Supplementary Material for

Kinetics of Phenol Escape from

the Insulin R_6 Hexamer

Adam Antoszewski,[†] Chatipat Lorpaiboon,[†] John Strahan,[†] and

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

E-mail: dinner@uchicago.edu

Nomenclature. When numbering the six phenols, we followed the conventions in the PDB file. This differs from the numbering convention in ref. 1, which instead assigns phenols 1, 2, and 3 to the "top" trimer, and phenols 4, 5, and 6 to the "bottom" trimer. Furthermore, our nomencalutre for protein chains also differs from ref. 1. In particular, for WT insulin, we choose the nomenclature such that both the A and B chain belong to the insulin monomer closest to phenol 4, which we use to study the phenol unbinding process for WT insulin. From there, the naming trends are identical to the PDB/literature naming convention, with chains ranging from A to L - only the starting point (which chain we designate as A) is shifted. For the mutant systems, we simulate the release of phenol 2, instead of phenol 4. For these systems, then, we followed a similar convention, and denote the monomer closest to phenol 2 as having the A and B chains.

Selecting starting structures for unbiased simulations. To ensure sampling in all areas of our CV space even with relatively short trajectories (40 ns), we initialize the unbiased simulations as follows. We first run a large number of adiabatic-bias molecular dynamics (ABMD) simulations to bias the system toward phenol release, as discussed in the main text. We then define a grid of $10 \times 10 \times 10$ points that covers the sampled regions and find the single frame from our driven database closest to each point (Figure S1). Although there are 1000 grid points, the same structure can be the closest to more than one grid point. In the WT case, we obtain 326 unique structures.

DGA details. DGA requires the use of a basis set. We used a basis set of 298 modified pairwise distances and one constant function as described in the main text. A summary of the distances is given in Table S1. For categories involving the phenol, we measure distances from both C_1 and C_4; thus there are $2 \times 22 = 44$ distances to the C_α atoms of the 22 residues in the binding pocket and $2 \times 4 = 8$ distances to the C_α atoms of the 4 gatekeeper residues for PW1 and PW4. The PW3 gatekeeper residues in Table S1 are A2, A19, and B25.
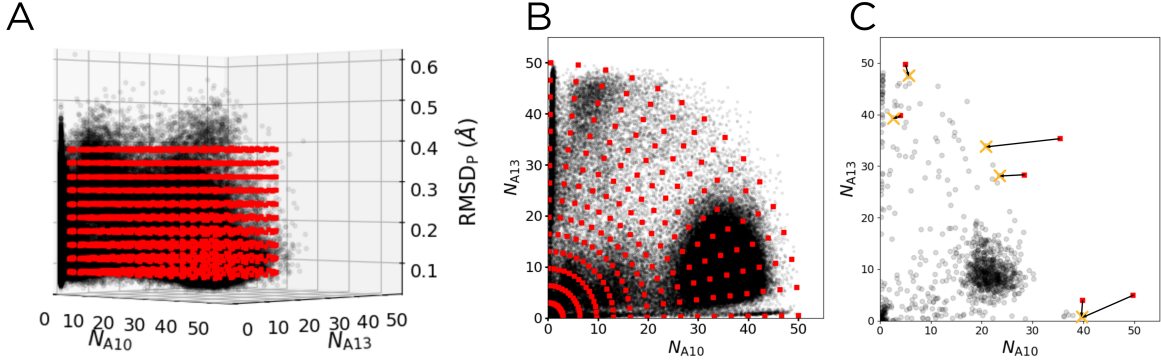
Figure S1: A schematic showing how we chose unique starting structures for the unbiased sampling. (A) 3D representation in the space of N_{A10}, N_{A13}, and RMSP$_P$. Data from our ABMD database is shown by the black dots, and our desired starting points are shown by the red squares. (B) A two-dimensional slice of (A), more clearly showing the r and θ dependence of our desired starting points. (C) A schematic illustrating how we select the closest structures to each desired point. The frame from the ABMD data set closest to each desired starting point is represented by the orange X. For clarity, we only display six desired starting points that lead to five unique starting structures.

Table S1: Description and number of pairwise distances used as inputs to make our DGA basis functions. Note that to make the eventual 299-dimensional set of basis functions, we also include the constant function.

Type of Distance	Number
Phenol C_1/C_4 atoms to Binding Pocket C_α atoms	44
Binding Pocket C_α atoms to Binding Pocket C_α atoms	231
Phenol C_1/C_4 atoms to PW1/PW4 gatekeeper C_γ atoms	8
PW1/PW4 C_γ atoms to PW1/PW4 gatekeeper C_γ atoms	6
Phenol C_1/C_4 atoms to PW3 gatekeeper C_α atoms	6
PW3 gatekeeper C_α atoms to PW3 gatekeeper C_α atoms	3
Total	298

In this work, we modified the guess functions $\psi_{\text{binding}}(x)$ and $\psi_{\text{unbinding}}(x)$ from their definitions in ref. 2, so $\psi_{\text{binding}}(x) = 1 - \psi_{\text{unbinding}}(x)$:

$$\psi_{\text{binding}}(x) = \frac{d^2_{\text{unbound}}}{d^2_{\text{bound}} + d^2_{\text{unbound}}} \tag{1}$$

$$\psi_{\text{unbinding}}(x) = \frac{d^2_{\text{bound}}}{d^2_{\text{bound}} + d^2_{\text{unbound}}} \tag{2}$$

These choices ensure that $q_{\text{unbinding}} = 1 - q_{\text{binding}}$. Above, d_{bound} and d_{unbound} are the smallest Euclidean distances in the 298-dimensional space of pairwise distances (Table S1) from any point in the reactive domain (i.e., outside the bound and unbound states) in the data set to

any point in the bound and unbound states, respectively.

Since DGA does not enforce the fact that committors are probabilities and thus must be between zero and one, it produces estimates between –0.2 and 1.2. We shift those below zero to zero and those above one to one before using them for further analysis (plotting and reactive current calculations).

In addition, one of the essential parameters for DGA is the lag time[2] (see also refs. 3 and 4). We calculated statistics for lag times ranging from 10 ps to 10 ns. For WT insulin, we found that the $q_{\text{unbind}} = 0.5$ surface was approximately constant as the lag time changed from 500 ps to 5 ns. For the mutant simulations, this was only the case for lag times greater than 1.25 ns. Similar behavior was observed for the relative weights of the six pathways (Figure S2).
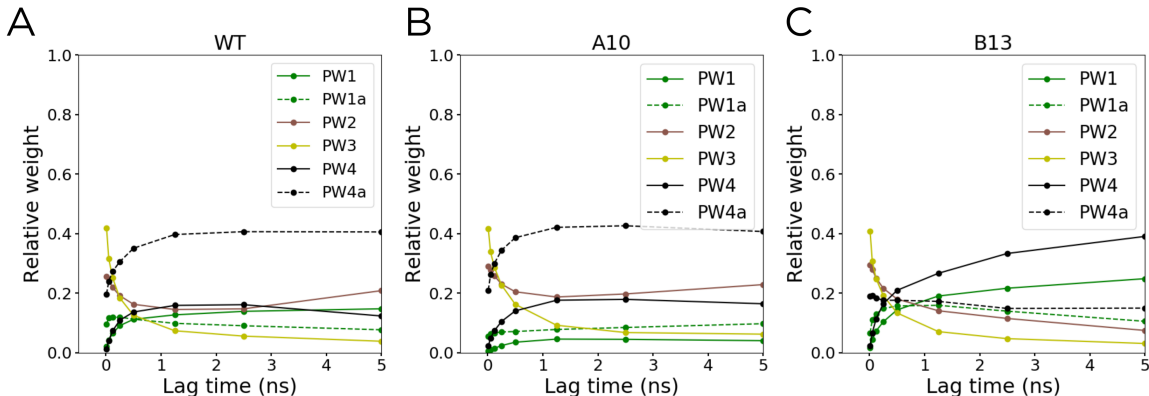

Figure S2: The relative weight of the six identified pathways as a function of lag time for (A) WT insulin, (B) A10 Ile → Val insulin, and (C) B13 Glu → Gln insulin.

Cooperativity of phenol release. We examined whether there was evidence in our ABMD simulations of positive intra-trimer cooperativity and negative inter-trimer cooperativity for phenolic binding, as suggested by stopped-flow spectroscopy[5] and isothermal titrating calorimetry.[6] In addition to the 276 simulations described in the main text, which favored increasing the distance d_n ($1 \leq n \leq 6$) between each phenol and the closest bound zinc ion, we also performed 140 simulations that instead favored decreasing the number of non-hydrogen contacts between the protein and each phenol. 10 simulations (each of length

5 ns) were run for each of 14 force constants evenly spaced between 3×10^{-9} to 16×10^{-9} kJ/mol. In general, fewer phenols were released, and none of the 140 simulations led to dissociation of all six phenols, compared to 77/276 for the simulations biasing on d_n.

Plots summarizing the sequences of release for the ABMD simulations are shown in Figure S3. For clarity, transitions observed in less than 10% of each set of simulations are not drawn. For both sets of simulations, the most common sequence of release is phenol 4, followed by phenol 6 (both in the same trimer, here labeled trimer 1). For the simulations biasing on the number of non-hydrogen contacts between phenol and protein (Figure S3B), the next most likely release is that of phenol 2, also in trimer 1. For the simulations biasing of the distance between the phenol and the nearest bound zinc, the most likely third release is either phenol 2 or phenol 1, which is in trimer 2. These data support the existence of negative inter-trimer cooperativity, as the phenols in trimer 1 are preferentially released over the phenols in trimer 2. Because we do not see any large-scale motions that correlate with the sequence of release, we presume that the cooperativity in the simulations reflects subtle differences between the two trimers in the starting structure, but which differences are most important is not apparent.
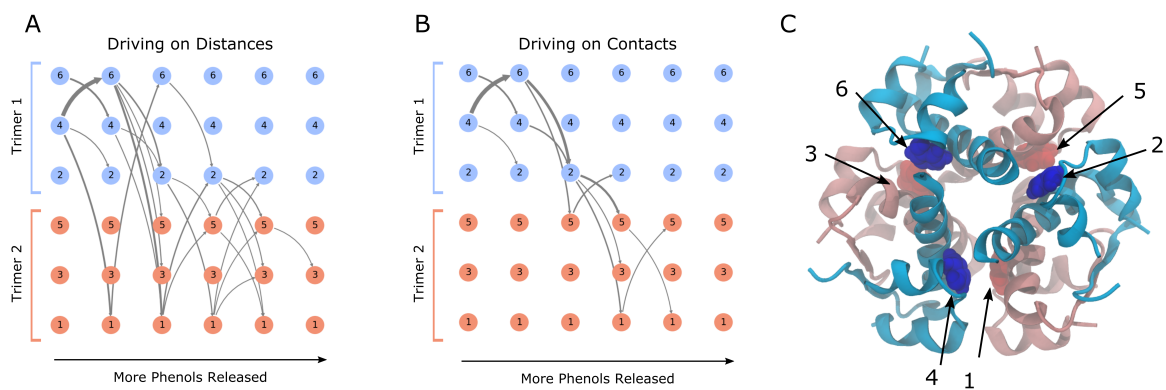

Figure S3: The sequence of phenol release for the ABMD simulations biasing on (A) the distance between each phenol and the closest bound zinc, and (B) the number of non-hydrogen contacts between phenol and protein. Phenols bound to trimer 1 and trimer 2 are represented by the blue and red circles, respectively. The size of an arrow represents the relative weight of the indicated transition. (C) The hexamer, colored as in (A) and (B), with phenols labeled.
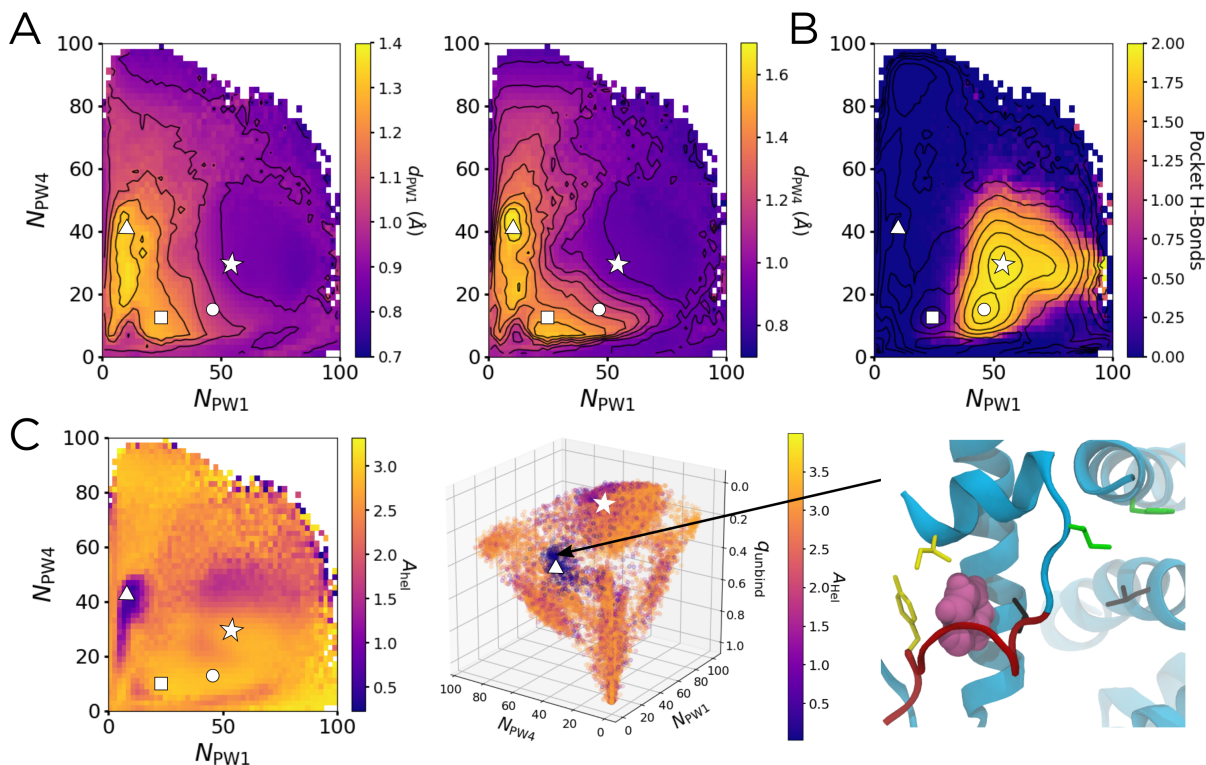
Figure S4: Averages of observables, taken from our unbiased data set, associated with different aspects of channel opening, and projected using N_{PW1}, N_{PW4}, and N_{PW3}. The star, circle, square, and triangle mark the same landmarks as in Figures 2 and 3 in the main text. (A) The average of d_{PW1} (left) and d_{PW4} (right) as a function of N_{PW1} and N_{PW4}, with contours shown every 0.1 Å. (B) The average of N_{HP}, with contours from the WT insulin PMF overlaid. (C) The average of A_{hel} as a function of N_{PW1} and N_{PW4} (left) and as a scatter plot in the space of N_{PW1}, N_{PW4}, and $q_{unbinding}$ (middle). A structural representation of the melted C-terminal A-chain α helix (red) along PW3 (triangle) is shown in the right panel. Gatekeeper residues are shown as in the main text.

Channel opening analysis. Using our unbiased data, we can characterize channel opening by describing d_{PW1} and d_{PW4}, the distances between the α carbons of the gatekeeper residues along PW1 and PW4, respectively (Figure S4A). We also calculate N_{HP}, 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} (Figure S4B). Finally, we calculate A_{hel}, the helicity of the C-terminal A-chain α helix, A13-A21 (Figure S4C). This is the effective number of six-residue segments in the selection in an idealized α helical conformation, based on RMSD;[7] a value of four corresponds to a fully structured C-terminal α helix, while a value of zero corresponds to a fully melted helix.

To further probe the stability of the phenolic escape channel, normal mode analysis was

performed on the crystal structure of the WT hexamer. We used WebNMA 3.3,[8] which creates an elastic network model from all of the α carbons in a protein, and solves for the normal modes of this oscillator system. The normalized squared displacements of these α carbons for the four lowest modes (excluding the translational/rotational modes and any identical modes due to the symmetry of the system) are shown in Figure S5, where the displacements have been averaged across all 6 monomers to provide a representative set of displacements.

The areas of highest flexibility for the dominant normal modes correlate well with the A-chain α helices and the B-chain β turn, shown in gray on the left and right, respectively. The spikes at B1 and B31 are attributed to the large flexibility of terminal residues. These data suggest an area of flexibility between the A chains and adjacent B-chain β turns in the hexameric structure, the same area that forms the phenolic binding pocket/escape channel.
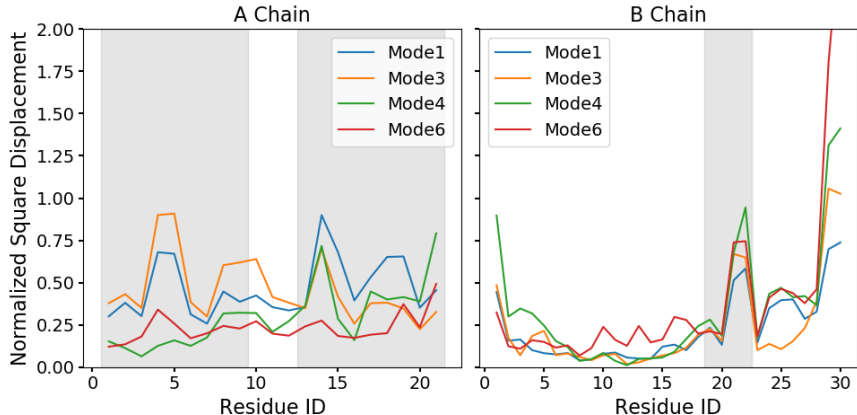

Figure S5: Squared displacements from normal mode analysis, where each mode has been normalized so that the sum of the displacements equals 100. The displacements are averaged across the six monomers. The gray areas mark relevant secondary structure elements: the N and C terminal α A-chain helices, A1-A9 and A13-21, respectively, and the B-chain β turn, B18-B22.

Comparing the REUS PMF and the DGA PMF. To validate the PMF generated by DGA, we ran Replica Exchange Umbrella Sampling (REUS) simulations and used them to compute an independent PMF. To generate starting structures for the REUS, we first created a 20×20 evenly spaced grid in the cylindrical space of (r, θ), where $N_{\mathrm{PW1}} = r \cos \theta$ and $N_{\mathrm{PW4}} = r \sin \theta$. Structures closest to each one of these grid points were drawn from our

unbiased simulation database and used to initialize the windows. Harmonic biases were then placed on N_{PW1} and N_{PW4}, with window strengths, k, set by applying eq. 4 in ref. 9, with a maximum possible $k = 3$ kJ/mol. The simulation and 2D exchange procedure of ref. 9 was also followed, simulating for a total of 2 ns per window, for an aggregate sampling time of 800 ns. The PMF was constructed from this sampling using the Eigenvector Method for Umbrella Sampling (EMUS)[10] extended to replica exchange data.[9] The resulting PMF and related statistics are shown in Figure S6.
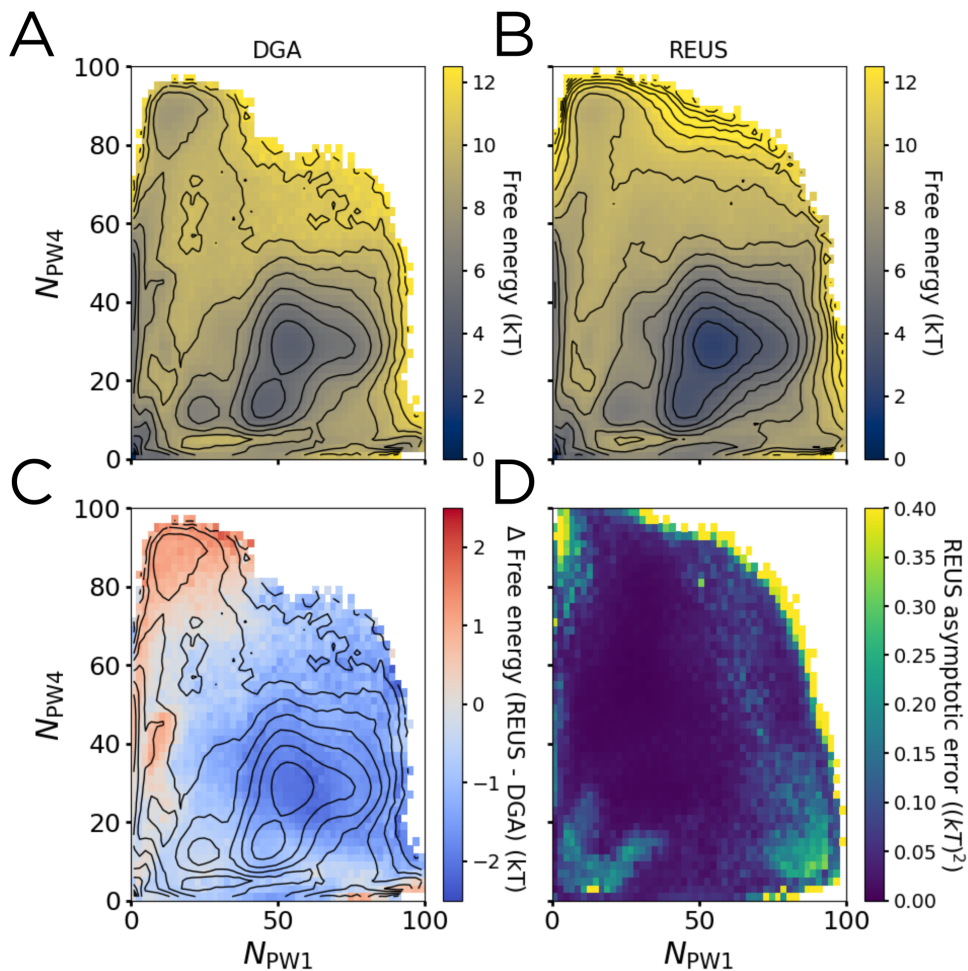

Figure S6: Comparison of PMFs generated using REUS and DGA. (A) DGA and (B) REUS PMFs with contours shown every 1 k_BT. (C) The difference of the two PMFs , subtracting (A) from (B), with contours from the DGA PMF superimposed to guide the eye. (D) The asymptotic variance of the REUS PMF.

The qualitative features of the DGA and REUS PMFs are quite similar, including stable and metastable basin positions. Relative to the REUS PMF, the DGA PMF appears to

overestimate the free energy of the bound state by approximately 2 k_BT and to underestimate the free energy in the upper left corner of the PMF (along PW4) by approximately 1-2 k_BT. However, we take the overall agreement to be an indication that our sampling is sufficient.
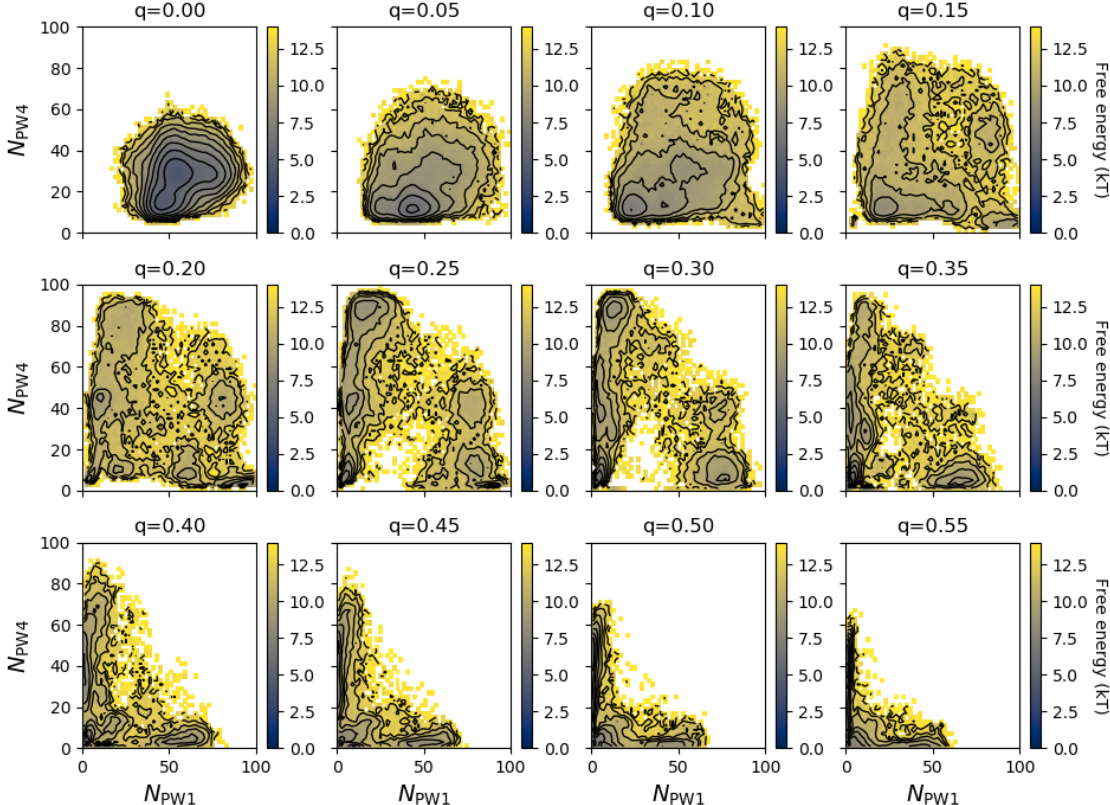

Figure S7: The PMF in the space of N_{PW1}, N_{PW4}, and q_{unbind}, shown at indicated slices of q_{unbind}. Contours spacing is 1 k_BT. The minimum free energies for the panels in the first row ($q = 0$ to $q = 0.15$) are 4.0, 5.7, 7.0, and 8.2 k_BT, from left to right.

We also projected the DGA-generated PMF into the space of N_{PW1}, N_{PW4}, and q_{unbind} as discussed in the main text. This PMF is shown for multiple different slices of q_{unbind} in Figure S7. In this representation, the free energy barrier for PW4 is the same as for the 2D projection (approximately 5-6 k_BT). In contrast, the free energy barrier for PW1 is somewhat higher (6-8 k_BT in the 3D case compared with 4-5 k_BT in the 2D case). This is not unexpected, as the 3D projection allows us to more easily separate PW1 from PW1a, and from the area of the PMF associated with the (energetically stable) His[F5] ring flip in the bound state. Despite these minor differences, we conclude that the free energy barriers for PW1 and PW4 are comparable.
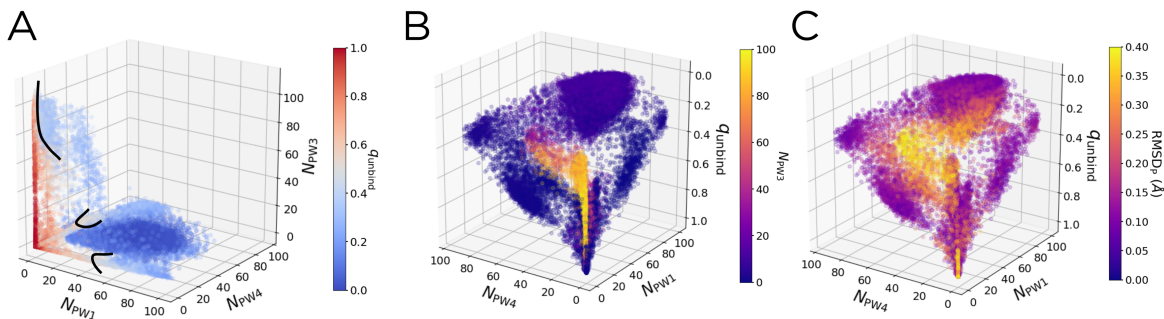
Figure S8: The committor and other statistics projected into three dimensions. (A) The unbinding committor q_{unbind} projected into the space of N_{PW1}, N_{PW4}, and N_{PW3}. The $q_{unbind} = 0.5$ transition state ensemble is highlighted by the black arcs. The large arc near $N_{PW3} \approx 80 - 100$ corresponds to the transition state along PW3. The two small arcs near $N_{PW1} \approx 60$ and $N_{PW4} \approx 60$ correspond to the transition states along PW1 and PW4, respectively. (B) The value of N_{PW3} projected into N_{PW1}, N_{PW4} ,and $q_{unbinding}$. (C) The value of $RMSD_P$ projected into the same space as (B).

Describing PW3 in 3D CV space. To help determine the transition state ensemble for PW3, we projected the unbinding committor q_{unbind} onto the space of N_{PW1}, N_{PW4}, and N_{PW3} (Figure S8). This shows that the $q_{unbind} = 0.5$ surface occurs where $N_{PW3} \approx 80 - 100$, the maximum value N_{PW3} obtains along PW3. This is in contrast to the transition states along PW1 and PW4, which occur when $N_{PW1} \approx 60$ or $N_{PW4} \approx 60$. These states correspond to phenol having already partially escaped from the crystallographic binding pocket, which occurs once the number of contacts with the corresponding gatekeeper residues have begun to decrease. In contrast, along PW3, the phenol has to push through a sterically occluded region of the A chain to reach the gatekeeper residues, meaning that the phenol has already partially escaped the binding pocket by the time it reaches them. Thus, despite the seemingly different locations of the $q_{unbind} = 0.5$ surface along PW1, PW3, and PW4, all of the transition states exist once the phenol has partially escaped from the binding pocket.

Furthermore, the projection of N_{PW3} into the space of N_{PW1}, N_{PW4}, and $q_{unbinding}$ (Figure S8) shows where this region of large N_{PW3} occurs in the projection used in the main text, further verifying the pathway definitions described there. Similarly, $RMSD_P$ projected onto the same space shows that the areas of elevated $RMSD_P$ correspond to PW1a, PW4a, and PW3, with the highest values corresponding to PW3. This is consistent with both the channel opening (along PW1a and PW4a) and the melting of the A-chain C-terminal α

helix (along PW3) previously noted.

Choice of the dividing surface.   For measuring the relative weights of different binding pathways, one needs to introduce a dividing surface between the bound and unbound states, and to partition that surface into patches corresponding to different pathways. The relative weight of each specific pathway is the sum of the current flowing normal through its patch on the surface, normalized by the total amount of current flowing normal to the full surface. The binding current, J_{bind}, was calculated with a lag time of 500 ps in the space of N_{PW1}, N_{PW4}, and q_{bind}. This was binned into a $50 \times 50 \times 50$ uniform grid, covering $0 \leq N_{\text{PW1}}, N_{\text{PW4}} \leq 100$, and $0 \leq q_{\text{bind}} \leq 1$. After binning in this CV space, the results were smoothed with a Gaussian filter, using a kernel with standard deviation of 4 bins.

We chose to use the binding statistics (J_{bind} and q_{bind}) to determine the relative weights of the six pathways, instead of using the unbinding statistics (J_{unbind} and q_{unbind}) as in the rest of our analysis. In the limit of infinite sampling, one would expect the binding and unbinding statistics to mirror each other exactly, as the dynamics are reversible. However, we found that finite sampling led to small but noticeable differences in the reactive currents (J_{bind} and J_{unbind}) that made the determination of the six pathways from J_{unbind} more difficult (Figure S9). In particular, the area of high reactive current along the N_{PW4} axis is much more diffuse in Figure S9B than Figure S9A, as PW4 and PW4a blur together when using the unbinding statistics. While slightly less visible due to the color scale, a similar blurring occurs for PW1 and PW1a when using the unbinding statistics instead of the binding statistics.

We use a plane of constant q_{bind} as the dividing surface. Based on visual inspection of Figure S10, we determined that $q_{\text{bind}} = 0.63$ provided the best separation of the pathways. We used it to define the patches for both WT and mutant insulins as follows:

- PW4: $N_{\text{PW4}} > 66$

- PW1: $N_{\text{PW1}} > 64$ and not the above

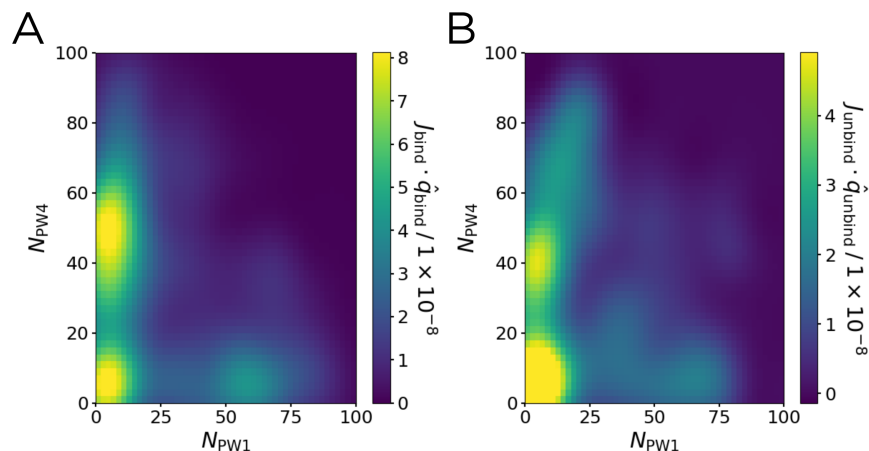- PW4a: $N_{\text{PW4}} > 32$ and not any of the above

Figure S9: Comparison of reactive currents for the unbinding and binding directions. In each case, we show the dividing surface that best separates the six pathways. (A) The q_{bind} component of J_{bind} at $q_{bind} = 0.67$. (B) The q_{unbind} component of J_{unbind} at $q_{unbind} = 0.33$.

- PW1a: $N_{PW1} > 40$ and not any of the above

- PW3: $N_{PW1} < 20$ and $N_{PW4} < 16$

- PW2: None of the above

These choices are consistent with our ABMD simulations as well. Using these patches, we calculated the relative weights of different pathways as a function of the value of q_{bind}, shown in Figure S11. Over the range of q_{bind} values shown, we find the results to be insensitive to the specific choice of dividing surface.

Table S2: The relative weights for phenol unbinding along our six identified pathways, measured after removing trajectories along the described pathways.

Data set	PW1 (%)	PW1a (%)	PW2 (%)	PW3 (%)	PW4 (%)	PW4a (%)
Full 692 trajectories	11.2	11.2	16.3	12.6	13.7	35.0
-25 PW1	11.3	9.7	16.6	13.2	13.9	35.2
-50 PW1	9.7	9.7	17.0	13.3	13.9	36.2
-25 PW4	10.7	11.6	15.9	13.1	14.1	34.5
-50 PW4	11.3	11.3	16.6	13.1	13.1	34.7

To probe the robustness of these relative weights, we measured them as we varied the sampling distribution. In particular, we identified 50 trajectories from our unbiased data set (692 total trajectories) which started along PW1/PW1a, and we identified a similar set of
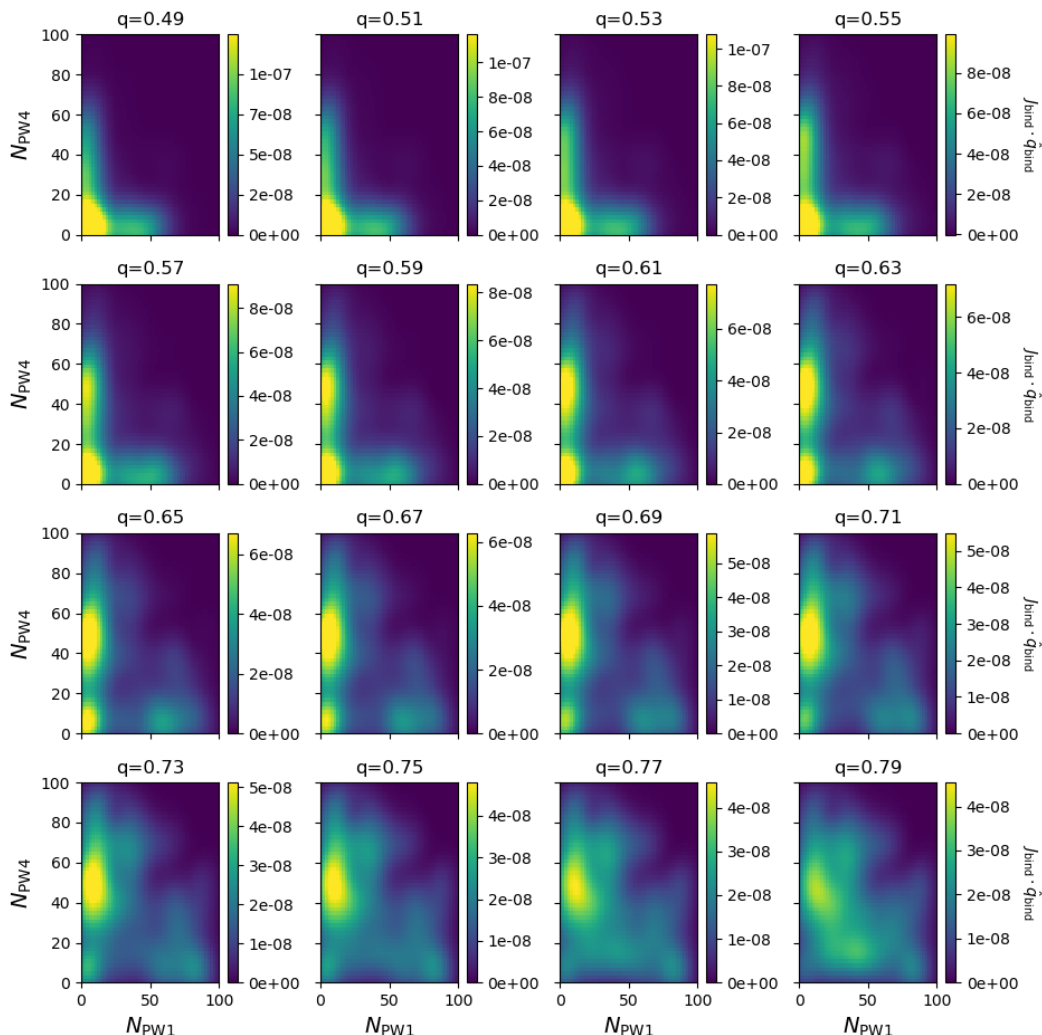
Figure S10: The q_{bind} component of J_{bind} at various different slices of q_{bind}.

50 trajectories that started along PW4/PW4a. We removed some of these trajectories and re-measured the relative weights for phenol unbinding along each of our six pathways (Supplemental Table S2). As we remove trajectories, from either PW1/PW1a or PW4/PW4a, the relative weights of the six observed pathways change very little (with all changes being $\leq 1.5\%$). This provides evidence that our sampling is robust enough to converge the relative weight estimates from DGA.

Mutant simulation details. In addition to that for the wildtype protein, we constructed models for two hexamers with one point mutation each (A10 Ile $\rightarrow$ Val, B13 Glu $\rightarrow$ Gln). In
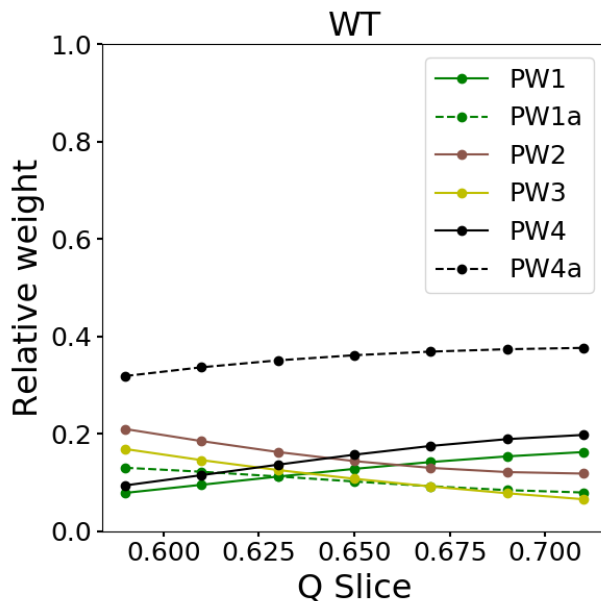
Figure S11: The relative weights for the six identified pathways as a function of the value of q_{bind} for the dividing surface.

each case, we used CHARMM-GUI (version 3.2) to modify the 1ZNJ crystal structure at the six sites that differed in sequence and then solvated the system following a similar procedure to that described for WT. Below, we will comment on any differences between the workflow described in the main text and the workflow we followed for each of these mutants.

For the A10 Ile $\rightarrow$ Val mutant, 54 K$^+$ and 44 Cl$^-$ additional ions were aded to achieve a neutral 150 mM KCl solution, for a total of 51,060 atoms. ABMD simulations were used only to generate initial structures for the unbiased simulations. 140 ABMD simulations of 5 ns each were originally run: 28 simulations for each of five force constants between $k = 6$ and $k = 14$ kJ / (mol nm). For two of these simulations, the binding pocket, which was open after we solvated the system, was observed to close. From each of these closed structures, 16 additional ABMD simulations were run at all five of the previously described force constants, plus $k = 16$ and $k = 18$ kJ / (mol/nm), as stronger force constants were needed to encourage dissociation once the channel had been closed. Finally, supplemental ABMD trajectories (28 at $k = 6$ and 28 at $k = 14$ kJ / (mol nm)) were run from one existing trajectory that sampled PW3 as identified by ref. 1. In total, this driven data set thus

consisted of $28 \times 5 + 32 \times 7 + 28 \times 2 = 420$ trajectories, each of length 5 ns.

The grid in the cylindrical space of our CVs (again with $N_{A10} = r \cos \theta$, $N_{A13} = r \sin \theta$, and $RMSD_P$) was changed to a $13 \times 13 \times 10$ grid in $(r, \theta, RMSD_P)$ space, in order to generate a similar number of unbiased simulations as we did previously; 356 unique structures were selected from the ABMD database. From each of these points, two 40 ns simulations were launched. To further sample PW3, as before, 20 structures from that pathway were identified from the ABMD database, and from each of those structures, two 40 ns simulations were launched. Thus, this unbiased data set consisted of 752 simulations, each of length 40 ns, for an aggregate simulation time of 30.08 μs.

For the B13 Glu $\rightarrow$ Gln mutant, 47 K$^+$ and 47 Cl$^-$ additional ions were added to achieve a neutral 150 mM KCl solution, for a total of 51,196 atoms. Since the mutated glutamine residues could form a series of hydrogen bonds with one another, PyMOL 2.3.0[11] was used to individually rotate each side chain to flip the carbonyl and amine groups, creating $2^6 = 64$ different conformations. We used the steepest descent algorithm to minimize the energy of each of these conformations until the maximum force felt by the system was below 1000 kJ/mol nm. The lowest energy conformation was selected for all further simulations of this mutant.

To generate initial structures for the unbiased simulations, 140 ABMD simulations of 5 ns each were originally run: 28 simulations for each of five force constants between $k = 15$ and $k = 23$ kJ / (mol nm). From these simulations, four structures were chosen that exhibited channel opening. For each of these structures, 40 additional ABMD simulations were run at $k = 23$ kJ / (mol nm), for a total of $28 \times 5 + 40 \times 4 = 300$ trajectories, each of length 5 ns. We used a $15 \times 15 \times 10$ grid in (r, θ, z) space to select 331 unique starting structures. In addition, 24 structures were chosen from the ABMD database that exhibited phenol release along PW3. From each of these 355 structures, two unbiased simulations of length 40 ns were launched, leading to a database of 710 trajectories (aggregate length 28.4 μs).

Unimolecular and bimolecular rate constant estimates. The results for unimolecular rates/equilibrium constants are presented as a function of lag time in Figure S12. To calculate the bimolecular association rate, we must account for diffusion. To do this, we adapt the approach of McCammon and coworkers.[12] We define two distances b and c ($b < c$), each measured from the center of mass of the two central Zn^{2+} ions. The distance b should be large enough that interactions between the phenol and hexamer can be considered centrosymmetric; based on our definition of the unbound state in the main text ($N_{PW1} < 2$, $N_{PW4} < 2$, and $N_{Prot} < 5$), we set b to be 3.3 nm. The distance c was set to 5.3 nm, similar to previous simulations.[13] For the statistics in this section, we redefine the unbound state to be the center of mass of the phenol at radii larger than c. The probability of diffusing from a sphere of radius c to a sphere of radius b is $\Omega = b/c$.[14] Using this, the association rate constant is $k'_{bind} = 4\pi Dbp$, where D is the diffusion constant and p is the probability of ultimately binding once the phenol first reaches a distance b. Adapting results from ref. 12, this quantity can be calculated from DGA using the relation:

$$p = \frac{q'_{bind}}{1 - \Omega(1 - q'_{bind})}, \tag{3}$$

where q'_{bind} is the binding committor with the bound state defined as in the main text and the unbound state defined as above. The factor p can also be used to correct the dissociation rate for re-binding: $k'_{unbind} = k'_{DGA}(1 - \Omega p)$, where k'_{DGA} is the unimolecular unbinding rate constant computed by DGA with the bound state as defined in the main text and the unbound state defined as above. Note that q'_{bind} and k'_{unbind} differ from q_{bind} and k_{unbind} in the main text owing to the redefinition of the unbound state.

The diffusion constant was calculated as a sum of the self-diffusion constants for the phenol and the hexamer determined separately, $D = D_{phenol} + D_{hexamer}$. The phenol was solvated and equilibrated using the same procedures as described for the full hexamer in the main text: the total number of atoms was 48,453, including 46 K^+ and 46 Cl^- ions. The
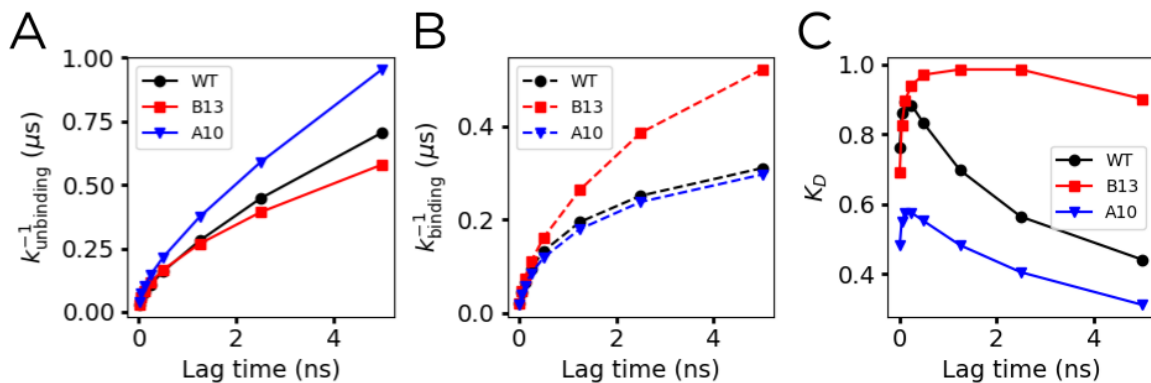
Figure S12: Unimolecular rate constants and their ratio at a range of DGA lag times for WT insulin and the two mutants, $Ile^{A10} \rightarrow Val^{A10}$ (A10 in the legend) and $Glu^{B13} \rightarrow Gln^{B13}$ (B13 in the legend). We show the inverse unbinding rate constant, $k_{unbinding}^{-1}$ (A), the inverse binding rate constant $k_{binding}^{-1}$ (B), and their ratio $K = k_{unbinding}/k_{binding}$ (C).
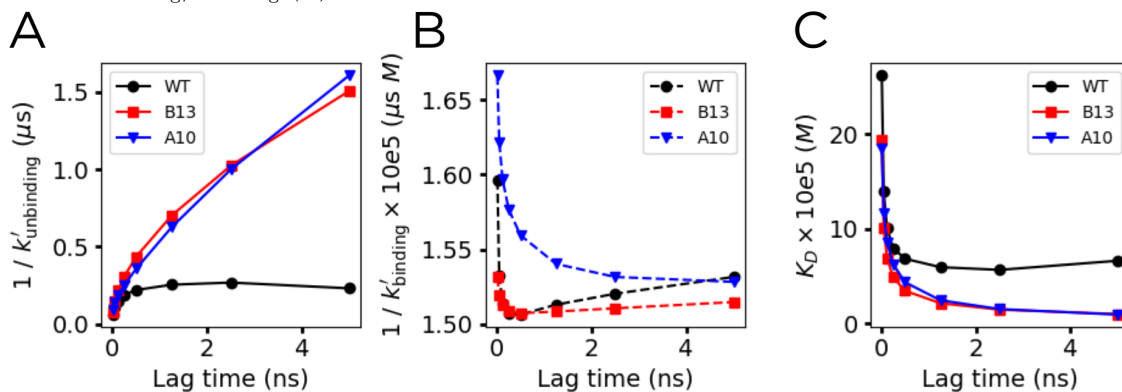

Figure S13: Bimolecular rate constant estimates as functions of DGA lag times for WT insulin and the two mutants, $Ile^{A10} \rightarrow Val^{A10}$ (A10 in the legend) and $Glu^{B13} \rightarrow Gln^{B13}$ (B13 in the legend). We show (A) the inverse unbinding rate constant, $1/k'_{unbind}$, (B) the inverse binding rate constant $1/k'_{bind}$, and (C) the dissociation constant, $K_D = k'_{unbind}/k'_{bind}$.

box size was $(7.839 \text{ nm})^3$. Both diffusion constants were determined by measuring the slope of the mean squared displacement of the center of mass of the relevant species, once it had achieved linearity. For this, the phenol was simulated for 20 ns at 303.15 K, and D_{phenol} was $2.7 \times 10^{-5} \text{ cm}^2/\text{s}$. The diffusion constant $D_{hexamer}$ was measured for each insulin species (WT, A10 Ile $\rightarrow$ Val, and B13 Glu $\rightarrow$ Gln) by analyzing a 40 ns trajectory with all six phenols bound, and D_{WT}, D_{A10}, and D_{B13} were 4×10^{-7}, 1×10^{-7}, and $2 \times 10^{-7} \text{ cm}^2/\text{s}$, respectively.

The bimolecular results are presented as a function of lag time in Figure S13, and for two representative lag times (as in the main text) in Table S3. The results are in good agreement

with measured values[15] given expected sources of error. First, the choice of b is likely too short for orientational effects to be negligible, which will tend to increase p. Second, diffusion in TIP3P water at 298 K is known to be a factor of 2.45 too high.[16] Third, both k'_{bind} and k'_{unbind} are products, compounding statistical uncertainties. Fourth, the experiments were performed with Co^{2+} and p-aminobenzoate as bound ligands, instead of Zn^{2+} and Cl^- as in these simulations.

Table S3: Bimolecular rate constant estimates: The inverse unbinding rate constant, $1/k'_{\text{unbind}}$, the inverse binding rate constant $1/k'_{\text{bind}}$, and the dissociation constant, $K_D = k'_{\text{unbind}}/k'_{\text{bind}}$. Ranges derive from taking lag times between 500 ps and 1.25 ns.

Statistic	WT	A10	B13
$1/k'_{\text{unbinding}}$ (μs)	0.22 - 0.25	0.36 - 0.63	0.44 - 0.70
$1/k'_{\text{binding}}$ (μs M)	$1.5 - 1.5 \times 10^{-5}$	$1.5 - 1.6 \times 10^{-5}$	$1.5 - 1.5 \times 10^{-5}$
K_D (M)	$6.0 - 6.8 \times 10^{-5}$	$1.5 - 2.6 \times 10^{-5}$	$3.5 - 2.2 \times 10^{-5}$
K_D experiment (M)[15]	18×10^{-5}	N/A	25×10^{-5}

Mutant DGA parameter choices. To the greatest extent possible, we used the same simulation parameters for DGA for WT and mutant insulins. Differences were as follows.

For the A10 Ile $\rightarrow$ Val mutant, the bound state was redefined to reflect the means and standard deviations of N_{PW1}, N_{PW4}, and RMSD_{P} in a 10 ns equilibrium simulation of the mutated structure with a closed channel. The means were 52.3, 29.8, and 0.076 Å, and the standard deviations were 5.18, 4.03, and 0.014 Å.

Comparing the PMFs for WT insulin and B13 Glu $\rightarrow$ Gln insulin, the most stable basin moves to where His^{F5} is flipped inward, facing the phenol (Figure S14). As a result, the bound state shifted; the corresponding CV means were 77.3, 26.6, and 0.096 Å, and the CV standard deviations were 5.92, 4.40, and 0.016 Å. For the B13 Glu $\rightarrow$ Gln mutant, we also used a longer lag time (1.25 ns compared with 500 ps for WT insulin) because the pathway weights converged more slowly (Figure S2). In turn, because the reactive current J_{bind} becomes noisier as the lag time increases, the standard deviation of the Gaussian filter for J_{bind} was increased from 4 to 5 bins.

Statistics for mutant insulins. Comparing the WT and A10 Ile $\rightarrow$ Val PMFs (Figure S14A), we see that this mutation stabilizes the basin corresponding to the partially open escape channel (see main text Figure 3 for the location of this region), making it lower in free energy than the normal bound state. At the same time, the mutation causes a shift in the binding reactive current from PW1, PW1a, and PW2 to PW3 and PW4 (Figure S14). The current flowing through PW4a generally shifts slightly upward in N_{PW4}. Overall, the total amount of reactive current flowing through the dividing surface is very similar to that for WT insulin, and as result the rates are similar as well (Figure S12).
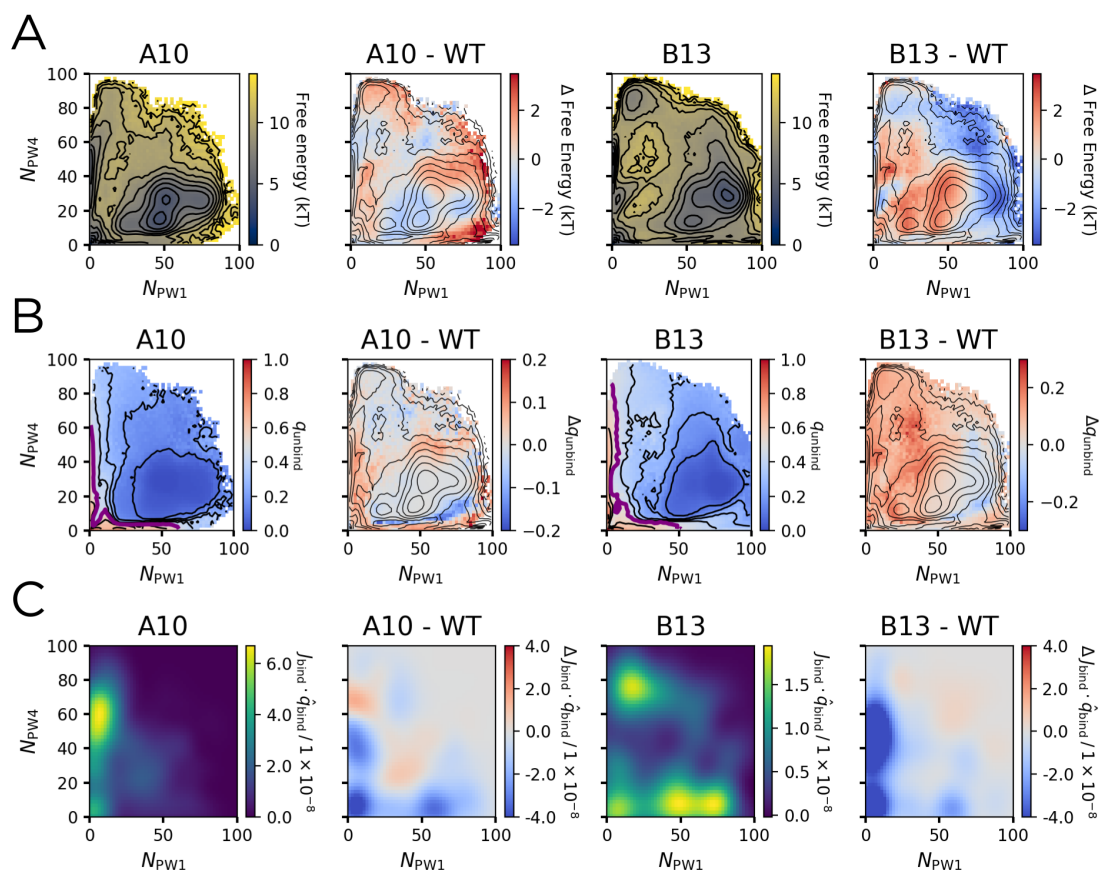

Figure S14: Comparing statistics for phenol escape between WT and mutant insulins. In each row, the first and third panels correspond to the Ile$^{A10} \rightarrow$ Val^{A10} mutant and Glu$^{B13} \rightarrow$ Gln^{B13} mutant, respectively. The second and fourth panels are the differences between the described mutant and WT insulin. (A) The potential of mean force, with contours shown every $k_B T$. For the differences, the contours from the WT insulin PMF are overlaid. (B) The average unbinding committor q_{unbind}, with contours shown every 0.1, and the $q_{\text{unbind}} = 0.5$ surface shown in purple. For the differences, the contours from the WT insulin PMF are overlaid. (C) The q_{bind} component of the binding reactive current J_{bind}, taken when $q_{\text{bind}} = 0.63$.
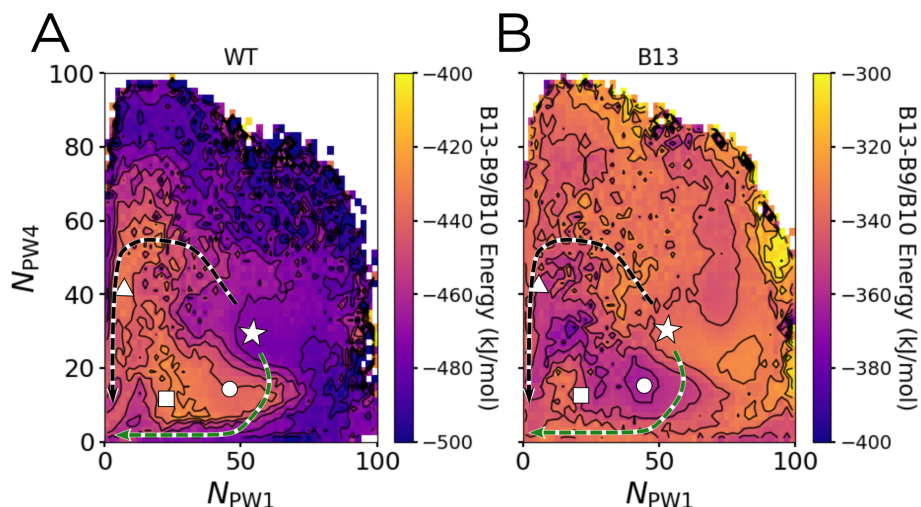
Figure S15: Interaction energies between B13 residues and the combination of $\text{Ser}^{B9}/\text{His}^{B10}$ as a function of N_{PW1} and N_{PW4}, with contours shown every 10 kJ/mol. Arrows representing PW1a and PW4a are overlaid in green and black, respectively. The star, circle, square, and triangle mark the same landmarks as in Figures 2 and 3 in the main text. This is shown for both (A) WT insulin, and (B) the B13 Glu → Gln mutant.

By contrast, B13 Glu → Gln insulin is dominated by a large-scale loss in J_{bind} through all six pathways. As a result, the binding rate is slower. Looking at the level of individual pathways, the largest area of reactive current loss corresponds to PW4a, and the areas corresponding to PW1 and PW4 lose comparatively little reactive current.

To understand the changes in the unimolecular ratio $K = k_{\text{unbinding}}/k_{\text{binding}}$ upon mutation, we calculated the interaction energies between the mutated residues and all 50 other protein residues, as well as the bound $\text{Zn}^{2+}/\text{Cl}^-$ ions, and the solvent (including the phenols). We then averaged these quantities across the free and bound state for both WT insulin and the relevant mutant, and measured $\Delta\Delta E = \Delta E_{\text{mutant}} - \Delta E_{\text{WT}}$, where $\Delta E = E_{\text{free}} - E_{\text{bound}}$. For the B13 Glu → Gln mutation, all such interactions where $|\Delta\Delta E| > 1$ kJ/mol are shown in Table S4.

The interactions with the largest magnitude of $\Delta\Delta E$ upon the B13 Glu → Gln mutation are B13 with Ser^{B9} and His^{B10}, as discussed in the main text. Phenol unbinding and channel opening are correlated, and in the unbound/open state, the B13 side chain is able to rotate to interact with the backbone residues of Ser^{B9} and His^{B10}. By making the B13 Glu →

Table S4: The change in interaction energies between the free and bound state for WT insulin and the B13 Glu → Gln mutant

Interaction	ΔE_{WT} (kJ/mol)	ΔE_{B13} (kJ/mol)	$\Delta\Delta E$ (kJ/mol)
B13 - Bound Zn^{2+}, Cl^-	–3.9	1.5	5.4
B13 - B13	3.9	5.1	1.3
B13 - B12	–0.9	0.4	1.3
B13 - B10	2.3	–3.5	–5.7
B13 - B9	14.9	–1.0	–15.9

Gln mutation, we replace a repulsive interaction between the carboxylate side chain and the backbone carbonyl with a energetically-favorable hydrogen bond between the amide side chain and the same backbone carbonyl. These interactions, which stabilize the unbound state upon mutation, partially explain why K increases upon mutation. We calculated a similar set of interactions for the A10 Ile → Val mutant, shown in Table S5. For this mutant, no single interaction is comparably dominant.

Table S5: The change in interaction energies between the free and bound state for WT insulin and the A10 Ile → Val mutant

Interaction	ΔE_{WT} (kJ/mol)	ΔE_{A10} (kJ/mol)	$\Delta\Delta E$ (kJ/mol)
A10-B1	–1.6	–0.4	1.2
A10-B5	0.8	1.9	1.1
A10-B2	2.9	1.6	–1.3
A10-Solvent	–0.5	–3.7	–3.2

To understand the effect of the B13 Glu → Gln mutation on the relative weights for the six pathways (and for PW1a and PW4a in particular), we projected the B13 electrostatic and Van der Waals interaction energies with Ser^{B9} and His^{B10} as functions of N_{PW1} and N_{PW4} for both WT insulin (Figure S15A) and the B13 Glu → Gln mutant (Figure S15B). The interaction energy decrease along PW1a (green) is approximately 20 kJ/mol, while the interaction energy decrease along PW4a (black) is 10 kJ/mol.
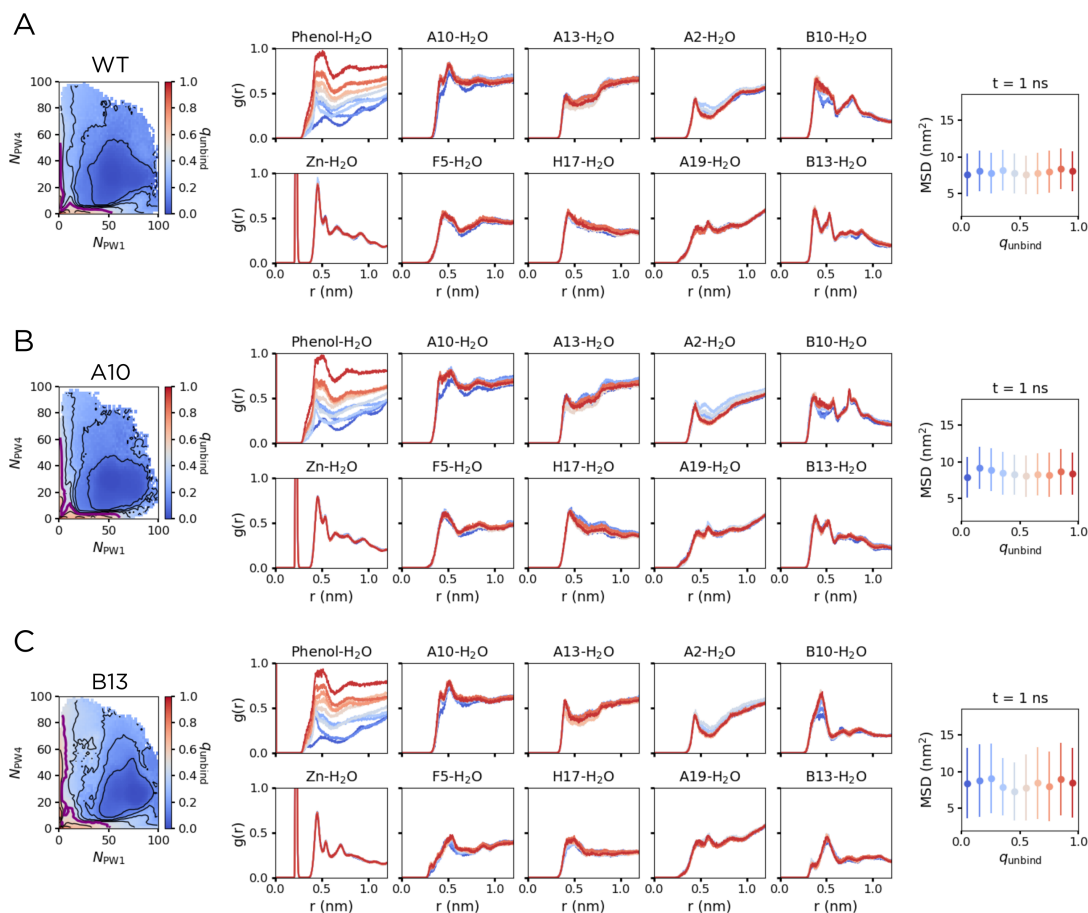
Figure S16: Solvation dependence on reaction progress. The (left) average committor, (middle) radial distribution functions for water around the specified species, and (right) mean square displacement (MSD) over 1 ns of waters in the central cavity of the hexamer for (A) WT insulin, (B) A10 Ile→Val insulin, and (C) B13 Glu→Gln insulin. Results are shown for 10 evenly sized bins for committor values between 0 and 1, with color given by the scale in the leftmost panel. We compute the radial distribution function, $g(r)$, from 15,000 structures in each committor value bin; we define r as the distance between the center of mass of the specified species/residue (including main chain atoms) and the center of mass of each water molecule. MSD values are for displacements over 1 ns from 5000 starting structures for each committor value bin.
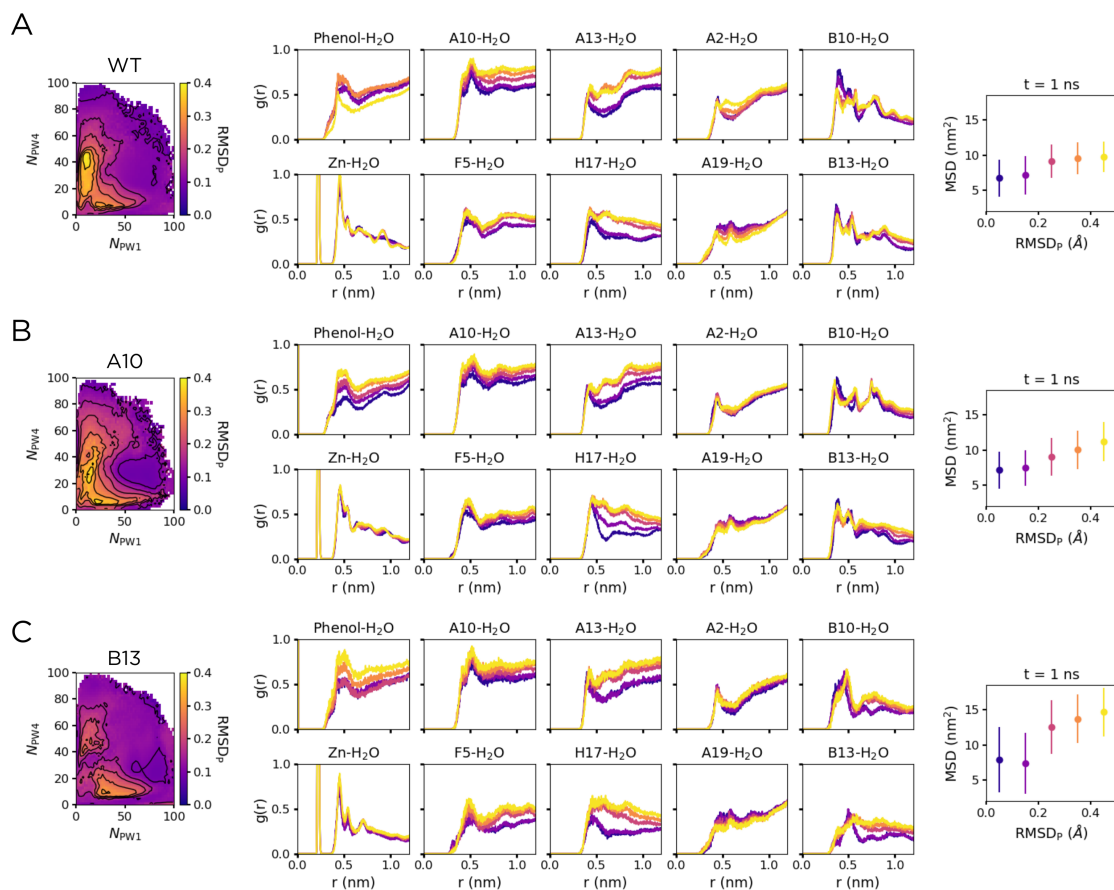
Figure S17: Solvation dependence on channel opening. The (left) average RMSD$_P$, (middle) radial distribution functions for water around the specified species, and (right) mean square displacement (MSD) over 1 ns of waters in the central cavity of the hexamer for (A) WT insulin, (B) A10 Ile→Val insulin, and (C) B13 Glu→Gln insulin. Results are shown for 5 evenly sized bins for RMSD$_P$ values between 0 and 0.5 Å, with color given by the scale in the leftmost panel. We compute the radial distribution function, $g(r)$, from 5,000 structures in each RMSD$_P$ value bin; we define r as the distance between the center of mass of the specified species/residue (including main chain atoms) and the center of mass of each water molecule. MSD values are for displacements over 1 ns from 3000 starting structures for each RMSD$_P$ value bin.

References

(1) 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.

(2) 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.

(3) Webber, R. J.; Thiede, E. H.; Dow, D.; Dinner, A. R.; Weare, J. Error Bounds for Dynamical Spectral Estimation. *SIAM J. Math. of Data Sci.* **2021**, *3*, 225–252.

(4) Lorpaiboon, C.; Thiede, E. H.; Webber, R. J.; Weare, J.; Dinner, A. R. Integrated Variational Approach to Conformational Dynamics: A Robust Strategy for Identifying Eigenfunctions of Dynamical Operators. *J. Phys. Chem. B* **2020**, *124*, 9354–9364.

(5) 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.

(6) 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.

(7) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197–2201.

(8) Tiwari, S. P.; Fuglebakk, E.; Hollup, S. M.; Skjærven, L.; Cragnolini, T.; Grindhaug, S. H.; Tekle, K. M.; Reuter, N. WEBnm@ v2.0: Web Server and Services for Comparing Protein Flexibility. *BMC Bioinformatics* **2014**, *15*, 427.

(9) 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.

(10) Thiede, E. H.; Van Koten, B.; Weare, J.; Dinner, A. R. Eigenvector Method for Umbrella Sampling Enables Error Analysis. *J. Chem. Phys.* **2016**, *145*, 084115.

(11) Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 2.3.0. 2015.

(12) Northrup, S. H.; Allison, S. A.; McCammon, J. A. Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. *J. Chem. Phys.* **1984**, *80*, 1517–1524.

(13) 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.

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

(15) 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.

(16) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.