

Kinetics of Phenol Escape from the Insulin R₆ Hexamer

Adam Antoszewski, Chatipat Lorpaboon, John Strahan, and Aaron R. Dinner*



Cite This: *J. Phys. Chem. B* 2021, 125, 11637–11649



Read Online

ACCESS |



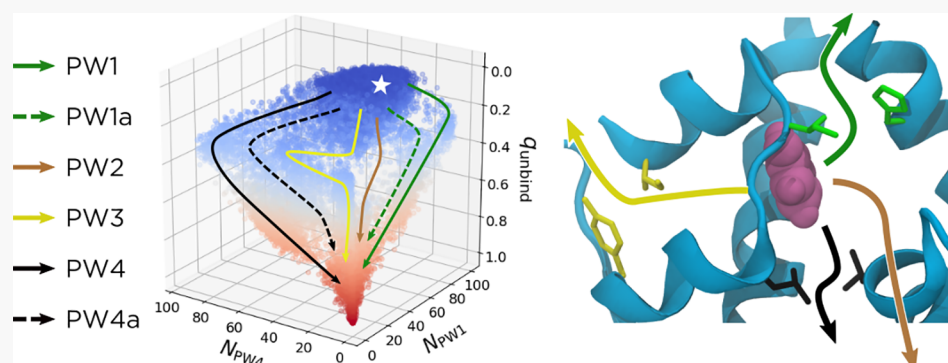
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Therapeutic preparations of insulin often contain phenolic molecules, which can impact both pharmacokinetics and shelf life. Thus, understanding the interactions of insulin and phenolic molecules can aid in designing improved therapeutics. In this study, we use molecular dynamics to investigate phenol release from the insulin hexamer. Leveraging recent advances in methods for analyzing molecular dynamics data, we expand on existing simulation studies to identify and quantitatively characterize six phenol binding/unbinding pathways for wild-type and A10 Ile → Val and B13 Glu → Gln mutant insulins. A number of these pathways involve large-scale opening of the primary escape channel, suggesting that the hexamer is much more dynamic than previously appreciated. We show that phenol unbinding is a multipathway process, with no single pathway representing more than 50% of the reactive current and all pathways representing at least 10%. We use the mutant simulations to show how the contributions of specific pathways can be rationally manipulated. Predicting the net effects of mutations is more challenging because the kinetics depend on all of the pathways, demanding quantitatively accurate simulations and experiments.

INTRODUCTION

The primary therapeutic for diabetes management is the protein hormone insulin. Both the pharmacokinetics and the shelf life of insulin depend on equilibria between conformational and oligomeric states. Insulin is typically a hexamer in therapeutic formulations, a dimer in the blood, and a monomer when bound to its receptor;^{1,2} therapeutic formulations require cold storage and transport to suppress off-pathway equilibria that lead to fibrillation.^{3–5} Much effort is directed toward modulating insulin's equilibria to achieve therapeutics with desired properties.⁶ For example, the widely used fast-acting diabetes therapeutics lispro (Pro^{B28} → Lys^{B28}, and Lys^{B29} → Pro^{B29})⁷ and aspart (Pro^{B28} → Asp^{B28})⁸ destabilize the insulin dimer. In contrast, slow-acting basal insulin analogues like glargine⁹ and detemir^{10,11} function by either decreasing insulin solubility or adsorption in the body.¹² Longer-acting insulin analogues are also being developed,^{13,14} but such formulations are often expensive and complicated. To enable the rational design of improved insulin mutants and analogues, better understanding of insulin equilibria at the molecular level is needed.

In the present study, we focus on the insulin hexamer. The hexamer exists in three conformational states that are designated T₆, T₃R₃, and R₆ for the contributing monomer conformational states.^{15–17} Each monomer consists of a 21-amino acid A chain joined to a 30-amino acid B chain by two interchain disulfide bonds (Cys^{A7}–Cys^{B7} and Cys^{A20}–Cys^{B19}). Each A chain contains an intrachain disulfide bond (Cys^{A6}–Cys^{A11}) and two α helices: Gly^{A1}–Ser^{A9} and Leu^{A13}–Asn^{A21}. Each B chain has a single α helix, and this secondary structure element differs in length depending on whether the monomer is in the T or R state. It spans Ser^{B9}–Cys^{B19} in the T state and Phe^{B1}–Cys^{B19} in the R state.¹⁸ The hexamer can be viewed as a trimer of dimers, with the dimer interface made up of the B chain α helix and a β strand (Phe^{B24}–Tyr^{B26}) from each monomer.

Received: July 22, 2021

Revised: September 22, 2021

Published: October 14, 2021



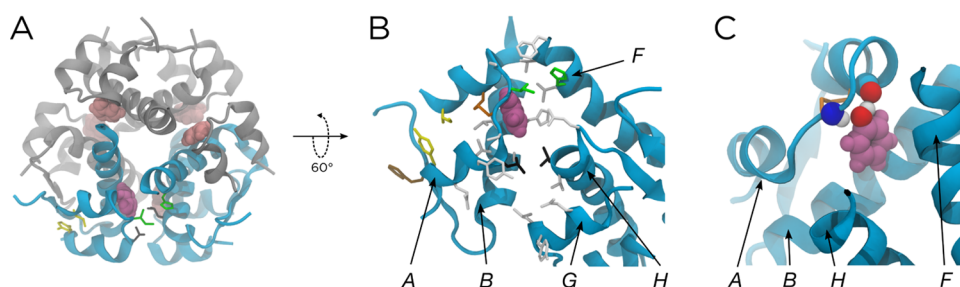


Figure 1. Solvated and equilibrated crystal structure for the R_6 insulin hexamer. (A) Full hexamer, with phenols shown in magenta/pink. The protein is colored to correspond with the other panels. (B) Cyan protein chains from panel (A), labeled as chains A, B, F, G, and H. For a full description of the protein nomenclature, see the [Supporting Information](#). Side chains that define the phenolic binding pocket are shown. Some specific side chains are highlighted as follows: Ile^{A10} and His^{F5} (green), Leu^{A13} and Leu^{H17} (black), Ile^{A2} and Tyr^{A19} (yellow), Cys^{A6} and Cys^{A11} (orange), and Phe^{B25} (brown). Other residues involved in the binding pocket and escape pathways are shown in white. We omit hydrogens for clarity. (C) Configuration in panel (B) represented to show the two hydrogen bonds formed by the -OH in the phenol, one with the backbone carbonyl oxygen of Cys^{A6} and another with the backbone amide NH of Cys^{A11}.

The T_6 and R_6 states of the hexamer have markedly different dissociation kinetics. The T_6 state dominates under physiological conditions² and dissociates to dimers in minutes.¹⁹ By contrast, the lifetime of the R_6 state is hours to days.¹⁹ Therapeutic formulations often contain phenol because it prolongs their shelf lives by shifting the equilibrium to R_6 ,²⁰ which presumably suppresses dissociation and, in turn, off-pathway equilibria. The phenol binds in pockets that are formed in R_6 between the A chains of one insulin dimer and the Gly^{B1}–Gly^{B8} segments from an adjacent dimer.¹⁵ The X-ray crystal structure of the phenol-bound human insulin R_6 hexamer (PDB ID 1ZNI)²¹ suggests that when bound, each phenol forms two hydrogen bonds, one with the backbone carbonyl oxygen of Cys^{A6} and one with the backbone amide NH of Cys^{A11}, and interacts with His^{F5}, among other residues in the binding pocket (Figure 1).^{15,22} These pockets are absent from T_6 .

Compared to the hexamer dissociation time scales, the phenol unbinding time scale is relatively fast. Existing NMR data suggest that phenol unbinding and rebinding occur on the submillisecond time scale.²³ This fast equilibrium is somewhat surprising, as the R_6 binding pockets bury the phenols nearly completely.^{15,24} In this case, the crystal structure is not sufficient to predict the time scales of phenol escape, and we must consider the dynamics of the protein.

NMR data indicate that Ile^{A10} serves as an essential “gatekeeper” residue, whose flexibility is required for the binding/unbinding of phenolic ligands.²³ Two subsequent molecular dynamics studies, in which external forces were used to accelerate phenol escape,^{25,26} identified three pathways for unbinding. Abrams and Vashisth²⁶ called these pathways PW1, PW2, and PW3, and we use this nomenclature throughout our paper. PW1 is a gate-pushing mechanism, where the phenol pushes through the gatekeeper residues Ile^{A10} and His^{F5} (green in Figure 1). By contrast, PW2 is a gate-hopping mechanism, with the phenol passing through an existing escape channel between Ile^{A10}/His^{F5} and Leu^{A13}/Leu^{H17} (black in Figure 1). Finally, PW3 is an unrelated pathway, where the phenol escapes through the A chain, interacting with Ile^{A2} and Tyr^{A19} (yellow in Figure 1).

While the pathways in these simulations suggest possible mechanisms of phenol unbinding, applying external forces can significantly bias the dynamics.²⁷ Consequently, it is difficult to determine the significance of the pathways and to estimate their kinetics. To address this issue, here, we exploit recent

advances in computational methods for estimating kinetic statistics from an ensemble of short, unbiased simulations^{28,29} to compute free energies, escape and rebinding probabilities (committers), and reactive currents for phenol unbinding from the R_6 hexamer. By combining these quantities, we are able to characterize six unbinding pathways quantitatively and determine their relative weights. All pathways contribute significantly to unbinding/binding rates, highlighting the importance of methods that naturally account for a diversity of mechanisms,^{28,29} in contrast to traditional rate theories.^{30,31}

In addition to providing quantitative insights, our simulations reveal qualitatively new dynamics. We observe a large-scale opening of the primary channel for phenol escape and delineate six pathways for phenol unbinding. Based on our observations, we identify and model two mutations that we expect to impact the dynamics. We find that one mutation (B13 Glu → Gln) encourages phenol unbinding, and the other (A10 Ile → Val) discourages it. The relative weights of our six pathways are dramatically affected for both mutations, with the preferred wild-type (WT) pathway becoming sparsely populated for the B13 Glu → Gln mutation. We thus demonstrate how molecular dynamics simulations can inform the design of therapeutics with improved properties.

METHODS

System Setup. All systems were prepared using CHARMM-GUI, version 3.0 for the wild-type (WT) simulations and version 3.2 for the mutant simulations.^{32–34} The crystal structure for wild-type (WT), phenol-bound human insulin R_6 hexamer was retrieved from the Protein Data Bank (PDB ID 1ZNI).²¹ All ions (including the two Zn²⁺ ions and two Cl[−] ions bound within the hexamer) and 331 crystallographic water molecules were retained. Six of the seven phenols in the structure were included in the simulations; following the procedure in ref 26, the seventh phenol, located adjacent to one of the bound Cl[−] ions, was deleted. Missing Thr residues at the C-termini of the six insulin B chains were added, along with hydrogen atoms. We refer to the A and B chains of the six insulin monomers as chains A through L, with nomenclature specifics given in the [Supporting Information](#). CHARMM parameter files for phenol were generated from the phenol structure data file (SDF) available from the RCSB and the CHARMM general force field.³⁵ Protonation states were chosen using the PROPKA3 algorithm.^{36,37} The edited crystal structure was solvated in an (8.2 nm)³ box of TIP3P water.³⁸

To neutralize the system at a concentration of 150 mM KCl, 53 K⁺ and 43 Cl[−] additional ions were added.³⁹ There was a total of 51,064 atoms.

Equilibration. All simulations were performed using GROMACS 2019.4⁴⁰ patched with PLUMED 2.5.3,^{41–43} using the CHARMM36m force field.^{44–46} Unless otherwise noted, simulations were carried out in the isochoric isothermal (NVT) ensemble at 303.15 K using a Langevin thermostat⁴⁷ with a 2 fs time step and a friction constant of 10 ps^{−1} applied to all atoms. This friction constant was chosen to be as weak as possible, while still maintaining the correct temperature, so as to minimally perturb the dynamics. The LINCS algorithm⁴⁸ was used to constrain all bonds to hydrogen. The particle-mesh Ewald method⁴⁹ was used to calculate electrostatic forces, accounting for periodic boundary conditions, with a cutoff distance of 1.2 nm. The Lennard-Jones interactions were smoothly switched off from 1.0 to 1.2 nm through the built-in GROMACS force-switch function. We used VMD⁵⁰ for molecular visualization.

To relax each system following its preparation, we used the steepest descent algorithm to minimize the energy until the maximum force felt by the system was below 1000 kJ/(mol nm). We then equilibrated the system for 100 ps in the NVT ensemble with a 1 fs time step, followed by 10 ns in the NPT ensemble at 1 bar using the Parrinello–Rahman barostat,⁵¹ with a 2 fs time step and a time constant of 5.0 ps. For the energy minimization and equilibration above, harmonic restraints were used to stabilize the positions of all nonhydrogen protein atoms. The system was equilibrated further for 1 ns in the NPT ensemble without position restraints, and the average box size was determined to be (7.96 nm)³. This box size was used for all further simulations. The system was equilibrated once more without position restraints for 3 ns in the NVT ensemble. The resulting equilibrated structure was used to initialize further simulations as described below.

Comparison of Phenols. Although, in principle, all six phenols are equivalent, in practice their behaviors may differ owing to asymmetries in the initial structure. To determine if this was the case, we used adiabatic-bias molecular dynamics (ABMD) implemented in PLUMED 2.5.3^{41–43,52} to drive their dissociation from the WT hexamer. This method uses a half-harmonic potential, moved in a ratchet-and-pawl-like mechanism, to trap natural fluctuations toward a specific target in collective variable (CV) space. It thus tends to bias the dynamics more gently than alternatives. Specifically, we used ABMD because we previously observed ABMD to successfully drive insulin dimer dissociation without melting otherwise stable α helices, in contrast to steered molecular dynamics.⁵³

Our first guesses for CVs to control phenol dissociation were six distances d_n , where $1 \leq n \leq 6$ denotes the identity of the phenol being released. Specifically, d_n is the distance between the geometric center of the nonhydrogen atoms of phenol n and the Zn²⁺ ion bound closest to it. We biased the six d_n s independently but simultaneously; since the timing of events in ABMD simulations depends on the specific fluctuations that occur, this led to simulations where varying numbers of the phenols dissociated with no externally imposed order. A total of 276 such simulations were performed (each run for 5 ns with structures saved every 10 ps): 20 simulations for each of nine equally spaced force constants ranging from 20 to 28 kJ/(mol nm), inclusive, and 48 simulations for both 29 and 30 kJ/(mol nm).

For these simulations, phenols were considered dissociated if the number of nonhydrogen atom contacts between them and the protein dropped below five. Phenol 4 was released in all 276 simulations, while phenol 3 was released in only 150/276 simulations. Furthermore, for the 77 simulations where all phenols were released, phenol 4 was released either first or second in 74/77 simulations, while phenol 3 was released either fifth or sixth in 42/77 simulations. This suggested that phenols 4 and 3 are the ligands that are most and least easily released, respectively. To ensure that our results were not specific to any particular feature of the initial structure, we investigated the release of each of these two phenols following the procedure described below. However, because we found that the results for them were nearly identical, we present only those for phenol 4.

Generation of Starting Structures. To generate starting structures for the unbiased trajectories used for later analyses, we again used ABMD, this time to explore various mechanisms of phenol release for a single phenol (phenol 4), as opposed to all six phenols at once. This time, the bias was placed on just d_4 . Initially, 120 simulations were performed (each run for 5 ns with structures saved every 5 ps): 30 simulations for each of the four force constants 12.5, 15.0, 17.5, and 20.0 kJ/(mol nm). Through visual analysis of these trajectories, three CVs were initially chosen to represent the dominant features observed: N_{A10} , N_{A13} , and RMSD_p. N_{A10} and N_{A13} are the number of nonhydrogen atom contacts between phenol 4 and either Ile^{A10} (green in Figure 1) or Leu^{A13} (black in Figure 1), respectively. RMSD_p is the distance root-mean-squared deviation (RMSD) between the α carbons of residues in the phenolic binding pocket compared to their position in the crystal structure. The distance RMSD is defined by first calculating the pairwise distances between all atoms in a selection for a reference structure, then calculating those same distances for each frame of the simulation, and finally calculating the root-mean-squared difference of these pairwise distances. The binding pocket residues, pictured in green, orange, black, and white in Figure 1, were defined to be the residues most often interacting with phenol, both in its crystallographic binding pocket and along the different observed pathways of dissociation. These 22 binding pocket residues, following the naming convention described in the Supporting Information, are as follows: Cys^{A6}, Ser^{A9}–Tyr^{A14}, Leu^{A16}, Glu^{A17}, His^{B10}, Leu^{B11}, Ala^{B14}, Phe^{F1}, Val^{F2}, His^{F5}, Leu^{F6}, Leu^{G13}, Tyr^{G14}, Glu^{G17}, Leu^{H17}, Val^{H18}, and Glu^{H21}. It should be noted that while N_{A10} and N_{A13} proved sufficient to categorize our initial sampling, we also present analysis based on other contact-based CVs that we subsequently developed to better capture the unbinding pathways.

When building this ABMD data set, we wanted to fully explore unbinding pathways. To do this, we wanted multiple trajectories to sample all populated areas in our CV space of N_{A10} , N_{A13} , and RMSD_p. By doing this, not only did we ensure that the starting structures sampled all relevant pathways, but we also decreased the possible correlation between starting structures for our unbiased simulations. From our initial set of ABMD simulations, we observed pathways similar to PW1, PW2, and PW3 as identified by Vashisth and Abrams in ref 26, although only a handful of trajectories followed PW1, and only one pathway followed PW3. To supplement the PW1 data, we ran an extra 30 ABMD simulations, each of length 5 ns with force constant $k = 12.5$ kJ/(mol nm), starting from a structure from a previous ABMD run that ended approximately halfway

along PW1. We supplement the PW3 data later in the unbiased data set. Furthermore, we saw a number of trajectories where RMSD_p jumped from approximately 0.5 Å to approximately 3–4 Å. As discussed in the [Results and Discussion](#) section, this corresponds to a channel-opening mechanism, where two dimers rotate away from one another, exposing more of the phenol directly to the solvent. To sample this behavior further, we identified two trajectories that led to channel opening but without phenol release and ran 30 extra ABMD simulations starting from each of these configurations. Each of these trajectories was 5 ns with a force constant of $k = 12.5$ kJ/(mol nm). In total, we thus supplemented our initial data set with 90 additional ABMD simulations to create a data set of 210 ABMD trajectories.

State Definitions. To measure binding and unbinding statistics, one needs to define both the bound and free states. The CVs we used to do this include RMSD_p , as previously defined, and N_{PW1} , N_{PW4} , N_{Prot} , and N_{HP} , which we define here. N_{PW1} and N_{PW4} can be viewed as extensions of the previously defined N_{A10} and N_{A13} , chosen to more fully describe the unbinding behavior in our simulations. N_{PW1} is the number of nonhydrogen atom contacts between phenol and residues Ile^{A10} and His^{F5} (green in [Figure 1](#)). Residue His^{F5}, like Ile^{A10}, helps us to define pathway 1 as in [ref 26](#). Similarly, N_{PW4} is the number of nonhydrogen atom contacts between phenol and residues Leu^{A13} and Leu^{H17} (black in [Figure 1](#)). Residue Leu^{H17}, along with Leu^{A13}, helps us to define pathway 4 (PW4), a pathway similar to but distinct from those previously characterized, which we discuss in the [Results and Discussion](#) section. Finally, N_{Prot} is the total number of nonhydrogen atom contacts between the phenol and the protein, while N_{HP} is the two-step rolling average of the number of hydrogen bonds between the phenol and both the backbone carbonyl of Cys^{A6} and the backbone amide NH of Cys^{A11}. For this CV, we defined a hydrogen bond as a distance of less than 4 Å between the nonhydrogen atoms of the donor (D) and acceptor (A) and a D–H–A angle less than 60° out of line.

The free state is defined so that $N_{\text{PW1}} < 2$, $N_{\text{PW4}} < 2$, and $N_{\text{Prot}} < 5$. The bound state is defined by $N_{\text{Prot}} > 540$, $N_{\text{HP}} > 1$, and $\sum_{\text{CV}} (\text{CV} - \mu_{\text{CV}})^2 / (2\sigma_{\text{CV}})^2 \leq 1$ for $\text{CV} = N_{\text{PW1}}$, N_{PW4} , and RMSD_p . Here, μ_{CV} and σ_{CV} represent the average and standard deviation of the corresponding CV as defined from a 10 ns unbiased simulation, respectively. For N_{PW1} , N_{PW4} , and RMSD_p , the means were 52.503, 31.924, and 0.0872 Å, and the standard deviations were 5.023, 5.264, and 0.0223 Å, respectively.

DGA Data Set. Our goal is to estimate equilibrium and dynamical quantities through the dynamical Galerkin approximation (DGA).^{28,29} Using this technique, dynamical statistics like the committor, reactive current, and reaction rate can be calculated from an ensemble of unbiased trajectories. The essential idea is that we cast quantities of interest as solutions to operator equations involving this transition operator, the operator that determines the statistics of the dynamics. In general, the solutions are subject to boundary conditions involving both the bound and free states, defined in the previous section. We approximate the solutions through a basis expansion, choosing our basis so that it captures all of the movements important for phenol binding/unbinding. This enables us to represent the action of the transition operator through its inner products with the basis functions; in turn, these can be estimated from averages over unbiased trajectories. This approach can be thought of as an extension

of Markov state models (MSMs)^{54–58} that directly yields statistics for a specified reaction.

To generate the unbiased trajectories needed for DGA, we first defined a $10 \times 10 \times 10$ grid in cylindrical coordinates that fully covers the area of CV space sampled by our ABMD data set ($N_{\text{A10}} = r \cos \theta$, $N_{\text{A13}} = r \sin \theta$, and RMSD_p). Specifically, r varied between 3 and 50 contacts, θ varied between 0 and 90°, and z varied between 0.08 and 0.38 Å. The structure from the ABMD data set closest to each of these 1000 grid points was found, leading to 326 unique structures (as the ABMD sampling is not uniform in this space, many grid points had duplicate structures; see [Supporting Figure S1](#)). From each one of these unique structures, we launched two independent 40 ns unbiased simulations and saved structures every 10 ps. Of these 652 trajectories, only three trajectories partially or fully sampled PW3, as defined by Vashisth and Abrams.²⁶ To supplement the sampling of that pathway, from these three trajectories, we selected a total of 20 structures that captured various features of PW3. From each of these 20 structures, two independent 40 ns simulations were launched. Thus, our total unbiased data set consisted of 692 trajectories of length 40 ns for an aggregate simulation time of 27.68 μs .

DGA Basis Choice. As described earlier, DGA solves operator equations for dynamical statistics by expanding them in terms of a set of basis functions. In this work, we chose the modified pairwise distance form described in [ref 29](#). To construct the appropriate pairwise distances, we first included those between two carbons on opposite sides of the escaping phenol and the α carbons of the 22 binding pocket residues described earlier (protein–ligand distances). We also included the pairwise distances between the α carbons themselves (protein–protein distances) to account for protein rearrangements. To account for movement along PW3, which involves residues not in the binding pocket, we added the protein–protein and protein–ligand distances between the phenol and the α carbons for residues Ile^{A2}/Tyr^{A19} (yellow in [Figure 1](#)) and Phe^{B25} (brown in [Figure 1](#)), as suggested by both [ref 26](#) and our preliminary results. Additionally, to capture movements of side chains along pathways 1 and 2, we added the protein–protein and protein–ligand distances between the phenol and C $_{\gamma}$ atoms from residues Ile^{A10}, Leu^{A13}, His^{F5}, and Leu^{H17}. Finally, we added one constant basis function, which improved the numerical solutions. Combining these distances led to a 299-dimensional basis set, a summary of which is shown in [Supporting Table S1](#).

For each point in our unbiased data set, we measured the minimum distance in the space of the 298 pairwise distances to points in both the bound and free states. Using these distances, we followed the procedure in [ref 29](#) to construct the smoothing function $h(x)$, the guess function(s) $\psi(x)$, and the smoothed basis functions $\phi_i(x)$. These basis functions were then orthogonalized using singular value decomposition. A key choice in applying DGA is the time interval for the transition operator, termed the “lag time”. We found most statistics to be insensitive to the lag time over ranges tested ([Supporting Figure S2](#)) and report results for a lag time of 500 ps unless otherwise noted because it yielded convergence of estimates with a minimum of apparent statistical error ([Supporting Figure S2](#)). The exceptions were rate constants and their ratios. Based on the convergence of our estimates for the relative weights of pathways ([Supporting Figure S2](#)), we use lag times between 500 ps and 1.25 ns for rates and their ratios. We

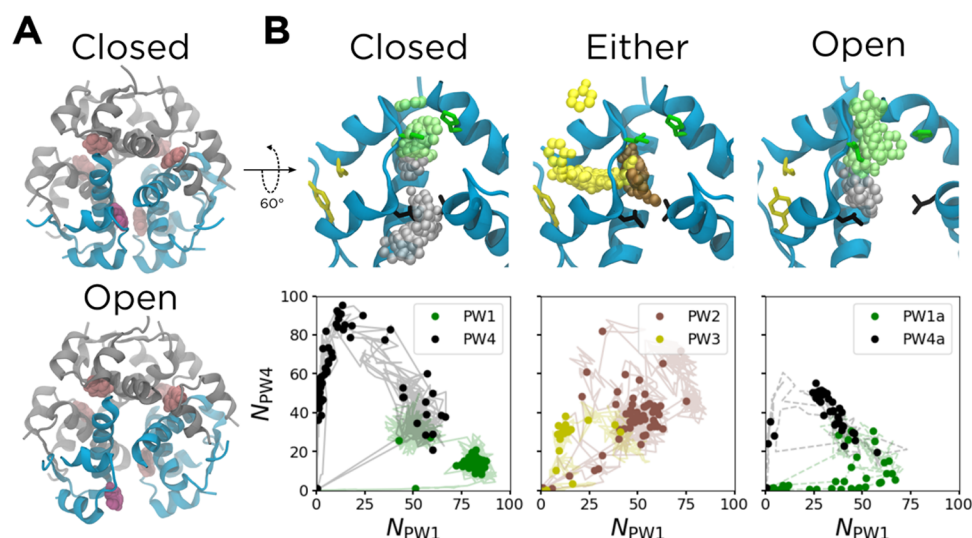


Figure 2. Results from the ABMD simulations. (A) Structure of the hexamer, with the phenolic escape channel closed (top) and open (bottom). The chains that form the phenolic binding pocket are shown in cyan; other chains are shown in gray. The released phenol is shown in purple, with other phenols shown in pink. (B) Six unbinding pathways shown both structurally (top) and as a function of N_{PW1} and N_{PW4} (bottom). The structures shown correspond to the solid data points and represent the k -medoids cluster centers along each pathway. The solid protein cartoons correspond to the starting structure for each set of driven simulations, and the translucent spheres are the nonhydrogen atoms of the phenols from the cluster centers, all aligned to the A chain backbone of the starting structure. The translucent lines in the bottom panels represent the data used to generate the k -medoids clusters. The cyan chains in the top panels are the same as those in panel (A). Nonhydrogen atoms of gatekeeper side chains along PW1 (green, Ile^{A10} and His^{F5}), PW3 (yellow, Ile^{A2} and Tyr^{A19}), and PW4 (black, Leu^{A13} and Leu^{H17}) are also shown in the top panels. The configuration of the escape channel is indicated above each panel. For PW2 and PW3, the channel can be either open or closed.

discuss further details of these calculations in the [Supporting Information](#).

Mutants. A similar workflow to the above was followed for both insulin mutants studied, A10 Ile $\rightarrow$ Val and B13 Glu $\rightarrow$ Gln. Details on system generation using CHARMM-GUI^{32–34} and workflow for these mutants are given in the [Supporting Information](#).

RESULTS AND DISCUSSION

The goal of this study is to probe the dynamics of phenol release from the R₆ insulin hexamer to understand how the release mechanism(s) could be altered by mutations. To do this, as described in detail in the [Methods](#) section, we employed a pipeline of multiple simulation techniques. First, we used an array of adiabatic-bias molecular dynamics (ABMD) simulations to create a large, diverse data set of driven dissociation events. From this, we discovered a small set of physically meaningful collective variables (CVs) that can be used to visualize the results: $RMSD_p$ describes the distance RMSD (referenced to the crystal structure) of the 22 residues determined to be in the phenolic binding pocket, while N_{PW1} , N_{PW3} , and N_{PW4} describe the number of nonhydrogen atom contacts between the released phenol and gatekeeper residues along pathway (PW) 1 (Ile^{A10} and His^{F5}, green in [Figure 1](#)), PW3 (Ile^{A2} and Tyr^{A19}, yellow in [Figure 1](#)), and PW4 (Leu^{A13} and Leu^{H17}, black in [Figure 1](#)), respectively. To escape via each of these pathways, the phenol must pass between the side chains of its two gatekeeper residues. We do not consider an analogous N_{PW2} because there are no specific gatekeeper residues for PW2.

We then ran short (40 ns) unbiased simulations starting from a selection of partially unbound structures from the ABMD data set and used dynamical Galerkin approximation (DGA)^{28,29} to estimate long-time statistics from the resulting short trajectories. Using this method, we computed free

energies, escape versus rebinding probabilities (committers), and reactive currents. Combining these statistics allows us to delineate six pathways and their transition states. DGA furthermore enables us to estimate the relative weights for the unbinding pathways and rates, providing insights into kinetics that go beyond the free energies by themselves. Below, we first describe the simulations for WT insulin, followed by those for the A10 Ile $\rightarrow$ Val and B13 Glu $\rightarrow$ Gln mutants.

Driven Simulations Reveal Channel Opening and Six Dissociation Pathways. When visually analyzing the ABMD simulations with driven phenol release, we were able to distinguish six unbinding pathways. Three of these pathways (PW1, PW2, and PW3) were previously reported by Vashisth and Abrams,²⁶ while the other three (PW1a, PW4, and PW4a) are reported here for the first time. An essential feature of two of these pathways (PW1a and PW4a) is a large-scale widening of the primary escape channel (indicated by cyan coloring in [Figure 2A](#)) that results from two adjacent insulin dimers twisting away from one another. This behavior is discussed in depth in the following section. In addition, while we explicitly discuss the driven ABMD simulations in this section, the same six pathways with the same molecular characteristics are further supported by the unbiased simulations and our quantitative analysis of them.

We tested a variety of CVs for their utility in visualizing the six unbinding pathways. As noted earlier, three of the best that we found were N_{PW1} , N_{PW3} , and N_{PW4} , the numbers of nonhydrogen contacts between the phenol and the gatekeeper residues for PW1, PW3, and PW4, respectively. We determined the gatekeeper residues based on the previous literature and our analysis of the ABMD trajectories. The gatekeeper residues for PW1 are Ile^{A10} and His^{F5}; these were identified by both Swegat et al.²⁵ and Vashisth and Abrams²⁶ and were confirmed by our own simulations. The gatekeeper residues for PW3 are Ile^{A2} and Tyr^{A19}; these were identified by

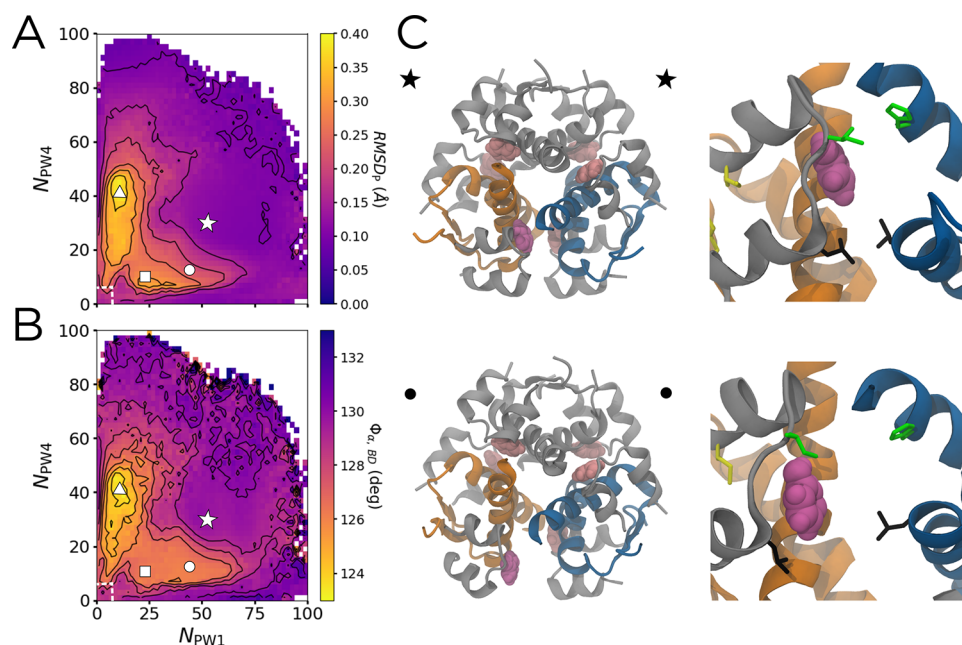


Figure 3. Measures of flexibility of the phenolic binding pocket. The bound state is marked by the white star, and the unbound state is marked by the white dashed lines. The circle and square represent the partially open escape channel with the phenol bound and partially unbound, respectively. The triangle represents a PW3 intermediate in which the A chain is partially melted. (A) Distance RMSD of the 22 binding pocket residues as a function of N_{PW1} and N_{PW4} . Contour spacing is 0.5 Å. (B) Pseudodihedral angle between the α helices at the dimer interface between chains B and D, $\Phi_{\alpha, BD}$, as a function of N_{PW1} and N_{PW4} . Contour spacing is 1°. (C) Hexameric (left) and binding pocket (right) structures showing the closed-to-open transition indicated by the stars and circles in panels (A) and (B), respectively. Chains B and D are shown in orange, while chains F and H are shown in blue.

Vashisth and Abrams²⁶ and confirmed by our own simulations. The gatekeeper residues for PW4 are Leu^{A13} and Leu^{H17}. Vashisth and Abrams²⁶ note these residues in conjunction with PW2, but we designate them as gatekeeper residues for PW4 in this study because the phenol passes between them only along PW4 (versus near them along PW2). We present most of our results in terms of N_{PW1} and N_{PW4} because they provide the best separation of all six pathways when projected to two dimensions.

To visualize these pathways, we first identified three ABMD trajectories that corresponded to each pathway other than PW3. We then identified the section of each trajectory that corresponded to phenol unbinding and used *k*-medoids clustering in the 298-dimensional set of pairwise distances described in the Methods section to identify 60 cluster centers per pathway. For PW3, we had only one ABMD trajectory (as noted in the Methods section, we addressed this issue through additional unbiased simulations), so we generated only 20 cluster centers. Regardless, the cluster centers were used to understand the characteristics of each observed pathway. Structural and CV representations of the six unbinding pathways are shown in Figure 2B.

As shown in the left panel of Figure 2B, PW1 consists of the green phenol moving upwards toward the green gatekeeper residues, Ile^{A10} and His^{F5}, then pushing through them and escaping into the solvent. This corresponds to the green loop in the CV representation. As discussed in the Methods section, the bound state exists where $N_{PW1} \approx 53$ and $N_{PW4} \approx 32$. Starting from there, PW1 loops down and outward to $N_{PW1} \approx 85$ and $N_{PW4} \approx 10$, corresponding to the phenol approaching the green gatekeeper residues. As the phenol pushes through these residues and escapes into the solvent, both N_{PW1} and N_{PW4} decrease to zero.

By contrast, PW4 proceeds in the opposite direction, with the gray phenol passing through the black gatekeeper residues, Leu^{A13} and Leu^{H17}. This leads to the black loop in the CV representation, where the system moves from the bound state to $N_{PW1} \approx 10$ and $N_{PW4} \approx 85$ as the phenol approaches the gatekeeper residues. As the phenol pushes through those residues and escapes into the solvent, both N_{PW1} and N_{PW4} decrease to zero. While PW1 was described in previous simulation studies,^{25,26} PW4 is presented for the first time in this work.

Both PW1 and PW4 occur with a closed escape channel, with each gatekeeper residue remaining in close proximity to its position in the crystal structure. However, very similar pathways can occur with an open escape channel, as the H and F chains twist away from the A chain. This large-scale protein motion, seen in Figure 2A, leads to the change in the binding pocket seen in the right panel of Figure 2B. Here, the separation between both the green residues and the black residues is increased, meaning they no longer function as gates. For example, along the PW1-like pathway, the phenol still interacts with Ile^{A10}, but His^{F5} is shifted far enough away to allow a layer of water between the side chain and the phenol. We refer to this pathway as PW1a. Similarly, along the PW4-like pathway, the phenol interacts with Leu^{A13} but not Leu^{H17}. We refer to this pathway as PW4a. In the CV representations, both of these pathways are associated with smaller loops in the bottom panels of Figure 2B, consistent with the contacts coming from only one gatekeeper residue. Both of these pathways are described for the first time in this work.

Finally, we also observed PW2 and PW3 in our driven simulation data set. These are illustrated in the middle panel of Figure 2B. PW2 is a gate-hopping mechanism, where the brown phenol escapes by hopping between the green and black

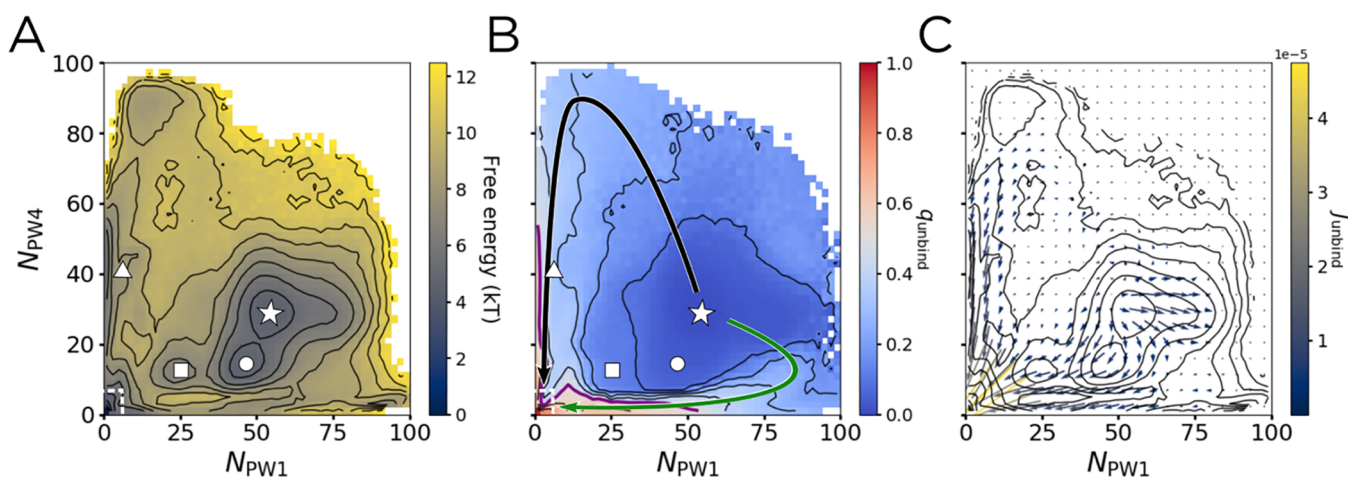


Figure 4. Potential of mean force (PMF), unbinding committor (q_{unbind}), and unbinding reactive current (J_{unbind}) projected on N_{PW1} and N_{PW4} . The points marked by the star, circle, square, and triangle are the same as in Figure 3, with the unbound state outlined by the dashed white box. (A) PMF, with contours spaced by $1 k_B T$. (B) q_{unbind} with contours spaced by 0.1 and the $q_{\text{unbind}} = 0.5$ surface marked in purple. Arrows showing PW1 (green) and PW4 (black) are overlaid. (C) J_{unbind} binned into a 22×22 grid spanning from 0 to 100 in both N_{PW1} and N_{PW4} . The results shown are smoothed with a Gaussian filter, using a kernel with a standard deviation of 1 bin. Contours are the same as in panel (A) to aid in comparison.

gatekeeper residues. In Figure 2B, this is moving out of the plane of the paper, directly toward the reader. In contrast, along PW3, the yellow phenol moves toward the yellow gatekeeper residues, Ile^{A2} and Tyr^{A19}, before pushing through them and escaping into the solvent. Both of these pathways were described by Vashisth and Abrams,²⁶ but we add an important clarification: along both of these pathways, the escape channel can be either closed or open. However, since PW2 involves an escape through a channel that exists in the crystal structure and PW3 involves escape through an unrelated channel, neither of these two pathways are particularly affected by the opening/closing of the escape channel. For a discussion of the sequence of release of the six phenols and its relation to the experimentally observed cooperativity of phenol binding to the insulin hexamer,^{59,60} see Supporting Figure S3 and the associated text in the Supporting Information.

Further Characterization of the Channel Opening.

The observed channel opening was not an artifact of the ABMD bias, as we also found it in many of our unbiased simulations seeded with a closed channel. Furthermore, by analyzing our unbiased simulations, we can correlate the channel opening with the rotation of the α helices on the dimeric interface, a conformational transition previously found to be accessible to the insulin dimer.^{53,61} In Figure 3A, we quantify channel opening by measuring the average of RMSD_p , which is a distance RMSD that quantifies how much the phenolic binding pocket structurally deviates from the binding pocket present in the crystal structure. This quantity was also compared to the average of $\Phi_{\alpha, \text{BD}}$, the pseudodihedral angle of the α helices on the dimeric interface of chains B and D (shown in orange in Figure 3B), as defined in ref S3.

The bound state, marked by the white star in Figure 3A and located at $N_{\text{PW1}} \approx 53$ and $N_{\text{PW4}} \approx 32$, is characterized by a relatively low average RMSD_p of approximately 0.1 Å. The phenolic binding pocket in this state is quite similar to that in the crystal structure. Across the space defined by N_{PW1} and N_{PW4} , RMSD_p is highly correlated with $\Phi_{\alpha, \text{BD}}$. The BD dimer, along with the dimer defined by the interface made of chains F and H (shown in blue in Figure 3B), makes up the primary escape channel for the phenol studied. The average of $\Phi_{\alpha, \text{FH}}$ is

nearly identical to that for $\Phi_{\alpha, \text{BD}}$ shown in Figure 3B, so it is omitted for clarity. Both have a value of approximately 132° in the bound state (Figure 3B). As discussed previously, this leads to both the green gatekeeper residues (Ile^{A10} and His^{F5}) and the black gatekeeper residues (Leu^{A13} and Leu^{H17}) remaining in close proximity, acting as closed gates that define the boundaries of the escape channel.

Moving from star to circle to square in Figure 3A, RMSD_p increases from 0.1 to 0.2 to 0.3 Å, corresponding to a 3 degree rotation of the α helices on the dimer interfaces. Structurally, this dimeric rotation leads to separation of the four gatekeeper residues that define the edges of the phenolic escape channel, as seen in the right panels in Figure 3C. Specifically, again moving from star to circle to square, the average separation of the α carbons of Leu^{A13} and Leu^{H17} increases from 0.8 to 1.2 to 1.4 Å, and the average separation of the α carbons of Ile^{A10} and His^{F5} increases from 0.8 to 0.9 to 1.2 Å (Supporting Figure S4A). While both the circle and square correspond to structures with a partially open escape channel, the square represents a larger channel opening that is paired with the breaking of the hydrogen bonds between the phenol and both the backbone carbonyl of Cys^{A6} and the backbone amide nitrogen of Cys^{A11}; although the phenol is still located in the binding pocket, it is more loosely bound (Supporting Figure S4A,B). Overall, this channel opening, found in both our biased and unbiased simulations, is also consistent with the displacements observed in the lowest-frequency modes of a normal mode analysis of an elastic network representation of the hexamer (Supporting Figure S5), suggesting that this flexibility is an intrinsic feature of the structure. The channel opening allows for increased penetration of the solvent closer to the phenol and decreased steric occlusion by the side chains.

Finally, the point marked by the triangle in Figure 3 has an RMSD_p between 0.5 and 1.0 Å higher than those for the points marked by the square and the circle. This corresponds to the melting of the A chain C-terminal α helix along PW3, as the phenol escapes through the A chain (Supporting Figure S4A,C). As this helix melts, the PW1 and PW4 gatekeeper residues separate even further (Figure 3B and Supporting Figure S4A); although this motion affects the binding pocket, it occurs after the phenol is no longer in it.

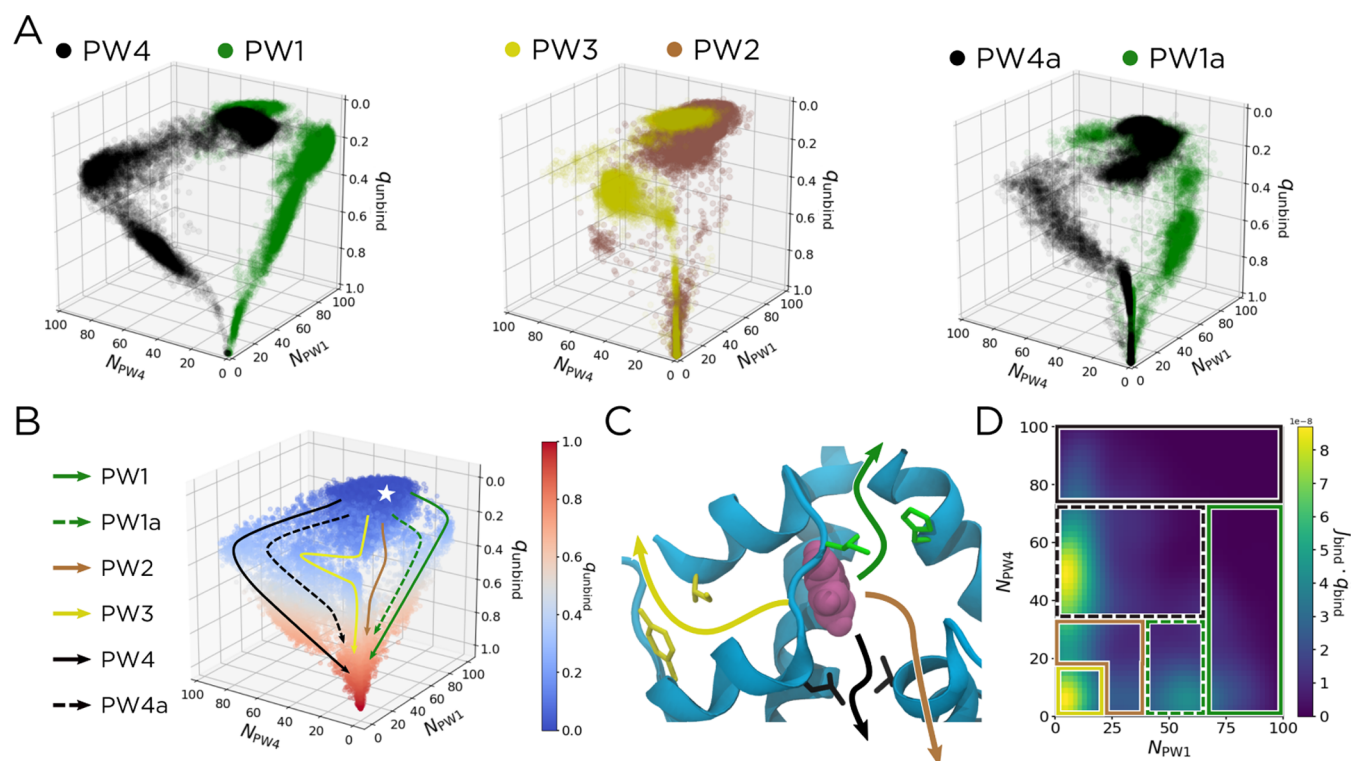


Figure 5. Pathways can be more readily distinguished in the space of N_{PW1} , N_{PW4} , and q_{unbind} . (A) Scatter plots of trajectories along each of our six pathways. From the unbiased data set, we identify four trajectories that correspond to each pathway, which are shown for PW1/PW4 (left), PW2/PW3 (middle), and PW1a/PW4a (right), mirroring the conventions used in Figure 2B. (B) Scatter plot of q_{unbind} for all of the unbiased data. Six pathways are overlaid and labeled. The bound state is represented by the white star. (C) Structural representation of the unbinding pathways in panel (B). The dashed arrows in panel (B) correspond to the similarly colored solid arrows in panel (C), except with the escape channel being opened as in Figure 3. (E) The q_{bind} component of J_{bind} , taken at $q_{bind} = 0.63$. The patches corresponding to each pathway are overlaid, using the coloring and line styles from panel (B).

Combining the Potential of Mean Force, Committor, and Reactive Current Enables Quantitative Characterization of Transition States and Intermediates. DGA can be used to estimate the relation between the sampled distribution and the stationary distribution and, in turn, potentials of mean force (PMFs). However, the true power of DGA is in the ability to combine these free energies with estimates for dynamical statistics like the committor, reactive current, and rate. The committor describes the probability of proceeding to a product state before returning to a reactant state. By explicitly defining bound and unbound states (see the Methods section), we can thus calculate the unbinding committor, q_{unbind} , that describes, at every point in our unbiased data set, the probability of proceeding to the unbound state before returning to the bound state. This is, by construction, the perfect reaction coordinate to track phenol unbinding, as it measures the likelihood of phenol unbinding regardless of pathway. By definition, transition states have $q_{unbind} = 0.5$. By projecting q_{unbind} into various CV spaces, we can distinguish molecular rearrangements (e.g., those giving rise to the metastable states that we identify on a PMF) that increase the probability of unbinding from those that do not.

Another useful quantity is the reactive current, which describes how trajectories that lead to phenol unbinding flow through each point in a CV space. Just like a diagram of fluxes between clusters of states in a Markov state model,^{54–58} a plot of the reactive current can give a sense of the populations of different pathways. Here, as established by recent work,²⁹ we represent the reactive current, J_{unbind} , as a vector field. An

advantage to this representation is that it can be compared directly with the PMF and committor as functions of the same CVs. By combining these statistics, we can characterize the transition states and intermediates along the six pathways. J_{unbind} can furthermore be used to determine their relative weights; integration of J_{unbind} yields the rate.

We show the free energy, committor, and reactive current projected onto N_{PW1} and N_{PW4} in Figure 4. The unbound state (bottom-left corner in Figure 4A) is set to be the zero of free energy. The DGA-generated PMF is in good agreement with one independently generated through replica-exchange umbrella sampling (Supporting Figure S6), suggesting that the unbiased trajectories cover the space sufficiently to draw quantitative conclusions. The bound state with energy $4.5 k_B T$ is marked by a white star at $N_{PW1} \approx 53$ and $N_{PW4} \approx 32$. As discussed in the Introduction section, in this state, the phenol makes one hydrogen bond with the backbone carbonyl of Cys^{A6} and one with the backbone amide NH of Cys^{A11}. We find that in this state, His^{F5} is rotated outward toward the solvent, with the ring pointing away from the phenol (see Figures 2B and 3C). The free energetic shoulder directly to the right of the star involves a side-chain ring-flip of this histidine, so that it is instead facing inward, toward the phenol, which increases the number of contacts between the phenol and the histidine ring. This is consistent with a previous ¹H-NMR study that suggested the presence of an unidentified aromatic ring-flip²³ and a computational study that proposed His^{H5} as one of the residues that could undergo such a flip.²⁶

The point marked by a circle in Figure 4A corresponds to a free-energy basin, confirming that the channel-opened state is energetically stable. Furthermore, this transition only involves a free-energy barrier of about $1 k_B T$, suggesting that these structures can readily interconvert. Corroborating this idea, all of these areas have an average q_{unbind} of less than 0.1 (Figure 4B), meaning that both the channel opening and the histidine ring-flip do not markedly increase the probability of escape before returning to the bound state. This suggests that while channel opening may play a part in phenol unbinding, such a motion alone is not enough to allow the phenol to escape. The point marked by the square, which corresponds to the phenol breaking its hydrogen bonds to the channel-opened structure, also lies in a free energetic basin, one separated from the channel-opened bound state by a barrier of about $3 k_B T$. This barrier is three times as high as the one for channel opening and also corresponds to a 0.1 increase in q_{unbind} . The breaking of the hydrogen bonds to Cys^{A6} and Cys^{A11} thus slightly increases the probability of successful unbinding, while the channel opening by itself does not.

As discussed earlier, this projection effectively separates PW1 and PW4, marked by the solid black and green arrows in Figure 4B. Comparing these arrows to the PMF in Figure 4A, one finds the barriers for PW1 and PW4 to be 4–5 and 5–6 $k_B T$, respectively, significantly lower than the 20–30 $k_B T$ barriers found by Vashisth and Abrams.²⁶ Our free-energy barriers change very little when projected into a three-dimensional space that further separates PW1 and PW4 (Supporting Figure S7). Vashisth and Abrams estimate free energies by applying Jarzynski's equality to steered molecular dynamics simulations, which we would expect to overestimate barriers because such simulations will only rarely sample trajectories with protein fluctuations that anticipate the phenol motion. By contrast, our unbiased trajectories, as well as our independent replica-exchange umbrella sampling (Supporting Figure S6), capture these dynamics.

Additionally, comparing Figure 4 to Figure 3, it is clear that there is very little opening of the escape channel inherent along either PW1 or PW4, as the average RMSD_p along these pathways stays below 0.1 Å. Instead, the phenol is directly pushing through the adjacent gatekeeper residues. However, this projection does not allow us to fully distinguish PW2, PW3, PW1a, and PW4a because they partially overlap (see Figure 2B). At the same time, the area through which they pass has the highest magnitude of reactive current (Figure 4C), suggesting that the dominant binding pathways lie in this region. To distinguish these pathways, we introduce a third dimension.

Three-Dimensional Projections Show That Competing Pathways Are Similarly Populated. Since the unbinding committor q_{unbind} tracks phenol unbinding across all pathways, we use it as the third dimension on which to project our data. From our unbiased data set, we visually identified four trajectories for each of our six unbinding pathways. These trajectories are shown in Figure 5A, with coloring consistent with Figure 2B. We also show the fully unbiased data set colored to show q_{unbind} and with arrows representing the six pathways overlaid in Figure 5B. A structural representation of these pathways is shown in Figure 5C.

As shown in Figure 5A, using q_{unbind} as the third dimension clearly separates the six pathways. The $q_{\text{unbind}} = 0.5$ surface (gray points in Figure 5B) is the transition-state ensemble, as

structures in this region have an equal probability of proceeding to the unbound state as they do of returning to the bound state. Notably, for PW1/PW1a and PW4/PW4a, this does not occur until after N_{PW1} and N_{PW4} have reached their respective maxima along each pathway. The transition states thus occur when the phenol is just outside the gatekeeper residues, and the numbers of contacts with those residues have started to decrease. In contrast, for PW3, the $q_{\text{unbind}} = 0.5$ surface occurs just as the phenol maximizes its contacts with Ile^{A2} and Tyr^{A19}, the PW3 gatekeeper residues (Supporting Figure S8). Since PW2 encompasses all gate-hopping mechanisms and the phenol can be closer to the gatekeeper residues for PW1 or those for PW4, its transition state is comparatively structurally diverse.

We can also use DGA to calculate the relative weight of each unbinding pathway. To do this, we adapt the method described in ref 29, using the binding reactive current J_{bind} projected into the space of N_{PW1} , N_{PW4} , and q_{bind} . For this calculation only, we used kinetic statistics for binding as opposed to unbinding, as this slightly improved our ability to distinguish the six pathways (Supporting Figure S9). We chose as a dividing surface the plane corresponding to $q_{\text{bind}} = 0.63$, as this best separates the reaction pathways (Supporting Figure S10). We then partitioned this surface into patches that encompass the pathways (Figure 5D). By binning the reactive current in the direction of q_{bind} across these patches, one can calculate the relative weight of each pathway as the ratio of the reactive current in each patch compared to the total reactive current flowing through the dividing surface. We provide exact definitions of the patches, as well as a discussion of the parameters used to calculate and smooth J_{bind} in the Supporting Information. Using this method, we determined the relative weights of the six identified pathways (Table 1). These relative weights are robust to changes in the sampling distribution (Supporting Table S2) and choice of dividing surface (Supporting Figure S11).

Table 1. Relative Weights of Unbinding Mechanisms for WT Insulin and the A10 Ile → Val and B13 Glu → Gln Mutations

Pathway	WT (%)	A10 (%)	B13 (%)
PW1	11.2	4.3	19.0
PW1a	11.2	7.8	16.0
PW2	16.3	16.7	14.1
PW3	12.6	8.7	7.1
PW4	13.7	18.4	26.7
PW4a	35.0	44.2	17.2

The pathway with the clear plurality of reactive current (35.0%) is PW4a (dashed black arrow), which corresponds to an unbinding mechanism with an open channel, where the phenol closely passes by Leu^{A13} as it leaves the binding pocket. While this is the most likely unbinding mechanism, all of the other pathways are significantly populated (ranging from 11 to 17%). Furthermore, the preference for mechanisms with an open escape channel is relatively mild, with the PW1/PW4 percentage being 24.9% compared to the PW1a/PW4a percentage being 46.2%. As a note, as defined in this study, PW2 and PW3 can have either a closed or open escape channel. So while channel opening does play a role in the overall phenol unbinding process, the majority (53.8%) of unbinding events occur through pathways that do not

necessarily involve channel opening. As we discuss further below, the manifestly multipathway nature of the unbinding process makes it challenging to predict how point mutations will impact phenol unbinding.

A10 and B13 Mutations Alter the Preferred Pathway and the Binding/Unbinding Rates. As discussed in the Introduction section, there is much interest in designing insulin mutants and analogues for the management of diabetes, including those that might slow the unbinding of phenol. To this end, we simulated two mutants: A10 Ile $\rightarrow$ Val and B13 Glu $\rightarrow$ Gln. The A10 Ile $\rightarrow$ Val mutation is one of the three sequence differences between human and bovine insulin; since Ile^{A10} is one of the gatekeeper residues for PW1/PW1a, we expected its mutation to affect the rate of phenol release. The B13 Glu $\rightarrow$ Gln mutation removes negative charges from the center of the hexamer and stabilizes the R state, as evidenced by the fact that it leads to the formation of T₃R₃ hexamers even in the absence of zinc.⁶² Given the stabilization of the R state, we expected the B13 Glu $\rightarrow$ Gln mutation to stabilize the R₆ hexamer and impact channel opening.

We performed simulations analogous to those above for WT human insulin, with each of these mutations separately, and determined their effects on the unimolecular rate constants of unbinding and binding, $k_{\text{unbinding}}^{-1}$ and k_{binding}^{-1} , as well as the corresponding ratio, $K = k_{\text{unbinding}}^{-1}/k_{\text{binding}}^{-1}$ (Table 2 and

Table 2. Inverse Unimolecular Unbinding Rate Constant $k_{\text{unbinding}}^{-1}$, the Inverse Unimolecular Binding Rate Constant k_{binding}^{-1} , and Their Ratio ($K = k_{\text{unbinding}}^{-1}/k_{\text{binding}}^{-1}$)^a

Statistic	WT	A10	B13
$k_{\text{unbinding}}^{-1}$ (μs)	0.16–0.28	0.21–0.37	0.17–0.27
k_{binding}^{-1} (μs)	0.13–0.20	0.12–0.18	0.16–0.26
K	0.70–0.83	0.40–0.48	0.97–0.99

^aRanges derive from taking lag times between 500 ps and 1.25 ns.

Supporting Figure S12). We also calculated the relative weights for the six unbinding pathways for each mutant (Table 1). Note that k_{binding}^{-1} is a unimolecular rate constant that does not include the contribution from diffusion, which we expect to be less sensitive to mutations, and consequently K differs from an experimentally measured dissociation constant in that it is a ratio of unimolecular rates. Rate constants that account for diffusion are discussed in the Supporting Information, with results in Supporting Figure S13 and Table S3. These estimates provide good quantitative agreement with available experimental values for dissociation constants,⁶³ indicating that DGA yields accurate results. Here, we focus on unimolecular rate constants because we expect them to have fewer sources of error. The binding and unbinding inverse unimolecular rate constants for WT insulin and the two mutants are in the range of 0.11–0.31 μs , which is consistent with the expected submillisecond phenol unbinding time scale predicted from the existing NMR data.²³

We first describe the A10 Ile $\rightarrow$ Val mutation. We find that the A10 Ile $\rightarrow$ Val mutation has almost no effect on the phenol binding rate constant while slightly decreasing the unbinding rate constant (increasing the time scale of unbinding). This, in turn, leads to a 42–43% decrease in the ratio K compared to WT insulin. To understand how this mutation, which decreases the size of the A10 side chain, inhibits phenol dissociation, we examine the relative weights for the six pathways (Table 1). Because A10 is a gatekeeper for PW1/

PW1a, we expected the A10 Ile $\rightarrow$ Val mutation to impact the weights of these pathways most strongly, and indeed this is the case. The decreases in the relative weights of these pathways are corroborated by increases in the free energies and decreases in the unbinding committors and reactive currents in associated regions (Supporting Figure S14). Intermediate structures along these pathways are less likely, and when they do occur, they are less likely to lead to the unbound state. The preferred binding pathway remains PW4a, although the relative weights of PW2 and PW4 are all higher than for WT.

Our calculations indicate that the B13 Glu $\rightarrow$ Gln mutation, while having little effect on phenol unbinding, slows phenol binding. As a result, there is a 19–39% increase in K compared to that of WT. As seen in Table 1, this mutation leads to a dramatic decrease in relative weight for PW4a, the preferred unbinding mechanism for both WT and A10 Ile $\rightarrow$ Val insulins. This is paired with the corresponding increase in the relative weight for both PW1 and PW4, the two mechanisms that explicitly do not include any channel opening and instead involve the phenol pushing through gatekeeper residues, as well as an increase in the relative weight for PW1a, which does involve channel opening. This agrees with our calculations that, after the mutation, the energetic benefit of channel opening is approximately 10 kJ/mol greater along PW1a than along PW4a (Supporting Figure S15). The mutation thus discourages PW4a more than PW1a. The overall shift in relative weight away from PW4a is further corroborated by the PMF, committor, and reactive current (Supporting Figure S14): areas of CV space along PW4a are 2 $k_{\text{B}}T$ higher in free energy compared to the same areas for WT insulin.

Beyond the shifting relative weights between pathways, the overall effect of the mutation is to destabilize the bound state and increase K . Molecularly, this can be explained by the more favorable interactions the mutated B13 residue can make with the rest of the protein in the free state, particularly with Ser^{B9} and His^{B10} (Supporting Table S4). This prediction agrees with existing experimental data for this mutated species. Dunn and co-workers⁶³ used UV/vis spectroscopy and a three-state allosteric model to determine the dissociation constant for WT insulin to be $(1.8 \pm 1) \times 10^{-4}$ M and that for the B13 mutant to be $(2.5 \pm 1) \times 10^{-4}$ M, meaning that the mutation caused a 39% increase. Assuming that the main impact of the mutation is on the protein dynamics and not diffusion, this is in agreement with our estimates, which predict a 19–39% increase in K upon B13 Glu $\rightarrow$ Gln mutation.

For both mutations, there are at least four pathways, each of which represents at least 10% of the overall reactive current, and no one mechanism ever makes up more than 50%, similar to our findings for WT insulin. As a result, the effects of a mutation on certain pathways can be compensated by those on others. For example, when the A10 Ile $\rightarrow$ Val mutant discouraged unbinding through PW1 and PW1a, the absolute amount of reactive current flowing through PW4 increased (Supporting Figure S14). Indeed, the multipathway nature of unbinding is the main takeaway from our simulations, and it suggests that future mutation and ligand design studies of the insulin R₆ hexamer need to target multiple pathways at once.

Solvation of Phenol and Its Binding Pocket. Given the interplay between the pathways and channel opening, we characterized the solvation of the phenol and select residues by calculating radial distribution functions of water around those species as a function of the committor value (Supporting Figure S16). In all cases (WT, A10 Ile $\rightarrow$ Val, B13 Glu $\rightarrow$

Gln), the phenol becomes more solvated as it leaves the binding pocket. In general, the radial distribution functions around gatekeeper and pocket residues do not change dramatically as a function of committor. However, we do observe slight changes for three residues: Ile^{A2} (a gatekeeper for PW3), Ile^{A10} (a gatekeeper for PW1/PW1a), and His^{B10} (a pocket residue that also defines the binding site for the Zn²⁺ ions). The increase in solvation of Ile^{A10} and His^{B10} comes early in the release process, whereas that for Ile^{A2} occurs as the phenol escapes.

We also computed radial distribution functions of water for different RMSD_p values, tracking channel opening (Supporting Figure S17). These data reveal that the gatekeeper residues near the phenol escape channel (Ile^{A10}, His^{F5}, Leu^{A13}, and Leu^{H17}) all become more solvated upon channel opening. The other residues considered are affected less by channel opening. Interestingly, while channel opening does not significantly affect the solvation of Glu^{B13} in WT insulin, it does lead to increased solvation for Glu^{B13} in the B13 Glu → Gln mutant insulin.

Finally, Bagchi and co-workers have argued that water molecules confined in the central cavity stabilize the hexamer.^{64,65} To examine whether the dynamics that we observe can facilitate exchange of water molecules between the cavity and the bulk, we defined the cavity waters as those within 6 Å of the center of mass of the two Zn²⁺ ions and calculated their mean square displacements (MSD) over 1 ns (right panels in Supporting Figures S16 and S17). For comparison, we measure an MSD over 1 ns of 35 nm² for bulk waters in our simulations. The MSD of the waters in the cavity shows very little dependence on committor and is generally less than 10 nm², meaning that these waters are confined regardless of the dissociation of phenol. By contrast, the MSD of waters in the cavity increases as RMSD_p increases, indicating that channel opening facilitates exchange of solvent between the cavity and the bulk.

CONCLUSIONS

Diabetes management can be improved through both the introduction of insulin analogues with modulated pharmacokinetics and delivery preparations that can facilitate transport and storage. Because phenol stabilizes the R₆ insulin hexamer, understanding the phenol unbinding mechanism can inform the design of improved therapeutics. Here, we use molecular dynamics simulations to investigate this mechanism for WT and two mutant insulins. We expand on existing simulation studies by Swegat et al.²⁵ and Vashisth and Abrams²⁶ to identify and quantitatively characterize six phenol binding/unbinding pathways. A number of these pathways involved large-scale opening of the primary escape channel, suggesting that the hexamer is much more dynamic than previously appreciated. Methods that we recently introduced^{28,29} enable us to determine the intermediates, transition states, and relative weights of the pathways. For WT insulin, a pathway in which the channel opens and phenol passes between Leu^{A13} and Leu^{H17} (PW4a) represents 35% of the reactive current, but each of the other pathways represents at least 10% of the reactive current. Phenol unbinding/binding is thus a multi-pathway process.

Our simulations of mutants show that it is possible to rationally control the prevalence of pathways and the overall unbinding kinetics. The A10 Ile → Val mutant reduced the contributions from pathways for which this residue is a

gatekeeper (PW1 and PW1a) and decreased phenol unbinding; the B13 Glu → Gln mutation stabilized the phenol-free state and thus led to increased phenol unbinding. However, because other pathways than those targeted can compensate, the overall effects on rates can be challenging to predict without quantitative simulations like those presented here. By combining computation and experiment, it may be possible to target multiple pathways to achieve larger shifts in kinetics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.1c06544>.

Further simulation details, parameter choices, cooperativity analysis, additional PMFs and kinetic statistics, solvent analysis, and rate constant estimates accounting for diffusion (PDF)

AUTHOR INFORMATION

Corresponding Author

Aaron R. Dinner – Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, United States; James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States; Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637, United States; orcid.org/0000-0001-8328-6427; Email: dinner@uchicago.edu

Authors

Adam Antoszewski – Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, United States; orcid.org/0000-0003-0097-5257

Chatipat Lorpaiaboon – Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, United States

John Strahan – Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jpcb.1c06544>

Notes

The authors declare no competing financial interest.

Scripts with an implementation of DGA are available at <https://github.com/dinner-group/insulin-hexamer>.

ACKNOWLEDGMENTS

The authors thank Luis Busto de Moner, Jonathan Weare, and Michael Weiss for helpful discussions and feedback. This work was supported by the National Institutes of Health award R35 GM136381. Computations were performed on resources provided by the University of Chicago Research Computing Center, the GM4 cluster supported by the National Science Foundation (NSF) Major Research Instrumentation award DMR-1828629, and the Extreme Science and Engineering Discovery Environment⁶⁶ (NSF award ACI-1548562) Bridges (PSC) computing nodes through allocation TG-MCB180007.

REFERENCES

- (1) DeFelippis, M. R.; Chance, R. E.; Frank, B. H. Insulin Self-Association and the Relationship to Pharmacokinetics and Pharmacodynamics. *Crit. Rev. Ther. Drug Carrier Syst.* **2001**, *18*, 201–264.

- (2) Weiss, M. A. In *Insulin and IGFs*; Litwack, G., Ed.; Vitamins and Hormones.; Academic Press: London, 2009; Vol. 80, Chapter 2, pp 33–49.
- (3) Brange, J.; Andersen, L.; Laursen, E. D.; Meyn, G.; Rasmussen, E. Toward Understanding Insulin Fibrillation. *J. Pharm. Sci.* **1997**, *86*, 517–525.
- (4) Hua, Q. X.; Weiss, M. A. Mechanism of Insulin Fibrillation: The Structure of Insulin Under Amyloidogenic Conditions Resembles a Protein-Folding Intermediate. *J. Biol. Chem.* **2004**, *279*, 21449–21460.
- (5) Ivanova, M. I.; Sievers, S. A.; Sawaya, M. R.; Wall, J. S.; Eisenberg, D. Molecular Basis for Insulin Fibril Assembly. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18990–18995.
- (6) Zaykov, A. N.; Mayer, J. P.; DiMarchi, R. D. Pursuit of a Perfect Insulin. *Nat. Rev. Drug Discov.* **2016**, *15*, 425–439.
- (7) DiMarchi, R.; Chance, R.; Long, H.; Shields, J.; Sliker, L. Preparation of an Insulin with Improved Pharmacokinetics Relative to Human Insulin through Consideration of Structural Homology with Insulin-Like Growth Factor I. *Horm. Res.* **1994**, *41*, 93–96.
- (8) Setter, S. M.; Corbett, C. F.; Campbell, R. K.; White, J. R. Insulin Aspart: A New Rapid-Acting Insulin Analog. *Ann. Pharmacother.* **2000**, *34*, 1423–1431.
- (9) Rosenstock, J.; Schwartz, S. L.; Clark, C. M.; Park, G. D.; Donley, D. W.; Edwards, M. B. Basal Insulin Therapy in Type 2 Diabetes: 28-Week Comparison of Insulin Glargine (HOE 901) and NPH Insulin. *Diabetes Care* **2001**, *24*, 631–636.
- (10) Havelund, S.; Plum, A.; Ribel, U.; Jonassen, I.; Vølund, A.; Markussen, J.; Kurtzhals, P. The Mechanism of Protraction of Insulin Detemir, a Long-Acting, Acylated Analog of Human Insulin. *Pharm. Res.* **2004**, *21*, 1498–1504.
- (11) Hermansen, K.; Davies, M.; Derezinski, T.; Martinez Ravn, G.; Clauson, P.; Home, P. A 26-Week, Randomized, Parallel, Treat-to-Target Trial Comparing Insulin Detemir With NPH Insulin as Add-On Therapy to Oral Glucose-Lowering Drugs in Insulin-Naive People With Type 2 Diabetes. *Diabetes Care* **2006**, *29*, 1269–1274.
- (12) Poon, K.; King, A. B. Glargine and Detemir: Safety and Efficacy Profiles of the Long-Acting Basal Insulin Analogs. *Drug, Healthcare Patient Saf.* **2010**, *2*, 213–223.
- (13) Pettus, J.; Santos Cavaioia, T.; Tamborlane, W. V.; Edelman, S. The Past, Present, and Future of Basal Insulins. *Diabetes/Metab. Res. Rev.* **2016**, *32*, 478–496.
- (14) Cheng, A.; Bailey, T. S.; Mauricio, D.; Roussel, R. Insulin Glargine 300 U/mL and Insulin Degludec: A Review of the Current Evidence Comparing These Two Second-Generation Basal Insulin Analogues. *Diabetes/Metab. Res. Rev.* **2020**, No. e3329.
- (15) Derewenda, U.; Derewenda, Z.; Dodson, E. J.; Dodson, G. G.; Reynolds, C. D.; Smith, G. D.; Sparks, C.; Swenson, D. Phenol Stabilizes More Helix in a New Symmetrical Zinc Insulin Hexamer. *Nature* **1989**, *338*, 594–596.
- (16) Choi, W. E.; Brader, M. L.; Aguilar, V.; Kaarsholm, N. C.; Dunn, M. F. Allosteric Transition of the Insulin Hexamer is Modulated by Homotropic and Heterotropic Interactions. *Biochemistry* **1993**, *32*, 11638–11645.
- (17) Jacoby, E.; Kruger, P.; Karatas, Y.; Wollmer, A. Distinction of Structural Reorganization and Ligand Binding in the T-R Transition of Insulin on the Basis of Allosteric Models. *Biol. Chem.* **1993**, *374*, 877–885.
- (18) Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. M. C.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D.; Sakabe, K.; et al. The Structure of 2Zn Pig Insulin Crystals at 1.5 Å Resolution. *Philos. T. Roy. Soc. B* **1988**, *319*, 369–456.
- (19) Hassiepen, U.; Federwisch, M.; Mülders, T.; Wollmer, A. The Lifetime of Insulin Hexamers. *Biophys. J.* **1999**, *77*, 1638–1654.
- (20) Berchtold, H.; Hilgenfeld, R. Binding of Phenol to R₆ Insulin Hexamers. *Biopolymers* **1999**, *51*, 165–172.
- (21) Turkenburg, J.; Whittingham, J.; Derewenda, U.; Derewenda, Z.; Dodson, E.; Dodson, G.; Smith, G.; Xiao, B. Structure Determination and Refinement of Two Crystal Forms of Native Insulins, 1998. www.rcsb.org/structure/1znj (accessed April 20, 2020).
- (22) Roy, M.; Brader, M. L.; Lee, R. W.; Kaarsholm, N. C.; Hansen, J. F.; Dunn, M. F. Spectroscopic Signatures of the T to R Conformational Transition in the Insulin Hexamer. *J. Biol. Chem.* **1989**, *264*, 19081–19085.
- (23) Jacoby, E.; Hua, Q. X.; Stern, A. S.; Frank, B. H.; Weiss, M. A. Structure and Dynamics of a Protein Assembly. ¹H-NMR Studies of the 36 kDa R₆ Insulin Hexamer. *J. Mol. Biol.* **1996**, *258*, 136–157.
- (24) Smith, G. The Phenolic Binding Site in T₃R₃ Insulin. *J. Mol. Struct.* **1998**, *470*, 71–80.
- (25) Swegat, W.; Schlitter, J.; Krüger, P.; Wollmer, A. MD Simulation of Protein-Ligand interaction: Formation and Dissociation of an Insulin-Phenol Complex. *Biophys. J.* **2003**, *84*, 1493–1506.
- (26) Vashisth, H.; Abrams, C. F. Ligand Escape Pathways and (Un)Binding Free Energy Calculations for the Hexameric Insulin-Phenol Complex. *Biophys. J.* **2008**, *95*, 4193–4204.
- (27) Hu, J.; Ma, A.; Dinner, A. R. Bias Annealing: A Method for Obtaining Transition Paths de Novo. *J. Chem. Phys.* **2006**, *125*, No. 114101.
- (28) Thiede, E. H.; Giannakis, D.; Dinner, A. R.; Weare, J. Galerkin Approximation of Dynamical Quantities Using Trajectory Data. *J. Chem. Phys.* **2019**, *150*, 244111.
- (29) 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. *J. Chem. Theory Comput.* **2021**, *17*, 2948–2963.
- (30) Ma, A.; Nag, A.; Dinner, A. R. Dynamic Coupling Between Coordinates in a Model for Biomolecular Isomerization. *J. Chem. Phys.* **2006**, *124*, No. 144911.
- (31) Hu, J.; Ma, A.; Dinner, A. R. A Two-Step Nucleotide-Flipping Mechanism Enables Kinetic Discrimination of DNA Lesions by AGT. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4615–4620.
- (32) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (33) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (34) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.
- (35) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible With the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *32*, 671–690.
- (36) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (37) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
- (38) 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.
- (39) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063.
- (40) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

- (41) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; et al. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16*, 670–673.
- (42) 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.
- (43) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (44) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (45) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (46) 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. *Nat. Methods* **2017**, *14*, 71–73.
- (47) Goga, N.; Rzepiela, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637–3649.
- (48) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (49) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (50) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (51) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (52) Marchi, M.; Ballone, P. Adiabatic Bias Molecular Dynamics: A Method to Navigate the Conformational Space of Complex Molecular Systems. *J. Chem. Phys.* **1999**, *110*, 3697–3702.
- (53) 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. *J. Phys. Chem. B* **2020**, *124*, 5571–5587.
- (54) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19011–19016.
- (55) 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.
- (56) Paul, F.; Wehmeyer, C.; Abualrous, E. T.; Wu, H.; Crabtree, M. D.; Schöneberg, J.; Clarke, J.; Freund, C.; Weikl, T. R.; Noé, F. Protein-Peptide Association Kinetics Beyond the Seconds Timescale from Atomistic Simulations. *Nat. Commun.* **2017**, *8*, No. 1095.
- (57) Paul, F.; Noé, F.; Weikl, T. R. Identifying Conformational-Selection and Induced-Fit Aspects in the Binding-Induced Folding of PMI from Markov State Modeling of Atomistic Simulations. *J. Phys. Chem. B* **2018**, *122*, 5649–5656.
- (58) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.
- (59) Birnbaum, D. T.; Dodd, S. W.; Saxberg, B. E. H.; Varshavsky, A. D.; Beals, J. M. Hierarchical Modeling of Phenolic Ligand Binding to 2Zn-Insulin Hexamers. *Biochemistry* **1996**, *35*, 5366–5378.
- (60) Birnbaum, D. T.; Kilcomons, M. A.; DeFelippis, M. R.; Beals, J. M. Assembly and Dissociation of Human Insulin and LysB28ProB29-Insulin Hexamers: A Comparison Study. *Pharm. Res.* **1997**, *14*, 25–36.
- (61) Feng, C.-J.; Sinitskiy, A.; Pande, V.; Tokmakoff, A. Computational IR Spectroscopy of Insulin Dimer Structure and Conformational Heterogeneity. *J. Phys. Chem. B* **2021**, *125*, 4620–4633.
- (62) Bentley, G.; Brange, J.; Derewenda, Z.; Dodson, E.; Dodson, G.; Markussen, J.; Wilkinson, A.; Wollmer, A.; Xiao, B. Role of B13 Glu in insulin assembly. *J. Mol. Biol.* **1992**, *228*, 1163–1176.
- (63) Bloom, C. R.; Choi, W. E.; Brzovic, P. S.; Ha Sheng-Tung Huang, J. J.; Kaarsholm, N. C.; Dunn, M. F. Ligand Binding to Wild-type and E-B13Q Mutant Insulins: A Three-state Allosteric Model System Showing Half-site Reactivity. *J. Mol. Biol.* **1995**, *245*, 324–330.
- (64) Mukherjee, S.; Mondal, S.; Deshmukh, A. A.; Gopal, B.; Bagchi, B. What Gives an Insulin Hexamer Its Unique Shape and Stability? Role of ten Confined Water Molecules. *J. Phys. Chem. B* **2018**, *122*, 1631–1637.
- (65) Mukherjee, S.; Deshmukh, A. A.; Mondal, S.; Gopal, B.; Bagchi, B. Destabilization of Insulin Hexamer in Water-Ethanol Binary Mixture. *J. Phys. Chem. B* **2019**, *123*, 10365–10375.
- (66) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; et al. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62–74.