THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 1883−1895

pubs.acs.org/JPCB

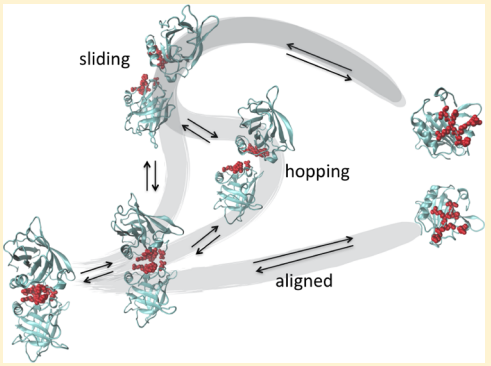
Unbiased Atomistic Insight into the Mechanisms and Solvent Role for Globular Protein Dimer Dissociation

Z. F. Brotzakis[†,‡,†] and P. G. Bolhuis*

Van't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam, Science Park 904, 1090 GD Amsterdam, The Netherlands

Supporting Information

ABSTRACT: Association and dissociation of proteins are fundamental processes in nature. Although simple to understand conceptually, the details of the underlying mechanisms and role of the solvent are poorly understood. Here, we investigate the dissociation of the hydrophilic β-lactoglobulin dimer by employing transition path sampling. Analysis of the sampled path ensembles reveals a variety of mechanisms: (1) a direct aligned dissociation (2) a hopping and rebinding transition followed by unbinding, and (3) a sliding transition before unbinding. Reaction coordinate and transition-state analysis predicts that, besides native contact and neighboring salt-bridge interactions, solvent degrees of freedom play an important role in the dissociation process. Bridging waters, hydrogen-bonded to both proteins, support contacts in the native state and nearby lying transition-state regions, whereas they exhibit faster dynamics in further lying transition-state regions, rendering the proteins more mobile and assisting in rebinding. Analysis of the structure and dynamics of the solvent molecules reveals that the dry native interface induces enhanced populations of both disordered hydration water near hydrophilic residues and tetrahedrally ordered hydration water nearby hydrophobic residues. Although not exhaustive, our sampling of rare unbiased reactive molecular dynamics trajectories enhances the understanding of protein dissociation via complex pathways including (multiple) rebinding events.



INTRODUCTION

Protein association and dissociation is essential for biologically relevant processes, such as cell signaling, DNA replication/transcription, cellular transport, immune response, gene editing,[1] as well as for protein aggregation and self-assembly into structures with desired properties, e.g., in food and colloids.[2,3] Moreover, knowledge of the kinetics and mechanisms of association is crucial for understanding and controlling biochemical network and cascade reactions of a processive or distributive nature.[4,5] Yet, this kinetics is poorly understood even on the dimer level and varies with the nature of the proteins.[6] Whereas hydrophobic association/dissociation occurs through the dewetting effect,[7,8] association of hydrophilic proteins involves wet dimer native interfaces[9,10] and has received significantly less attention, even though 70% of the protein−protein interfacial residues are hydrophilic.[11] The widely studied hydrophilic dimers Barnase−Barstar or acetylcholinesterase−fasciculin associate through a diffusion-limited reaction wherein the slow step is finding the transient encounter complex, a process accelerated by water-assisted electrostatic steering between the charged hydrophilic interfaces.[12−22] However, little is known about hydrophilic dimers whose association is slower and not in the electrostatic steering regime, including many small proteins.[23] Several theoretical models stress the importance of water-mediated interactions assisting binding in the absence of steering. Ben-Naim[24] proposed that hydrogen-bond bridging water interactions between the two proteins are maximized toward the native dimer. Northrup et al.[23] state that water assists binding through stabilization of a diffusion encounter complex, which increases the rebinding probability. Association via rebinding increases for stronger isotropic (dispersion) interactions, which smooth the rugged energy landscape due to anisotropic (charged) interactions.[25,26] In fact, water could also play such a smoothing role, e.g., by screening anisotropic salt bridges at the interface.

Various experimental studies on globular dimeric proteins, including NMR, PRE, and sedimentation experiments,[10,14,27−29] yielded (indirect) information on stable states, potential transition states, and intermediates. However, atomistic details on association/dissociation pathways allowing direct insight into the mechanisms of dimer formation are still lacking. Although molecular dynamics (MD) simulations can, in principle, provide such a detailed insight, straightforward MD using all atom force fields is impractical as timescales of dissociation and association are on the order of milliseconds to seconds. Only recently, a handful of MD studies addressed hydrophilic dimers (mostly on Barnase−Barstar).[12,17,18,30] Using Markov State Modeling (MSM) techniques, Plattner et al.[17] were able to assess the formation of the Barnase−

Received: October 13, 2018
Revised: January 30, 2019
Published: February 4, 2019

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

Barstar dimer. However, although very powerful, MSM techniques do not have direct access to the full dissociation transition due to the high barriers involved, which in turn induce long dwell times in the bound states. In contrast, the Transition Path Sampling (TPS) methodology bypasses the long dwell times in the stable unbound and bound states by focusing on the reactive association/dissociation trajectories directly.[31] TPS harvests a collection of unbiased molecular dynamics trajectories connecting two predefined stable states. The resulting path ensemble contains all pertinent dynamical and mechanistic information.

Here, we apply TPS to the rare dissociation/association transition ($k_{off} \leq 0.1\ s^{-1}$) of the widely experimentally studied β-lactoglobulin (β-lac) globular protein dimer. We obtain atomistic insight into the mechanism of this transition by analyzing the dynamically unbiased rare pathways and by extracting the best low-dimensional models for the pertinent reaction coordinates (RC). This analysis includes the role of the solvent during the dissociation/association process itself, something that is only possible because we have access to the reactive transition paths. In addition, our study provides general insight into the dissociation mechanism for proteins that do not bind through steering interactions.[6,32]

We find that dissociation of the β-lac dimer from its native state occurs along a variety of multistep routes with transient intermediates, namely, a direct aligned unbinding route and an indirect dissociation route through sliding or hopping to misaligned configurations before unbinding. The first dissociation bottleneck or transition-state ensemble (TSE) appears in all paths and is associated with the breaking of native contacts and the formation of hydrogen bond bridging water-mediated interactions. There is a secondary bottleneck related to the solvation of the persisting salt bridge R40−D33. In the aligned mechanism, the number of waters at the interface and the distance r_{40-33} are pertinent ingredients of the RC. In the sliding mechanism, the secondary bottleneck involves also a relative rotation of the proteins, which induces a loss of native contacts while still forming the R40−D33 directional salt bridge. Here, the distance r_{40-33} and the rotation angle ϕ are important ingredients of the RC. Finally, the hopping mechanism involves a rebinding before dissociating, with the protein−protein distance as the relevant RC. As the TPS pathways are microscopically reversible, we can interpret the transition paths both in the forward (dissociation) as well as in the backward (association) direction. However, we stress that these paths do not represent the complete, full association/dissociation process but only those that occur within a maximum allowed limited time, as specified by the path ensemble.

Indeed, the presence of multiple sequential barriers causes paths to become long, hampering the transition path sampling. To avoid the additional dwell time between the first barrier, the breaking of the contacts, and the secondary solvation barrier, we also sampled the transition path ensemble for this secondary barrier using a more relaxed nonspecifically bound state definition (abbreviated B) using the desolvated (dry) contact area. The trajectories in this path ensemble are significantly shorter but also exhibit mechanisms involving direct dissociation, a sliding mechanism where first the dry area decreases by a sliding movement, before dissociation, and a hopping dissociation route where the proteins first rebind before completely dissociating.

Since we find that solvent is an ingredient of the RC, e.g., the number of interfacial waters or hydrogen-bond bridging water occurring in the transition-state ensemble, we further investigate the structure and reorientation dynamics of water during the dissociation process. Here, we find that water at the native dimer interface comprises of a disordered slow population due to formation of long-lived hydrogen-bonded bridging water and a tetrahedral population, reorienting slower than bulk, which is typical of hydrophobic solvation and thus characteristic of β-lac's mixed polar−apolar interface. This finding also extends to water populations in the transition state, which contain more mobile hydrogen-bond bridging waters, enabling enhanced rotational mobility of the protein dimer with respect to a completely dry contact surface.

The remainder of the paper is organized as follows. In the following section, we review the MD simulation settings and TPS algorithms as well as the analysis methods and tools. Next, we present and discuss the results on the specific unbinding transition and nonspecific binding transition. This is followed by an analysis of the hydration structure and dynamics. We end with concluding remarks.

METHODS

Molecular Dynamics. In this study, we performed molecular dynamics (MD) simulations using Gromacs 4.6.7 with GPUs.[33] All potential energy interactions were defined using amber99sb-ildn and TIP3P force fields.[34,35] We obtained the β-lactoglobulin (β-lac) PDB structure from the Protein Data Bank (PDB:2AKQ) and placed it in a dodecahedral simulation box, which was energy-minimized in vacuum using the conjugate gradient method. After solvation of the box with 20 787 water molecules and a second energy minimization, we performed a 10 ps NPT short equilibration of water under ambient conditions with the protein position restrained. The solvated system was equilibrated for 1 ns in ambient conditions in the NPT ensemble and thereafter was subjected to a long 200 ns NPT simulation. All bonds were constrained using the Lincs algorithm. We used a 1 nm cutoff for nonbonded van der Waals interactions. The electrostatic interactions were treated by the particle mesh Ewald algorithm, with a Fourier spacing of 0.12 nm and a 1 nm cutoff for the short-range electrostatic interactions. The updating frequency of the neighbor list was 10 fs, with a cutoff of 1 nm, and the time step was 2 fs.[34] In the NPT simulations, the temperature was controlled using the velocity-rescaling thermostat[36] with a coupling time constant of 0.2 ps. The pressure was controlled using the Parrinello−Rahman barostat[37] with a coupling time constant of 1.0 ps.

Short NVE MD simulations were performed to characterize the water structure and dynamics. No position restraints were imposed, and to prevent energy drift, we used a switching function for the nonbonded interactions from 0.8 to 1.0 nm. The pair lists were updated every 5 fs with a cutoff of 1.2 nm, and the time step was 1 fs. The frequency of the energy calculation was 10 fs.

TPS Spring Shooting Algorithm. Transition path sampling[38,39] harvests an ensemble of rare trajectories that lead over a high free-energy barrier, connecting two predefined stable states. Starting from an initial reactive path, TPS performs a random walk in trajectory space by selecting a time frame, changing the momenta slightly, and shooting off a new trial trajectory forward and backward in time by integrating the equations of motion. Acceptance or rejection of the trial trajectory is done according to the Metropolis rule,[38,39] in

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

which for the standard two-way shooting move under a fixed path length, the trial move is accepted if the trial path connects the two stable states. If not, the trial path is rejected.

The more efficient one-way flexible shooting algorithm[39,40] samples the minimal length pathways between stable states and has been previously used in other protein systems.[41,42] The one-way shooting method has several drawbacks. First, it requires more shots to decorrelate paths (although not more computer time). Second, it suffers in efficiency for asymmetric barriers, which occur, for instance, when the system on one side of the main barrier is trapped in an intermediate state, whereas it can easily reach the stable state on the other side. This means the paths on the trapped side become much longer. When uniform one-way shooting is used, this asymmetry leads to many more shooting attempts on one side of the barrier with respect to the other, increasing the inefficiency.

The spring shooting algorithm is especially developed for use with the one-way algorithm.[43] It only differs in the way the shooting point is selected. Instead of uniform random selection, the spring shooting shifts the shooting point index with respect to the last successful shooting point, not in a symmetric but in an asymmetric way according to an acceptance criterion

$$P_{\mathrm{acc}}^{\mathrm{sp}}[\tau \to \tau'] = \min\left[1, \frac{\exp(sk\tau')}{\exp(sk\tau)}\right] = \min[1, e^{sk\Delta\tau}] \quad (1)$$

where $\Delta\tau = \tau' - \tau$ is the number of shifted frames from the previous shooting point τ, k denotes a force constant determining the magnitude of the bias, and $s \in \{-1, 1\}$ is determined by the direction of shooting; i.e., $s = -1$ for forward shooting, and $s = 1$ for backward shooting. The spring shooting algorithm thus treats the forward and backward shooting move as different types of moves. As a large $\Delta\tau$ either yields an exponentially small acceptance ratio or is likely to produce a failed shot, in practice, we limit the choice of $\Delta\tau$ between the interval $[-\Delta\tau_{\max}, \Delta\tau_{\max}]$, analogous to the maximum allowed displacement in a regular MC translational move. When the trial shooting point falls outside the current path, the acceptance probability becomes zero and the move is rejected. The remainder of the shooting move is identical to that in the uniform one-way shooting algorithm. For a detailed description of the algorithm, see the SI and ref 43.

The advantage of this approach is that unfavorable shooting points are discarded without extra cost. Pathways are decorrelated as much as possible, without wasting time creating partial paths that do not contribute to the decorrelation. Note also that the algorithm rejects trial paths that become longer than $L_{\max}$, which is set to prevent memory or storage problems or as an indication that the path generation went awry, e.g., became trapped in a long-lived intermediate state.

Defining the Stable States and Creating the Initial Path. To define the stable states, we performed a 200 ns MD simulation in the NPT ensemble at ambient conditions, during which the β-lac dimer remained in its native bound state (see Figure S1). The native contacts were identified as those residue pairs that stayed within a minimum heavy atom distance of 0.4 nm for at least 90% in the 200 ns NPT trajectory. As shown in Table S1, only eight residue pairs are shown to fulfill this criterion (150−146, 148−148, 146−150, 148−147, 147−148, 149−146, 146−149, and 33−33). These contacts are between residues of the β sheets of the I-strand, and the AB-loops of the protein and have been also verified to be important for the stability of the dimer by experiments.[29] These eight residue pairs, as well as four native hydrogen bonds (between backbone NH and CO of residues 146−150, 148−148, and 150−146) define the stable native contact state (N). Note that residue pairs 33−40 and 40−33 discussed in the Results and Discussion section have smaller occupancy than 70% and hence are not a part of the native state. The unbound state (U) requires the minimum distance between the two proteins to be greater than 1 nm ($r_{\min} > 1$ nm). Finally, the nonspecific dissociation B ⇆ U transition requires the definition of a bound state B. To be as nonspecific as possible, we only required for this bound state that the protein−protein interfacial area is >2 (nm^2). All definitions are summarized in Table 1.

To obtain the initial path for TPS, we enforced the dissociation from the native bound state using metadynamics at 300 K employing the PLUMED package[44] with the above-mentioned MD settings. As the resulting trajectory is strongly biased, we launched unbiased MD trajectories from particular frames, performed at a slightly elevated temperature of 330 K to avoid getting trapped in long-lived near-native intermediate states. Concatenating a trajectory returning to the native state with one going forward to the unbound state yields the desired unbiased initial TPS path. For completeness, we checked that performing the simulations at 330 K does not perturb the protein conformations significantly. See the SI for more details.

TPS Simulation Settings. We performed TPS simulations of the native state to the unbound state (N ⇆ U) at $T = 330$ K and $P = 1$ atm in the NPT ensemble using home-written scripts encoding the spring shooting scheme. The maximum path length for this transition was set to $L_{\max} = 7000$, which with frames saved every 10 ps, translates into a maximum path duration of $t_{\max} = 70$ ns. The spring shooting move parameters were set to $k = 5$ and $\Delta\tau_{\max} = 200$. The spring constant ensures the shooting points remain close to the top of the very asymmetric barrier of β-lac dissociation. The $\Delta\tau_{\max}$ was chosen small compared with the maximum path length allowed (3%) so that the shooting point rejection as well as the whole trial move rejection was kept to a minimum. For the B ⇆ U transition, the spring constant k is 0.1 and the $\Delta\tau_{\max}$ is 70 frames.

Analysis of the Path Ensemble. Home-written scripts analyzed the path sampling results to produce the path tree, the least changed path (LCP), the path length distribution, and the path density. We construct the path density by choosing two order parameters (e.g., protein−protein minimum distance vs native patch vector angle) and binning each frame of each trajectory in the path ensemble to a 2D grid. Every path can only contribute to a specific bin once, even if visited multiple

Table 1. State Definitions for the Native State (N), the Unbound State (U), and Any Bound State (B) as a Function of Native Contacts (NC), Native Hydrogen Bonds (N H Bonds), Protein−Protein Minimum Distance ($r_{\min}$), and Dry Contact Area

	NC	N H bonds	$r_{\min}$	dry area
native (N)	≥8	≥4		
unbound (U)	0	0	>1 nm	0
bound (B)				2 nm^2

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

times. Note that accepted paths can occur multiple times in the ensemble, depending on whether the next trial move has been rejected. The least changed path, consisting of the stretches between successive alternating forward/backward shooting points, acts as an approximation for the transition-state ensemble.[43] For the path length distribution, each accepted path of a different length L is histogrammed according to its weight in the path ensemble.

Since the protein orientation degrees of freedom might be important during the dissociation transition/association, we calculate the relative orientation of the two proteins, characterized by an angle ϕ (see Figure S7 for a graphical illustration)

Water plays an important role in dissociation/association. To address the solvent degrees of freedom, for each configuration, we count the number of waters residing inside a cylindrical tube between the two proteins. The tube's base centers are defined by the center of mass of each protein, a radius r = 1.4 nm or r = 1.1 nm and length l being the center-of-mass distance between the two proteins.

Reaction Coordinate Analysis by Likelihood Maximization. The reaction coordinate is an invaluable description of a complex transition as it can predict the progress of the reaction. "Transition Path Theory (TPT) states that the perfect reaction coordinate is the commitment probability (committor or p-fold) $p_B(x)$ as it gives the probability for a configuration x to reach the final state B.[45] Although the committor $p_B(x)$ is mathematically well defined, it is excessively expensive to compute, and being a highly dimensional function, not very insightful as to which are the relevant slow degrees of freedom. Peters and Trout developed a likelihood maximization (LM) method to extract the best (linear) model for the reaction coordinate on the basis of an approximation of the committor function using the shooting point data from TPS.[46,47] Each trial shooting point in this data set can be regarded as drawn from the committor distribution.[46,47] Using as input the N forward (or backward) shooting point configurations x_{sp} of the accepted trajectories ending in the final state B ($x_{sp} \rightarrow$ B) and the shooting points of the rejected trajectories ending in state A ($x_{sp} \rightarrow$ A), the method defines the likelihood that a model reaction coordinate r can reproduce the observed data

$$L = \prod_{x_{sp}\rightarrow B} p_B(r(x_{sp})) \prod_{x_{sp}\rightarrow A} (1-p_B(r(x_{sp}))) \quad (2)$$

where the committor $p_B(r)$ is a function of the reaction coordinate r. The committor function is modeled/parametrized as

$$p_B(r(x_{sp})) = \frac{1}{2} + \frac{1}{2}\tanh(r[q(x)]) \quad (3)$$

where the reaction coordinate $r(q(x))$ is approximated by a linear combination of m collective variables $q(x)$ as follows

$$r(q(x)) = a_0 + \sum_i^m a_i q_i(x) \quad (4)$$

The LM analysis serves as a screening tool for linear combinations of candidate CVs and returns the one that best parametrizes the committor probability, given a good data set obtained from the TPE. Adding an additional CV in the analysis, i.e., increasing m by one, should lead at least to an increase of $\delta L_{min} = \frac{1}{2}\ln N$ in the maximum likelihood, for the RC to be deemed a significant improvement.[46,47] The spring shooting algorithm is naturally suited for use with this LM approach, since it gives access to shooting points close to the transition-state ensemble. We use the candidate CVs listed in Table S2.

Water Structure and Dynamics Analysis. To obtain insight into the role of the solvent, we analyzed the structural and orientation dynamics of water in the dissociation transition, with the same protocol as that in refs 48, 49.

As the TPS simulations are done at 330 K, we first analyze the water structure and dynamics at this temperature as follows. From a decorrelated reactive path in the TPS ensemble, we selected three different frames belonging to the native, the transition state TSR_2 (native contacts = 2, $\phi = 60°$), and the unbound state, respectively. For each of them, a 1 ns NPT run at 330 K is performed with the protein position restrained, followed by 10 individual NVE simulations starting at different positions of the short 1 ns NPT simulations (see Molecular Dynamics section for details). Frames were saved every 100 fs to obtain sufficient data for analysis of the water dynamics. To identify whether water structure and dynamics changes significantly from 330 to 300 K, we repeated the same analysis at 300 K.

From the short NVE trajectories, we computed the reorientation decay time τ for each water molecule in the hydration layer of the proteins from an orientational correlation function (see the SI for details). The decay times of the individual water molecules allow us to establish the relation between the water structure and dynamics. Note that the decay times here are systematically shorter than those in refs 48, 49 due to the use of the TIP3P water model.

In addition, we characterize the tetrahedral structure of water around amino acids on the basis of the probability distribution $P(\theta)$ of the minimum water−water OOH angle θ (see Figure S8) for all water−water pairs within 3.5 Å from each other and solvating the amino acid.[50−52] The distribution $P(\theta)$ of these angles takes on a bimodal distribution with a minimum at 30°, distinguishing between tetrahedral water population (angles lower than 30°) and a perturbed H-bond network, mostly occurring around hydrophilic groups (angles higher than 30°). The tetrahedral structure parameter S is defined as the integral of $P(\theta)$ up to θ = 30°.[50−52] Water around hydrophobic groups has a larger S due to smaller H-bond angles θ, inducing stronger water−water bonds, with larger energy fluctuations and therefore a positive change in heat capacity of the solvating water. In contrast, the introduction of a hydrophilic group around water strains the water−water H-bond angle and shifts the angle distribution to higher values and hence a lower S, thus decreasing the water−water bond energy and fluctuations, which decreases the heat capacity of solvation.[50−52] Throughout the text, we will associate tetrahedral structured water with a larger S value (high tetrahedral water population) and unstructured water with a lower S value (low tetrahedral water population). Unstructured water coinciding with slow reorientation dynamics (as characterized by τ > 4 ps) will be labeled as disordered water.

We computed the distribution $P(\theta)$ and extracted a structural order parameter S for each molecule hydrating an amino acid separately. We bin the τ−S pair for each residue in a 2D histogram, to investigate a possible correlation between water tetrahedral structuring and reorientation dynamics (see the SI for more information).

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

Hydrogen-Bond Bridge Survival Correlation Function. The hydrogen-bond bridge correlation function (eq 5) is a correlation function that traces the decay time of a hydrogen-bond bridge between two intermolecular protein residues.

$$C_{\text{Bridging}}(\tau) = \frac{1}{C_{\text{Norm}}} \int \sum_{i,j}^{N_{\text{res}}} \sum_{N=1}^{N_{\text{wat}}} (\mathbb{1}_{\text{Brid}}^{N,i,j}(t) \cdot \mathbb{1}_{\text{Brid}}^{N,i,j}(t+\tau))\, \mathrm{d}t \tag{5}$$

where τ is time, C_{Norm} is a normalization constant, such that $C_{\text{Bridging}}(0) = 1$, i and j are running over the residue number of proteins A and B, respectively, N_{res} is the total number of residues per protein and N_{wat} is the total amount of waters in the simulation box. $\mathbb{1}_{\text{Brid}}^{N,i,j}(t)$ is an indicator function at time t, which is unity if water N is hydrogen-bond bridging residues i and j and zero otherwise.

RESULTS AND DISCUSSION

TPS of the Specific Dissociation Transition. We performed several TPS runs at 330 K employing the spring shooting algorithm for the transition between the unbound state (U) and the native bound dimer state (N). In total, we performed 560 shooting trial moves, of which 18.3% was accepted with total aggregate and accepted path simulation times of 23.7 and 3.43 μs, respectively. Decorrelation was tested using path trees (see SI). The average path duration was 33.6 ns. The path length distribution in Figure 1 is broad and includes a significant population of longer paths. To shed light on this bimodal distribution, we analyze the transition path ensemble (TPE), first by inspection.

All trajectories in the TPE fall into three categories, representing three qualitatively distinct dissociation mechanisms: (1) In the aligned mechanism, proteins directly separate to the unbound state without (much) rotation. (2) In the hopping mechanism, proteins first dissociate and then rebind to a low dry surface nonspecific configuration before fully dissociating. (3) In the sliding mechanism, proteins first rotate and slide out of the native state to a higher dry interface nonspecifically bound transition state, before fully dissociating. Figure 2 illustrates these three mechanisms schematically. Switches between these different mechanisms frequently occur in the path sampling (see Table S3). We also observe trajectories exhibiting a convolution of the above mechanisms, in which the proteins slide and rotate away from the native state but instead of dissociating, return to the native state, followed by an aligned dissociation.

The observed probability per mechanism is 22%, 18%, and 60% for the aligned, hopping, and sliding mechanisms,

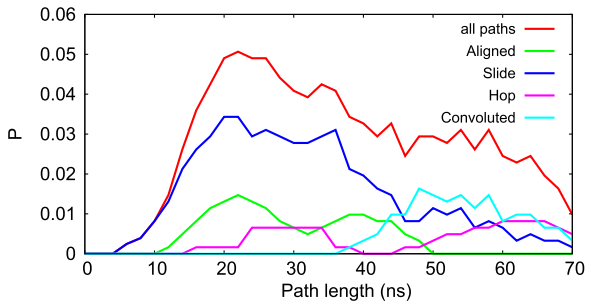


Figure 1. Path length distribution of the N ⇆ U path ensemble and for the different observed mechanisms.

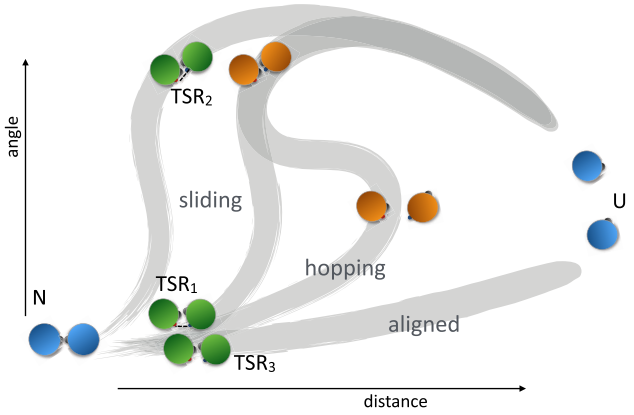


Figure 2. Cartoon network of transitions and respective TSRs during the full dissociation/association process. We identify three types of mechanisms: aligned, hopping and sliding transitions.

respectively. The sliding mechanism is most abundant in the TPE. The aligned mechanism is least prevalent, most likely due to the lower orientation entropy. The longest transition paths tend to be the sliding, hopping, and convoluted mechanisms, because of the presence of transiently formed intermediates. These findings are in agreement with the existing literature on sliding and rebinding mechanisms, wherein dimers exhibit early intermediates that effectively increase the binding rate.[17,19,25,26] However, the nature of interactions participating in such mechanisms has not been addressed in detail. To gain further insight, we define several collective variables (CVs) that are important for the dissociation mechanisms. Besides the number of native contacts NC and the minimum distance r_{min}, we define the angle of rotation ϕ where $\phi = 0°$ corresponds to the fully aligned dimer. Further, we identify specific contacts. Many trajectories in the path ensembles exhibit the transient but relatively long-lived (occupancy >10 ns and heavy atom distance < 0.4 nm) R40−D33 double salt bridge contact between the carbonyl groups of D33 of one protein and the amide groups of R40 of the other protein. Other long-lived intermolecular contacts are preserved in a large number of paths, although not throughout the entire TPE (see Table 2). Note the lack of symmetry due to imperfect sampling. We highlight a configuration containing the three most occurring contacts (R40−D33, H146−S150, and I29−S150) in the path ensembles in Figure 3. The presence of such long-lived polar, apolar, and charged interactions signifies that β-lac dissociation/association is not a one-step process and transient interactions occur during the pathway. Such transient interactions have been observed in various other protein dimers, e.g., the Barnase−Barstar dimer, the insulin dimer, and Ras-Raf-RBD, RNase-Hi-SSB-Ct, and TYK2-pseudokinase complexes.[17,19] In the case of β-lac, Sakurai et al.[29] showed

Table 2. Number of Paths in which Individual Intermolecular Contacts between Amino Acid of Protein A and Amino Acid of Protein B Occur with a Lifetime Higher than 10 ns, for Each Path Ensemble

#aa$_A$−#aa$_B$	TPE	#aa$_A$−#aa$_B$	TPE	#aa$_A$−#aa$_B$	TPE
H146−S150	77	D33−R40	47	I29−F151	32
R40−D33	71	S150−H146	44	S150−I29	28
I29−S150	67	I29−Q155	44	H146−N152	24
D28−N152	53	Q155−I29	35	K141−D130	22

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

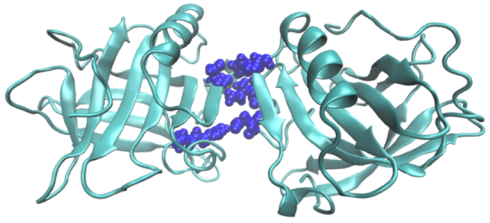


Figure 3. Structure of an on pathway transiently formed intermediate. In blue are highlighted the long-lived contacts H146−S150, R40−D33, and I29−S150.

that the R40−D33 interactions and contacts of H146, R148, and S150 are essential for the dimerization, since mutating any of these residues drastically reduces the binding affinity. Indeed, here, we also observe these contacts (see Tables 2 and S4). Furthermore, in a recent microsecond MD study, Bello et al.[53] noticed the importance of R40−D33 in the formation of a protein−protein interface. Such contacts have also been found in various crystal structures.[54] As interactions of the H146, R148, and S150 residues are included in the definition of the native contacts, we focus here on the salt bridge distance $r_{R40-D33}$. (Note that we could as well have focused on the symmetric salt-bridge D33−R40.) We will address the importance of the R40−D33 interaction in the TSE discussion below.

Since the role of water in hydrophilic association[10,23,24] is crucial, yet elusive, we define several solvent-based CVs: A_{dry}, the dry surface area of contact between the proteins; N_{HB}, the number of hydrogen-bonded bridging waters; and $N_{tube,}$, the number of waters in a tube of a certain radius r between the proteins' centers of mass. Figure 4 shows the path density as a function of various CVs for both the full transition path ensemble as well as for the least changed path ensemble (see Methods and the SI). The LCP approximately samples the transition barrier region, serving as a proxy of the transition-state ensemble.[43] The path densities of the full TPE show that all paths pass through configurations with partially formed native contacts ($1 < NC < 4$), but the LCP indicates that in fact, the TSE is split into several regions, which we denote as transition-state regions (TSR). TSR_1 occurs at $\phi < 40°$ and $1 < NC < 4$, whereas TSR_2 is located at $\phi > 50°$ and $0 < NC < 2$. At the same time, the TSRs are characterized by a substantial number of hydrogen-bonded bridging waters. Although this CV is not found to be among the most pertinent ingredients of the RC discussed later, the formation/breaking of hydrogen-

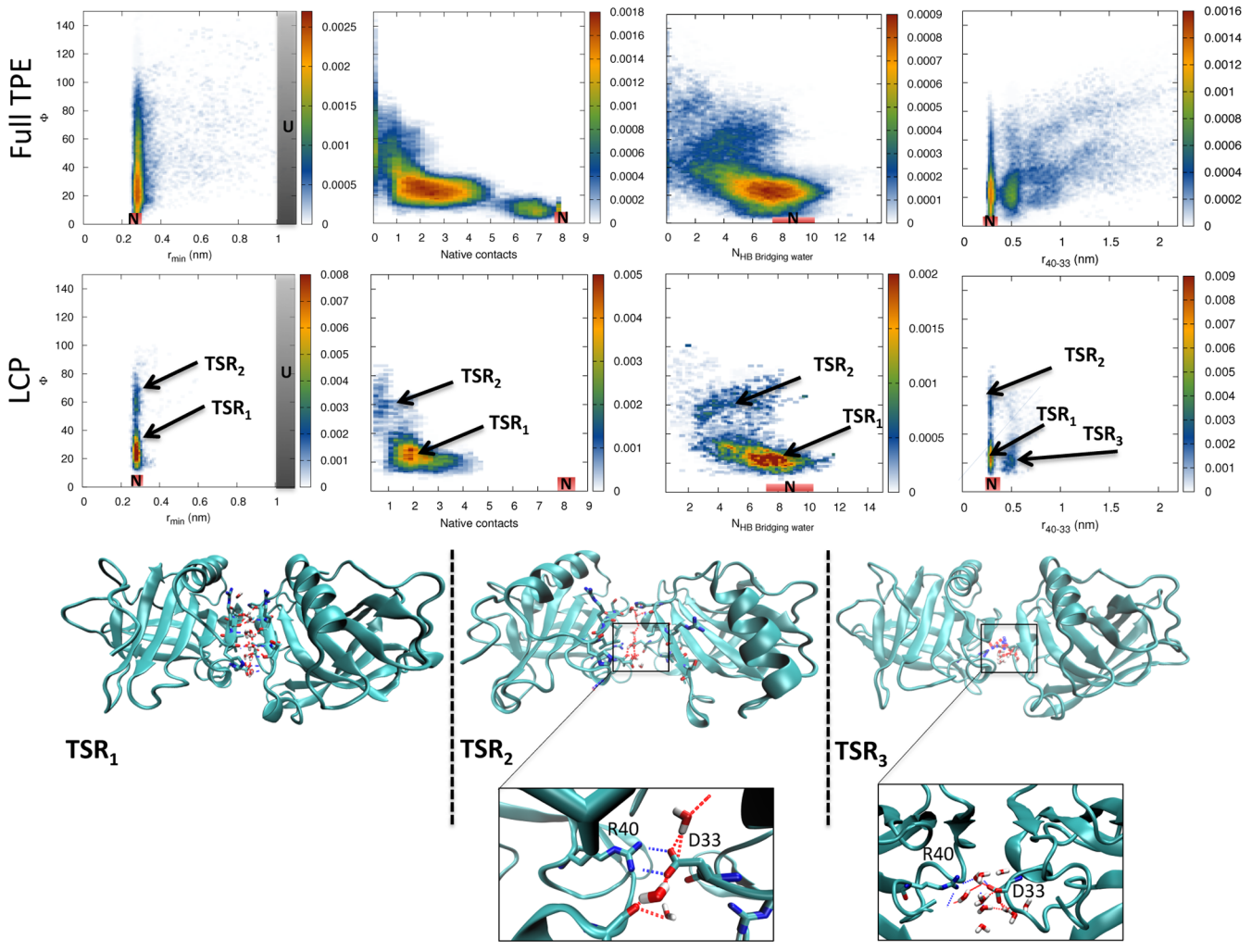


Figure 4. Path density plots of ϕ as a function of protein−protein minimum distance (column 1), native contacts (column 2), number of hydrogen-bond bridging waters (column 3), and R40−D33 distance (column 4) for the full transition path ensemble (full TPE) and least-changed path ensemble (LCP), respectively. The bottom row shows snapshots of configurations of the TSR_1, TSR_2, and TSR_3.

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

bond bridging waters between the proteins represents a dynamical bottleneck in the association/dissociation process. This observation is in agreement with the prediction of Ben-Naim and Northrup[23,24] that water-mediated interactions drive or characterize the hydrophilic association. The path density and LCP in the $\phi - r_{40-33}$ plane suggest the presence of a third TSR_3 characterized by a hydrogen-bonded water between residues R40 and D33. Several representative configurations from these TSRs are highlighted in Figure 4. The TSR_3 bottleneck suggests that water solvation of R40−D33 helps the proteins escape a very strong and directional salt bridge interaction during the dissociation process.

As the specific dissociation TPE is a superposition of the three possible mechanisms, separately plotting the TPE for each route in Figure S11 enables a discussion of each route individually. For each of these subensembles, we also performed a reaction coordinate analysis (see Methods and the SI). In the aligned mechanism, the native contacts break, while the angle ϕ stays below ($\phi < 50°$), followed by solvation of the dry contact surface and unbinding of the proteins. The paths pass through TSR_1 where the proteins have partially formed native contacts ($1 < NC < 5$) and through TSR_3, which involves the breaking of the salt bridge between R40 and D33, forming a water-mediated interaction. The dynamical bottleneck upon binding is thus the correct alignment, local rearrangement, and formation of the native contacts, as well as expelling the water at the interface, and, in particular, at the R40−D33 contact, and increasing the number of hydrogen-bond bridging waters. RC analysis shows that the solvation of the protein−protein interface (N_{tube11}) is the most important RC for the aligned mechanism (see Table 3 and Figure 5). However, the salt bridge distance $r_{R40-D33}$ is deemed a good additional reaction order parameter, only barely missing the threshold for significance, perhaps due to the presence of TSR_3 (see Table 3).

In the hopping mechanism, the dimer first breaks the native contacts to become dissociated with zero dry contact surface area but with several hydrogen-bond bridging waters between the protein surfaces. Then, the proteins rebind to form a nonspecifically bound configuration, characterized by transient nonspecific contacts, a large angle $\phi > 50°$, a small dry contact surface area of 1−2 nm^2, and some hydrogen-bond bridging waters. Finally, the dimer completely unbinds. The paths pass through TSR_1 and TSR_3 as proteins dissociate and rebind. The rebinding is a diffusive process and thus a matter of chance. We do not see a special role for the hydrogen bonding networks, although of course when they are there, they could induce a higher probability of rebinding by establishing stabilizing bonds between the protein interfaces. RC analysis for the hopping mechanism indicates that the protein−protein center-of-mass distance is the most pertinent collective variable (see Table 3).

Table 3. LM Analysis for the N ⇆ U Transition Based on the Forward Shooting Points for (a) the Aligned Paths, (b) the Sliding Paths, and (c) the Hopping Paths[a]

n	ln L	reaction coordinate
	Aligned Paths δL_{min} = 2.04 (59 Shooting Points)	
1	−33.24	$-5.24 + 0.09N_{tube11}$
2	−31.46	$-5.14 + 3.35r_{40-33} + 0.06N_{tube11}$
3	−31.32	$23.24 - 9.8\text{dist} + 2.95r_{40-33} + 0.14N_{tube11}$
	Sliding Paths δL_{min} = 2.54 (161 Shooting Points)	
1	−104.93	$-3.41 + 0.02N_{tube14}$
2	−102.28	$-1.59 + 1.67r_{40-33} + 0.02\phi$
3	−102.09	$-2.44 + 1.84r_{40-33}$ + 0.27 native contacts + 0.03ϕ
	Hop Paths δL_{min} = 1.74 (33 shooting points)	
1	−18.95	$-15.17 + 4.30\text{dist}$
2	−17.83	$-35.21 + 12.61\text{dist} - 0.05N_{tube14}$
3	−17.25	$-78.05 + 19.44\text{dist} + 1.98\text{native contacts} + 27.50r_{min}$

[a] δL_{min} denotes the minimum required increase in likelihood when adding an additional CV.

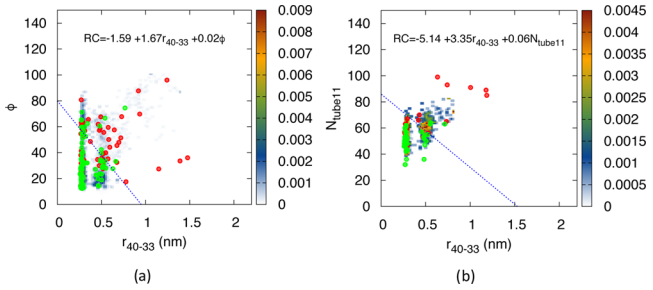


Figure 5. Plot of forward shooting points and the predicted dividing surface RC = 0 (blue dashed line) for (a) the sliding and (b) aligned paths. Red points end in U and green points in N. Note the split nature of the TSE.

In the sliding mechanism, proteins first break their native contacts by sliding and rotating to a misaligned configuration before unbinding. During sliding, a dry contact surface area is preserved (2−5 nm^2) with some hydrogen-bond bridging waters present. Paths can pass through either TSR_1, TSR_2, or TSR_3 with roughly equal probabilities. The R40−D33 salt bridge is still present in TSR_2. RC analysis indicates that the relevant CV describing the bottleneck is the salt bridge distance as well as the rotation angle ϕ. Thus, either water can assist breaking of directional interactions (R40−D33) present in TSR_1 before rotating or vice versa in TSR_2. Note that (a part of) the sliding trajectories can be viewed as hopping trajectories in which the proteins remain in contact until the final dissociation step.

The above findings can be compared with the existing literature on the binding affinity of β-lac. On one hand, they indicate the importance of the formation of the R40−D33 contact in the binding affinity.[29,53] On the other hand, the presence of hydrogen-bond bridging waters in the natively bound state and in the TSR1 bottleneck corroborates the findings of a combined MD isothermal titration calorimetry study, which attributed the observed negative change in heat capacity upon binding not only to the desolvation of the protein surface upon binding but also to the presence of dynamically slow water at the interface.[10]

TPS of the Nonspecific Dissociation Transition. In the previous section, we focused on the specific dissociation/association from and to the native dimer state. In this section, we focus on the nonspecific dissociation/association of this dimer by studying the B ⇆ U transition. Here, the nonspecific bound state B is defined in a much less strict sense by requiring a minimal dewetted dry contact surface area of 2 nm^2 (see Table 1). We performed a TPS run at 300 K employing the spring shooting algorithm for the nonspecific dissociation/association mechanism between the unbound state (U) and any bound state (B). In total, we performed 531 shooting trial moves, of which 37% was accepted, with total aggregate and accepted path simulation times of 854 and 398 ns, respectively.

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

The path length distribution in Figure 6 shows that the average path length is now 4.3 ns, an order of magnitude shorter than that in the N ⇆ U transition. A partial path tree is shown in Figure S12. Upon inspection, the path ensemble shows three distinct mechanisms that are very similar to the N ⇆ U case, namely, a direct dissociation mechanism, one that involves a hopping and rebinding mechanism, and the sliding mechanism, depicted in a cartoon representation in Figure 7. In the direct mechanism, proteins transit in around 1 ns between an unbound and bound state by quickly increasing the dry contact surface area. In the hopping mechanism, proteins solvate the dry contact surface area, (partially) unbinding and rebinding first before fully dissociating. The observed probabilities of these mechanisms are 18%, 71%, and 11% for the direct, hopping, and sliding mechanisms, respectively. The hopping paths are not only the most abundant, they tend to be the longest, because they undergo a two-step process. The reason the hopping transition is more abundant than the N ⇆ U transition is that the B ⇆ U transition has a much less strict bound state, which can be easily reached via hopping. In the sliding mechanism, bound proteins first slide to a configuration with a low dry contact surface area before dissociating.

To analyze the TPE of the nonspecific unbinding process B ⇆ U, we computed several collective variables. Besides the minimum distance r_{min} and the angle of rotation ϕ, we measure the minimum salt bridge distance, the dry contact surface area A_{dry}, the number of waters in a cylinder of a certain radius r between the proteins' centers of mass $N_{tube,r}$ and the number of hydrogen-bond bridging water $N_{HB\ bridging\ water}$. Figure 8 shows the path density for both the LCP and the entire transition path ensemble for several combinations of these CVs. As these path densities are convolutions of the three different mechanisms, we also present multiple TPE path density plots for each individual mechanism in Figure S13. The LCP shows two major TSRs, one (TSR_4) at r_{min} = 0.3 and a dry area of around 1−1.5 nm^2, and one (TSR_5) at r_{min} = 0.5 and a dry area of 0.15 nm^2. TSR_5 is partly due to direct paths that show a simultaneous decrease in dry surface area, increased solvation, increased salt bridge distance upon dissociation, and drastic decrease of the hydrogen-bond bridging water (see Figure S13). RC analysis of the direct mechanism (see Table 4) indicates that it involves formation of salt bridges and (to a lesser extent) attaining the proper orientation ϕ. As the direct mechanism only has to pass the barrier TSR_5, this mechanism exhibits much shorter transition paths than the hopping transition (see Figure 6). Visual inspection of configurations in TSR_5 (see Figure 9) indicated that although the two proteins are separated by a solvent layer without any contacts formed, their charged and polar residue-dominated interface surfaces are correctly aligned, suggesting that from an association perspective, the proteins can quickly form a dry area of more than 2 nm^2.

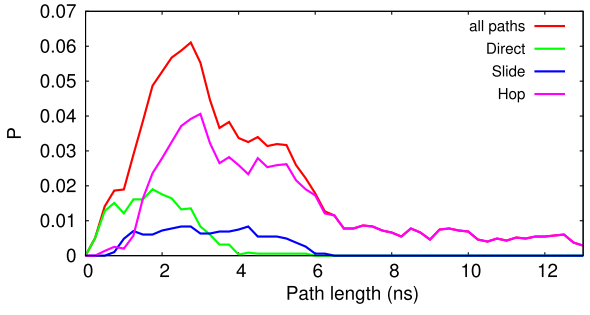


Figure 6. Path length distribution of the B ⇆ U transition path ensemble with respect to the different underlying mechanisms. In red are all paths of the path ensemble, and in green, blue, and magenta, the ones corresponding to the direct, slide, hop dissociation/association, respectively.

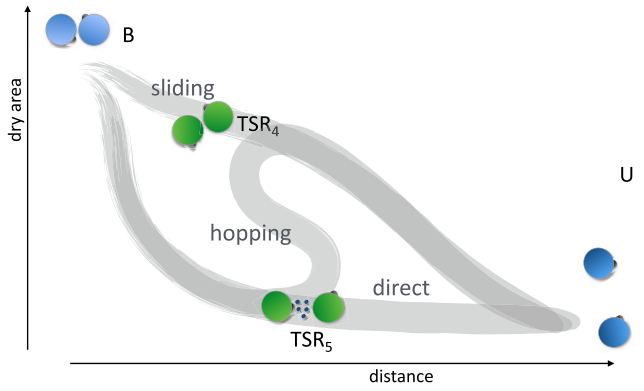


Figure 7. Cartoon network of transitions and respective TSRs during the nonspecific dissociation/association process. We identify three types of mechanisms: direct, hopping, and sliding mechanism.

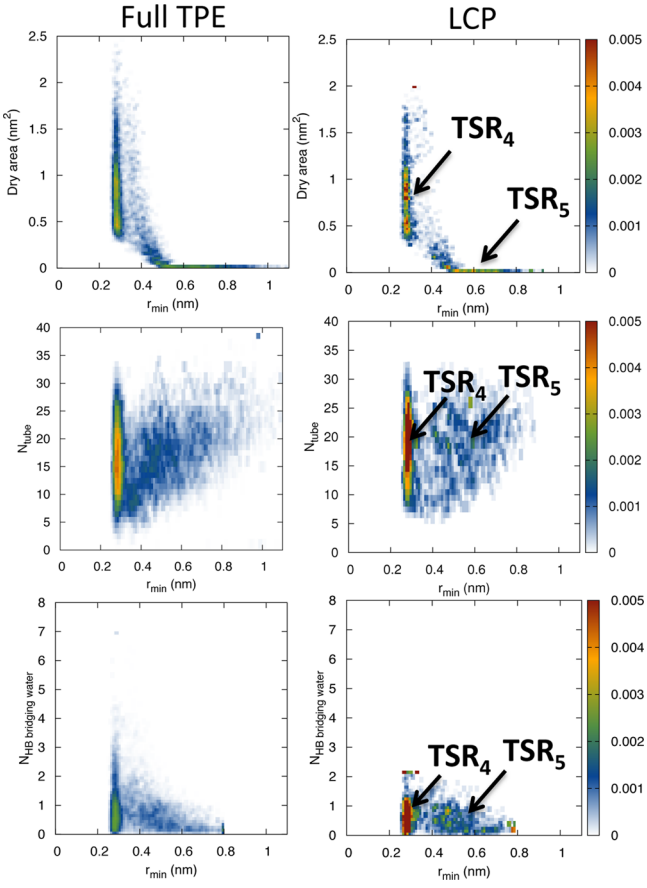


Figure 8. Transition path ensemble and least changed path density plots of the number of dry area vs protein−protein minimum distance (up), number of waters in the tube of radius 0.8 nm vs protein−protein minimum distance (middle), and the number of hydrogen-bond bridging waters vs protein−protein minimum distance (bottom) for the TPE (left) and LCP (right).

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

Table 4. LM Analysis for the B ⇆ U on the Basis of the Forward Shooting Points for (a) Direct Paths, (b) Sliding Paths, and (c) Hopping Paths

n	$\ln L$	reaction coordinate
	Direct Paths δL_{min} = 2.07 (63 Shooting Points)	
1	−39.39	$0.49 - 0.24N$
2	−36.87	$0.95 + 2.85r_{min}$ (salt bridges) $- 0.10\phi$
3	−36.47	$0.98 - 0.15N + 2.10r_{min}$ (salt bridges) $- 0.07\phi$
	Sliding Paths δL_{min} = 1.56 (23 Shooting Points)	
1	−14.00	0.93 − 1.63dry area
2	−12.96	$4.46 - 4.64$dry area $- 5.80r_{min}$
3	−12.38	$17.11 - 1.02N - 11.33r_{min}$ (salt bridges) $+ 0.22\phi$
	Hopping Paths δL_{min} = 2.55 (164 Shooting Points)	
1	−108.42	−8.50 + 1.97dist
2	−106.85	$-6.90 - 0.78$dry $HF_{area} + 1.66$dist
3	−105.86	$-12.91 - 0.89$dry $HF_{area} + 1.93$dist + 9.61S14

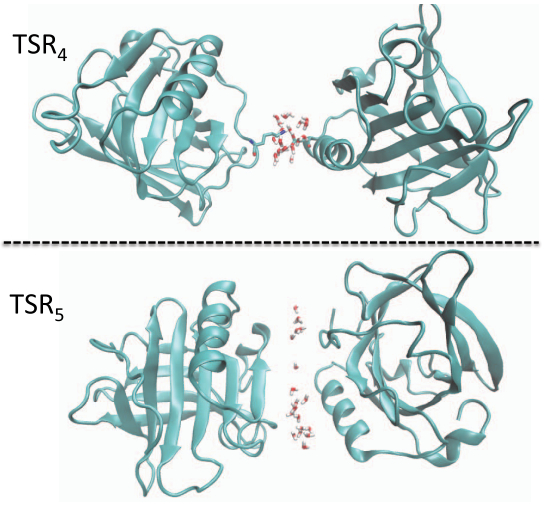


Figure 9. Snapshots of configurations in TSR_4 and TSR_5.

The path density for the sliding mechanism exhibits a strong peak at contact r_{min} = 0.3 and dry area around 1−1.5 nm^2, characteristic for TSR_4. RC analysis of the sliding mechanism (see Table 4) shows that the pertinent degree of freedom is indeed the dry surface area. Moreover, this dry area involves formation of salt bridges, as shown by the r_{min} (salt-bridge) vs dry area plot of Figure S13c. Figure 9 suggests the TSR_4 comprises formation of a few salt bridges in a wet interface without any other contacts formed, inducing a small crevice region around it. Hence, the proteins have to sequentially slide each other to reach the bound state B. Hopping paths can exhibit both TSR_4 and TSR_5, clear by the presence of two peaks in the path densities of Figure S13. Thus, the hopping transition can involve a barrier at contact, as well as a barrier at separated distances. The difference between hopping and sliding is even more clearly visible in Figure S14, which shows the LCP plot for each of these mechanisms. Indeed, whereas the sliding LCP shows basically just one peak at r_{min} = 0.3 and dry area around 1−1.5 nm^2, the hopping LCP exhibits a second peak at $r_{min} \approx 0.6$ and zero dry area. RC analysis of the hopping transition shows that the CM distance of the proteins is the most relevant collective variable (see Table 4). Although this makes sense as hopping requires complete separation before rebinding, it is probably due to the mixed influence of the two TSRs. Again, we can view the sliding mechanism as rebinding without the proteins truly separating. Finally, we looked at the hydrophilic nature of the dry contact area, which although did not make the threshold, was also deemed a reasonable RC for the hopping mechanics. Plotting the LCP as a function of the dry interfacial area and the hydrophilic minus hydrophobic dry interfacial area in Figure S14 makes clear that the nature of the dry interface is indeed hydrophilic, which is expected for a hydrophilic protein such as β-lac.

Note that hopping paths are in majority in the path ensemble for the nonspecific B−U transition (see Figure 6), whereas the N ⇆ U path ensemble showed more sliding paths (see Figure 1). We can rationalize this as follows. In the N ⇆ U transition, paths have to end in the N state, which causes an entropic bottleneck for the direct binding and the hopping mechanism. Sliding requires a two dimensional search over the surface of the protein, which does not suffer from this entropic bottleneck as much. Hence, the sliding paths are in the majority. In the B ⇆ U transition, the bound state is not specific and can easily be reached by the direct, sliding, and hopping mechanisms. Hopping is now very likely as from a hydrated transition state, the proteins can easily rebind. Our findings of direct, sliding, and rebinding paths for the nonspecific dissociation for the B ⇆ U suggest that these atomistic mechanisms of protein unbinding, also occurring in the specific N ⇆ U, are general dissociation mechanisms. Indeed, recent numerical work highlights these mechanisms using simplified models of protein binding.[25,26] These findings are also fully compatible with previous MD simulations.[6,55]

Structure and Dynamics of Hydration Water during Dissociation. As we find that water is an ingredient in the RC for both the B ⇆ U transition (for example, in desolvation barrier TSR_5) as well as the N ⇆ U transition, and by the presence of hydrogen-bond bridging contacts between proteins both at TSR_1, TSR_2, and TSR_3, we further investigate the structure and reorientation dynamics of water during the dissociation process. We focus on the hydration of three states during the N ⇆ U transition: native, transient transition state (TSR_2), and unbound state.

In particular, we identify the structural parameter S and reorientational decay times τ as in our previous study[49] (see Methods and the SI) for interfacial waters in the three states taken from a reactive dissociation path from the N ⇆ U transition path ensemble. We run short NVE MD simulations at 330 K and 300 K and analyze the water structure and dynamics. The results are qualitatively the same at 300 K and 330 K (see Figures 10 and S15, respectively). In the native state, water around the native contact region, comprising residues D33, H146, I147, R148, L149, and S150, exhibits two hydration populations. The first is a more tetrahedrally structured ($S > 0.4$ and $\tau < 4$) water population, labeled tetrahedral water, and the second is a more slowly reorienting, less structured water population denoted as disordered slow ($S < 0.4$ and $\tau > 4$). Note that the bulk water structure parameter is S = 0.38 at 300 K, substantially lower than the tetrahedral population. Upon protein association, the tetrahedral water population increases, as S shifts from 0.4 to 0.5 (Figure 10a−c). The tetrahedral water, although faster than the disordered water, is still relatively slowed down with respect to the bulk water, for which $\tau_b \approx 0.9$ ps for TIP3P. Thus, although the tetrahedral water population is slower than bulk water, it has a bulklike component, similar to that found in ref 48. Since the tetrahedral water population also lives near hydrophobic groups, we computed the $S-\tau$ correlation plot for water

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

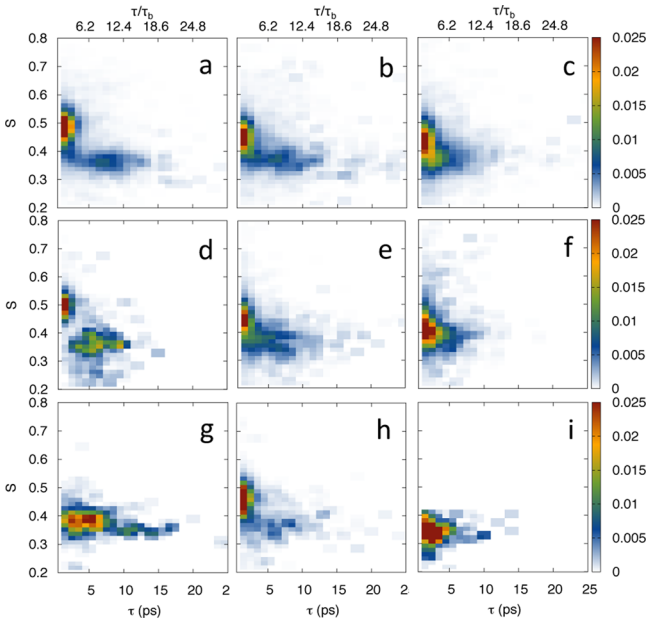


Figure 10. Two-dimensional histograms of the reorientation time τ (slowdown τ/τ_b) versus structural parameter S or water molecules residing at the native interface from NVE simulations performed in the (a) native, (b) transition state (TSR_2), and (c) unbound states at 300 K. Structural parameter-reorientation time correlation for the native state hydrophobic residues I147 and L149 at (d) native, (e) transition state, and (f) unbound configurations. Structure reorientation correlation for the hydrophilic residue D33 at (g) native, (h) transition state, and (i) unbound configurations.

around the hydrophobic amino acids I147 and L149 (Figure 10d–f). Although these hydrophobic amino acids do not exhibit much tetrahedral water structure in the unbound state, they do so in the native interface where these amino acids are opposite to each other. Moreover, these hydrophobic amino acids also show substantial disordered slow water populations, due to the influence of the hydration state of the neighboring charged H146, R148, and (one of the) D33 residues. In contrast, some amino acids, such as the charged residue D33, exhibit only the disordered hydration state in the N and the U state (Figure 10g–i). Indeed, the slow population around D33 increases upon association as the formation of the hydrophilic interface provides polar and charged neighboring amino acids, e.g., R40, as well as a more excluded volume (dry interface) environment.[48] However, in TSR2, there is a temporary increase of tetrahedral water, because the second D33 has not (yet) formed a contact with R40 but is exposed to hydrophobic residues I147, I29, and L149. Moreover, the smaller dry area of TSR2 compared to N excludes the solvent around the second D33 less, yielding faster reorientation dynamics and more tetrahedral structure.[48]

Native contacts are supported by water-mediated interactions: bridging waters hydrogen-bonded to both proteins. In Figure 11, we plot the hydrogen-bond bridge survival correlation function, which decays slower for waters in the hydrophilic contact-rich native state interface compared to waters in the hydrophilic contact-poorer near-native state (TSR2), in accordance with findings of a slow water population at the interface found in ref 10. The faster decay in the transition state indicates proteins are more mobile, as expected. This can enhance (re)binding and dissociation kinetics. The slow decay in the native state reflects the presence of long-lived disordered water around charged and polar amino acids in the interface. Indeed, the average residence time of the disordered water population increases upon binding, namely, 36 ps in the unbound state and 69 ps in the native state, which is much longer than that of the tetrahedral waters, with residence times 17.2 and 16.3 ps, respectively (see Figure 11b).

Thus, water at the native dimer interface consists of a disordered slow population due to formation of long-lived hydrogen-bonded bridging water, and a tetrahedral population, reorienting still slower than bulk, which is typical of hydrophobic solvation and thus characteristic of the β-lac's mixed polar–apolar interface. This conclusion also extends to water in the transition state, which has more mobile hydrogen-bond bridging waters enabling enhanced protein rotation. Although the β-lac protein dimer is not in the electrostatically steered regime, the slow reorientation of the disordered slow waters could be related to the reduction in dielectric permittivity, as observed by Ahmad et al.[12] for the Barnase–Barstar complex. Indeed, this permittivity was found to be small even at larger distances, indicating small fluctuations in the dipole moment, in turn, implying slower water reorientation dynamics.[12]

As the TPS paths were performed at 330 K, we also analyzed the water structure and dynamics for this elevated temperature. To check whether lowering the temperature to room temperature would severely alter the results, we repeated the

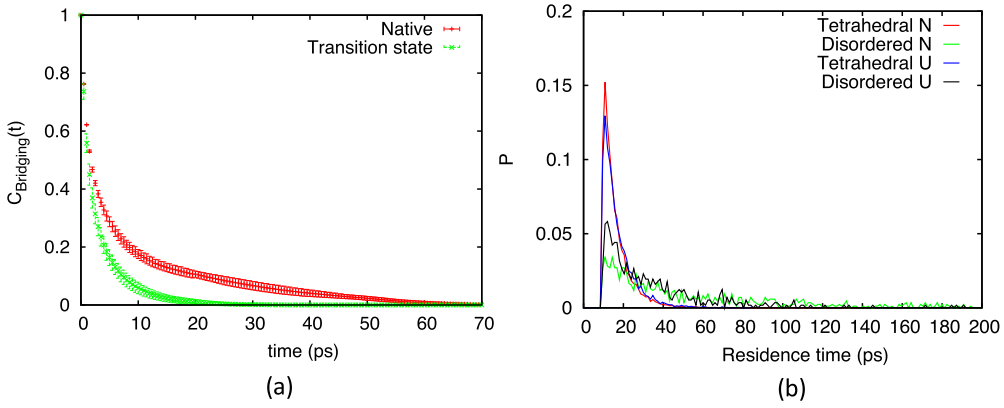


Figure 11. (a) Hydrogen-bond bridge survival correlation function between any amino acid pair of the interface for native (red) and transition state (green) at 300 K. (b) Distribution of residence times for the tetrahedral and disordered water population.

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883–1895

analysis for 300 K. Our findings indicate that the hydration states of water do not change qualitatively upon lowering the temperature from 330 to 300 K (see Figures S15−S17). Note that in the above discussion we have used the 300 K results.

CONCLUSIONS

In this study, we performed extensive TPS simulations of specific and nonspecific dissociation of the hydrophilic β-lactoglobulin dimer. This resulted in ensembles of unbiased dynamical transition paths that are inaccessible with standard MD. Analysis of these sampled path ensembles revealed that specific dissociation can occur either via a direct aligned transition, a hopping and rebinding transition followed by unbinding, or via a sliding transition before unbinding.

For nonspecific dissociation, the trajectories are significantly shorter but also exhibit mechanisms involving direct dissociation, a sliding mechanism where first the dry area decreases by a sliding movement, before dissociation, and a hopping dissociation route where the proteins first rebind before completely dissociating. This finding suggests that the mechanism of direct dissociation, sliding, and rebinding are general dissociation mechanisms. Theoretical and numerical work on simple models shows that indeed protein association is influenced by rebinding.[25,26] The sliding mechanism can thus be viewed as rebinding to a nonspecific state without full solvation.

Employing reaction coordinate and transition-state analysis, we found that in the transition states regions, only a small fraction (∼25%) of the native contacts are present. This conclusion is in agreement with the recent straightforward simulations by the DE Shaw group.[19] In addition, we found evidence for an important role of the D33−R40 salt bridge, also implicated by experiments.[29] Moreover, we investigated the role of the solvent in the dissociation process by assessing the structure and dynamics of the solvent molecules. This analysis revealed that the dry native interface induces enhanced populations of both disordered hydration water near hydrophilic residues and hydration water with higher tetrahedrality, mainly nearby hydrophobic residues. In addition, we found that the hydrogen-bond bridging waters decay faster in the transition state, than in the native state. Thus, whereas water assists binding by forming hydrogen bond bridging water interactions at the natively bound and near-native TSR1 transition states, as predicted by theoretical models,[24] its dynamic nature at the (further lying) transition-state region TSR2 assists protein diffusion on the surface, leading to enhanced kinetics and (re)binding probability.[23]

In summary, the rare unbiased reactive molecular dynamics trajectories shows in full detail how proteins can dissociate via complex pathways including (multiple) rebinding events. Our results give an unbiased dynamical view of the mechanism of protein−protein dissociation in explicit solvent, as well as insight into the structural and dynamical role of the solvent in this process. The atomistic insight obtained assists in further understanding and control of the dynamics of protein−protein interaction including the role of solvent. We expect that our predictions can be experimentally tested, e.g., with spectroscopic techniques such as NMR or vibration sum frequency generation.

Finally, we remark that our approach does not provide a complete full sampling of the dissociation or, because of the time reversible nature of the dynamics, the association process. The main reason for this is the limited allowed duration of the pathways. Restricting the paths to a maximum time excludes all paths that show many hoppings between intermediates. On the other hand, the fact that there is still a substantial number of successful dissociation paths accepted shows that these long rebinding paths might not be so important for dissociation. Nevertheless, we stress that for a full kinetic description of the entire dissociation and association process, much more sampling is needed. A viable way might be to combine the path sampling approach with the MSM methods of ref 17.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b10005.

Spring shooting TPS algorithm, creating the initial path for TPS, assessing the conformational stability of β-lac at 330 K, analysis of the path ensemble, water reorientation dynamics, structure−water reorientation correlation, TPS of the specific dissociation transition, hydration states of water (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: p.g.bolhuis@uva.nl. Phone: +31-205256447.

ORCID

Z. F. Brotzakis: 0000-0001-7024-1430

P. G. Bolhuis: 0000-0002-3698-9258

Present Addresses

‡Institute of Computational Science, Universitá della Svizzera Italiana (USI), Via Giuseppe Buffi 13 Lugano, CH-6900 Lugano, Ticino, Switzerland (Z.F.B.).

†Department of Chemistry and Applied Bioscience, ETH Zürich, c/o USI Campus, Via Giuseppe Buffi 13 Lugano, CH-6900, Lugano, Ticino, Switzerland (Z.F.B.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research leading to these conclusions was funded by the NanoNextNL Programme, a micro and nanotechnology consortium of the Government of the Netherlands and 130 partners. We acknowledge support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) for the use of supercomputer facilities.

REFERENCES

(1) Lodish, L.; Berk, A.; Kaiser, C. A.; Krieger, M.; Bretscher, A.; Ploegh, H.; Amon, A. M.; Martin, K. C. *Molecular Cell Biology*, 8th ed; W.H. Freeman: Boston, 2016.

(2) Plakoutsi, G.; Bemporad, F.; Calamai, M.; Taddei, N.; Dobson, C. M.; Chiti, F. Evidence for a Mechanism of Amyloid Formation Involving Molecular Reorganisation within Native-like Precursor Aggregates. *J. Mol. Biol.* **2005**, *351*, 910−922.

(3) Wang, Y.; Wang, Y.; Breed, D. R.; Manoharan, V. N.; Feng, L.; Hollingsworth, A.; Weck, M.; Pine, D. J. Colloids with Valence and Specific Directional Bonding. *Nature* **2012**, *491*, 51−55.

(4) Takahashi, K.; Tanase-Nicola, S.; ten Wolde, P. R. Spatio-Temporal Correlations can Drastically Change the Response of a MAPK Pathway. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2473−2478.

(5) Vijaykumar, A.; Bolhuis, P. G.; ten Wolde, P. R. The Intrinsic Rate Constants in Diffusion-Influenced Reactions. *Faraday Discuss.* **2016**, *195*, 421−441.

(6) Schreiber, G.; Haran, G.; Zhou, H. X. Fundamental Aspects of Protein-Protein Association Kinetics. *Chem. Rev.* **2009**, *109*, 839−860.
(7) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−7.
(8) Huang, X.; Margulis, C. J.; Berne, B. J. Dewetting-Induced Collapse of Hydrophobic Particles. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11953−11958.
(9) McLain, S. E.; Soper, A. K.; Daidone, I.; Smith, J. C.; Watts, A. Charge-Based Interactions between Peptides Observed as the Dominant Force for Association in Aqueous Solution. *Angew. Chem., Int. Ed.* **2008**, *47*, 9059−9062.
(10) Bello, M.; Martiniano, P. E.; Gerardo, A. R.; et al. Energetics of protein homodimerization: Effects of water sequestering on the formation of b -lactoglobulin dimer. *Proteins* **2008**, *70*, 1475−1487.
(11) Ansari, S.; Helms, V. Statistical Analysis of Predominantly Transient Protein-Protein Interfaces. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 344−355.
(12) Ahmad, M.; Gu, W.; Geyer, T.; Helms, V. Adhesive Water Networks Facilitate Binding of Protein Interfaces. *Nat. Commun.* **2011**, *2*, No. 261.
(13) Spaar, A.; Dammer, C.; Gabdoulline, R. R.; Helms, V.; et al. Diffusional Encounter of Barnase and Barstar. *Biophys. J.* **2006**, *90*, 1913−1924.
(14) Schreiber, G.; Fersht, A. R. Rapid, Electrostatically Assisted Association of Proteins. *Nat. Struct. Biol.* **1996**, *3*, 427−431.
(15) Selzer, T.; Albeck, S.; Schreiber, G. Rational Design of Faster Associating and Tighter Binding Protein Complexes. *Nat. Struct. Biol.* **2000**, *7*, 537−541.
(16) Doench, J. G.; Hartenian, E. G.; Daniel, B.; Tothova, Z.; Hegde, M.; Smith, I.; Sullender, M.; Ebert, B. L.; Xavier, R. J.; Root, D. E. Rational Design of Highly Active sgRNAs for CRISPR-Cas9-Mediated Gene Inactivation. *Nat. Biotechnol.* **2014**, *32*, 1262−1267.
(17) Plattner, N.; Doerr, S.; de Fabritiis, G.; Noé, F. Complete Protein-Protein Association Kinetics in Atomic Detail Revealed by Molecular Dynamics Simulations and Markov Modelling. *Nat. Chem.* **2017**, *9*, 1005−1011.
(18) Ulucan, O.; Jaitly, T.; Helms, V. Energetics of Hydrophilic Protein - Protein Association and the Role of Water. *J. Chem. Theory Comput.* **2014**, *10*, 3512−3524.
(19) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-Level Characterization of Protein-Protein Association. *Biophys. J.* **2018**, *114*, No. 557a.
(20) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. Computer Simulation of Protein-Protein Association Kinetics: Acetylcholinesterase-Fasciculin. *J. Mol. Biol.* **1999**, *291*, 149.
(21) Radić, Z.; Kirchhoff, P. D.; Quinn, D. M.; McCammon, J. A.; Taylor, P. Electrostatic Influence on the Kinetics of Ligand Binding to acetylcholinesterase distinctions between active center ligands and fasciculin. *J. Biol. Chem.* **1997**, *272*, 23265−23277.
(22) Elcock, A. H.; Sept, D.; McCammon, J. A. Computer Simulation of Protein-Protein Interactions. *J. Phys. Chem. B* **2001**, *105*, 1504−1518.
(23) Northrup, S. H.; Erickson, H. P. Kinetics of Protein-Protein Association Explained by Brownian Dynamics Computer Simulation. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3338−42.
(24) Ben-Naim, A. On the Driving Forces for Protein-Protein Association. *J. Chem. Phys.* **2006**, *125*, No. 024901.
(25) Newton, A. C.; Kools, R.; Swenson, D. W. H.; Bolhuis, P. G. The Opposing Effects of Isotropic and Anisotropic Attraction on Association Kinetics of Proteins and Colloids. *J. Chem. Phys.* **2017**, *147*, 155101−155112.
(26) Vijaykumar, A.; Ten Wolde, P. R.; Bolhuis, P. G. Rate Constants for Proteins Binding to Substrates with Multiple Binding Sites using a Generalized Forward Flux Sampling Expression. *J. Chem. Phys.* **2018**, *148*, No. 124109.
(27) Tang, C.; Iwahara, J.; Clore, G. M. Visualization of Transient Encounter Complexes in Protein-Protein Association. *Nature* **2006**, *444*, 383−386.
(28) Frisch, C.; Fersht, A. R.; Schreiber, G. Experimental Assignment of the Structure of the Transition State for the Association of Barnase and Barstar. *J. Mol. Biol.* **2001**, *308*, 69−77.
(29) Sakurai, K.; Goto, Y. Manipulating Monomer-Dimer Equilibrium of Bovine β-Lactoglobulin by Amino Acid Substitution. *J. Biol. Chem.* **2002**, *277*, 25735−40.
(30) Pan, A. C.; Jacobson, D.; Borisov, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-Level Characterization of Protein-Protein Association. *Biophys. J.* **2018**, *114*, 557a.
(31) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(32) Mercadante, D.; Melton, L. D.; Norris, G. E.; Loo, T. S.; Williams, M. aK.; Dobson, R. C. J.; Jameson, G. B. Bovine β-Lactoglobulin is Dimeric under Imitative Physiological Conditions: Dissociation Equilibrium and Rate Constants over the pH Range of 2.5-7.5. *Biophys. J.* **2012**, *103*, 303−12.
(33) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−54.
(34) 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* **2010**, *78*, 1950−8.
(35) 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.
(36) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(37) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(38) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **1998**, *108*, 1964−1997.
(39) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling. *Adv. Chem. Phys.* **2002**, *123*, 1−86.
(40) Bolhuis, P. G. Transition Path Sampling on Diffusive Barriers. *J. Phys. Condens. Matter* **2003**, *113*, S113−S120.
(41) Vreede, J.; Juraszek, J.; Bolhuis, P. G. Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2397−402.
(42) Juraszek, J.; Bolhuis, P. G. Rate Constant and Reaction Coordinate of Trp-cage Folding in Explicit Water. *Biophys. J.* **2008**, *95*, 4246−57.
(43) Brotzakis, Z. F.; Bolhuis, P. G. A One-Way Shooting Algorithm for Transition Path Sampling of Asymmetric Barriers. *J. Chem. Phys.* **2016**, *145*, 164112−12.
(44) 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.
(45) E, W.; Vanden-Eijnden, E. Towards a Theory of Transition Paths. *J. Stat. Phys.* **2006**, *123*, 503−523.
(46) Peters, B.; Trout, B. Obtaining Reaction Coordinates by Likelihood Maximization. *J. Chem. Phys.* **2006**, *125*, No. 054108.
(47) Peters, B.; Beckham, G. T.; Trout, B. L. Extensions to the likelihood maximization approach for finding reaction coordinates. *J. Chem. Phys.* **2007**, *127*, No. 034109.
(48) Brotzakis, Z. F.; Groot, C. C. M.; Brandeburgo, W. H.; Bakker, H. J.; Bolhuis, P. G. Dynamics of Hydration Water around Native and Misfolded α-Lactalbumin. *J. Phys. Chem. B* **2016**, *120*, 4756−4766.
(49) Brotzakis, Z. F.; Voets, I. K.; Bakker, H. J.; Bolhuis, P. G. Water Structure and Dynamics in the Hydration Layer of a Type III Antifreeze Protein. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6996.

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895

(50) Madan, B.; Sharp, K. Heat Capacity Changes Accompanying Hydrophobic and Ionic Solvation: A Monte Carlo and Random Network Model Study. *J. Phys. Chem.* **1996**, *100*, 7713−7721.
(51) Sharp, K. A.; Vanderkooi, J. M. Water in the Half Shell: Structure of Water, Focusing on Angular Structure and Solvation. *Acc. Chem. Res.* **2010**, *43*, 231−239.
(52) Gallagher, K. R.; Sharp, K. A. Analysis of thermal hysteresis protein hydration using the random network model. *Biophys. Chem.* **2003**, *105*, 195−209.
(53) Bello, M.; Fragoso-Vázquez, M. J.; Correa Basurto, J. Energetic and Conformational Features Linked to the Monomeric and Dimeric States of Bovine BLG. *Int. J. Biol. Macromol.* **2016**, *92*, 625−636.
(54) Gutiérrez-magdaleno, G.; Bello, M.; Portillo-téllez, M. C.; Rodríguez-romero, A.; García-hernández, E. Ligand binding and self-association cooperativity of b-lactoglobulin. *J. Mol. Recognit.* **2013**, 67−75.
(55) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun.* **2015**, *6*, No. 7653.

DOI: 10.1021/acs.jpcb.8b10005
J. Phys. Chem. B 2019, 123, 1883−1895