure for the HREX-MD simulations, which were performed without any restraints on the protein.

All HREX-MD simulations were performed using GROMACS v. 4.6.7,[81,82] interfaced with the PLUMED v. 2.1 plugin.[83] All bonds were constrained with the P-LINCS algorithm[84] in the simulations. The cutoff for the short-range nonbonded interactions was set to 10 Å, and the electrostatic interactions were calculated using the PME[78] method, in combination with periodic boundary conditions. All HREX-MD simulations were performed in an NPT ensemble, which was maintained using the same thermostat and barostat as used in the NPT equilibration. Loops 6 (Pro166-Ala176) and 7 (Tyr208-Ser211) of both chains were chosen as the "hot" region for the HREX-MD simulations. The charges in the "hot" region were scaled by $\sqrt{\lambda}$, the Lennard-Jones parameter by λ, and the proper dihedral angles were scaled by either λ or $\sqrt{\lambda}$, depending on whether both the first and the fourth of the dihedral, or just one of the two atoms of the dihedral, were in the "hot" region. The simulations were performed using 8 replicas, with the scaling factor, λ, exponentially ranging from 1.0 to 0.6. This corresponds to an effective temperature range from 300 to 500 K. Exchanges between neighboring replicas were attempted every 2 ps, and configurations were saved every 2 ps. A 4 fs integration step was used in the simulations of apo-TIM, as virtual sites were used for hydrogens, while a 2 fs time step was used in simulations of the TIM Michaelis complex with DHAP. Each replica was sampled for 200 ns in the case of the apo-TIM simulations and for 300 ns in the case of the DHAP-TIM simulations. This resulted in an average exchange rate of 20−30% in both simulations. Only the replica without scaling ($\lambda = 1.0$) was used for further analysis.

Bias-Exchange Metadynamics Simulations. In the BE-METAD approach,[58,59] a set of collective variables (CVs) that are expected to be relevant to the process under investigation are chosen for the simulations. A number of MD simulations (replicas) are then run in parallel, with each replica being biased by a history dependent Gaussian-type bias[58] acting on one of the selected CVs. During the BE-METAD simulations, the sampling is promoted by attempting swaps of the bias potentials between pairs of replicas at fixed time intervals. The probability of the swap is determined based on eq 1:

$$p_{ij} = \min\{1, \exp(\beta[V_G^i(X^i, t) + V_G^j(X^j, t) - V_G^i(X^j, t) - V_G^j(X^i, t)])\}$$

(1)

where X^i and X^j are the configurations of replica i and j, while V_G^i and V_G^j are the metadynamics bias potentials acting on replica i and j, respectively. A native contacts parameter (Q), ranging from 1 for the closed state using 1NEY as reference to 0 for the open state without contacts between loops 6 and 7, was used to define the CVs and was calculated using the following equation:[85]

$$Q(X) = \frac{1}{S} \sum_{(i,j) \in S} \frac{1}{1 + \exp[\beta(r_{ij}(X) - \lambda r_{ij}^0)]} \tag{2}$$

where X is a conformation along the reaction coordinate, $r_{ij}(X)$ is the distance between atoms i and j in conformation X, r_{ij}^0 is the distance between atoms i and j in the reference conformation, S is the number of all pairs of heavy atoms (i,j), β is set as 5 Å^{-1}, and λ is chosen as 1.4 for a tight determination of native contacts. For the 7 CVs corresponding to the 7 biased replicas, the distances between the heavy backbone atoms (N, C_α, C, O) of each of the residues Ala169-Ala175 of loop 6 and the backbone atoms of all residues of loop 7, that is, Tyr208-Ser211, were used to train the native contacts, using λ = 1.4 and without the 4.5 Å cutoff proposed in the original paper as some of the distances defined as CVs were greater than 4.5 Å. The reference distances for the BE-METAD simulations were taken from chain A of PDB ID 1NEY.[16,35] The 7 CVs used for these simulations are listed in Table S4. The eighth ("unbiased") replica was not biased by a metadynamics potential.

All BE-METAD simulations were performed using GROMACS v 5.1.4[81,82,86] interfaced with PLUMED v. 2.3.0.[83] A 2 fs integration step was used in the simulations, and all bonds were constrained with the P-LINCS algorithm.[84] The crystal structure of yTIM in complex with DHAP (PDB ID: 1NEY)[16,35] was used as the starting structure for the TIM-DHAP simulations, while the coordinates of the substrates were simply removed from the PDB structure for the apo-TIM simulations. The same protocol as used for the HREX-MD simulations was applied for both the equilibration and production BE-METAD runs but using a time step of 2 fs for both apo-TIM and DHAP-TIM. A Gaussian-shape biasing potential was added to each of the 7 CVs every 2 ps with an initial height of 0.1 kcal mol^{-1}, which was gradually decreased on the basis of adaptive biasing[87] by a factor of 10. The width of the Gaussians was set at 0.05. Exchanges between replicas were attempted every 2 ps, and each replica was sampled for 500 ns with configurations saved every 2 ps. This resulted in an average exchange rate of ~40%.

Computation of Markov State Models. For further elucidating the loop motions and kinetics, Markov state models (MSM)[51−53] were constructed from the conventional MD data with the PyEMMA Python library (v. 2.4).[88] The first stage of building an MSM is to discretize the conformational space of the molecule and reduce the trajectory to a sequence of transitions between discrete states. Obtaining good state space discretization is key to obtaining an MSM that is both descriptive and predictive.[51] To this end, we describe the conformation of TIM in terms of the distances between the C_α atoms in loops 6 and 7. Next, we apply PCA to reduce the dimensionality of our system characterization from 120 interatomic distances to 2 collective coordinates along which the largest motions take place. PCA is performed taking both the DHAP-TIM and apo-TIM data into consideration, so that both systems can be represented on a common set of collective coordinates and compared with each other. Subsequent operations, however, are applied to the two sets of data individually. K-means clustering is applied to these reduced systems to divide their respective conformation spaces into 100 "microstates". A granular 100-state MSM can then be built for both the apo- and DHAP-TIM systems by counting the transitions between the microstates, constructing a matrix of transition counts at a specified lag-time, and normalizing it by the total number of transitions emanating from each state to obtain the transition probability matrix. A lag time of 100 ns was chosen for our MSM because the time scales of the transition matrix were found to converge sufficiently by this interval. To aid our understanding, we coarse-grained the 100-state MSM into a hidden Markov model,[89] using the PCCA+ algorithm[90] to assign each microstate from the granular MSM a probability of belonging to a macrostate, thereby implementing a fuzzy lumping together of kinetically similar microstates.

Empirical Valence Bond Simulations. We recently performed extensive EVB simulations of the wild-type and mutant TIM-catalyzed deprotonation of substrates DHAP and GAP as well as substrate pieces glycoaldehyde (GA) and glycoaldehyde + phosphite ion (HP$_i$).[26,27] These studies focused only on modeling the initial rate-limiting deprotonation of the substrate and/or substrate piece, because the subsequent proton transfer steps are expected to be fast. We demonstrated that these models are able to capture the catalytic effect of not just the wild-type enzyme but also a range of key active site mutants with high quantitative accuracy. We have used the same simulation protocol to study the impact of loop 6 dynamics on the energetics of the TIM-catalyzed deprotonation of DHAP in the present work, with the exception of the fact that a shorter 20 ns equilibration was used (rather than 40 ns as in our previous work),[26] due to the much larger number of simulations involved. In brief, we extracted 10 structures at each of Q = 1.00, 0.99, 0.98, 0.97, 0.96, 0.95, 0.90, 0.85, 0.80, 0.70, 0.60, 0.50, 0.40 from our BE-METAD simulations of DHAP-TIM, to a total of 130 starting structures over all Q values considered. Note that we stopped at Q = 0.40 because this corresponded to a fully open conformation of loop 6. These structures were effectively randomly selected, with the only exception that we took into account the distance between the C_δ atom of Glu165 and the reacting carbon atom of the substrate as a selection criterion when extracting these structures, selecting only those in which the distance between these two atoms are <4.0 Å (the distance in the crystal structure, PDB ID: 1NEY,[16,35] is 3.83 Å). This was necessary because in many snapshots Glu165 moved out of the active site (in agreement with crystal structures which indicate a conformational change of this side chain out of the active site in structures of TIM with loop 6 in an open conformation),[4] and clearly those conformations will be noncatalytic. By selecting structures in which Glu165 points into the active site, even as the loop starts to open, we obtain a lower limit for the activation free energy for that loop conformation, assuming an idealized conformation for the Glu165 side chain.

We then performed 20 ns of equilibration on each structure, using the protocol described in refs 26 and 63, followed by 5 EVB simulations performed using 51 EVB mapping frames of 200 ps each, from the end point of each equilibration, with the starting points for each EVB simulation generated by performing an additional 110 ps of equilibration using five different random seeds. This led to a total of 50 individual EVB trajectories for structures extracted at each Q value and thus a total of 650 EVB trajectories over all Q values. We also performed additional EVB simulations on the crystal structure (PDB ID: 1NEY)[16,35] with loop 6 unrestrained, as a control. In the case of the crystal structure, we performed 10 initial equilibrations using 10 different random seeds and spawned five new EVB simulations from each equilibrated structure, by assigning new random seeds, as with the structures extracted from the BE-METAD simulations. Each individual EVB trajectory was sampled from the transition state in either reactant or product directions, as in our previous work, and each individual trajectory was a total of 10.2 ns, leading to a total simulation time of 510 ns over all 50 trajectories per loop conformation. This corresponded to a cumulative total of 2.8 μs equilibration and 7.14 μs of EVB simulation time over all systems.

All the EVB simulations were performed as described in detail in our previous work, the only exception to this being the sphere size used in our simulations, which was extended from a radius of 20 to 24 Å from the simulation center (the reacting carbon of substrate DHAP), in order to fully capture the flexibility of loop 6. The size of the water droplet used in our simulations was also extended from 20 to 24 Å accordingly. As with our previous work, all titratable residues with side chains that fell within the inner 85% of the sphere (i.e., within a distance of 20 Å from the reacting carbon of substrate DHAP) were ionized, and all residues in the outer 15% of the simulation sphere were kept neutral and restrained in place by a 10 kcal mol^{-1} Å^{-2} harmonic restraint (in the outer 15% of the simulation sphere) or a 200 kcal mol^{-1} Å^{-2} harmonic restraint (fully outside the simulation sphere). Note that neutralizing these titratable residues is a standard practice in such simulations in order to avoid system instabilities by having charged residues outside the simulation sphere.[26,27,91−96] A full list of all ionized residues and the protonation states of all histidine residues in this system can be found in Table S5. All EVB parameters used in this work are provided in the Supporting
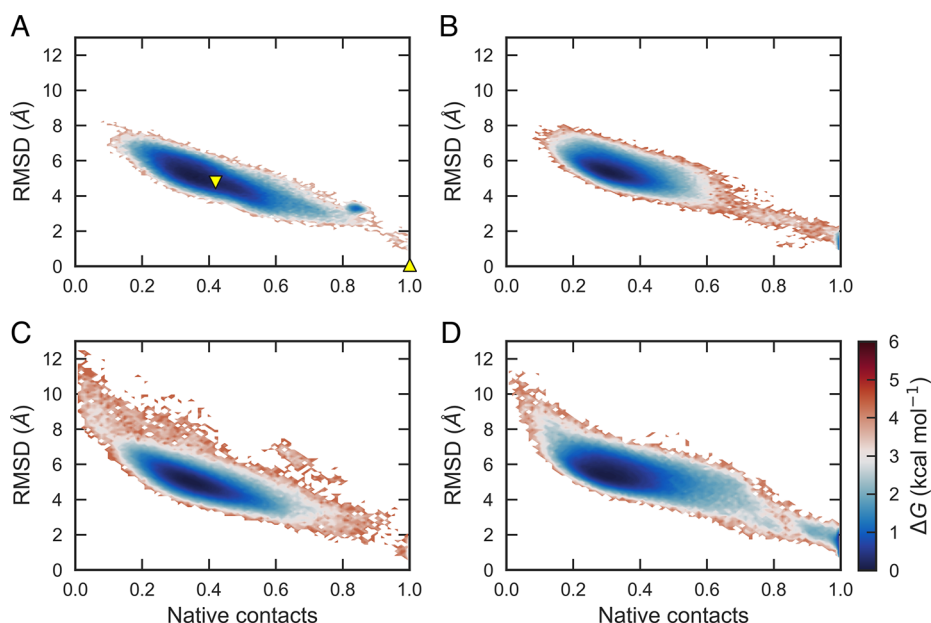
Figure 3. Relative free energy surfaces (T = 300 K) calculated from (A, B) HREX-MD and (C, D) BE-METAD simulations of (A, C) apo-TIM and (B, D) DHAP-TIM. The surfaces are defined in terms of the native contacts (Q) between loops 6 and 7 and the RMSD of the loop 6 backbone. The open and closed states of the loops are projected from chains A of crystal structures 1YPI (▼) and 1NEY (▲), respectively.

Information of ref 26. For full protocols and simulation details, please see refs 26 and 63.

Analysis. In order to define the closed and open states of loop 6, we calculated root-mean-square deviations (RMSD) of this loop, using the conformation of this loop in the structure of DHAP-TIM with loop 6 in a closed conformation (PDB ID: 1NEY)[16,35] as the reference conformation. The RMSD of the backbone of loop 6 was calculated after the alignment of all protein backbone atoms (apart from those from loop 6) to the reference conformation. In the BE-METAD simulations, the contacts between loops 6 and 7 were divided into 7 CVs (termed $Q1$ to $Q7$) and used for the different biased replicas (Table S4). The native contacts parameter, Q, with values from $Q1$ to $Q7$ (see eq 2) was used to distinguish the closed (Q approaching 1) and open ($Q < 0.45$) states of loop 6 of TIM. As shown in Table S6, the RMSD and Q value of loop 6 in its open conformation (from PDB ID: 1YPI)[35,41] are 4.77 Å and 0.419, respectively. Finally, the pK_as of all ionizable residues in the protein at different Q values obtained from our EVB Michaelis complexes (1000 snapshots extracted from 50 independent trajectories per Q value) were estimated using PROPKA 3.1.[97] The pK_a of the catalytic base, Glu165, was also estimated for crystal structures of TIM in complex with DHAP and PGA, as described in the Results and Discussion.

■ RESULTS AND DISCUSSION

Modeling TIM Loop Dynamics Using Conventional Molecular Dynamics Simulations. The classical picture of loop movement in TIM is that of a two-state rigid-body motion,[4,7−9,13,14,37] and this system is in fact often used as a prototypical example of such motion in biologically relevant systems.[98] This makes TIM a particularly important system to understand the role of loop motions in enzyme catalysis and protein function.

As our starting point, we performed a cumulative 48.4 μs of conventional MD simulations, as described in the Methodology section and summarized in Table S3. These involved simulations of yTIM from different starting structures of both the open and closed conformations of TIM loop 6 (residues 166−176), which were obtained from crystallography, with and without the substrate DHAP bound. The "open" and

"closed" conformations of loop 6 are defined as described in the Analysis section. Interestingly, as can be seen in Figures S3−S6, in all simulations, and irrespective of the starting conformation of the loop, loop 6 starts opening up substantially within the first 10−20 ns of these simulations and never fully closes again during the 400 ns of simulation time for the conventional MD simulations, including in the simulations of the DHAP-TIM complex. We also performed shorter test simulations with multiple different versions of the AMBER force field[65,99] as well as with the OPLS-AA force field[100,101] in order to rule out force field artifacts; we observed loop opening in all cases. The simulation time in the present study is substantially longer than in previous MD studies of TIM loop dynamics,[7,8,39,40,102] and it is plausible that similar observations of loop opening would have been made in those studies had the simulation time been extended.

Williams and McDermott have performed solid-state NMR measurements to study TIM loop movement in both apo-TIM and TIM in complex with substrate and transition state analogs.[9] Based on their experimental data, they predicted an activation free energy of 12 kcal mol^{-1} for the transition from open to closed conformations of loop 6 of apo-TIM (based on the rate at 10 °C), with a free energy difference of 1.8 kcal mol^{-1} in favor of the open conformation of the TIM loop in the apo form of the enzyme. They estimated that the presence of substrate and transition state analogues would alter this equilibrium slightly to favor the closed conformation of TIM loop 6 by up to 2.8 kcal mol^{-1} in the presence of the PGA trianion (and a barrier reduction from ∼12 to 9.2 kcal mol^{-1} for this transition). In light of the (from a simulation perspective) slow rate of change between the states and the small free energy difference between them, it is perhaps unsurprising that unbiased MD simulations struggle to capture the closure of loop 6 on reasonable simulation time scales. The key question then becomes how to capture this loop movement in a meaningful way during atomistic molecular simulations.
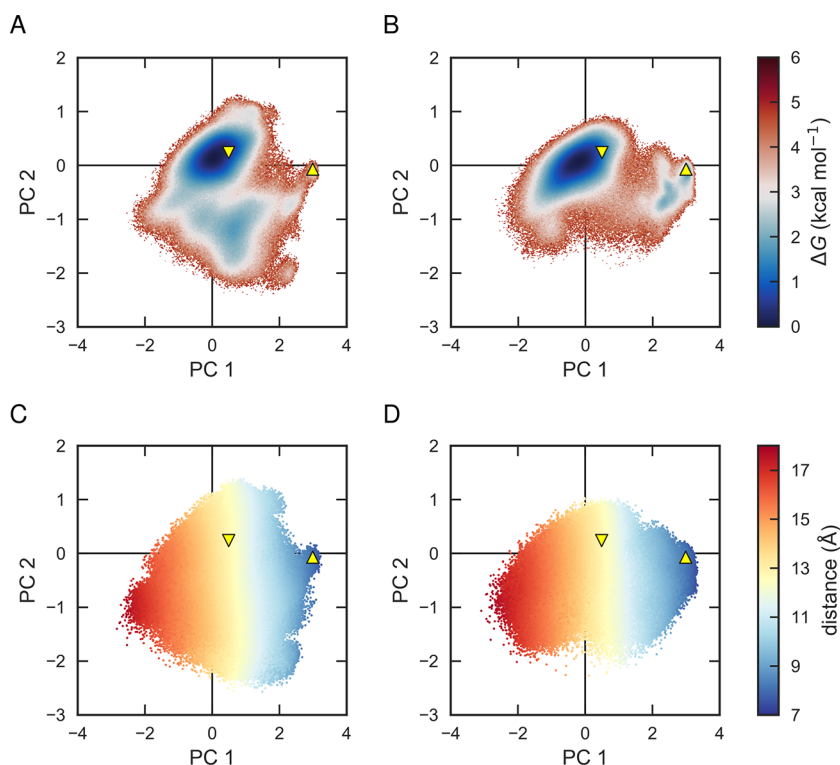
Figure 4. Free energy surfaces (at T = 300 K), defined in terms of the first two principal components, PC1 and PC2, obtained from conventional MD simulations of (A) apo-TIM (60 independent simulations totaling 22.0 μs) and (B) TIM in complex with DHAP (66 independent simulations totaling 26.4 μs, see Table S3). The distribution of distances (Å) between loops 6 and 7, defined as the average distance between the C_α atoms belonging to the two loops, along PC1, and PC2 are shown for (C) apo-TIM and (D) DHAP-TIM. The open and closed states of the loops are projected from chains A of crystal structures 1YPI (▼) and 1NEY (▲), respectively.

Using Enhanced Sampling Approaches To Capture the "TIM-Closed" State. In order to capture TIM loop motion, including the closed state of loop 6, we first performed HREX-MD) simulations, as described in the Methodology section and shown in Figures 3, S7, and S8. However, as with our conventional MD simulations, the HREX-MD simulations fail to capture the closed state of loop 6. That is, while the closed conformation of loop 6 appears to be very briefly sampled in the DHAP-TIM Michaelis complex (Figures 3 and S8), this is due to the fact that the simulations were initiated from the closed state of loop 6, and, even here, the loop opens within 25 ns of simulation time and does not close again. Again, the latter is not very surprising, given the fact that replica exchange MD simulations were shown to achieve about an order of magnitude of sampling speedup compared to conventional MD,[103] which is still not sufficient to reach the ~100 μs needed for loop 6 to close, as estimated by experiment.[14]

In contrast, however, BE-METAD simulations (Figures 3 and S9−S11) can capture a closed state of loop 6 in the presence of DHAP. That is, while the simulations of apo-TIM indicate only one minimum on the free energy surface projected along Q and RMSD, which corresponds to the open state of loop 6 with Q = 0.34 and an RMSD of 5.0 Å (Figure 3C), two minima are observed in the presence of DHAP (Figure 3D), one for the open state at (Q = 0.29, RMSD = 5.36 Å) and one for the closed state (Q = 1.00, RMSD = 1.91 Å). These states are essentially identical in free energy ($\Delta\Delta G$ = 0.1 kcal mol^{-1} in favor of the closed state, which falls within the noise of the simulations), yet the minimum for the open state involves a large catchment basin,

while the basin leading to the minimum of the closed state is small. This implies that loop 6 can adopt many different open states (with Q values ranging between 0.1 and 0.8 and RMSD values ranging between 3 and 8 Å), while the closed conformation is dominated by a single conformation of loops 6 and 7.

The convergence of the BE-METAD simulations was assessed by following the time evolution of the free energy profiles along each of the 7 CVs. The corresponding one-dimensional profiles for the last 200 ns of simulation time using a 50 ns time interval are shown in Figures S12 and S13 for the simulations of apo- and DHAP-TIM, respectively. In addition, as can be seen in Figures S14 and S15, the loops experience diffusive dynamics along the CVs, which is indicative of dynamics on a flattened energy landscape, wherein the biasing potentials compensate for the underlying free energy surface. Based on this, we consider the BE-METAD simulations to be converged for both apo- and DHAP-TIM.

Characterizing the Different Conformational States of Loop 6. Up to this point, our focus has been on capturing the closed state of loop 6 in our simulations. However, as it has been argued previously that loop movement in TIM is a two-state rigid-body movement,[4,7−9,13,14,37] it is also interesting to examine the dynamics of the open state of TIM. To achieve this, we returned to our conventional MD simulations and characterized the motion of loop 6 during the simulations using PCA, subsequently projecting the free energies for the enzyme along the most dominant motions, that is, PC1 and PC2.
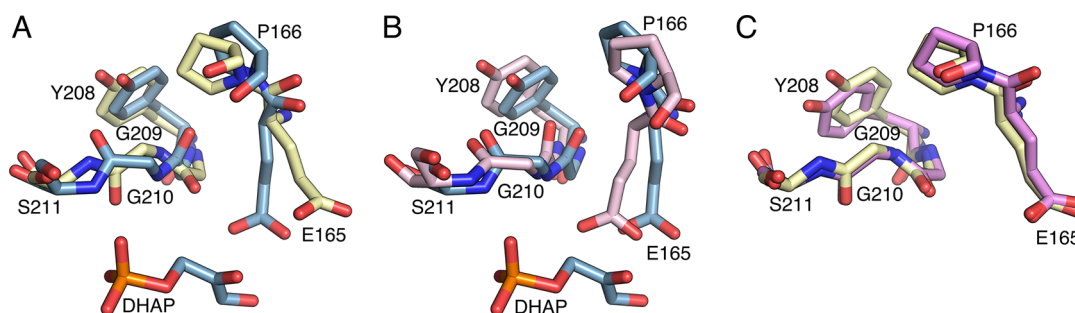
Figure 5. A comparison of the conformations taken by loop 7 (the YGGS motif) in structures of TIM at different conformations of loop 6, indicating the coupled conformational changes observed in loop 7 upon movement of loop 6.[4] (A) An overlay of the crystallographic structures of loop 7 in structures with closed (PDB ID: 1NEY,[16,35] blue) and open (PDB ID: 1YPI,[35,41] yellow) conformations of loop 6. (B) An overlay of the crystallographic (PDB ID: 1NEY,[16,35] blue) and MD structures (MSM centroid, pink) of loop 7 in structures with closed conformations of loop 6. (C) An overlay of the crystallographic (PDB ID: 1YPI,[35,41] yellow) and MD (MSM centroid, magenta) structures of loop 7 in structures with open conformations of loop 6.
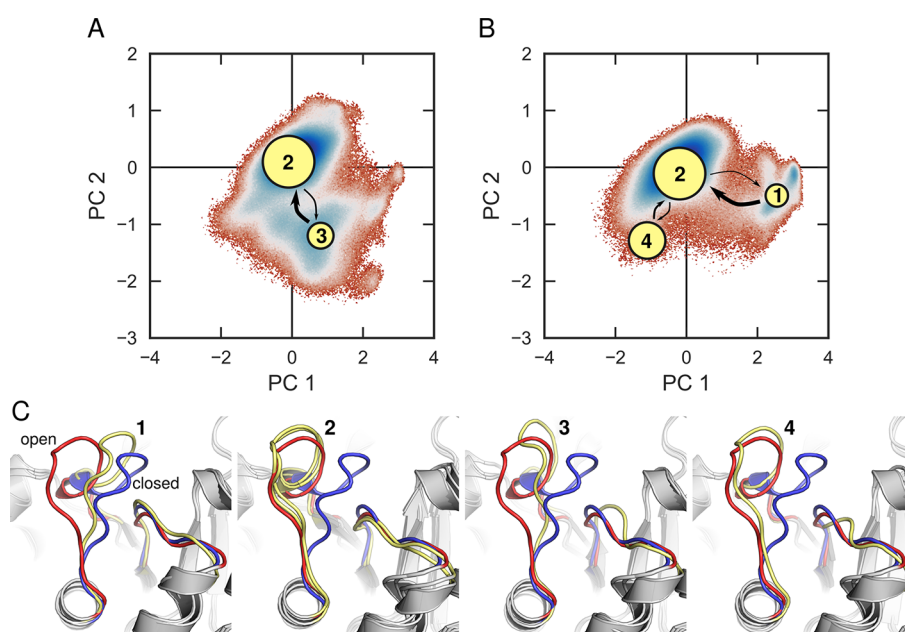

Figure 6. MSMs of (A) apo-TIM and (B) DHAP-TIM superimposed onto the corresponding free energy surfaces. The area of the nodes representing the metastable states corresponds to the population of the nodes, while the thickness of arrows corresponds to the transition probabilities. (Areas and thicknesses do not scale linearly with populations and transition probabilities, which are given in Table S7). (C) Overlays of representative structures for each metastable state sampled in simulations of apo-TIM and DHAP-TIM, with the numbering corresponding to the states shown in (A) and (B). The crystal structure for the open conformation of the loop (PDB ID: 1YPI)[35,41] is shown in red, and the closed conformation of the loop (PDB ID: 1NEY)[16,35] is shown in blue, and the representative structure(s) from the MSM centroid are shown in yellow. Note that metastable state 2 is virtually identical for apo- and DHAP-TIM, as indicated by the two structures shown in panel (C2). The corresponding % native contacts (Q) and RMSD values for the five centroids are listed in Table S6. As can be seen in (C) and (D), in the case of apo-TIM, both centroids correspond to open conformations of loop 6, whereas in the case of DHAP-TIM, one closed and two open states are sampled in the simulations.

In order to facilitate direct comparison between the two systems, we performed the PCA on the combined simulation data obtained from conventional MD. We then calculated the free energy from the simulations of apo-TIM and DHAP-TIM separately, but along the same principal components (Figure 4A,B, respectively). In Figure 4C,D, the interloop distance, defined as the average distance between the C_α atoms belonging to the two loops, is projected onto the corresponding free energy landscape.

These plots clearly demonstrate that, in both systems, the main conformational motions of the loop take place along the first PC. On the other hand, the free energy surfaces also reveal considerable loop motion along the second PC, especially in

the open state of loop 6. That is, this figure shows that while PC1 is the motion that mainly distinguishes between the open and closed states of loop 6, PC2 indicates that several conformational pathways exist for loop opening and closure.

Inspection of conformations along PC2 indicates a simultaneous inward/outward motion of loops 6 and 7, where loop 6 becomes slightly twisted as one-half of it (residues 169−173) tries to move closer toward loop 7, which simultaneously turns away from loop 6. This is in agreement both with computational studies using coarse-grained simulations that suggest a concerted motion for these loops,[30] and with experimental observations[4,13,16,41,62] (based on structural information) of likely concerted loop motion for loops 6 and 7,

due to a ligand-induced conformational change of the YGGS motif of loop 7 (Figure 5), which in turn causes the Glu165-Pro166 dipeptide to rotate, resulting in the placement of Glu165 in a catalytically competent conformation in the TIM active site.[4] Specifically, ligand binding causes a 90° rotation of the Gly209-Gly210 peptide plane, which further leads to a 90° rotation of O (Gly209). This creates a steric clash with the Pro166 side chain, which triggers the conformational movement along the Glu165-Pro166 plane that places Glu165 in the active site. These movements are also coupled to a substantial 180° rotation along the Gly210-Ser211 peptide bond.[4] Figure 5 shows a comparison of crystallographic structures and snapshots from our simulation, illustrating that our simulations sample an approximation to these substantial conformational changes, in good agreement with structural information.[16,41] Therefore, our extensive atomistic simulations, in line with previous coarse-grained simulations by Katebi and Jernigan,[30] show that motion of loops 6 and 7 is concerted. We now also characterize this concerted motion in terms of both thermodynamics and kinetics. In addition, although these loops do sample multiple conformational states, they are not independently flexible, but rather the conformational space sampled by the loops is interdependent, restricting which combined conformations are available.

As discussed in the previous section, the conformations sampled for the open state in the simulations populate a large basin on the free energy surface. On the other hand, the closed state is not stably sampled in our conventional MD simulations. Consistent with this, simulations that started from the closed state quickly evolved toward the open state, while the reverse was never observed (also true for the HREX-MD simulations). For the DHAP-TIM complex, conformations were sampled leading toward the crystal structure of the closed state (PC1 > 2 and interloop distances ≤10 Å), which were, however, not observed for apo-TIM. In the absence of the substrate DHAP, the minimal interloop distances that were obtained were around 10 Å, which corresponds to a semi-open (or semi-closed) state. This is due to the fact that the closed state is stabilized by H-bonds between the substrate and loops 6 and 7 in DHAP-TIM, which are missing in the ligand-free enzyme.

Finally, in order to further elucidate the dynamics of loop 6 as well as the time scales of these motions, we built MSMs from the extensive conventional MD data and superimposed them onto the calculated free energy surfaces (Figure 6). The large 7:1 ratio between the first and second slowest time scales for the MSM constructed from simulations of apo-TIM results in only two metastable states. The loop dynamics in the presence of DHAP can be characterized by three different metastable states, of which state 1 represents a semi-closed, yet not very stable state (<1% population, see Figure 6 and Table S6). State 2 is the same for apo- and DHAP-TIM and is conformationally similar to the open state as determined by crystallography (see, e.g., PDB ID: 1YPI).[35,41] It is also the most stable state, which is populated 90.9% of the time in simulations of apo-TIM and 63.8% of the time in simulations of DHAP-TIM. Both systems can adopt another open state, which is, however, different between the simulations of apo- and DHAP-TIM. As these states are not identical, we denote these states as state 3 for apo-TIM and state 4 for DHAP-TIM. The main motion for the interconversion between the two different open states in both systems is observed along PC2. State 3 in apo-TIM, which has a population of 9.1%, is

represented by structures in which loops 6 and 7 are closer to the TIM barrel than is the case in state 2. This is also reflected in the somewhat smaller interloop distances of 11−13 Å in state 3 vs 13−15 Å in state 2 (Figure 6). State 4 is the most open metastable state for DHAP-TIM, with interloop distances of 15−17 Å and a population of about 36.1%. While there is still some loop motion along PC2 in the presence of DHAP, this motion is suppressed compared to the corresponding motion in the ligand-free enzyme, as the differently populated conformational areas belonging to states 3 and 4 reveal (Figure 6). This plausibly indicates that, in the presence of ligand, the sampling is more confined to the open-closed transition pathway, which in turn may indicate a ligand gated conformational change, as has been suggested in several experimental studies (e.g., refs 13, 17, 22, and 27).

Another difference between DHAP-TIM and apo-TIM is the time scales for the interconversion between the metastable states. As already mentioned above, the opening of loop 6 happens very quickly in our simulations, as evidenced by the, on average, 55 ns that DHAP-TIM needs for the transition from the semi-closed state 1 to the open state 2. The transition back to state 1, however, appears to be a slow process with a predicted time scale of 39.2 μs. This is by a factor of 2 faster than the experimentally determined time scale for this motion,[14] although the experimentally measured values are associated with full closure of the loop, which we do not observe in our simulations. In light of this, the predicted time scale for the closure of loop 6 can be considered to be in good agreement with the experimental value, with a predicted activation barrier of ~11.5 kcal mol⁻¹, compared to an experimental estimate of 12 kcal mol⁻¹. However, loop opening should happen much more slowly than observed in our simulations, as the rate of loop opening is predicted to be only ~10-fold faster than loop closing in the apoenzyme and up to 100-fold slower than loop closing when TIM is in complex with a substrate or transition state analogue.[13,14] This overestimate of the rate of loop opening is probably due to an underestimation of the stability of the closed state of loop 6. The calculated interconversion here between the two open states of DHAP-TIM is fast with time scales of 168 ns for the transition from state 2 to state 4 and of 95 ns for the transition back to state 2. The 10-fold higher population of state 2 compared to state 3 in apo-TIM is also reflected in the transition rates: the time scale for the transition from state 2 to state 3 is 4.91 μs, whereas it takes only 0.49 μs for the back-transition. Therefore, our simulations show fast switching of loop conformations between different open states of the enzyme, which has also been reported in experiments of the apoenzyme where time scales of 10⁵−10⁶ s⁻¹ for small-angle motions of loop 6 were found.[14] In summary, these findings show that TIM loop motion is clearly more complex than a simple rigid-body movement involving two well-defined conformational states and that the "open" conformation of loop 6 samples multiple distinct conformational substates.

Exploring the Link Between Loop 6 Dynamics and the Reaction Catalyzed by TIM. Having extensively studied the conformational dynamics of loop 6, the obvious remaining question is how this affects the catalytic activity of the enzyme and whether the structures sampled along the minimum energy pathway from our BE-METAD simulations represent a transition toward a catalytically competent closed conformation of the enzyme. We have previously modeled the deprotonation of substrates DHAP and GAP by TIM in the
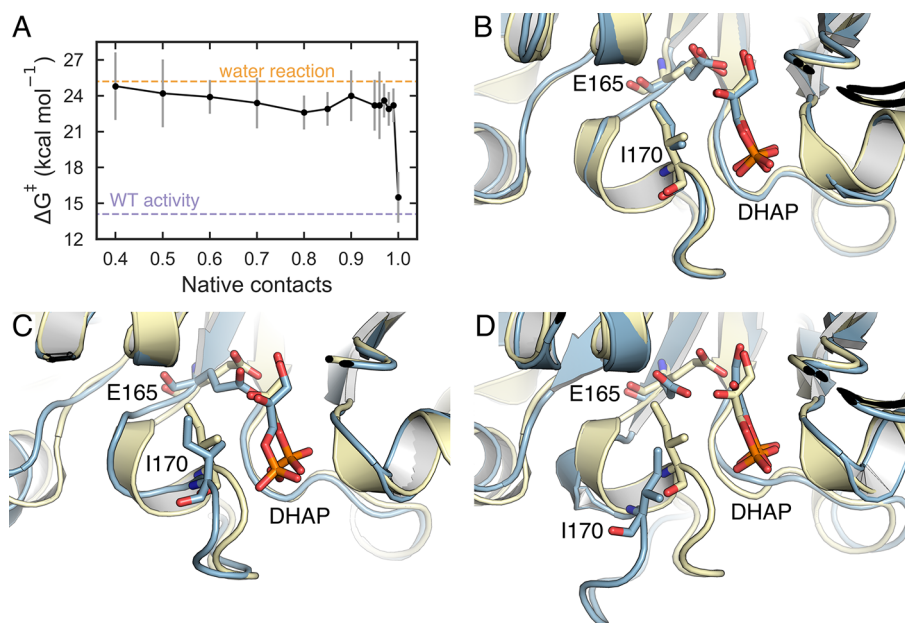
Figure 7. (A) Activation free energies obtained from EVB simulations of TIM at different conformations of loop 6, based on structures extracted from our BE-METAD simulations, as described in the main text. Data points show averages and standard deviations (as error bars) over 50 individual EVB trajectories per Q value. The corresponding raw data are shown in Table S8. (B−D) Comparison of the crystal structure of TIM with loop 6 in a closed conformation (PDB ID: 1NEY,[16,35] yellow) with representative starting structures from our BE-METAD simulations (blue), at native contacts of (B) $Q = 1.00$, (C) $Q = 0.99$, and (D) $Q = 0.40$. For structural overlays of all 10 structures used for the EVB simulations at each Q value, see Figure S16. For a description of how the structures were selected, see the main text.

closed state of the enzyme,[26] using the EVB approach and the same PDB structure (PDB ID: 1NEY)[16,35] that was also used in this work. Both we and others have demonstrated the importance of a very precise network of interactions in the closed state in facilitating catalysis by this enzyme.[26,104,105] In addition, previous work[26,106] has shown that mutations that allow even a single water molecule to enter the active site in the vicinity of the general base, Glu165, negatively impact the thermodynamic barrier to substrate deprotonation and that there is in turn a near-linear correlation between the reaction free energy (thermodynamic barrier, $\Delta G°$) and activation free energy (kinetic barrier, $\Delta G^{\ddagger}$).

In the present work, we performed EVB simulations of the deprotonation of substrate DHAP, starting from (1) the fully closed state of the enzyme (i.e., the crystal structure) with no restraint on loop 6 (see also ref 26) as well as (2) evenly distributed snapshots of the protein taken from our BE-METAD simulations over a range of different Q values from 1.0 to 0.4; 1.0 and 0.4 are reasonable approximations of the crystallographic "closed" and "open" conformations of loop 6 (see Figure 7). We extracted snapshots at 0.05 intervals of Q (see Methodology section) as well as additional snapshots in 0.01 increments between $Q = 1.0$ and $Q = 0.95$, to capture how quickly loop opening has an impact on the calculated energetics. In the case of calculations initiated from conformations of the loop other than the initial crystal structure, a 1 kcal mol^{-1} Å^{-2} restraint was placed on all backbone heavy atoms of loop 6 residues (Val167-Ala176) in order to maintain a Q value close to that extracted from the BE-METAD simulations.

We note that in the crystallographically observed open conformation of loop 6 (see, e.g., PDB ID: 1YPI),[35,41] there is a substantial displacement of Glu165 toward a noncatalytic conformation in which it points out of the active site.[13,41]

While we also sampled this conformation in our BE-METAD simulations, we selected structures at different Q values with this side chain pointing into the active site in a catalytically competent conformation (as described in the Methodology section), although this is not likely at smaller Q values, as loops 6 and 7 open substantially. It does, however, provide a lower limit to the activation free energy, assuming ideal positioning of Glu165. We note also that our structures for $Q = 1.0$ were extracted from the first 300 ps of the unbiased replica from the BE-METAD simulation.

From our data (Figure 7 and Table S8), it can be seen that the EVB simulations performed for the crystal structure and conformations at $Q = 1.0$, which were selected early enough in the simulation to be virtually identical to the crystal structure, give excellent agreement with the experimental value for the deprotonation of DHAP (calculated activation free energies of 15.2 ± 0.2 and 15.5 ± 0.3 kcal mol^{-1}, respectively, compared to the experimental value of 14.1 kcal mol^{-1}).[21] Already at $Q = 0.99$, however, only a small perturbation to the structure, the activation free energy, jumps to 23.2 ± 0.2 kcal mol^{-1}, which is similar to the experimentally estimated activation free energy for the non-enzymatic reaction (25.2 kcal mol^{-1}),[10,107] indicating that loop displacement has created an environment very similar to simply performing the reaction in water. This 8 kcal mol^{-1} increase upon opening of the loop that we observe computationally is also in qualitative agreement with the effect of an experimental loop deletion mutant, where deletion of residues 170−173 of loop 6 (Ile170, Gly171, Thr172, Gly173) led to a ~6.4 kcal mol^{-1} increase in activation free energy.[108] In addition, the overall reaction free energy increases from 6.8 ± 0.4 to 16.9 ± 0.3 kcal mol^{-1}, which is again similar to the estimate for the non-enzymatic reaction (18.9 kcal mol^{-1}),[10] in line with our previously observed linear relationship between kinetic and thermodynamic barriers to the reaction[26] (i.e.,

between the activation and reaction free energies). These values then remain fairly similar across the structures selected for the simulations, even compared to $Q = 0.4$ at which point the loop is open (for comparison, the fully open crystal structure, PDB ID: 1YPI,[35,41] is at $Q = 0.42$).

Finally, a comparison of the structures at $Q = 1.0$ and $Q = 0.99$ shows that even at this point there is a substantial displacement of loop 6, which leads to flooding of the active site with water molecules from ~8 at $Q = 1.0$ to as high ~14 at lower Q values (see Figure 7 and Tables S9 and S10) as well as a displacement in substrate position and key active site side chains, which remains the case with decreasing Q values (see Table S11, although the impact of this substrate displacement on reacting geometries is minimal, as highlighted in Table S12). It is notable that the largest displacement upon loop opening (apart from displacements of the reacting atoms, substrate DHAP and side chain of Glu165) is observed in the side chain of Ile170 (Figure 7 and Table S11). Ile170 is already subtly displaced at $Q = 0.99$ and then further displaced at $Q = 0.4$ (Figure 7B–D). Both we,[26] and prior to that Richard et al.,[106] have shown the importance of the active site residues Ile170 and Leu230 in forming a "hydrophobic clamp" that regulates the basicity of the Glu165 side chain by blocking the access of solvent to the active site. The shift in position of the Ile170 side chain in conjunction with (small) displacements of loop 6 is therefore notable.

In summary, as can be seen from these simulations, achieving the crystallographic closed conformation is extremely challenging. As our metadynamics simulations bias only the backbone, even when the native contacts shown in Figure 7 are as high as $Q = 0.99$, the structure is not the fully (crystallographically observed) closed state: loop 6 remains partially open, and the protein is in a catalytically incompetent conformation. The negative impact of this on the activation free energy is perhaps unsurprising, as the role of active site architecture and solvent exclusion in catalysis by TIM and related enzymes has been discussed in detail elsewhere.[26] However, these data demonstrate that despite the overall flexibility of the TIM loop, and the fact that it appears to sample many different conformations, it must be fully closed over the active site for efficient catalysis. This is necessary to provide the key interactions for efficient transition-state stabilization, while simultaneously creating a hydrophobic cage that excludes water molecules from the active site, thus contributing to the elevated pK_a of the catalytic base, Glu165, which is in turn necessary to promote the deprotonation of the substrate. This pK_a has been experimentally estimated to be as high as ~10 in the presence of the phosphoglycolate (PGA) trianion.[109] We have performed qualitative estimates of the pK_a of this residue at different Q values using PROPKA 3.1.[97] We have, in addition, estimated the pK_a of this residue in PDB IDs: 1NEY[16,35] and 4TIM,[35,110] which are structures of TIM in complex with DHAP and PGA, respectively (see Figure S18 and Table S14). PROPKA underestimates the pK_a of Glu165 in the presence of PGA (calculated pK_a of 7.4 compared to ~10 from experiment),[109] but it should predict trends in its pK_a, indicating how the pK_a changes with changing loop conformation. These calculations show that the pK_a of Glu165 is highest when the loop is fully closed and drops as loop opening increasingly exposes the active site to solvent. No other ionizable residues show changes in pK_a that follow a trend with Q values that could lead to changes in protonation states upon loop opening or closure. We note that the

importance of creating a hydrophobic cage for the reaction of TIM to occur has been discussed at great length elsewhere in the literature (see, e.g., refs 26, 104, and 106). Closure of the active site to provide such a hydrophobic environment (reducing solvation of the catalytic base) to favor proton abstraction by the active site base is also observed in other enzymes, such as ketosteroid isomerase,[111] and is essential for efficient reaction.

Finally, the conformational differences we observe between apo-TIM and DHAP-TIM highlight the role of ligand binding in helping to attain this catalytically competent conformation. This provides further evidence that the large energetic changes observed when comparing the reaction of the full substrate GAP and that of the substrate piece glycoaldehyde in the absence and presence of phosphite dianion[27] are not due to chemical activation, but rather due to ligand binding shifting the conformational equilibria between the open and closed conformations of loop 6 of this enzyme (see also the discussion in ref 98).

■ CONCLUSIONS

Enzyme conformational changes, in particular of active site loops, are a common feature of the regulation of enzyme activity as well as enzyme functional evolution.[112−115] Triosephosphate isomerase in particular has been a prototype system for understanding the importance of loop conformational changes in enzyme function. In addition, understanding how TIM is regulated is of interest from a biocatalysis perspective, as the TIM barrel is one of the most versatile and evolvable protein scaffolds.[11,15,24,32] The conventional image of loop motion in TIM has been that of a simple two-state rigid-body motion,[4,7−9,13,14,37] which is ligand-gated, with ligand binding shifting the equilibrium of the closed-state of loop 6 from a marginally stable conformation to the preferred, catalytically competent conformation of the enzyme.[22]

In the present study, we have performed detailed simulations of the conformational dynamics of loop 6 and the role of this dynamics in the TIM-catalyzed deprotonation of substrate DHAP, using conventional MD, Hamiltonian replica exchange, bias-exchange metadynamics, and empirical valence bond simulations, making this to our knowledge the most comprehensive computational study of TIM conformational dynamics to date. Through comparing these different techniques, we demonstrate that highly sophisticated sampling methods are necessary to obtain a meaningful description of this loop closure using atomistic simulations. Taken together, our simulations demonstrate for the first time that, contrary to the simple two-state rigid-body motion often assumed in the literature, loop 6 is dynamic and flexible and can sample multiple kinetically distinguishable conformational states. Furthermore, the nature of these states is altered by the presence of the substrate. However, despite sampling these multiple conformational states, the enzyme is only fully active when in a structure that approximates the crystallographically observed completely closed state of the enzyme (as shown in our EVB simulations). This is because loop 6 closure is necessary in order to create a hydrophobic cage that elevates the pK_a of Glu165 such that it is able to deprotonate the substrate. This insight impacts both design of experiments to analyze this loop motion (for example, additional sites may need to be labeled in NMR studies) as well as the choice of appropriate reaction coordinates in computational studies; simple geometric reaction coordinates such as specific

pseudodihedrals, as commonly used in many computational studies,[8,102] are unlikely to capture the overall conformational dynamics of the loop. This will also pose challenges for the further refinement of tools for automatic prediction of loop dynamics, such as, for example, ref 116.

Our EVB simulations, which complement our earlier studies of this enzyme,[26,27] demonstrate that complete loop closure is absolutely critical for efficient catalysis by TIM, which is, in turn, driven by a combination of a network of highly specific hydrogen bonding and electrostatic interactions, combined with solvent exclusion from the active site. We also demonstrate that even slight displacements of the loop away from its catalytically competent conformation can have substantial negative impact on the calculated activation free-energies, due to breaking these interactions and allowing solvent to penetrate the active site.

So, is loop motion in TIM a simple open and shut case? Our simulations altogether show that yes it is, in the sense that full closure of the loop is essential for catalytic activity, but they also show that the motion of the loop is complex. Our study shows that a conformational change that was thought to be simple is in fact complex, sampling multiple states along the reaction coordinate. We emphasize again that, while our study has focused on TIM as a model system, such ligand-gated conformational changes are certainly not unique to TIM and are in fact common. It will be important in the future to characterize the motion of functional loops and their effects on activity in other enzymes. The binding of ligands to other enzymes (structurally and functionally unrelated to TIM) such as orotidine 5′-monophosphate decarboxylase[49] (ODCase) and glycerol 3-phosphate dehydrogenase (GPDH)[25,50] suggests an evolutionary role for the regulation of loop dynamics. In contrast, highly promiscuous enzymes frequently possess large and "sloppy" active sites with fewer specific binding interactions, which is thought to allow them to adapt to different substrates for efficient catalysis.[112−114,117]

From a computational perspective, the results here demonstrate that simulating loop dynamics is challenging, requiring extensive conformational sampling and specialized computational approaches. The complexity of what had been thought to be a simple loop motion is a sobering reminder of the challenges posed by simulations of protein conformational changes and the need for enhanced sampling methods, on time scales of at least several hundred ns, in order to be able to characterize these motions. The results here demonstrate, however, the essential role of atomically detailed simulations in understanding protein conformational behavior, and in informing and complementing experiments and the interpretation of experimental data.[118] Finally, simulations of ligand-gated conformational changes and their role in catalysis provide a promising route for understanding the catalytic activity of many enzymes and assisting protein engineering and artificial enzyme design.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b09378.

> Additional simulation details and analysis (PDF)

> Loop 6 closure along the native contacts (Q) reaction coordinate based on snapshots extracted from our BE-METAD simulations (MOV)

AUTHOR INFORMATION

Corresponding Author

*lynn.kamerlin@kemi.uu.se

ORCID ⓘ

Qinghua Liao: 0000-0002-2260-8493
Dušan Petrović: 0000-0002-1834-7358
Adrian J. Mulholland: 0000-0003-1015-4567
Marc W. van der Kamp: 0000-0002-8060-3359
Birgit Strodel: 0000-0002-8734-7765
Shina Caroline Lynn Kamerlin: 0000-0002-3190-1173

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Knowles, J. R. To Build an Enzyme. *Philos. Trans. R. Soc., B* **1991**, *332*, 115−121.

(2) Knowles, J. R. Enzyme Catalysis: Not Different, Just Better. *Nature* **1991**, *350*, 121−124.

(3) Knowles, J. R.; Albery, W. J. Perfection in Enzyme Catalysis: The Energetics of Triosephosphate Isomerase. *Acc. Chem. Res.* **1977**, *10*, 105−111.

(4) Wierenga, R. K.; et al. Triosephosphate Isomerase: a Highly Evolved Biocatalyst. *Cell. Mol. Life Sci.* **2010**, *67*, 3961−3982.

(5) Richard, J. P. A Paradigm for Enzyme-Catalyzed Proton Transfer at Carbon: Triosephosphate Isomerase. *Biochemistry* **2012**, *51*, 2652−2661.

(6) Alagona, G.; Desmeules, P.; Ghio, C.; Kollman, P. A. Quantum Mechanical and Molecular Mechanical Studies on a Model for the Dihydroxyacetone Phosphate-Glyceraldehyde Phosphate Isomerization Catalyzed by Triose Phosphate Isomerase (TIM). *J. Am. Chem. Soc.* **1984**, *106*, 3623−3632.

(7) Brown, F. K.; Kollman, P. A. Molecular Dynamics Simulations of ″Loop Closing″ in the Enzyme Triose Phosphate Isomerase. *J. Mol. Biol.* **1987**, *198*, 533−546.

(8) Joseph, D.; Petsko, G.; Karplus, M. Anatomy of a Conformational Change: Hinged ″Lid″ Motion of the Triosephosphate Isomerase Loop. *Science* **1990**, *249*, 1425−1428.

(9) Williams, J. C.; McDermott, A. E. Dynamics of the Flexible Loop of Triose-Phosphate Isomerase: The Loop Motion Is Not Ligand Gated. *Biochemistry* **1995**, *34*, 8309−8319.

(10) Åqvist, J.; Fothergill, M. Computer Simulation of the Triosephosphate Isomerase Catalyzed Reaction. *J. Biol. Chem.* **1996**, *271*, 10010−10016.

(11) Wierenga, R. K. The TIM-Barrel Fold: A Versatile Framework for Efficient Enzymes. *FEBS Lett.* **2001**, *492*, 193−198.

(12) Cui, Q.; Karplus, M. Quantum Mechanical/Molecular Mechanical Studies of the Triosephosphate Isomerase-Catalyzed Reaction: Verification of Methodology and Analysis of Reaction Mechanisms. *J. Phys. Chem. B* **2002**, *106*, 1768−1798.

(13) Rozovsky, S.; Jogl, G.; Tong, L.; McDermott, A. E. Solution-State NMR Investigations of Triosephosphate Isomerase Active Site Loop Motion: Ligand Release in Relation to Active Site Loop Dynamics. *J. Mol. Biol.* **2001**, *310*, 271−280.

(14) Rozovsky, S.; McDermott, A. E. The Time Scale of the Catalytic Loop Motion in Triosephosphate Isomerase. *J. Mol. Biol.* **2001**, *310*, 259−270.

(15) Nagano, N.; Orengo, C. A.; Thornton, J. M. One Fold with Many Functions: The Evolutionary Relationships between TIM Barrel Families Based on their Sequences, Structures and Functions. *J. Mol. Biol.* **2002**, *321*, 741−765.

(16) Jogl, G.; Rozovsky, S.; McDermott, A. E.; Tong, L. Optimal Alignment for Enzymatic Proton Transfer: Structure of the Michaelis Complex of Triosephosphate Isomerase at 1.2-Å Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 50−55.

(17) Desamero, R.; Rozovsky, S.; Zhadin, N.; McDermott, A.; Callender, R. Active Site Loop Motion in Triosephosphate Isomerase: T-Jump Relaxation Spectroscopy of Thermal Activation. *Biochemistry* **2003**, *42*, 2941−2951.

(18) Massi, F.; Wang, C.; Palmer, A. G. Solution NMR and Computer Simulation Studies of Active Site Loop Motion in Triosephosphate Isomerase. *Biochemistry* **2006**, *45*, 10787−10794.

(19) Amyes, T. L.; Richard, J. P. Enzymatic Catalysis of Proton Transfer at Carbon: Activation of Triosephosphate Isomerase by Phosphite Dianion. *Biochemistry* **2007**, *46*, 5841−5854.

(20) Malabanan, M. M.; Amyes, T. L.; Richard, J. P. A Role for Flexible Loops in Enzyme Catalysis. *Curr. Opin. Struct. Biol.* **2010**, *20*, 702−710.

(21) Malabanan, M. M.; Go, M. K.; Amyes, T. L.; Richard, J. P. Wildtype and Engineered Monomeric Triosephosphate Isomerase from Trypanosoma Brucei: Partitioning of Reaction Intermediates in D$_2$O and Activation by Phosphite Dianion. *Biochemistry* **2011**, *50*, 5767−5779.

(22) Malabanan, M. M.; Amyes, T. L.; Richard, J. P. Mechanism for Activation of Triosephosphate Isomerase by Phosphite Dianion: The Role of a Ligand-Driven Conformational Change. *J. Am. Chem. Soc.* **2011**, *133*, 16428−16431.

(23) Reyes, A. C.; Zhai, X.; Morgan, K. T.; Reinhardt, C. J.; Amyes, T. L.; Richard, J. P. The Activating Oxydianion Binding Domain for Enzyme-Catalyzed Proton Transfer, Hydride Transfer and Decarboxylation: Specificity and Enzyme Architecture. *J. Am. Chem. Soc.* **2015**, *137*, 1372−1382.

(24) Goldman, A. D.; Beatty, J. T.; Landweber, L. F. The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *J. Mol. Evol.* **2016**, *82*, 17−26.

(25) Amyes, T. L.; Malabanan, M. M.; Zhai, X.; Reyes, A. C.; Richard, J. P. Enzyme Activation through the Utilization of Intrinsic Dianion Binding Energy. *Protein Eng., Des. Sel.* **2017**, *30*, 159−168.

(26) Kulkarni, Y. S.; Liao, Q.; Petrović, D.; Krüger, D. M.; Strodel, B.; Amyes, T. L.; Richard, J. P.; Kamerlin, S. C. L. Enzyme Architecture: Modeling the Operation of a Hydrophobic Clamp in Catalysis by Triosephosphate Isomerase. *J. Am. Chem. Soc.* **2017**, *139*, 10514−10525.

(27) Kulkarni, Y. S.; Liao, Q.; Byléhn, F.; Amyes, T. L.; Richard, J. P.; Kamerlin, S. C. L. Role of Ligand-Driven Conformational Changes in Enzyme Catalysis: Modeling the Reactivity of the Catalytic Cage of Triosephosphate Isomerase. *J. Am. Chem. Soc.* **2018**, *140*, 3854−3857.

(28) Walden, H.; Bell, G. S.; Russell, R. J. M.; Siebers, B.; Hensel, R.; Taylor, G. L. Tiny TIM: A Small, Tetrameric, Hyperthermostable Triosephosphate Isomerase. *J. Mol. Biol.* **2001**, *306*, 745−757.

(29) Walden, H.; Taylor, G. L.; Lorentzen, E.; Pohl, E.; Lilie, H.; Schramm, A.; Knura, T.; Stubbe, K.; Tjaden, B.; Hensel, R. Structure

and Function of a Regulated Archaeal Triosephosphate Isomerase Adapted to High Temperature. *J. Mol. Biol.* **2004**, *342*, 861−875.

(30) Katebi, A. R.; Jernigan, R. L. The Critical Role of the Loops of Triosephosphate Isomerase for Its Oligomerization, Dynamics, and Functionality. *Protein Sci.* **2014**, *23*, 213−228.

(31) Brändén, C.-I. The TIM Barrel—The Most Frequently Occurring Folding Motif in Proteins. *Curr. Opin. Struct. Biol.* **1991**, *1*, 978−983.

(32) Sterner, R.; Höcker, B. Catalytic Versatility, Stability and Evolution of the $(\beta\alpha)_8$-Barrel Enzyme Fold. *Chem. Rev.* **2005**, *105*, 4038−4055.

(33) Åqvist, J. Cold Adaptation of Triosephosphate Isomerase. *Biochemistry* **2017**, *56*, 4169−4176.

(34) van der Kamp, M. W.; Prentice, E. J.; Kraakman, K. L.; Connolly, M.; Mulholland, A. J.; Arcus, V. L. Dynamical Origins of Heat Capacity Changes in Enzyme-Catalyzed Reactions. *Nat. Commun.* **2018**, *9*, 1177.

(35) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(36) Xiang, J.; Sun, J.; Sampson, N. S. The Importance of Hinge Sequence for Loop Function and Catalytic Activity in the Reaction Catalyzed by Triosephosphate Isomerase. *J. Mol. Biol.* **2001**, *307*, 1103−1112.

(37) Kempf, J. G.; Jung, J.-y.; Ragain, C.; Sampson, N. S.; Loria, J. P. Dynamic Requirements for a Functional Protein Hinge. *J. Mol. Biol.* **2007**, *368*, 131−149.

(38) Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D. S.; Milo, R. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **2011**, *50*, 4402−4410.

(39) Cansu, S.; Doruker, P. Dimerization Affects Collective Dynamics of Triosephosphate Isomerase. *Biochemistry* **2008**, *47*, 1358−1368.

(40) Kurkcuoglu, Z.; Bakan, A.; Kocaman, D.; Bahar, I.; Doruker, P. Coupling between Catalytic Loop Motions and Enzyme Global Dynamics. *PLoS Comput. Biol.* **2012**, *8*, e1002705.

(41) Lolis, E.; Alber, T.; Davenport, R. C.; Rose, D.; Hartman, F. C.; Petsko, G. A. Structure of Yeast Triosephosphate Isomerase at 1.9 Å Resolution. *Biochemistry* **1990**, *29*, 6609−6618.

(42) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498−2504.

(43) Doncheva, N. T.; Klein, K.; Domingues, F. S.; Albrecht, M. Analyzing and Visualizing Residue Networks of Protein Structures. *Trends Biochem. Sci.* **2011**, *36*, 179−182.

(44) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(45) Shehu, A.; Kavraki, L. E. Modeling Structures and Motions of Loops in Protein Molecules. *Entropy* **2012**, *14*, 252−290.

(46) Sultan, M. M.; Kiss, G.; Pande, V. S. Towards Simple Kinetic Models of Functional Dynamics for a Kinase Subfamily. *Nat. Chem.* **2018**, *10*, 903−909.

(47) Richard, J. P.; Zhai, X.; Malabanan, M. M. Reflections on the Catalytic Power of a TIM-Barrel. *Bioorg. Chem.* **2014**, *57*, 206−212.

(48) Vijayabaskar, M. S.; Vishveshwara, S. Insights into the Fold Organization of TIM Barrel from Interaction Energy Based Structure Networks. *PLoS Comput. Biol.* **2012**, *8*, e1002505.

(49) Richard, J. P.; Amyes, T. L.; Reyes, A. C. Orotidine 5′-Monophosphate Decarboxylase: Probing the Limits of the Possible for Enzyme Catalysis. *Acc. Chem. Res.* **2018**, *51*, 960−969.

(50) He, R.; Reyes, A. C.; Amyes, T. L.; Richard, J. P. Enzyme Architecture: The Role of a Flexible Loop in Activation of Glycerol-3-phosphate Dehydrogenase for Catalysis of Hydride Transfer. *Biochemistry* **2018**, *57*, 3227−3236.

(51) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144.

(52) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386−2396.

(53) Wang, W.; Cao, S.; Zhu, L.; Huang, X. Constructing Markov State Models to Elucidate the Functional Conformational Changes of Complex Biomolecules. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1343.

(54) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(55) Amadei, A.; Linssen, A. B. M.; de Groot, B. L.; van Aalten, D. M. F.; Berendsen, H. J. C. An Efficient Method for Sampling the Essential Subspace of Proteins. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615−625.

(56) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193−226.

(57) Bussi, G. Hamiltonian Replica Exchange in GROMACS: A Flexible Implementation. *Mol. Phys.* **2014**, *112*, 379−384.

(58) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(59) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.

(60) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.

(61) Kamerlin, S. C. L.; Warshel, A. The Empirical Valence Bond Model: Theory and Applications. *WIREs Comput. Mol. Sci.* **2011**, *1*, 30−45.

(62) Davenport, R. C.; Bash, P. A.; Seaton, B. A.; Karplus, M.; Petsko, G. A.; Ringe, D. Structure of the Triosephosphate Isomerase-Phosphoglycolohydroxamate Complex: An Analog of the Intermediate on the Reaction Pathway. *Biochemistry* **1991**, *30*, 5821−5826.

(63) Amrein, B. A.; Steffen-Munsberg, F.; Szeler, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L. CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCrJ* **2017**, *4*, 50−64.

(64) 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.

(65) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.

(66) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(67) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.

(68) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(69) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalensko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *AMBER 2016*; University of California: San Francisco, CA, 2016.

(70) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.;

Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.

(71) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.

(72) Götz, 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. Theory Comput.* **2012**, *8*, 1542−1555.

(73) Salomon-Ferrer, R.; Götz, 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. Theory Comput.* **2013**, *9*, 3878−3888.

(74) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(75) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-Methylamide. *Biopolymers* **1992**, *32*, 523−535.

(76) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409−1419.

(77) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

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

(79) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(80) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(81) Berendsen, H. J. C.; Van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(82) van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(83) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(84) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.

(85) Best, R. B.; Hummer, G.; Eaton, W. A. Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.

(86) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(87) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(88) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation,

Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.

(89) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and Hidden Markov Models for Calculating Kinetics and Metastable States of Complex Molecules. *J. Chem. Phys.* **2013**, *139*, 184114.

(90) Weber, M.; Kube, S. In *Robust Perron Cluster Analysis for Various Applications in Computational Life Science*; R. Berthold, M., Glen, R. C., Diederichs, K., Kohlbacher, O., Fischer, I., Eds.; Springer: Berlin, Heidelberg, 2005; pp 57−66

(91) Warshel, A.; King, G. Polarization Constraints in Molecular Dynamics Simulation of Aqueous Solutions: The Surface Constraint All Atom Solvent (SCAAS) Model. *Chem. Phys. Lett.* **1985**, *121*, 124−129.

(92) King, G.; Warshel, A. A Surface Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions. *J. Chem. Phys.* **1989**, *91*, 3647.

(93) Sham, Y. Y.; Warshel, A. The Surface Constrained All Atom Model Provides Size Independent Results in Calculations of Hydration Free Energies. *J. Chem. Phys.* **1998**, *109*, 7940.

(94) Barrozo, A.; Duarte, F.; Bauer, P.; Carvalho, A. T. P.; Kamerlin, S. C. L. Cooperative Electrostatic Interactions Drive Functional Evolution in the Alkaline Phosphatase Superfamily. *J. Am. Chem. Soc.* **2015**, *137*, 9061−9076.

(95) Amrein, B. A.; Bauer, P.; Duarte, F.; Janfalk Carlsson, Å.; Naworyta, A.; Mowbray, S. L.; Widersten, M.; Kamerlin, S. C. L. Expanding the Catalytic Triad in Epoxide Hydrolases and Related Enzymes. *ACS Catal.* **2015**, *5*, 5702−5713.

(96) Blaha-Nelson, D.; Krüger, D.; Szeler, K.; Ben-David, M.; Kamerlin, S. C. L. Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes. *J. Am. Chem. Soc.* **2017**, *139*, 1155−1167.

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

(98) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, L. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116*, 6391−6423.

(99) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(100) 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.* **1996**, *118*, 11225−11236.

(101) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(102) Derreumaux, P.; Schlick, T. The Loop Opening/Closing Motion of the Enzyme Triosephosphate Isomerase. *Biophys. J.* **1998**, *74*, 72−81.

(103) Pan, A. C.; Weinreich, T. M.; Piana, S.; Shaw, D. E. Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J. Chem. Theory Comput.* **2016**, *12*, 1360−1367.

(104) Richard, J. P.; Amyes, T. L.; Goryanova, B.; Zhai, X. Enzyme Architecture: On the Importance of Being in a Protein Cage. *Curr. Opin. Chem. Biol.* **2014**, *21*, 1−10.

(105) Zhai, X.; Reinhardt, C. J.; Malabanan, M. M.; Amyes, T. L.; Richard, J. P. Enzyme Architecture: Amino Acid Side-Chains That Function To Optimize the Basicity of the Active Site Glutamate of Triosephopshate Isomerase. *J. Am. Chem. Soc.* **2018**, *140*, 8277−8286.

(106) Richard, J. P.; Amyes, T. L.; Malabanan, M. M.; Zhai, X.; Kim, K. J.; Reinhardt, C. J.; Wierenga, R. K.; Drake, E. J.; Gulick, A. M. Structure-Function Studies of Hydrophobic Residues that Clamp a Basic Glutamate Side Chain During Catalysis by Triosephosphate Isomerase. *Biochemistry* **2016**, *55*, 3036−3047.

(107) Richard, J. P. Acid-Base Catalysis of the Eliminaton and Isomerization Reactions of Triose Phosphates. *J. Am. Chem. Soc.* **1984**, *106*, 4926−4936.

(108) Pompliano, D. L.; Peyman, A.; Knowles, J. R. Stabilization of a Reaction Intermediate as a Catalytic Device: Definition of the Functional Role of the Flexible Loop in Triosephosphate Isomerase. *Biochemistry* **1990**, *29*, 3186−3194.

(109) Malabanan, M. M.; Nitsch-Velasquez, L.; Amyes, T. L.; Richard, J. P. Magnitude and Origin of the Enhanced Basicity of the Catalytic Glutamate of Triosephosphate Isomerase. *J. Am. Chem. Soc.* **2013**, *135*, 5978−5981.

(110) Noble, M. E. M.; Verlinde, C. L. M. J.; Groendijk, H.; Kalk, K. H.; Wierenga, R. K.; Hol, W. G. J. Crystallographic and Molecular Modeling Studies on Trypanosomal Triosephosphate Isomerase: A Critical Assessment of the Predicted and Observed Structures of the Complex with 2-phosphoglycerate. *J. Med. Chem.* **1991**, *34*, 2709−2718.

(111) van der Kamp, M.; Chaudret, R.; Mulholland, A. J. QM/MM Modelling of Ketosteroid Isomerase Reactivity Indicates that Active Site Closure is Integral to Catalysis. *FEBS J.* **2013**, *280*, 3120−3131.

(112) Tokuriki, N.; Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **2009**, *324*, 203−207.

(113) Pabis, A.; Risso, V. A.; Sanchez-Ruiz, J. M.; Kamerlin, S. C. L. Cooperativity and Flexibility in Enzyme Evolution. *Curr. Opin. Struct. Biol.* **2018**, *48*, 83−92.

(114) Petrović, D.; Risso, V. A.; Kamerlin, S. C. L.; Sanchez-Ruiz, J. M. Conformational Dynamics and Enzyme Evolution. *J. R. Soc., Interface* **2018**, *15*, 20180330.

(115) Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of Conformational Dynamics in the Evolution of Novel Enzyme Function. *Chem. Commun.* **2018**, *54*, 6622−6634.

(116) Gu, Y.; Li, D.-W.; Brüschweiler, R. Decoding the Mobility and Time Scales of Protein Loops. *J. Chem. Theory Comput.* **2015**, *11*, 1308−1314.

(117) Ben-David, M.; Elias, M.; Filippi, J. J.; Duñach, E.; Silman, I.; Sussman, J. L.; Tawfik, D. S. Catalytic Versatility and Backups in Enzyme Active Sites: The Case of Serum Paraoxonase 1. *J. Mol. Biol.* **2012**, *418*, 181−196.

(118) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W., Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *WIREs Comput. Mol. Sci.* **2018**, e1393.