Analysis of the dynamics of a complex, multipathway reaction: Insulin dimer dissociation

Kwanghoon Jeong,[†] Spencer C. Guo,[†] Sammy Allaw,[†,§] and Aaron R. Dinner[*,†,‡,¶]

†*Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, USA*
‡*James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA*
¶*Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637, USA*
§*Current address: Indiana University School of Medicine, Indianapolis, Indiana 46202, USA*

E-mail: dinner@uchicago.edu

Abstract

The protein hormone insulin forms a homodimer that must dissociate to bind to its receptor. Understanding the kinetics and mechanism of dissociation is essential for rational design of therapeutic analogs. In addition to its physiological importance, this dissociation process serves as a paradigm for coupled (un)folding and (un)binding. Based on previous free energy simulations, insulin dissociation is thought to involve multiple pathways with comparable free energy barriers. Here, we analyze the mechanism of insulin dimer dissociation using a recently developed computational framework for estimating kinetic statistics from short-trajectory data. These statistics indicate that the likelihood of dissociation (the committor) closely tracks the decrease in the number of (native and nonnative) intermonomer contacts and the increase in the number of water contacts at the dimer interface; the transition state with equal likelihood of association and dissociation corresponds to an encounter complex with relatively few native contacts and many nonnative contacts. We identify four pathways out of the dimer state and quantify their contributions to the rate, as well as their exchange, by computing reactive fluxes. We show that both the pathways and their extents of exchange can be understood in terms of rotations around three axes of the dimer structure. Our results provide insights into the kinetics of insulin analogues and, more generally, how to characterize complex, multipathway processes.

Introduction

The kinetics of molecular association and dissociation govern many processes in chemistry and biology ranging from reactions of a few atoms to changes in cell state. The most basic theories of association treat molecules as spheres undergoing Brownian motion;[1] these theories remain relevant today because they can be extended to account for molecular structure by factoring the rate into terms for reaching an encounter complex and for dynamics within that complex.[1–4] Despite the success of this approach, it remains difficult to apply when the dynamics in the encounter complex involve significant intramolecular rearrangement, as is often the case for macromolecules such as proteins. In this case, simulations generally cannot directly bridge the separation of timescales between the fastest motions of the molecules (typically bond fluctuations) and the dynamics of interest. A common

approach to estimating kinetics when such a separation in timescales exists is to identify a low-dimensional representation and apply transition state theory (TST) and its elaborations that account implicitly for dynamics in the remaining degrees of freedom,[5,6] but whether the assumptions underlying this approach are valid generally remains unclear (e.g., see Ref. 7 for discussion). Even if they are, this strategy provides limited microscopic insight into the dynamics of the degrees of freedom beyond the reaction coordinates.

These issues are exemplified by studies of the association and dissociation of the monomers in the insulin homodimer. Each monomer consists of two polypeptide chains that are linked by disulfide bridges: a 21-amino acid A-chain and a 30-amino acid B-chain. In the dimer (Figure 1), the secondary structure of the A-chain consists of two α-helices (A1–A9 and A13–A20); that of the B-chain consists of an α-helix (B9–B19), followed by a β-turn (B20–B23), and a β-strand (B24–B26). At the dimer interface, the B-chain α-helices from the two monomers pack against each other, and the B24–B26 segments form an anti-parallel β-sheet. Both experiments[8,9] and simulations[10–12] indicate that the A-chain N-terminal helix (A1–A9), the B-chain N-terminal strand (B1–B7), and the B-chain C-terminal residues (B24–B30) become disordered when mutations and/or low pH are used to stabilize the monomer relative to the dimer.

Dinner and co-workers previously showed that the potential of mean force (PMF) as a function of collective variables (CVs) characterizing the contacts between the interfacial α-helices and β-strands supports a diversity of dissociation pathways ranging from ones in which the α-helices separate first to ones in which the β-strands separate first.[15] Along the limiting path in which the α-helices separate first, the α-helices twist apart, and water penetrates the hydrophobic core before the β-strands slide apart, detach from the core, and become disordered. Along the alternative limiting path, the β-strands twist apart first while remaining attached to the core, and the monomers separate largely in their dimeric structures. The paths were estimated to have comparable free energy barriers of around 14 kcal/mol in the dissociation direction and 1–2 kcal/mol in the association direction. These simulations provide a unified perspective of diverse dynamics in earlier simulations.[11,16–19]

In particular, Bagchi and co-workers inferred a path in which the interfacial α-helices separate first from the PMF as a function of the distance between the centers of mass of the monomers and the number of contacts between them.[17] Based on this two-dimensional PMF and its minimum free-energy path, they subsequently estimated the dissociation rate using a variety of theories; with corrections to the TST rate, they obtained a value of 0.4 μs^{-1}.[7] They identified the transition to the first metastable intermediate with twisted interfacial α-helices and a largely intact interfacial β-sheet[19] to be the rate limiting step in TST. The existence of this intermediate is consistent with a twisted dimeric state found in a Markov state model (MSM), though the MSM predicts longer timescales of transition to and from it ($\sim$20 μs).[20] These timescales estimated for the initial events in dissociation are on the order of those detected by temperature-jump two-dimensional infrared (2D IR) spectroscopy of the amide I vibrational modes, which suggestst that monomer disordering is in the 5–250 μs time range and dissociation is in the 250–1000 μs time range.[21] On the other hand, time-resolved X-ray scattering measurements suggest dissociation intermediates appear within 1 μs but cannot resolve their structural features.[22]

Given the heterogeneity of possible paths dissociation can take, it is interesting that the rate theory estimate is in reasonable agreement with the experimental data. Bagchi and co-workers report that their rate estimates are quite sensitive to their chosen coordinates,[7] making it important to be able to evaluate coordinates with respect to their ability to describe the dynamics objectively. A principled way of doing so is to compare coordinates to the committor,[23] which is the probability that from a given microscopic state a trajectory goes to a product state before a reactant state.[6,24]

Here, we use a generalization of MSMs termed

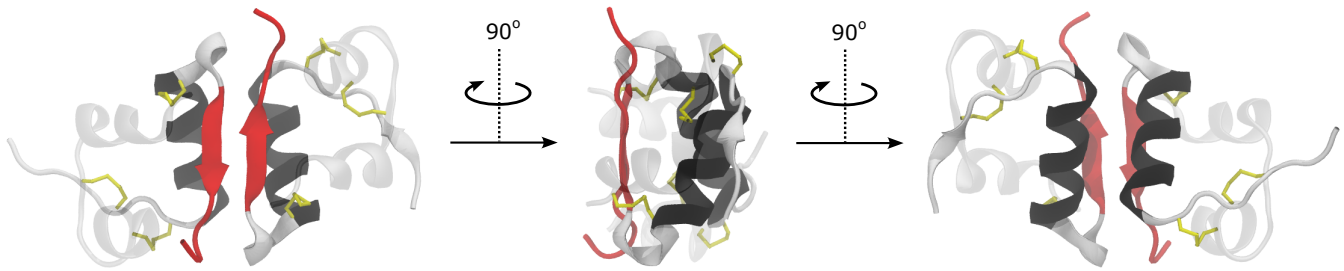

Figure 1: Experimental insulin dimer structure (PDB ID 3W7Y[13]) from three different points of view. The A-chains are translucent, while the B-chains are opaque. The interfacial β-strand and B-chain C-terminal residues (B24–30) are red, and the interfacial α-helices (B9–19) are black. Cysteine disulfide bonds are shown in yellow. We use VMD[14] to visualize molecular structures throughout the paper.

dynamical Galerkin approximation (DGA)[25,26] to estimate the committor for insulin dimer dissociation. We show that the committor correlates with the number of interfacial contacts (consistent with the choices of CVs in Refs. 15 and 17) and the total number of interfacial water molecules. We use the committor, together with the equilbrium probability, to compute the reactive current, which enables us to quantify the importance of competing pathways and estimate the rate of dissociation. We show that the heterogeneous paths out of the dimer state and the fluxes between them can be understood in terms of three axes around which the monomers rotate, providing a simple perspective on this complex reaction.

Methods

Here we describe the statistics that we use in our analysis and how we compute them. In brief, we run many short, unbiased simulations and use them to define a set of Markov states and the probabilities of transition between them. Because we draw the initial structures for these simulations from previous equilibrium simulations, we can use the earlier results to estimate the equilibrium probability (π) over the states, which provides information about relative stabilities of the Markov states. From the matrix of transition probabilities, we can compute the committor (q), which, as defined above, is the probability of reaching a product state before a reactant state from a given Markov state. Because it provides information about the likelihood of completing a stochastic reaction, the committor serves as a measure of reaction progress, and we define the transition state as the ensemble of states with equal likelihood of next going to the reactant (A) and product (B) states ($q_A = q_B = 0.5$). From the equilibrium probability and the committor, we can compute the reactive current, which tracks how probability flows within the ensemble of reactive trajectories; in turn, we integrate over the reactive current along the committor to obtain the reactive flux (R_{AB}; we denote the fractional contribution to the flux of state i by F_i). The reactive flux divided by the fraction of time the system spends having last come from state A is the rate constant (k_{AB}). Figure 2 summarizes our computational workflow.

Simulation data

In principle, both π and q can be obtained from DGA using an arbitrary sample distribution.[25–28] We initially generated a sample distribution by running unbiased dynamics from conformations sampled by adiabatic-bias molecular dynamics (ABMD),[29] following previous studies.[26,30] However, we found that the resulting PMF was inconsistent with that obtained from earlier replica-exchange umbrella sampling simulations (REUS),[15] which were converged with the aid of an error estimator.[31] Therefore, we instead selected conformations from the umbrella sampling simulations as the starting points for unbiased dynamics simulations, as we describe below.

The system setup was the same as in the

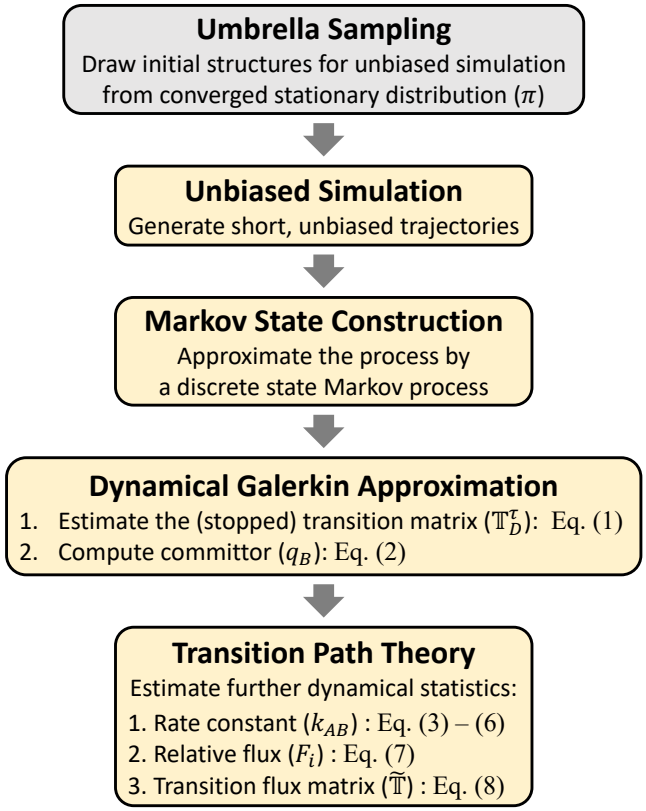

Figure 2: Schematic of the workflow. Gray box shows calculations performed in Ref. 15, and yellow boxes show calculations in the present study.

REUS simulations, and we refer the reader to Ref. 15 for details. To summarize, the system was constructed from a crystal structure of human insulin (PDB ID 3W7Y[13]) and modeled in aqueous solution with cubic periodic boundary conditions. The protein, 48 K^+ ions, and 44 Cl^- ions were represented by the CHARMM36m[32] force field, and 15,532 water molecules were represented by the TIP3P model;[33] the total number of atoms was 48,260. After equilibration, the simulation box size was fixed at $(7.82\,\text{nm})^3$.

From each of the 748 REUS simulations, we sampled 24 structures equally distributed along the 5 ns trajectory, resulting in $24 \times 784 = 18{,}816$ initial structures. From each sampled structure, we initialized a 5 ns trajectory using OpenMM 7.7,[34] yielding an aggregate simulation time of 94.08 µs. Simulations were performed in the canonical (isothermal-isochoric) ensemble at 303.15 K using the LFMiddle integrator[35] with a time step of 2 fs and a friction constant of $0.083\,\text{ps}^{-1}$. The particle-mesh Ewald method was used to calculate non-bonded forces with a cutoff distance of 1.2 nm; the non-bonded interactions were smoothly switched off from 1.0 to 1.2 nm through the built-in OpenMM force-switch function. The lengths of bonds to hydrogens were constrained using the SHAKE algorithm.[36] Structures were saved every 5 ps.

Markov state model (MSM)

We used the simulation data described above to define discrete states and estimate the probabilities of transition between them. In this section, we describe variables that we use to characterize the system and states that we define in terms of them.

Definition of the dimer and monomer states

We used the following variables to define the dimer and (separated) monomer states based on previous studies:

1. RMSD_{int} measures the root mean square deviation (RMSD) of interfacial C^α atoms from their positions in the crystal structure. We define the interfacial C^α atoms to be those within 10 Å of any C^α atom of the other monomer in the crystal structure.[18]

2. Φ_α is a pseudodihedral angle that measures twist of the interfacial α-helices.[15,20] It is defined by the geometric centers of Ser^{B9}–Leu^{B11}, Val^{B12}–Tyr^{B16}, $\text{Val}^{\text{B}'12}$–$\text{Tyr}^{\text{B}'16}$, and $\text{Ser}^{\text{B}'9}$–$\text{Leu}^{\text{B}'11}$ (the primes distinguish one monomer from the other).

3. $\bar{\alpha}$ measures the separation of the interfacial α-helices.[15] It is the average distance between C^α atoms of the following pairs of residues: Ser^{B9}-$\text{Tyr}^{\text{B}'16}$, Ser^{B9}-$\text{Glu}^{\text{B}'13}$, Val^{B12}-$\text{Tyr}^{\text{B}'16}$, Glu^{B13}-$\text{Glu}^{\text{B}'13}$, Glu^{B13}-$\text{Ser}^{\text{B}'9}$, Tyr^{B16}-$\text{Val}^{\text{B}'12}$, and Tyr^{B16}-$\text{Ser}^{\text{B}'9}$.

4. $\bar{\beta}$ measures the separation of the interfacial β-strands.[15] It is the average distance between C^α atoms of the following pairs of residues: Phe^{B24}-$\text{Tyr}^{\text{B}'26}$, Phe^{B25}-$\text{Phe}^{\text{B}'25}$, and Tyr^{B26}-$\text{Phe}^{\text{B}'24}$.

5. N_c measures the number of pairs of non-hydrogen atoms from separate monomers that are within 4.5 Å of each other.

6. N_{sw} measures the number of shared waters. The shared waters are ones with oxygen atoms that are simultaneously within 4 Å of at least one interfacial non-hydrogen atom from each monomer. Here, we define the interfacial atoms to be those within 4 Å of a non-hydrogen atom from the other monomer in the crystal structure.

Given these features, we define the dimer state (A) as structures with $\mathrm{RMSD}_{\mathrm{int}} < 2$ Å, $120° < \Phi_\alpha < 135°$, $\bar{\alpha} < 0.8$ nm, and $\bar{\beta} < 0.55$ nm. We define the monomer state (B) as structures with $\mathrm{RMSD}_{\mathrm{int}} > 10$ Å, $N_c = 0$, and $N_{sw} = 0$.

Features used to define states

We divide the remainder of the conformation space into $k-2$ states so that there are k states with the dimer and monomer states. Unless otherwise indicated, we use $k = 600$. We discuss the choice of the number of states in the Supporting Information (text and Figure S2). We cluster the data with k-means using 45 features: 26 local features and 19 global features.

The local features correspond to contacts between 26 residue pairs (Table S1 and Figure S1). These are based on the residues used to define $\bar{\alpha}$ and $\bar{\beta}$, as well as residues in the β-turn and close to the C-terminus. We consider more residues in B21–B29 than in the α-helices because this segment is more heterogeneous as dissociation proceeds. For each residue pair, we compute the distance between the C^α atoms, x_i, and then take as the feature $\tanh((x_i - \mu_i)/2\sigma_i)$, where μ_i and σ_i are the mean and standard deviation of x_i over the entire dataset.

We found that the local features were insufficient to characterize conformations at later stages of dissociation, when most contacts are broken. Consequently, we also include global features. The global features consist of half the distance between the monomer centers of mass, $R_{\mathrm{com}}/2$, and the sine and cosine of nine angles:

1. Two pseudodihedral angles, Φ_α and Φ_β, measuring twist of the interfacial α-helices and β-strands, respectively; Φ_α is defined above, and Φ_β is defined by the C^α atoms of $\mathrm{Tyr}^{\mathrm{B26}}$, $\mathrm{Phe}^{\mathrm{B25}}$, $\mathrm{Phe}^{\mathrm{B'25}}$, and $\mathrm{Tyr}^{\mathrm{B'26}}$;

2. Two angles, γ and γ', which measure splaying of the interfacial α-helices (see Figure 7 below); γ is the angle defined by the geometric centers of $\mathrm{Val}^{\mathrm{B12}}$–$\mathrm{Tyr}^{\mathrm{B16}}$, $\mathrm{Phe}^{\mathrm{B25}}$, and $\mathrm{Phe}^{\mathrm{B'25}}$, and γ' is the corresponding angle for the other monomer;

3. Five Euler angles characterizing the relative orientations and rotations of the monomers.[37]

As defined above (Table S1), the different types of features have comparable magnitudes, and we use them in the k-means algorithm without further scaling. Because trajectories tend toward the stable states, we used only the 50 frames from 255 ps–500 ps of the 5 ns trajectories when defining the cluster centers to ensure sufficient resolution outside the stable states. We used MDAnalysis[38,39] to compute all collective variables.

Estimation of equilibrium statistics

As a consequence of the fact that we initialize the unbiased simulations from previous umbrella sampling simulations, the equilibrium probabilities of the conformations sampled by the unbiased simulations are proportional to the weights, $w(\mathbf{x})$, of their associated initial conformations. After normalizing these weights so they sum to one, we compute the PMF over Markov states as

$$G(S_i) = -k_{\mathrm{B}}T \log\left(\sum_{\mathbf{x}\in S_i} w(\mathbf{x})\right)$$
$$\Delta G(S_i) = G(S_i) - G(A)$$

where S_i denotes Markov state i, $\mathbf{x}$ is a structure in S_i, k_{B} is Boltzmann's constant, and T is temperature.

Estimation of dynamical statistics

From the MSM, we compute the committor, reactive current, and rate. The central quantity in an MSM is the matrix of state-to-state transition probabilities ($\mathbb{T}$). A key aspect of DGA that distinguishes it from a traditional MSM is that the transition transition matrix is constructed with statistic-specific boundary conditions. That is,

$$(\mathbb{T}_D^\tau)_{ij} = \mathbf{P}[\mathbf{X}(\min(\tau, T_{D^\mathsf{c}})) \in S_j \mid \mathbf{X}(0) \in S_i], \tag{1}$$

where $\mathbf{X}(t)$ is the structure at time t, S_i and S_j are Markov states, $T_{D^\mathsf{c}} = \min\{t \geq 0 \mid \mathbf{X}(t) \in D^\mathsf{c}\}$ is the time of first exit from the domain D (here, $D = (A \cup B)^\mathsf{c}$, but we consider other choices below), and τ is a lag time that must be chosen to be sufficiently long that the dynamics are approximately Markovian in the variables used to define the states. The minimum in (1) effectively stops trajectories when they enter states A and B. We do not symmetrize $\mathbb{T}_D^\tau$ to enforce microscopic reversibility.

We denote the committor that is the probability of next going to state B rather than state A from structure $\mathbf{x}$ by $q_B(\mathbf{x})$. By definition, $q_A = 1 - q_B$, and by microscopic reversibility, q_A is also the probability of last coming from state A rather than state B. The committor satisfies the equation

$$q_B(\mathbf{x}) = \begin{cases} \mathbb{T}_D^\tau q_B(\mathbf{x}) & \mathbf{x} \in D \\ \mathbb{1}_B(\mathbf{x}) & \mathbf{x} \in D^\mathsf{c}, \end{cases} \tag{2}$$

where $\mathbb{1}_B(\mathbf{x})$ is an indicator function that is one if $\mathbf{x}$ is in state B and zero otherwise. We solve for q_B with a modified algorithm that treats non-Markovian effects through the addition of memory terms.[40] Here, we show committors estimated using a lag time of 2 ns and 1 memory term, which we choose to balance expressivity and statistical error (see discussion in Supporting Information and Figures 4, S2, and S3).

Given π, q_A, and q_B, we use transition path theory (TPT)[41] to compute the reactive current projected onto a vector of CVs, θ:

$$J_{AB}^\theta(s) = \int \mathbf{J}_{AB} \cdot \nabla\theta(\mathbf{x}) \delta(\theta(\mathbf{x}) - s) d\mathbf{x}, \tag{3}$$

where

$$\mathbf{J}_{AB} \cdot \nabla\theta(\mathbf{x}) = \lim_{\Delta t \to 0^+} \pi(\mathbf{x}) q_A(\mathbf{x}) q_B(\mathbf{x}) \times \frac{\mathbf{E}[(\theta(\mathbf{X}(\Delta t)) - \theta(\mathbf{X}(-\Delta t)) \mid \mathbf{X}(0) = \mathbf{x}]}{2\Delta t}. \tag{4}$$

Integrating the reactive current along the committor (i.e., setting $\theta = q_B$ in (3) and (4)) gives an estimate of the number of reactive trajectories per unit time, R_{AB}, or the net reactive flux:

$$R_{AB} = \int_0^1 J_{AB}^{q_B}(s)\, \mathrm{d}s. \tag{5}$$

Dividing R_{AB} by the fraction of time the system spends having last come from state A gives the TPT rate constant:

$$k_{AB} = \frac{R_{AB}}{\int \pi(\mathbf{x}) q_A(\mathbf{x})\, \mathrm{d}\mathbf{x}}. \tag{6}$$

In practice, we use finite-lag time estimators for $J_{AB}^\theta(s)$, R_{AB}, and k_{AB}[26,42]—we translate these estimators to our notation and provide further details concerning their use in the Supporting Information.

To estimate the contribution of a state to the net reactive flux, we compute the dynamical statistics with an extra state excluded from the domain,[43] $D := (A \cup B \cup C_i)^\mathsf{c}$ where C_i is any (coarse-grained) state of interest. The interpretation of the committor $q_B^{C_i}$ then becomes the probability to next go to state B rather than $A \cup C_i$, and $q_A^{C_i}$ becomes the probability to last come from state A rather than $B \cup C_i$. We can solve for the committor $q_{C_i}^B$ using (2) with $D := (A \cup B \cup C_i)^\mathsf{c}$. Similarly, we can solve for $q_A^{C_i}$ using the same equation but with A and B switched. We then compute the number of reactive trajectories from A to B without entering C_i per unit time, which we label $R_{AB\setminus C_i}$, by inserting $q_A^{C_i}$ and $q_B^{C_i}$ into (4) and (5). We then define the fractional contribution of each coarse-grained state to the net flux as

$$F_i = 1 - \frac{R_{AB\setminus C_i}}{R_{AB}}. \tag{7}$$

The procedure above can be viewed as a simplified version of the approach for computing

history-dependent TPT statistics in Ref. 44.

In the **Results**, we aggregate Markov states into eight coarse-grained states. To systematically evaluate how the coarse-grained states evolve, we compute a coarse-grained transition flux matrix, $\tilde{\mathbb{T}}$, where $\tilde{\mathbb{T}}_{ij}$ is the probability for a structure starting in state C_i to go to C_j without visiting any other states. More specifically, the elements of $\tilde{\mathbb{T}}$ are

$$\tilde{\mathbb{T}}_{ij} = \mathbf{P}\left[T_{C_j} = T_{\mathcal{C}\setminus C_i} \mid \mathbf{X}(0) \in C_i\right] \tag{8}$$

where

$$\mathcal{C} = \bigcup_{i=1}^{10} C_i \tag{9}$$

and C_i represents the dimer, monomer, and coarse-grained states.

Results

Our analysis of insulin dimer dissociation is organized as follows. We first present the PMFs, committor, and reactive currents as functions of the CVs used to control the sampling. While we project these quantities to two dimensions for visualization, we compute them for 600 Markov states defined in 45 dimensions, as described in **Methods**, and we use them without projection to estimate the rate of dissociation. To dissect the mechanism, we identify CVs that correlate with the committor and relate them to the dimer structure. Then, we consider CVs that do not correlate with the committor to characterize the diversity of reaction paths and their associated fluxes.

The inverse rate is approximately 100 ms

As described in **Methods**, we draw the initial conformations for our unbiased molecular dynamics simulations from umbrella sampling simulations with restraints in two CVs: the average occupancies of seven contacts between the interfacial α-helices ($\bar{\alpha}_\mathrm{c}$) and three contacts between the interfacial β-strands ($\bar{\beta}_\mathrm{c}$).[15] We show the PMF, committor, and reactive currents as functions of these CVs in Figure S4 and on the average distances for the atom pairs defining these CVs ($\bar{\alpha}$ and $\bar{\beta}$) in Figure 3.

The DGA/MSM PMFs closely resemble those from umbrella sampling,[15] indicating that the initial conformations are indeed consistent with the equilibrium distribution. Focusing on Figure 3, the committor increases smoothly from the dimer state in the lower left corner to the monomer state in the upper right corner. The lines of constant committor values (isocommittors) are diagonal, and the arrows representing the reactive currents fan out from the dimer state. Both of these features support the idea that the reaction can follow diverse paths, with the interfacial α-helices or β-strands separating in either order or concomitantly.

We integrate the flux along the committor to estimate the rate, as described in Methods. As we vary the lag time, the estimate appears to converge to the order of $10\,\mathrm{s}^{-1}$ for the rate, or 100 ms for its inverse (Figure 4), markedly slower than both experimental temperature-jump studies performed in ethanol[21,22] and previous computational estimates.[7] We discuss these studies further in the **Discussion**.

As a first step toward understanding the factors that contribute to the rate, we plot the potential of mean force as a function of the committor in Figure 6a (top left). There is a barrier of about 13 kcal/mol at $q_B \approx 0.14$ (blue shading), and the potential of mean force is relatively flat between $q_B \approx 0.14$ and $q_B \approx 0.9$. We relate these features to CVs in the following section, where we also describe how information is represented within the plots in Figure 6a.

The committor correlates with the total number of interfacial contacts

With a view toward structurally interpreting the committor, we computed its correlation with a large number of CVs (Figure 5).[23,27,45] We found that the CVs used to control sampling in previous studies $\bar{\alpha}_\mathrm{c}$ and $\bar{\beta}_\mathrm{c}$ in Ref. 15 and the center-of-mass distance, R_com, and number of C^α intermonomer contacts, Q, in Ref. 17 are highly correlated with the com-

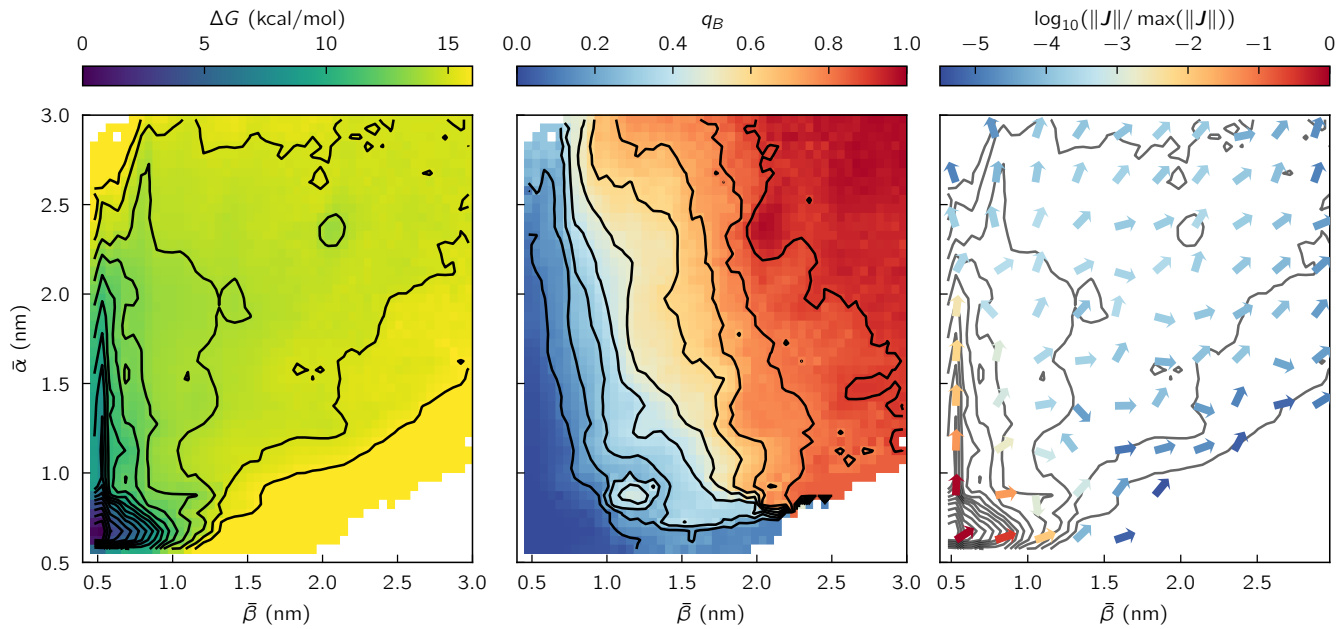

Figure 3: Projections of reactive statistics. The potential of mean force ΔG (left), committor q_B (middle), and reactive current $\boldsymbol{J} = (J^{\bar{\beta}}_{AB}, J^{\bar{\alpha}}_{AB})$ (right, arrows; contours reproduce ΔG from left) as functions of $\bar{\alpha}$ and $\bar{\beta}$. (left and right) Contours are drawn every 1 kcal/mol. (middle) Contours are drawn every 10% increase of probability.

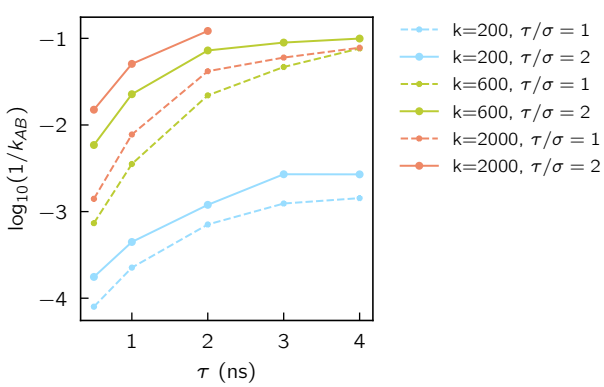

Figure 4: Dependence of inverse reaction rate estimates (in s) on hyperparameters: the number of Markov states, k, the lag time, τ, and the number of memory terms, $\tau/\sigma - 1$, where σ is the time interval used to compute the memory terms.

mittor. In addition to those CVs (and the closely related CVs $\bar{\alpha}$ and $\bar{\beta}$), the CVs with the highest correlation include the total number of intermonomer contacts, N_{c}, the total number of water molecules at the dimer interface, N_{w}, which increases as the dimer interface becomes exposed, and the number of shared water molecules at the dimer interface, N_{sw}, which decreases as the monomers separate.

Although the correlation coefficient is useful for identifying CVs that track monotonically with the committor, it can obscure nonmonotonic relationships, as well as more subtle relationships (e.g., how quickly a variable changes as the reaction progresses). Consequently, we also show plots in which we plot the average value of a variable for each Markov state as a function of its committor value; we term these evolution profiles. To highlight Markov states with the highest fluxes, we make the symbol sizes proportional to the fluxes, and we also indicate the flux-weighted quantile of the variable for the Markov states at each value of the committor (using bins of 0.1; see Figures 6 and S5).

Figure 6a shows the evolution profile for N_{c}, which is strongly correlated with the committor. There are two phases: there is a relatively rapid drop in N_{c} from the dimer state to

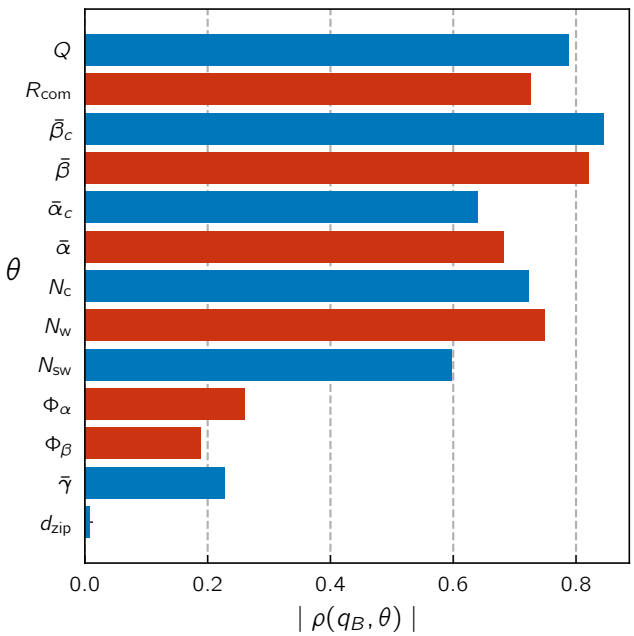

Figure 5: Pearson correlation coefficient between the committor and physical CVs of interest over the entire data set (18,816,000 structures). Red and blue colors correspond to positive and negative correlation coefficients, respectively.

$q_B \approx 0.14$, followed by a slow, steady decrease to zero at $q_B = 1$. The trends in the N_{c} profile are mirrored in the N_{w} profile (Figure S5); the two CVs track each other almost perfectly because water molecules fill the space vacated by protein (Figure S6). The initial rapid phase in the N_{c} profile coincides with the rise in free energy, while the subsequent slow phase coincides with the plateau in free energy.

To dissect this behavior, we separate the intermonomer contacts into native and nonnative ones. To determine the native contacts, we examine the percentage of structures in the dimer state that each (atom-atom) contact is occupied. Because we find there is no natural cutoff for these occupations, we take the native contacts to be those occupied in at least 1% of the structures in the dimer state (Figures 6b and S7; other cutoffs gave qualitatively similar results). We see that the drop in N_{c} from the dimer state to $q_B \approx 0.14$ results from an even more rapid drop in native contacts offset by a rise in nonnative contacts. While most of the rise in nonnative contacts is in this initial phase, the number of nonnative contacts peaks around $q_B \approx 0.5$, which we define as the transition state since states with this committor value have an equal probability of next going to the dimer and monomer states. Thus, the transition state is an encounter complex with about 20 native contacts and 50 nonnative contacts. However, this simple description masks structural diversity that we elaborate in the next section.

The diversity of pathways can be understood in terms of three axes of rotation based on the dimer structure

Given that the CVs that track closely with the committor are compatible with a diversity of structures at each stage of the reaction[15] (Figures 3, S4, and S5), we wanted to determine whether insulin dissociates by distinguishable pathways, or the reactive path ensemble is better viewed as a continuum. Answering this question and, more generally, developing a structural understanding of the mechanism can aid in interpreting the effects of mutations and designing new insulin analogues, as well as connecting our results to structural ideas about molecular recognition.[15,30,46] To this end, we sought to identify structural CVs that are orthogonal to the committor and to use them to characterize the reactive path ensemble.

Based on the results of previous studies,[15,17,20] we considered CVs that track twisting of the interfacial secondary structure elements. The twisting of the α-helices relative to each other can be quantified by a pseudodihedral angle, Φ_α, defined by the geometric centers of the backbone atoms of segments $\mathrm{Ser}^{\mathrm{B9}}$–$\mathrm{Leu}^{\mathrm{B11}}$, $\mathrm{Val}^{\mathrm{B12}}$–$\mathrm{Tyr}^{\mathrm{B16}}$, $\mathrm{Val}^{\mathrm{B'12}}$–$\mathrm{Tyr}^{\mathrm{B'16}}$, and $\mathrm{Ser}^{\mathrm{B'9}}$–$\mathrm{Leu}^{\mathrm{B'11}}$. Similarly, the twisting of the β-strands relative to each other can be quantified by a pseudodihedral angle, Φ_β, defined by the C^α atoms of $\mathrm{Tyr}^{\mathrm{B26}}$, $\mathrm{Phe}^{\mathrm{B25}}$, $\mathrm{Phe}^{\mathrm{B'25}}$, and $\mathrm{Tyr}^{\mathrm{B'26}}$. Note that these definitions differ slightly from those in Ref. 15, so that the dimer state corresponds to the *trans* state of the pseudodihedral angle in each case. These CVs are correlated ($\rho(\Phi_\alpha, \Phi_\beta) = 0.65$; Figure S8) and characterize rotations around axes that

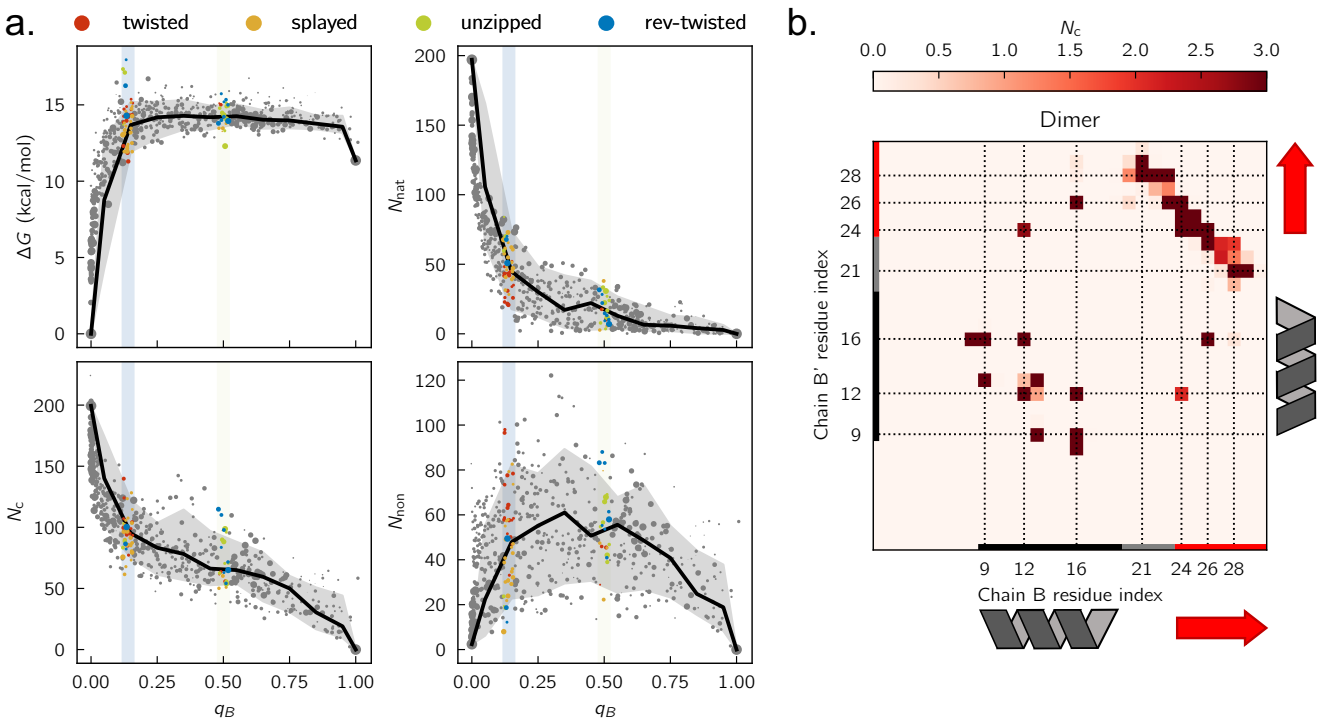

Figure 6: Comparison of the free energy profile to the evolution of the number of contacts. **a.** CVs as functions of the committor: (top left) free energy, (bottom left) total number of intermonomer contacts, (top right) native contacts, and (bottom right) nonnative contacts. The size of each symbol is proportional to the reactive flux passing through the associated Markov state. Blue and yellow bars mark $q_B \approx 0.14$ and $q_B \approx 0.5$. Symbol colors at $q_B \approx 0.14$ and $q_B \approx 0.5$ indicate coarse-grained state assignments. We divide the horizontal axes into intervals of 0.1 and sort the states within each interval by their vertical axis values; the black line marks the value corresponding to 50% of the flux, and the gray shaded region shows the [10%, 90%] interval of the flux (i.e., 40% on either side of the line). **b**. Average residue-residue contact map of the dimer state. See Figure S7 for the average atom-atom contact map.

are orthogonal to the dimer interface, parallel to the line connecting the centers-of-mass of the monomers (labeled the y-axis in Figure 7).

Motivated by this observation, we defined two other axes based on the dimer structure. We defined the x-axis as the C_2 rotation axis of the dimer and the z-axis as the cross product of the x- and y-axes, which is parallel to the β-strands. We then defined CVs that characterize rotations around the x- and z-axes. Because the monomers are not rigid bodies, one might expect more than one variable to be required to define the dynamics around each axis (as is the case for the y-axis, where we track both Φ_α and Φ_β). Nevertheless, we found that only one additional CV for each axis was sufficient. To describe the rotation around the z-axis, we define three angles: γ is the angle of the geometric centers of the segments $\mathrm{Val}^{\mathrm{B12}}$–$\mathrm{Tyr}^{\mathrm{B16}}$, $\mathrm{Phe}^{\mathrm{B25}}$, and $\mathrm{Phe}^{\mathrm{B'25}}$; γ' is the corresponding angle for the other monomer; and $\bar{\gamma} = \gamma + \gamma'$. To describe the rotation around the x-axis, we combine four C^α-C^α distances: $d_{\mathrm{zip}} = (d_{\mathrm{B21-B'29}} + d_{\mathrm{B16-B'26}}) - (d_{\mathrm{B29-B'21}} + d_{\mathrm{B26-B'16}})$. Figure 7 illustrates the axes and the CVs that characterize rotation around them.

We clustered the Markov states into coarse-grained states based on Φ_α, Φ_β, $\bar{\gamma}$, and $|d_{\mathrm{zip}}|$. By varying the number of clusters, we found that four coarse-grained states were sufficient to capture the qualitative features of the pathways. Each is distinguished from the others by a shift in one of the CVs relative to its value in the dimer state: the twisted state has low Φ_α, the splayed state has high $\bar{\gamma}$, the unzipped state has high $|d_{\mathrm{zip}}|$, and the rev-twisted state has high Φ_β (Figure S8). However, we emphasize that each such shift can be accompanied by others owing to correlations between the CVs.

Figures 7, S9, and S10 show structures of the coarse-grained states at different stages of the reaction, Figures 8 and S7 show their contact maps, and Figure S11 shows their solvation. At $q_B \approx 0.14$, both the twisted and splayed states exhibit a rotation in the positive direction around the y-axis and in turn relatively low Φ_α and Φ_β (Figure S8; see Figure S9 for an alternative view of the splayed state that makes clear the $+y$-rotation). This rotation shifts contacts of the C-terminal residues (B26–B30) of each monomer from the β-turn (B′21) to the C-terminal end of the α-helix (B′16) of the other monomer in a symmetric fashion. By contrast, the unzipped and rev-twisted states show detachment of the B′26–B′30 residues of one monomer (in Figures 7 and 8 we show data only for states with $d_{\mathrm{zip}} > 0$ to make this asymmetry apparent; states with $d_{\mathrm{zip}} < 0$ are similar but the monomers are switched).

The twisted and splayed states are differentiated by whether they exhibit further rotation around the y-axis or rotation around the z-axis. In the twisted state, further rotation around the y-axis leads to contacts between B26–B30 and B′12–B′13 in the interfacial α-helix of the other monomer, and it brings the N-termini of the interfacial α-helices together to make a B9–B′9 contact. In the splayed state, rotation around the z-axis results in a loss of contacts between the interfacial α-helices. Similarly, the unzipped state and rev-twisted states (which again occupy the same region of the $\Phi_\alpha\Phi_\beta$-plane; Figure S8) are differentiated by the rotation around the x-axis. In the unzipped state, this rotation results in a complete loss of contacts for the detached B′26–B′30 residues at $q_B \approx 0.14$.

At $q_B \approx 0.5$, the coarse-grained states are more heterogeneous. In the twisted state, only contacts between the β-turn region (B16 and B21) and the C-terminus (B′26–B′30) or the N-terminus of the α-helix (B′9) of the other monomer persist. In the splayed state, the contacts between the interfacial α-helices are completely lost, and one of the two groups of contacts between the C-terminal residues (B26–B30) and the β-turn (B21) is broken while the other remains intact, mirroring the effects of the x-rotation discussed above for the unzipped state at $q_B \approx 0.14$. Notably, both the twisted and splayed states show asymmetric contact maps.

In the unzipped state at $q_B \approx 0.5$, the C-terminus (B26–B30) of one monomer (left in Figure 7) inserts between the β-trun side of α-helix (B′16) and that of β-strand residues (B′24) of the other monomer, bringing B28–B30 in contact with the β-turn (B′21–B′23) as well.

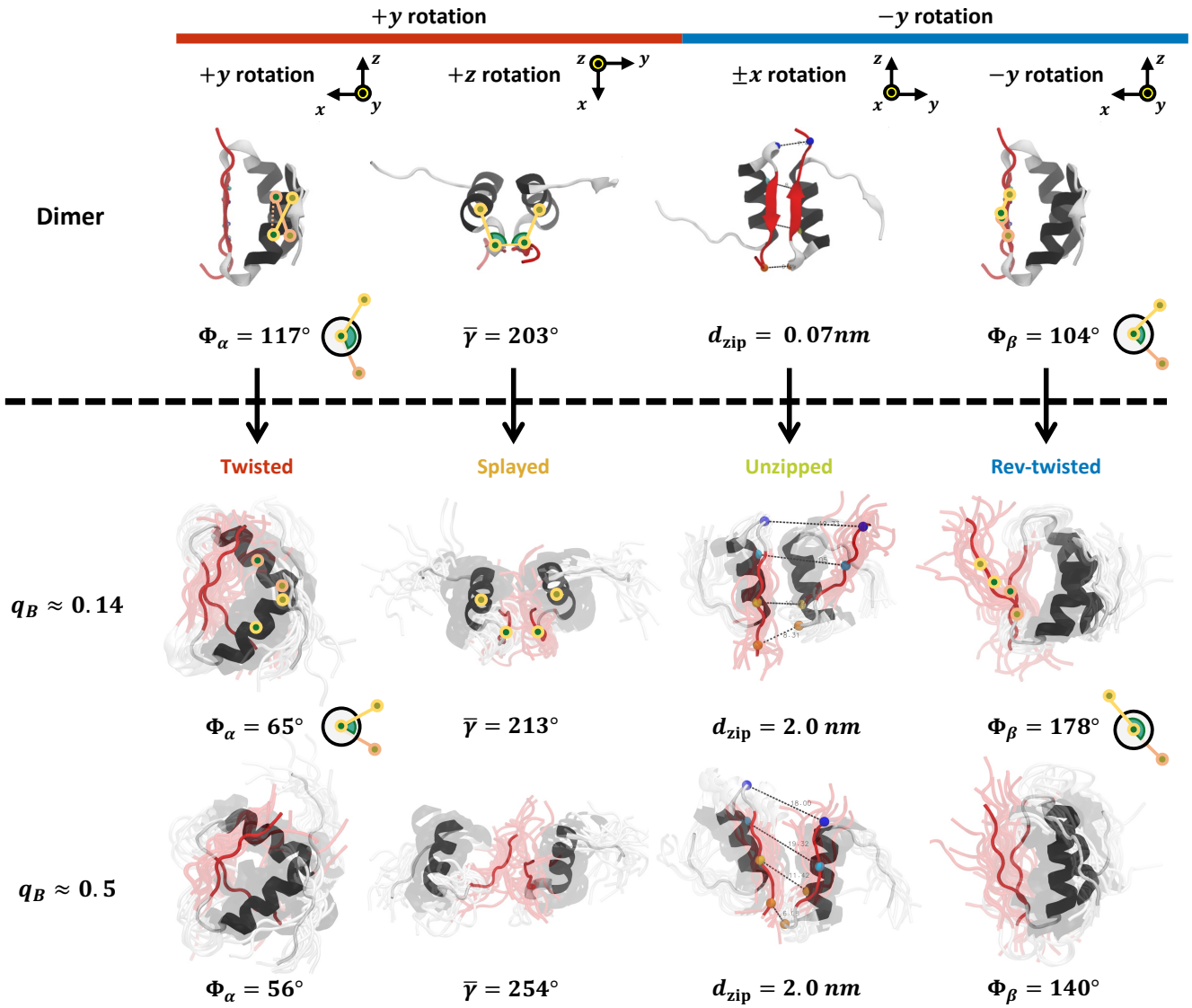

Figure 7: Coarse-grained states. (top row) Collective variables used to define the states and their associated rotations. (bottom two rows) Representative structures at $q_B \approx 0.14$ and $q_B \approx 0.5$ as indicated. To obtain the structures shown, we randomly sampled ten structures (translucent) from each of the Markov states associated with each coarse-grained state and then iteratively aligned the structures to their average structure until the backbone RMSDs to the average structure (opaque) converged. See Figures S9 and S10 for alternative views of the coarse-grained states.

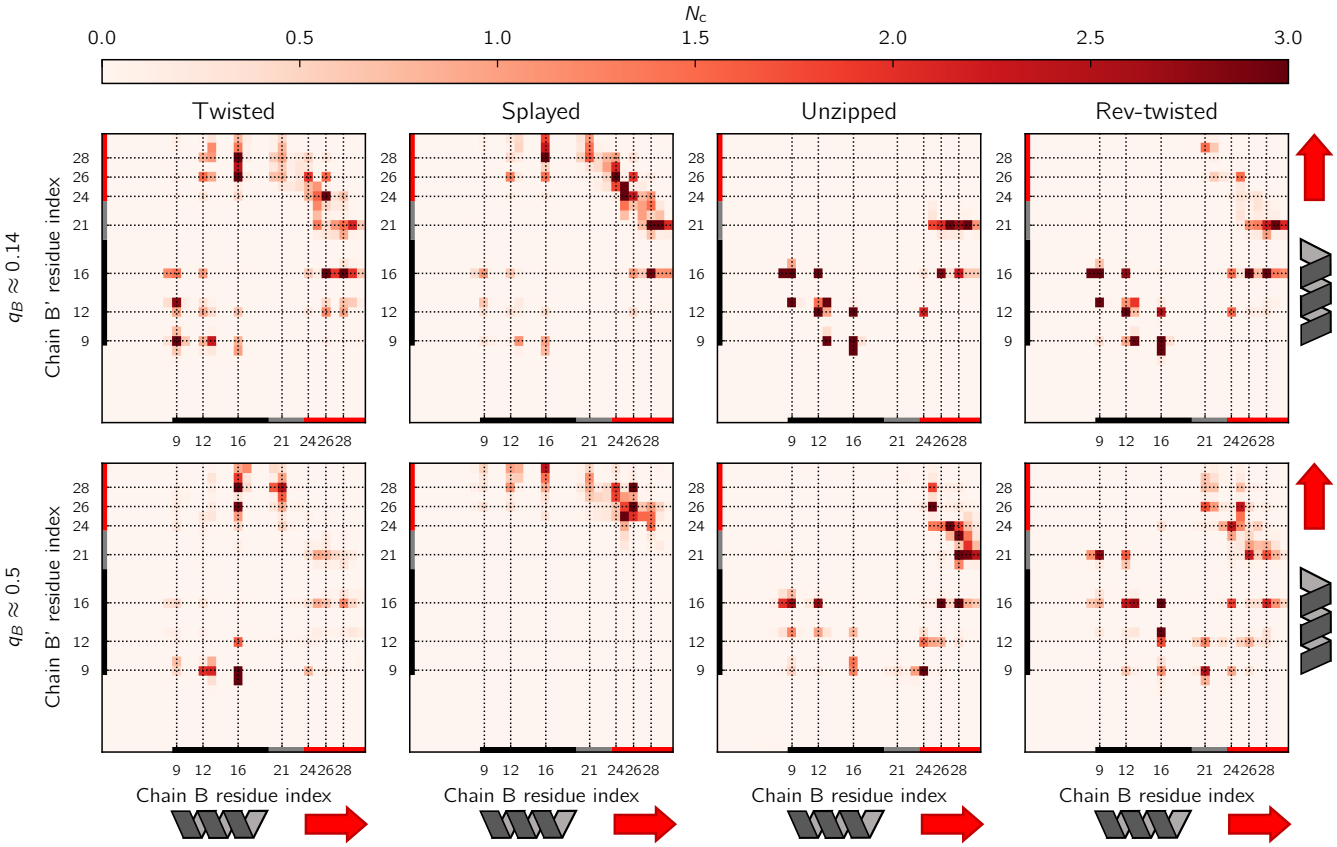

Figure 8: Average residue-residue contact map of each of the coarse-grained states. See Figure S7 for the average atom-atom contact map.

The other monomer's C-terminus (B′24–B′28) is heterogeneous in conformation but samples contacts with B25 of the former. The rev-twisted state rotates the opposite way around the y-axis as the twisted state, and this results in a nonnative contact between the C-termini of the interfacial α-helices at B16–B′16 (in contrast to the B9–B′9 contact for the twisted state at $q_B \approx 0.14$) as well as nonnative contacts that arise from the β-turn (B21) of one monomer inserting between the C-terminal part of the β-strand (B′26-B′28) and the N-terminal part of the α-helix (B′9 and B′12) of the other monomer. Notably, the contact map of the rev-twisted state at $q_B \approx 0.5$ is symmetric.

There is extensive mixing between selected pathways

To quantify the contributions of the pathways, we computed the fractional fluxes through the coarse-grained states as described in Methods (Figure 9a). At $q_B \approx 0.14$, the coarse-grained states that involve $+y$ rotation (splayed and twisted) make larger contributions to the net flux than the coarse-grained states that involve $-y$ rotation (unzipped and rev-twisted). This reflects the fact that the former's free energies are lower (compare the red and orange symbols with the blue and green ones at $q_B \approx 0.14$ in Figure 6a). At $q_B \approx 0.5$, the ordering is reversed, presumably owing to mixing between pathways.

To characterize the mixing, we computed the coarse-grained transition flux matrix as described in Methods (Figure 9c). One can consider both mixing within each committor value ($q_B \approx 0.14$ and $q_B \approx 0.5$) and from each to the other. We focus first on the transitions from $q_B \approx 0.14$ to $q_B \approx 0.5$ (Figure 9b and red box in Figure 9c). The coarse-grained states that involve only rotations around the y-axis (twisted and rev-twisted states) flow from $q_B \approx 0.14$ to $q_B \approx 0.5$ without branching. By contrast, the states that involve multiple rotation axes (splayed state with $+y$ and $+z$ rotation and un-

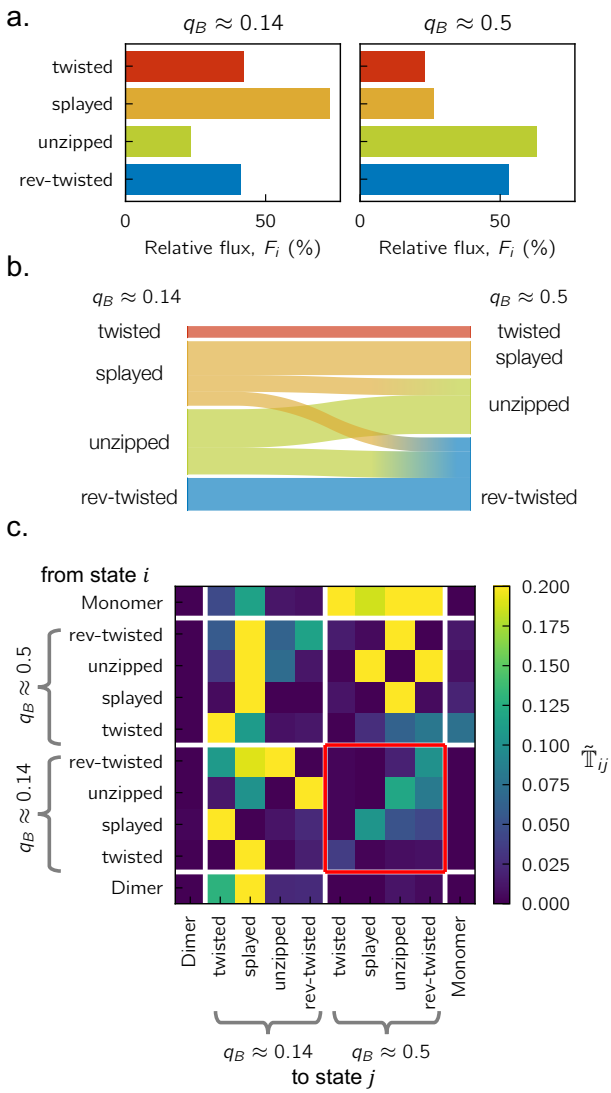

Figure 9: Quantifying the contributions of pathways and their mixing. **a**. Fractional contribution of the coarse-grained states to the net flux. **b**. Sankey diagram of mixing between coarse-grained states. The width of each bar is proportional to the corresponding element in the transition flux matrix $\tilde{\mathbb{T}}_{ij}$. **c**. Transition flux matrix between the coarse-grained states. Red box indicates the entries used to plot the Sankey diagram in **b**.

zipped state with $-y$ and $+x$ rotation) branch to other states. Furthermore, there is flow from the $+y$-rotation states to the $-y$-rotation states but not vice versa, consistent with the reordering in Figure 9a.

Within $q_B \approx 0.14$ (lower left block of Figure 9), states with the same direction of y rotation exchange more readily with each other than with states with the opposite direction of y rotation. Interestingly, the twisted and splayed states with $+y$ rotation at $q_B \approx 0.14$ are more likely to proceed to $q_B \approx 0.5$ (red box) than to exchange with the unzipped and rev-twisted states with $-y$ rotation at $q_B \approx 0.14$. On the other hand, the states with $-y$ rotation show comparable flux to $q_B \approx 0.5$ and exchange with the states with $+y$ rotation at $q_B \approx 0.14$. At $q_B \approx 0.5$ (upper right block of Figure 9), the twisted state appears relatively isolated, while the remaining three states are connected via the unzipped state, which shares structural features with both the splayed state ($+x$ rotation) and the rev-twisted state ($-y$ rotation).

Put together, our analysis of the structures and fluxes shows that pathways can be distinguished but there is extensive mixing between them, both within each committor value and from one committor value to another. This mixing can be rationalized in terms of the rotation axes that we define based on the dimer structure.

Discussion

Here, we construct an MSM to characterize insulin dimer dissociation statistically. The heterogeneity of pathways that we document makes this a challenging reaction to study. One key aspect of our approach was initializing the unbiased dynamics simulations used to construct the MSM from structures drawn from converged umbrella sampling simulations.[15,31] Using a framework that generalizes MSM to solve operator equations with statistic-specific boundary conditions,[25,26] we computed the equilibrium probabilities and committor values of Markov states and the fluxes through them. From these we obtained a rate and a compre-

hensive view of the dissociation mechanism.

The estimated timescale of dissociation (approximately 100 ms) is two orders of magnitude slower than the 250–1000 µs timescale ascribed to monomer separation in T-jump studies, which use both elevated temperature and ethanol to accelerate dissociation,[21,22] in contrast to the simulations. Accordingly, the dissociation constant was shown to decrease by two orders of magnitude in 20% EtOD,[47] which would be sufficient to account for the difference in k_{AB} (under the assumption that k_{BA} does not change). The 2DIR experiments, moreover, measure relaxation of a spectral feature assigned to interfacial β-sheet disruption, which would be considered an intermediate state in the simulations.

Our estimated rate is also markedly slower than estimates from Bagchi and co-workers using TST-based theories. The higher barrier that we obtain (compared with their reported values of 7.6–9.3 kcal/mol[7,17]) likely accounts for much of the difference. At the same time, the corrections to TST that they employ to account for dynamics orthogonal to the reaction coordinate are also unlikely to be able to fully capture the dynamics that we describe. Indeed, as the resolution of the model (number of Markov states) becomes lower in Figure 4, we estimate the rate to be faster. TST-based theories applied to a two-dimensional potential of mean force resolve far fewer details than even our coarsest MSM (200 states) and are correspondingly faster still.

We found that the committor correlates closely with CVs that quantify the total number of intermonomer contacts and the total solvent exposure of the dimer interface. Structures in the transition state ensemble, despite their diversity, generally have a relatively small percentage (<20%) of native contacts and a highly hydrated interface, consistent with general trends obtained from long, unbiased simulations[18] and earlier experimental measurements.[1,48,49] The loss of native contacts at $q_B \approx 0.14$ frees the participating residues to make new contacts, and this accounts for the concommitant gain of nonnative contacts (Figure 6a and 8). The fact that we do not see contacts involving residues that are far from the dimer interface supports the idea that encounter complexes can successfully associate only if they first make contact near the native interface,[4,18,50] but our sampling of states that are very different from the dimer may also be incomplete.

We can quantitatively compare the likelihood of association for different encounter complexes based on their contacts. For example, as their committor values indicate, the rev-twisted state at $q_B \approx 0.5$ is more likely to dissociate (or less likely to associate) than the twisted state at $q_B \approx 0.14$ even though their numbers of native and nonnative contacts are comparable (Figure 6a). This makes clear that details beyond the numbers of native and nonnative contacts are important.

Consistent with this idea, we identified four coarse-grained states that define dissociation pathways. We showed that these pathways can be understood in terms of three orthogonal axes of rotation based on the dimer structure. Along all pathways, the initial event is a rotation around the y-axis (Figure 7), consistent with twisted states previously reported.[15,17,19,20] Subsequently, $+z$-rotations can lead to splaying of the interfacial α-helices and x-rotations can lead to unzipping of the interfacial β-strands. These dynamics are consistent with low-frequency normal modes.[19]

Despite the diversity of pathways that we observe, overall, less disorder in the separated monomers than an earlier study of the insulin monomer[10] and as suggested by experiments.[21,47,51] However, these studies employed mutations and/or low pH to limit self-association. The only significant disorder that we observe over the course of the reaction is the detachment of the C-terminal β-strand (as measured by Ψ'_{d}, following Ref. 15, and $\Psi_{\mathrm{B'25B'30d}}$, following Ref. 10; shown in Figures S5 and S12, respectively). In Ref. 15, the detachment angle was averaged over the monomers and appeared to vary more over the α-path when projected on $\bar{\alpha}$ and $\bar{\beta}$. In our case, we resolve the contributions from the monomers. While we see evidence of detachment in both Ψ_{d} and Ψ'_{d} along the pathways involving $+y$-rotation (i.e., the

twisted and splayed states; Figure S5), which we associate with the α-path, we observe that Ψ'_{d} (Figure S5) and $\Psi_{\mathrm{B'25B'30d}}$ (Figure S12) exhibit extreme values at $q_B \approx 0.14$ along the pathways involving $-y$-rotation (i.e., the unzipped and rev-twisted states), which we associate with the β-path; the latter feature may be obscured when averaged. Interestingly, consistent with Ref. 15, the detachment is more pronounced at intermediate stages of the reaction than in the monomer state.

As the total number of contacts decreases over the course of the reaction, the number of water molecules in contact with the monomers increases. At no point do we see any evidence of a dewetting transition (Figure S6), as observed for some systems,[52,53] including the insulin dimer.[54] In the case of the twisted and splayed states, the initial $+y$-rotation brings $\mathrm{Lys}^{\mathrm{B29}}$ close to the hydrophobic core ($\mathrm{Val}^{\mathrm{B12}}$, $\mathrm{Tyr}^{\mathrm{B16}}$, $\mathrm{Phe}^{\mathrm{B24}}$, and $\mathrm{Tyr}^{\mathrm{B26}}$, and the corresponding residues from the other monomer); in the case of the unzipped and rev-twisted states, $-y$ rotation promotes β-strand detachment. Both the presence of charged residues and the absence of an extended hydrophobic surface are expected to suppress dewetting.[55–57] That said, dewetting may be very sensitive to details of the simulation conditions, and our choice of features may bias our analysis towards solute rather than solvent dynamics;[58] further analysis of the solvent in our simulations is warranted in the future.

The dynamics of B28-B29 are also important for understanding fast-acting therapeutic insulin analogues. Insulin lispro[59] ($\mathrm{Pro}^{\mathrm{B28}}\rightarrow$Lys and $\mathrm{Lys}^{\mathrm{B29}}\rightarrow$Pro) and insulin aspart[60] ($\mathrm{Pro}^{\mathrm{B28}}\rightarrow$Asp) are thought to accelerate dimer dissociation by destabilizing C-terminal native contacts.[61–63] In our simulations, we observe that these residues make many nonnative contacts as the reaction progresses, suggesting that a full understanding of the effects of these mutations on the kinetics requires consideration of nonnative interactions as well. A further complication is that mutations can shift the importance of competing pathways, as previously shown for dissociation of phenol from the insulin hexamer.[30] We expect such a shift to be possible in the case of dimer dissociation given the extensive mixing that we observe between pathways in the present study.

The discussion above points to the importance of being able not only to estimate the rate accurately but also the fluxes associated with pathways. In the present work, we quantify the flux through and exchange between pathways by redefining the boundary conditions when constructing the transition matrix, but history-dependent committors and reactive currents can be defined formally.[44] It would be interesting to analyze insulin dimer dissociation within that framework. Nevertheless, the present study, along with others,[26,27,30,40,64,65] shows that the theoretical frameworks and computational tools needed to treat complex reactions with many competing pathways quantitatively are now taking shape.

Acknowledgement The authors thank Adam Antozewski, Chatipat Lorpaiboon, John Strahan, Andrei Tokmakoff, and Jonathan Weare for useful discussions. This work was supported primarily by National Institutes of Health (NIH) award R35 GM136381. S.C.G. acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. 2140001. This work was completed with computational resources administered by the University of Chicago Research Computing Center, including Beagle-3, a shared GPU cluster for biomolecular sciences supported by the NIH under the High-End Instrumentation (HEI) grant program award 1S10OD028655-0.

Supporting Information Available

Extended discussion and supporting tables and figures.

References

(1) Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental aspects of protein-protein

association kinetics. *Chemical Reviews* **2009**, *109*, 839–860.

(2) Northrup, S. H.; Allison, S. A.; McCammon, J. A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *The Journal of Chemical Physics* **1984**, *80*, 1517–1524.

(3) Gabdoulline, R. R.; Wade, R. C. Simulation of the diffusional association of barnase and barstar. *Biophysical Journal* **1997**, *72*, 1917–1929.

(4) Plattner, N.; Doerr, S.; Fabritiis, G. D.; Noé, F. Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature Chemistry* **2017**, *9*, 1005–1011.

(5) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-rate theory: Fifty years after Kramers. *Reviews of Modern Physics* **1990**, *62*, 251–341.

(6) Peters, B. *Reaction rate theory and rare events*; Elsevier, 2017.

(7) Acharya, S.; Mondal, S.; Mukherjee, S.; Bagchi, B. Rate of Insulin Dimer Dissociation: Interplay between Memory Effects and Higher Dimensionality. *Journal of Physical Chemistry B* **2021**, *125*, 9678–9691.

(8) Hua, Q.-x.; Jai, W.; Weiss, M. A. Conformational Dynamics of Insulin. *Frontiers in Endocrinology* **2011**, *2*.

(9) Weiss, M. A.; Hua, Q. X.; Lynch, C. S.; Frank, B. H.; Shoelson, S. E. Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective deuterium and carbon-13 labeling with application to protein design. *Biochemistry* **1991**, *30*, 7373–7389.

(10) Busto-Moner, L.; Feng, C. J.; Antoszewski, A.; Tokmakoff, A.; Dinner, A. R. Structural Ensemble of the Insulin Monomer. *Biochemistry* **2021**, *60*, 3125–3136.

(11) Zoete, V.; Meuwly, M.; Karplus, M. A Comparison of the Dynamic Behavior of Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active Monomer. *Journal of Molecular Biology* **2004**, *342*, 913–929.

(12) Gorai, B.; Vashisth, H. Progress in Simulation Studies of Insulin Structure and Function. *Frontiers in Endocrinology* **2022**, *13*.

(13) Sakabe, N.; Sakabe, K.; Sasaki, K.; Murayoshi, M. 0.92A structure of 2Zn human insulin at 100K. www.rcsb.org/structure/3w7y, 2013.

(14) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38, 27–28.

(15) Antoszewski, A.; Feng, C. J.; Vani, B. P.; Thiede, E. H.; Hong, L.; Weare, J.; Tokmakoff, A.; Dinner, A. R. Insulin Dissociates by Diverse Mechanisms of Coupled Unfolding and Unbinding. *Journal of Physical Chemistry B* **2020**, *124*, 5571–5587.

(16) Zoete, V.; Meuwly, M.; Karplus, M. Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 79–93.

(17) Banerjee, P.; Mondal, S.; Bagchi, B. Insulin dimer dissociation in aqueous solution: A computational study of free energy landscape and evolving microscopic structure along the reaction pathway. *Journal of Chemical Physics* **2018**, *149*.

(18) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-level characterization of protein–protein association. *Proceedings of the National Academy of*

Sciences of the United States of America **2019**, *116*, 4244–4249.

(19) Mondal, S.; Mukherjee, S.; Acharya, S.; Bagchi, B. Unfolding of Dynamical Events in the Early Stage of Insulin Dimer Dissociation. *Journal of Physical Chemistry B* **2021**, *125*, 7958–7966.

(20) Feng, C. J.; Sinitskiy, A.; Pande, V.; Tokmakoff, A. Computational IR Spectroscopy of Insulin Dimer Structure and Conformational Heterogeneity. *Journal of Physical Chemistry B* **2021**, *125*, 4620–4633.

(21) Zhang, X. X.; Jones, K. C.; Fitzpatrick, A.; Peng, C. S.; Feng, C. J.; Baiz, C. R.; Tokmakoff, A. Studying Protein-Protein Binding through T-Jump Induced Dissociation: Transient 2D IR Spectroscopy of Insulin Dimer. *Journal of Physical Chemistry B* **2016**, *120*, 5134–5145.

(22) Rimmerman, D.; Leshchev, D.; Hsu, D. J.; Hong, J.; Kosheleva, I.; Chen, L. X. Direct Observation of Insulin Association Dynamics with Time-Resolved X-ray Scattering. *Journal of Physical Chemistry Letters* **2017**, *8*, 4413–4418.

(23) Ma, A.; Dinner, A. R. Automatic Method for Identifying Reaction Coordinates in Complex Systems. *The Journal of Physical Chemistry B* **2005**, *109*, 6769–6779.

(24) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the transition coordinate for protein folding. *The Journal of chemical physics* **1998**, *108*, 334–350.

(25) Thiede, E. H.; Giannakis, D.; Dinner, A. R.; Weare, J. Galerkin approximation of dynamical quantities using trajectory data. *Journal of Chemical Physics* **2019**, *150*.

(26) Strahan, J.; Antoszewski, A.; Lorpaiboon, C.; Vani, B. P.; Weare, J.; Dinner, A. R. Long-Time-Scale Predictions from Short-Trajectory Data: A Benchmark Analysis of the Trp-Cage Miniprotein. *Journal of Chemical Theory and Computation* **2021**, *17*, 2948–2963.

(27) Guo, S. C.; Shen, R.; Roux, B.; Dinner, A. R. Dynamics of activation in the voltage-sensing domain of Ciona intestinalis phosphatase Ci-VSP. *Nature Communications* **2024**, *15*.

(28) Wu, H.; Nüske, F.; Paul, F.; Klus, S.; Koltai, P.; Noé, F. Variational Koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *The Journal of Chemical Physics* **2017**, *146*, 154104.

(29) Marchi, M.; Ballone, P. Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems. *The Journal of Chemical Physics* **1999**, *110*, 3697–3702.

(30) Antoszewski, A.; Lorpaiboon, C.; Strahan, J.; Dinner, A. R. Kinetics of Phenol Escape from the Insulin R_6 Hexamer. *Journal of Physical Chemistry B* **2021**, *125*, 11637–11649.

(31) Thiede, E. H.; Van Koten, B.; Weare, J.; Dinner, A. R. Eigenvector method for umbrella sampling enables error analysis. *The Journal of Chemical Physics* **2016**, *145*, 084115.

(32) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* **2017**, *14*, 71–73.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(34) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.;

Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13*, e1005659.

(35) Zhang, Z.; Liu, X.; Yan, K.; Tuckerman, M. E.; Liu, J. Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. *The Journal of Physical Chemistry A* **2019**, *123*, 6056–6079.

(36) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327–341.

(37) Gumbart, J. C.; Roux, B.; Chipot, C. Standard binding free energies from computer simulations: What is the best strategy? *Journal of Chemical Theory and Computation* **2013**, *9*, 794–802.

(38) Michaud-Agrawal, N.; Denning, E.; Woolf, T.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of computational chemistry* **2011**, *32*.

(39) Richard J. Gowers,; Max Linke,; Jonathan Barnoud,; Tyler J. E. Reddy,; Manuel N. Melo,; Sean L. Seyler,; Jan Domański,; David L. Dotson,; Sébastien Buchoux,; Ian M. Kenney, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016; pp 98–105.

(40) Lorpaiboon, C.; Guo, S. C.; Strahan, J.; Weare, J.; Dinner, A. R. Accurate estimates of dynamical statistics using memory. *The Journal of Chemical Physics* **2024**, *160*, 084108.

(41) E, W.; Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annual review of physical chemistry* **2010**, *61*, 391–420.

(42) Vani, B. P.; Weare, J.; Dinner, A. R. Computing transition path theory quantities with trajectory stratification. *The Journal of Chemical Physics* **2022**, *157*.

(43) Miron, P.; Beron-Vera, F. J.; Helfmann, L.; Koltai, P. Transition paths of marine debris and the stability of the garbage patches. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2021**, *31*, 033101.

(44) Lorpaiboon, C.; Weare, J.; Dinner, A. R. Augmented transition path theory for sequences of events. *The Journal of Chemical Physics* **2022**, *157*, 094115.

(45) Hu, J.; Ma, A.; Dinner, A. R. A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proceedings of the National Academy of Sciences* **2008**, *105*, 4615–4620.

(46) Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **2010**, *35*, 539–546.

(47) Ganim, Z.; Jones, K. C.; Tokmakoff, A. Insulin dimer dissociation and unfolding revealed by amide i two-dimensional infrared spectroscopy. *Physical Chemistry Chemical Physics* **2010**, *12*, 3579–3588.

(48) Schreiber, G. Kinetic studies of protein–protein interactions. *Current Opinion in Structural Biology* **2002**, *12*, 41–47.

(49) Frisch, C.; Fersht, A. R.; Schreiber, G. Experimental assignment of the structure of the transition state for the association of barnase and barstar1. *Journal of Molecular Biology* **2001**, *308*, 69–77.

(50) Tang, C.; Iwahara, J.; Clore, G. M. Visualization of transient encounter complexes in protein–protein association. *Na-*

ture **2006**, *444*, 383–386, Publisher: Nature Publishing Group.

(51) Huus, K.; Havelund, S.; Olsen, H. B.; van de Weert, M.; Frokjaer, S. Thermal Dissociation and Unfolding of Insulin. *Biochemistry* **2005**, *44*, 11171–11177, PMID: 16101301.

(52) ten Wolde, P. R.; Chandler, D. Drying-induced hydrophobic polymer collapse. *Proceedings of the National Academy of Sciences* **2002**, *99*, 6539–6543, Publisher: Proceedings of the National Academy of Sciences.

(53) Liu, P.; Huang, X.; Zhou, R.; Berne, B. J. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* **2005**, *437*, 159–162.

(54) Banerjee, P.; Bagchi, B. Dynamical control by water at a molecular level in protein dimer association and dissociation. *Proceedings of the National Academy of Sciences* **2020**, *117*, 2302–2308.

(55) Cheng, Y.-K.; Rossky, P. J. The effect of vicinal polar and charged groups on hydrophobic hydration. *Biopolymers* **1999**, *50*, 742–750.

(56) Choudhury, N.; Pettitt, B. M. The dewetting transition and the hydrophobic effect. *Journal of the American Chemical Society* **2007**, *129*, 4847–4852.

(57) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S. Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function. *The Journal of Physical Chemistry B* **2012**, *116*, 2498–2503.

(58) Raddi, R. M.; Voelz, V. A. Markov State Model of Solvent Features Reveals Water Dynamics in Protein-Peptide Binding. *The Journal of Physical Chemistry B* **2023**,

(59) Howey, D. C.; Bowsher, R. R.; Brunelle, R. L.; Woodworth, J. R. [Lys(B28), Pro(B29)]-Human Insulin: A Rapidly Absorbed Analogue of Human Insulin. *Diabetes* **1994**, *43*, 396–402.

(60) Brange, J.; Ribel, U.; Hansen, J. F.; Dodson, G.; Hansen, M. T.; Havelund, S.; Melberg, S. G.; Norris, F.; Norris, K.; Snel, L. et al. Monomeric insulins obtained by protein engineering and their medical implications. *Nature* **1988**, *333*, 679–682.

(61) Brems, D. N.; Alter, L. A.; Beckage, M. J.; Chance, R. E.; DiMarchi, R. D.; Green, L. K.; Long, H. B.; Pekar, A. H.; Shields, J. E.; Frank, B. H. Altering the association properties of insulin by amino acid replacement. *Protein Engineering, Design and Selection* **1992**, *5*, 527–533.

(62) Whittingham, J. L.; Edwards, D. J.; Antson, A. A.; Clarkson, J. M.; Dodson, G. G. Interactions of Phenol and m-Cresol in the Insulin Hexamer, and Their Effect on the Association Properties of B28 Pro → Asp Insulin Analogues. *Biochemistry* **1998**, *37*, 11516–11523, PMID: 9708987.

(63) Chaudhary, Y.; Bhimalapuram, P. Insulin aspart dimer dissociation in water. *The Journal of Chemical Physics* **2022**, *156*, 105106.

(64) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *Journal of the American Chemical Society* **2019**, *141*, 6519–6526.

(65) Strahan, J.; Lorpaiboon, C.; Weare, J.; Dinner, A. R. BAD-NEUS: Rapidly converging trajectory stratification. *The Journal of Chemical Physics* **2024**, *161*, 084109.