

# Insulin Dissociates by Diverse Mechanisms of Coupled Folding and Binding

Adam Antoszewski,<sup>†</sup> Chi-Jui Feng,<sup>†</sup> Bodhi P. Vani,<sup>†</sup> Erik H. Thiede,<sup>‡,P</sup> Lu  
Hong,<sup>§</sup> Jonathan Weare,<sup>||</sup> Andrei Tokmakoff,<sup>†,⊥,#</sup> and Aaron R. Dinner<sup>\*,†,⊥,#</sup>

<sup>†</sup>*Department of Chemistry, The University of Chicago, Chicago, Illinois*

<sup>‡</sup>*Department of Computer Science, The University of Chicago, Chicago, Illinois*

<sup>P</sup>*Department of Statistics, The University of Chicago, Chicago, Illinois*

<sup>§</sup>*Graduate Program in Biophysical Sciences, The University of Chicago, Chicago, Illinois*

<sup>||</sup>*Courant Institute of Mathematical Sciences, New York University, New York, New York*

<sup>⊥</sup>*James Franck Institute, The University of Chicago, Chicago, Illinois*

<sup>#</sup>*Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois*

E-mail: dinner@uchicago.edu

**EMUS Asymptotic Error With Replica Exchange.** To derive the expression for asymptotic variances of averages given REUS data, we follow the following procedure. The notation corresponds to that in ref. 1.

**Assumption 0.1.** *We assume:*

1. *Sampling over all windows is performed by a single Markov chain  $\Xi$  whose state space is  $L$  copies of the molecular phase space.*
2. *A central limit theorem holds for the convergence of each sample average  $\bar{v}$  estimated in EMUS to its true value  $v$ , with asymptotic covariance matrix  $\Sigma$ . The entries in this matrix are given by*

$$\Sigma_{ij} = \frac{1}{2} \int_{-\infty}^{\infty} \mathbb{E} [v_i(X_t) v_j(X_0)] dt. \quad (1)$$

3. *The biasing functions  $\psi_i$  are chosen so that  $F$  is irreducible.*

Note that if we sort the averages  $v$  by the window in which the average was calculated, then unlike in ref. 1, the matrix  $\Sigma$  is not block diagonal. Rather it has nonzero off-diagonal blocks:

$$\Sigma = \left[ \begin{array}{c|c|c} \Sigma_{11} & \Sigma_{12} & \dots \\ \hline \Sigma_{21} & \Sigma_{22} & \dots \\ \hline \vdots & \vdots & \ddots \end{array} \right].$$

Under these assumptions, we can still apply Lemma VII.2 in ref. 1 to derive a central limit theorem for EMUS with replica exchange.

**Theorem 0.2.** *As in ref. 1, let  $B$  be a quantity of interest and  $dB/d\bar{v}$  be the partial derivative of  $B$  with respect to each sampled average. Under the assumptions above,*

$$\sqrt{N} (B(\bar{v}) - B(v)) \xrightarrow{d} N(0, \sigma^2), \quad (2)$$

where

$$\sigma^2 = \frac{\partial B^T}{\partial \bar{v}} \Sigma \frac{\partial B}{\partial \bar{v}}. \quad (3)$$

To estimate  $\sigma^2$  using sampled data, we write

$$\sigma^2 = \sum_{i,j} \frac{\partial B}{\partial \bar{v}_i} \Sigma_{ij} \frac{\partial B}{\partial \bar{v}_j} \quad (4)$$

$$\begin{aligned} &= \sum_{i,j} \frac{\partial B}{\partial \bar{v}_i} \int_{-\infty}^{\infty} \mathbb{E} [v_i(X_t) v_j(X_0)] dt \frac{\partial B}{\partial \bar{v}_j} \\ &= \int_{-\infty}^{\infty} \mathbb{E} \left[ \left( \sum_i \frac{\partial B}{\partial \bar{v}_i} v_i(X_t) \right) \left( \sum_j \frac{\partial B}{\partial \bar{v}_j} v_j(X_0) \right) \right] dt \end{aligned} \quad (5)$$

This is the the integrated covariance of the trajectory  $\sum_i \frac{\partial B}{\partial \bar{v}_i} v_i(X_t)$ .

**C-terminal detachment is correlated with solvation of the A-chain N-terminal segment and formation of nonnative contacts.** The detachment of the C-terminal segment of the B chain (Figure S10), discussed extensively in the main text, is also correlated with the increase in SASA of Gly<sup>A1</sup>-Val<sup>A3</sup> (Figure S11). Gly<sup>A1</sup>-Val<sup>A3</sup> are residues that are in contact with Pro<sup>B28</sup>-Ala<sup>B30</sup> in the dimer, and these contacts are sacrificed as detachment occurs along the  $\alpha$  path. This is consistent with the known binding behavior of the insulin monomer to the insulin receptor, as these Gly<sup>A1</sup>-Val<sup>A3</sup> residues are thought to form part of the binding interface<sup>2,3</sup> and are not in contact with Pro<sup>B28</sup>-Ala<sup>B30</sup> when bound to the receptor.<sup>4,5</sup>

Beyond the loss of contacts between Pro<sup>B28</sup>-Ala<sup>B30</sup> and Gly<sup>A1</sup>-Val<sup>A3</sup> along the  $\alpha$  path, one can also characterize the number of contacts that the B-chain C-terminal segment makes with other residues. A subset of these are shown in Figure S5. Namely, from the left panels, one sees that along the  $\alpha$  path (black), the native contacts of Pro<sup>B28</sup>-Ala<sup>B30</sup> with Gly<sup>B'20</sup>-Gly<sup>B'23</sup> and Gly<sup>A1</sup>-Val<sup>A3</sup> are lost. The rotation of the interface associated with both the  $\alpha$  and  $\beta$  paths moves the B-chain C-terminal segment away from these both Gly<sup>B'20</sup>-Gly<sup>B'23</sup> and Gly<sup>A1</sup>-Val<sup>A3</sup>.

The right plots, on the other hand, show that some specific nonnative interactions are

formed along the  $\alpha$  path, but not along the  $\beta$  path. Along the  $\alpha$  path, the B-chain C-terminal segment starts to interact with the Tyr<sup>B'16</sup> on the  $\alpha$  helix of the other monomer when  $\bar{\alpha} \approx 1.15$  nm. In the same region, Ser<sup>B9</sup> and Ser<sup>B'9</sup>, serines on opposite interfacial helices, start to interact and form a nonnative contact. Interestingly, comparing this to the core SASA shown in the main text Figure 5A, one sees that the core SASA increases sharply after the Ser<sup>B9</sup> residues lose contact with one another, around  $\bar{\alpha} \approx 1.35$  nm. This is also the area where significant detachment starts to occur along the  $\alpha$  path. It is possible that the combination of the interfacial rotations and detachment exposes the hydrophobic core of the dimer.

**Relation to available insulin therapeutics.** To replicate the biphasic activity of insulin *in vivo*, both fast-acting and slow-acting insulin therapeutics have been developed. Two of the most common fast acting therapeutics, lispro (Pro<sup>B28</sup>Lys<sup>B29</sup>  $\rightarrow$  Lys<sup>B28</sup>Pro<sup>B29</sup>, Humalog<sup>6</sup>) and aspart (Pro<sup>B28</sup>  $\rightarrow$  Asp<sup>B28</sup>, NovoLog<sup>7</sup>) involve mutation of C-terminal residues of the B chain to destabilize the dimer and favor the monomer. This presumably allows for more rapid and reliable uptake of glucose after injection, as the monomer is the species that binds to the receptor. These mutations have been hypothesized to reduce the interaction of Pro<sup>B28</sup>-Ala<sup>B30</sup> of one monomer with the  $\beta$  turn of the other monomer (Gly<sup>B'20</sup>-Gly<sup>B'23</sup>), either through reduced van der Waals attractive forces (lispro and aspart) or the addition of a repulsive electrostatic interaction (aspart).<sup>8</sup> To the best of our knowledge, no computational work has yet been done to explore these hypotheses as they relate to the mechanism of dissociation. In the previous section, it was described how, along both pathways, the native contacts between Pro<sup>B28</sup>-Ala<sup>B30</sup> and both Gly<sup>B'20</sup>-Gly<sup>B'23</sup> and Gly<sup>A1</sup>-Val<sup>A3</sup> are broken. It has been suggested<sup>6,8</sup> that one or both of the therapeutic mutations could reduce the energetic penalty for breaking these native contacts, specifically the contacts with Gly<sup>B'20</sup>-Gly<sup>B'23</sup>. To explicitly investigate this in our simulations of wildtype insulin, the sum of interaction energies of Pro<sup>B28</sup>-Ala<sup>B30</sup> and both Gly<sup>B'20</sup>-Gly<sup>B'23</sup> and Gly<sup>A1</sup>-Val<sup>A3</sup> was computed (Figure

S6).

**$\beta$  turn angle and disorder of the  $\beta$  turn also varies between dissociation path.**

We defined the  $\beta$  turn angle,  $\Phi_t$ , as the angle between the geometrical centers of the backbone atoms in Glu<sup>B13</sup>, Gly<sup>B20</sup>, and Phe<sup>B24</sup> to characterize the  $\beta$  turn. A larger  $\beta$  turn angle corresponds to a wider turn. We also calculated the RMSD of the  $\beta$  turn from its conformation in the dimer by fitting the conserved  $\alpha$  helix of the B chain to each monomeric unit, then measuring the RMSD to the solvated crystal structure. The average of both the  $\beta$  turn angle and the RMSD are shown in Figure S9.

The  $\beta$  turn angle increases initially and is much wider along the  $\alpha$  path than along the  $\beta$  path. The RMSD also increases as the  $\alpha$  helices separate. This increase in RMSD happens before the  $\beta$  sheets are broken for the  $\alpha$  path, and after the  $\beta$  sheets are broken for the  $\beta$  path. The different trends in  $\beta$  turn angle and RMSD between paths suggest that these residues on the  $\beta$  turn might be useful candidates for isotopic labeling for 2DIR experiments aimed at distinguishing dissociation pathways.

**FTIR to identify labels and solvated species.** Here we describe our simulations of FTIR spectra to map the effects of isotopic labels with considerably less computational cost than simulations of 2DIR spectra. As noted in the main text, we computed 50 such spectra along each of the two limiting paths, and each was converted into a difference spectrum by subtracting the unlabeled insulin spectrum at that point along the path from the labeled spectrum. The intensities of these difference spectra were converted into a colormap, with orange being a positive change, and blue being a negative change. This allows one to see the peaks associated specifically with the isotope label. These were then stacked and viewed as a function of path progress along both the  $\alpha$  and  $\beta$  paths. Thus, when a red feature shifts to a lower frequency, this corresponds to a redshift in the simulated labeled spectra. These series of spectra, for the Phe<sup>B24</sup> and Glu<sup>B13</sup> isotope-labeled insulins, are shown in Figure S2. We simulated all possible constructs with a single interfacial residue isotopically labeled

(between Ser<sup>B9</sup> and Ala<sup>B30</sup>). The labels on our identified  $\alpha$  and  $\beta$  contacts showed the most significant redshifts; of these Phe<sup>B24</sup> and Glu<sup>B13</sup> exemplified these effects.

From these spectra, we identified the ranges of path progress where the redshift (and correlated loss in intensity) first occurs, presumably due to either solvation or change in local secondary structure. These areas are marked areas in Figure S2. The left panels show the SASA of the backbone carbonyl group associated with the isotopically labeled residue (either Phe<sup>B24</sup> or Glu<sup>B13</sup>). One sees that these marked regions begin once the SASAs for the corresponding backbone carbonyl groups increase from the dimer (with SASAs near 0) to 60% of the monomeric average. One can thus conclude that the redshifting and peak broadening in these FTIR spectra correlates with the solvation of the backbone.

## References

- (1) 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.
- (2) Kristensen, C.; Kjeldsen, T.; Wiberg, F. C.; Schäffer, L.; Hach, M.; Havelund, S.; Bass, J.; Steiner, D. F.; Andersen, A. S. Alanine Scanning Mutagenesis of Insulin. *J. Biol. Chem.* **1997**, *272*, 12978–12983.
- (3) Thorsøe, K. S.; Schlein, M.; Steensgaard, D. B.; Brandt, J.; Schluckebier, G.; Naver, H. Kinetic Evidence for the Sequential Association of Insulin Binding Sites 1 and 2 to the Insulin Receptor and the Influence of Receptor Isoform,. *Biochemistry* **2010**, *49*, 6234–6246.
- (4) Scapin, G.; Dandey, V. P.; Zhang, Z.; Prosise, W.; Hruza, A.; Kelly, T.; Mayhood, T.; Strickland, C.; Potter, C. S.; Carragher, B. Structure of the Insulin Receptor-Insulin Complex by Single-Particle Cryo-EM Analysis. *Nature* **2018**, *556*, 122–125.
- (5) Weis, F.; Menting, J. G.; Margetts, M. B.; Chan, S. J.; Xu, Y.; Tennagels, N.; Wohl-

- fart, P.; Langer, T.; Müller, C. W.; Dreyer, M. K. et al. The Signalling Conformation of the Insulin Receptor Ectodomain. *Nat. Commun.* **2018**, *9*, 4420.
- (6) Ciszak, E.; Beals, J. M.; Frank, B. H.; Baker, J. C.; Carter, N. D.; Smith, G. Role of C-terminal B-chain Residues in Insulin Assembly: The Structure of Hexameric LysB28ProB29-Human Insulin. *Structure* **1995**, *3*, 615–622.
- (7) 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.
- (8) 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*, 201–64.
- (9) Moradi, M. Codes and Scripts, Theoretical and Computational Biophysics Group. <https://www.ks.uiuc.edu/mahmoud/codes.html>. (accessed July 15, 2019).
- (10) Banerjee, P.; Mondal, S.; Bagchi, B. Insulin Dimer Dissociation in Aqueous Solution: A Computational Study of Free Energy Landscape and Evolving Microscopic Structure Along the Reaction Pathway. *J. Chem. Phys.* **2018**, *149*, 114902.
- (11) Banerjee, P.; Mondal, S.; Bagchi, B. Effect of Ethanol on Insulin Dimer Dissociation. *J. Chem. Phys.* **2019**, *150*, 084902.

Table S1: Types and descriptions of contacts with high differential SASA between dimer and monomer, with primes indicating intermonomer interactions.

Class	Residue Pairs	Description
$\beta$ Contacts	Phe <sup>B24</sup> :Tyr <sup>B'26</sup> , Phe <sup>B25</sup> :Phe <sup>B'25</sup> , Tyr <sup>B26</sup> :Phe <sup>B'24</sup>	Phe <sup>B24</sup> and Tyr <sup>B26</sup> form hydrogen bonds at the $\beta$ sheet interface side chains make up part of the hydrophobic core of the dimer. Phe <sup>B25</sup> is also part of the $\beta$ sheet, but its side chain is not part of the hydrophobic core.
$\alpha$ Contacts	Ser <sup>B9</sup> :Glu <sup>B'13</sup> , Ser <sup>B9</sup> :Tyr <sup>B'16</sup> , Val <sup>B12</sup> :Tyr <sup>B'16</sup> , Glu <sup>B13</sup> :Ser <sup>B'9</sup> , Glu <sup>B13</sup> :Glu <sup>B'13</sup> , Tyr <sup>B16</sup> :Ser <sup>B'9</sup> , Tyr <sup>B16</sup> :Val <sup>B'12</sup>	Ser <sup>B9</sup> , Val <sup>B12</sup> , Glu <sup>B13</sup> , Tyr <sup>B16</sup> are residues from the $\alpha$ -helical portion of the B chain that are at least partially buried in the dimer. Val <sup>B12</sup> and Tyr <sup>B16</sup> are part of the hydrophobic core.
Cross Contacts	Val <sup>B12</sup> :Phe <sup>B'24</sup> , Tyr <sup>B16</sup> :Tyr <sup>B'26</sup> , Phe <sup>B24</sup> :Val <sup>B'12</sup> , Tyr <sup>B26</sup> :Tyr <sup>B'16</sup>	Buried residues across the dimer interface that pair $\alpha$ helices with $\beta$ strands and vice versa.
Intermittent Contacts	Val <sup>B12</sup> :Glu <sup>B'13</sup> , Glu <sup>B13</sup> :Val <sup>B'12</sup> , Glu <sup>B21</sup> :Pro <sup>B'28</sup> , Gly <sup>B23</sup> :Thr <sup>B'27</sup> , Thr <sup>B27</sup> :Gly <sup>B'23</sup> , Pro <sup>B28</sup> :Glu <sup>B'21</sup>	Thr <sup>B27</sup> and Pro <sup>B28</sup> are relatively disordered residues adjacent to the $\beta$ -sheet that form intermittent contacts with $\beta$ turn residues Glu <sup>B21</sup> and Gly <sup>B23</sup> . Val <sup>B12</sup> and Glu <sup>B13</sup> are on the interfacial $\alpha$ helices and also only interact intermittently.
Internal Contacts	Val <sup>B12</sup> :Tyr <sup>B26</sup> , Glu <sup>B13</sup> :Phe <sup>B24</sup>	Internal contacts within a monomer.



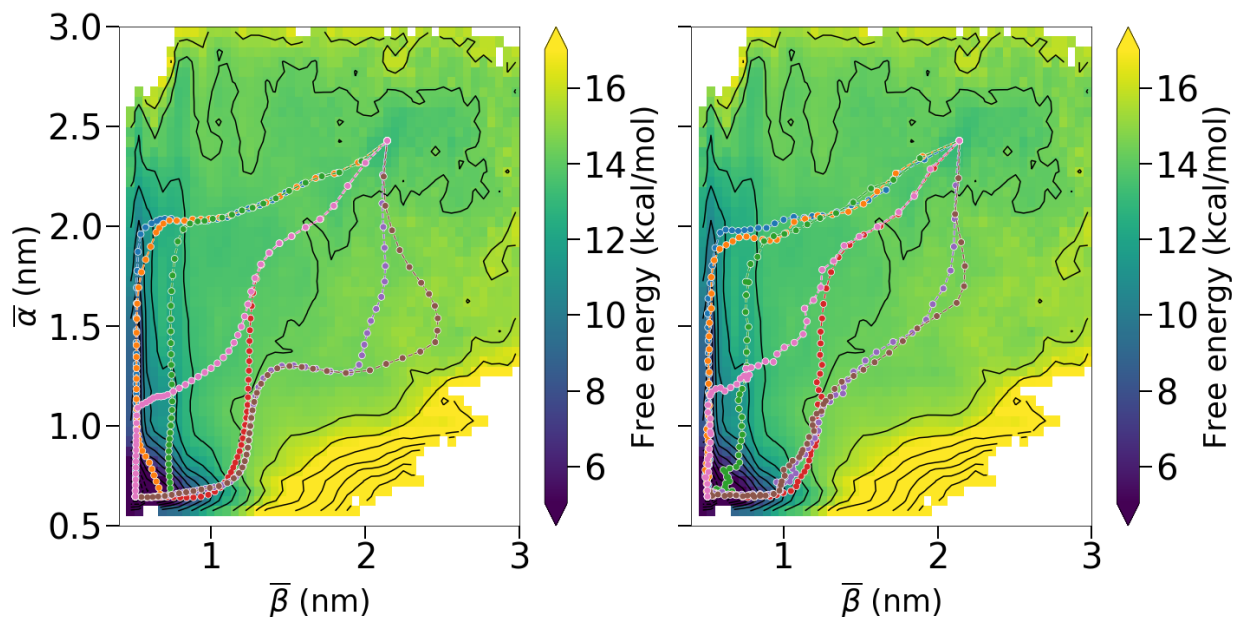


Figure S1: Using the string method to confirm stability of dissociation paths. (Left) Minimum free energy paths identified by using the LFEP algorithm from ref. 9. All the paths shown have a maximum free energetic barrier within  $4 k_B T$  of the others. These paths were used as initializations for the string method. (Right) The converged strings after further refinement with the string method. Comparing with the left panel, the orange path has collapsed to the  $\alpha$  path, and the purple and brown paths have shifted slightly from their initial positions. Note that while the purple path, which represents the  $\beta$  path, has shifted slightly in CV space, this does not affect the molecular trends discussed in the main text. Overall, the stability of these paths provides evidence that the averaging reducing the 10 interfacial distance dimensions to  $\bar{\beta}$  and  $\bar{\alpha}$  does not sacrifice mechanistic information.

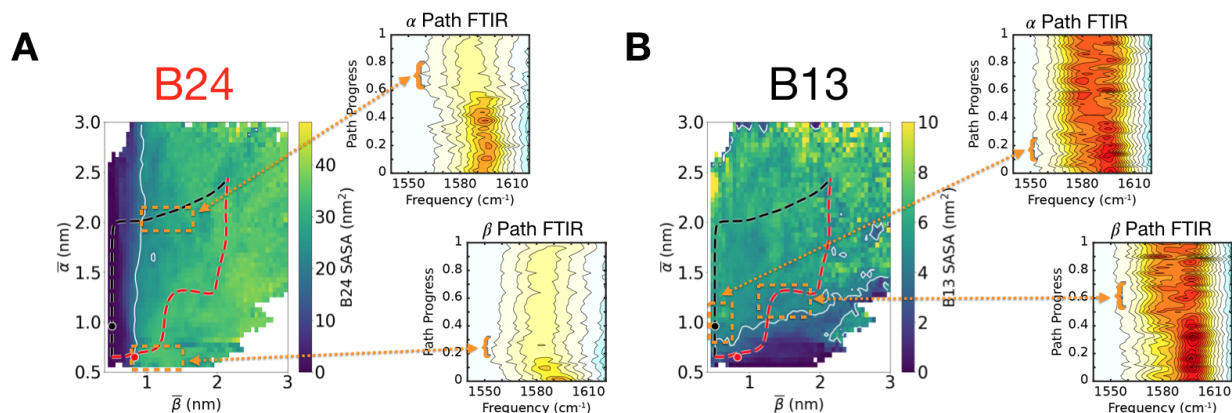


Figure S2: Relation between backbone solvation and simulated FTIR spectra. (A) The backbone carbonyl SASA for Phe<sup>B24</sup>, with the white contour showing where the SASA increases to 60% of the monomeric average (contour at 19.7 nm<sup>2</sup>, monomeric average at 32.8 nm<sup>2</sup>). The two Phe<sup>B24</sup>-labeled FTIR plots correspond to simulations along the  $\alpha$  (top) and  $\beta$  paths (bottom), with the y-axis being the path progress, or the fractional distance along each specific path. Each value of y corresponds to one FTIR simulation - a FTIR spectra was generated by combining 20 simulations started from a point at that specific value of path progress, and a difference was taken between that isotope-labeled simulated spectrum and the corresponding unlabeled simulated spectrum. 50 such difference spectra were created, and stacked such that the color represents the intensity of the difference. For each path, the spectra that first demonstrated the expected redshift were identified, and then those areas of path space were selected as the solvated species for future study (orange boxes on the SASA graph). (B) Similar graphs for the Glu<sup>B13</sup>-labeled insulin, showing the backbone carbonyl SASA for Glu<sup>B13</sup>. The contour is again shown where the SASA increases to 60% of the monomeric average (contour at 4.1 nm<sup>2</sup>, monomeric average of 6.8 nm<sup>2</sup>).

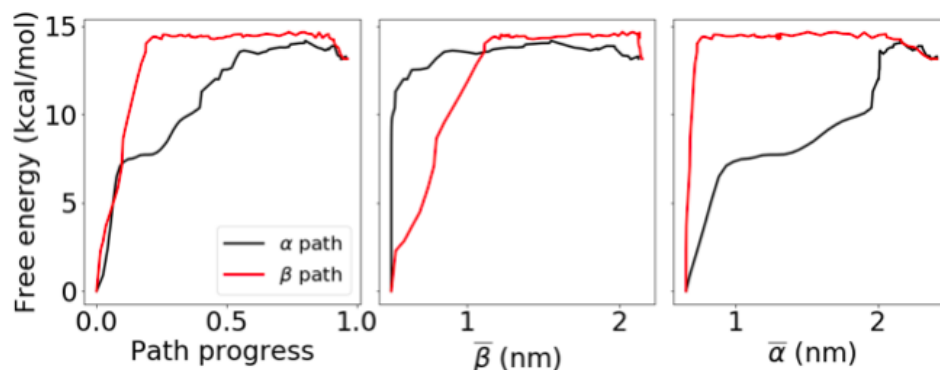


Figure S3: 1D cuts of our  $\alpha$  and  $\beta$  paths as functions of path progress (left),  $\bar{\beta}$  (middle), and  $\bar{\alpha}$  (right).

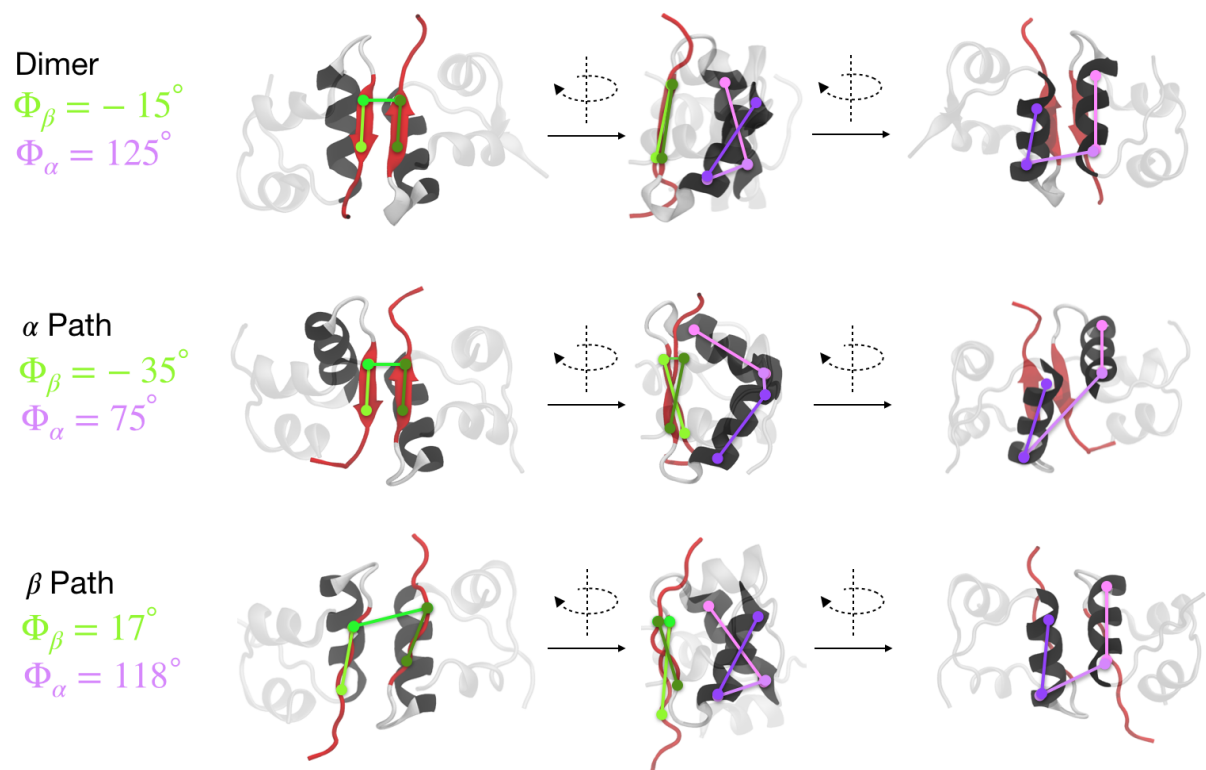


Figure S4: Structures representing the dimer (top), initial steps along the  $\alpha$  path (middle), and initial steps along the  $\beta$  path (bottom), with lines superimposed to show the interfacial pseudodihedral angles,  $\Phi_{\beta}$  (green) and  $\Phi_{\alpha}$  (purple). For these lines, a darker color in the side projection means the residues are in front while a lighter color means they are behind.

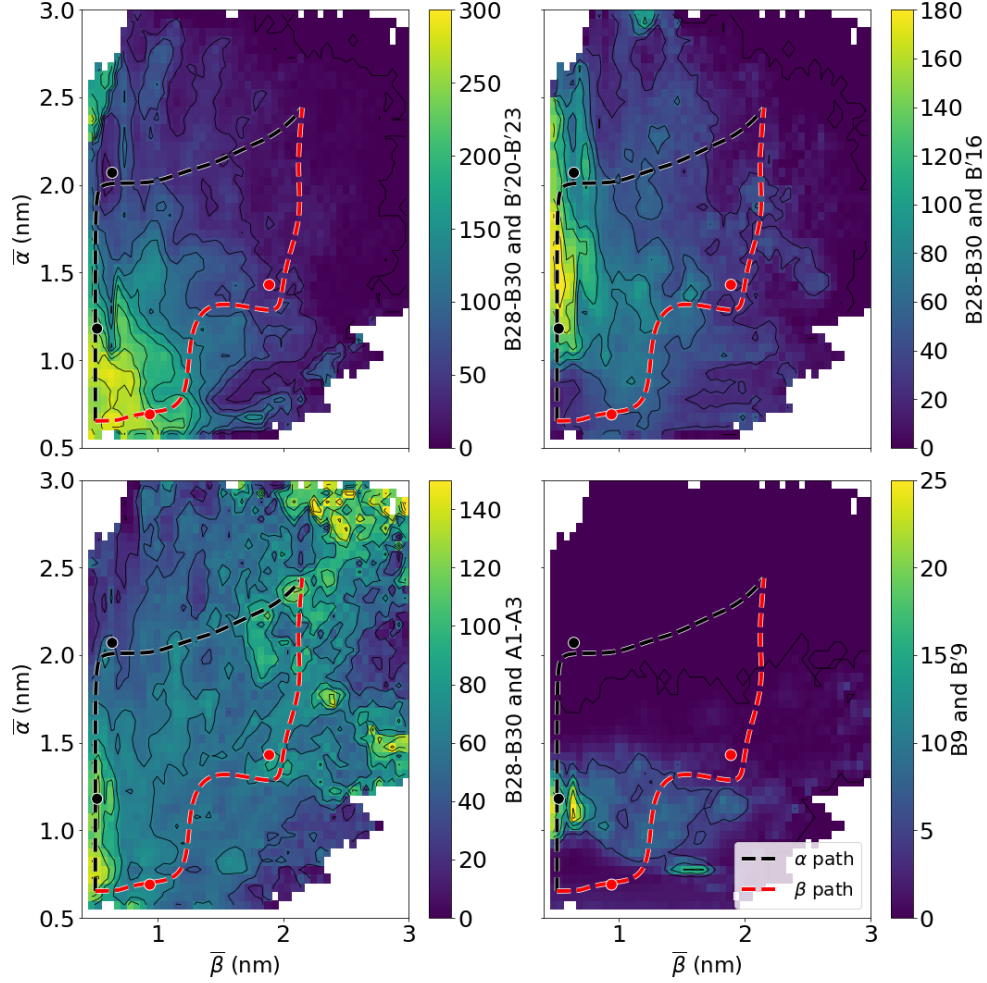


Figure S5: Averages of the number of native (left) and nonnative (right) contact pairs, with the specific pair given by the scale bar labels, as a function of  $\bar{\alpha}$  and  $\bar{\beta}$ . The left plots show that along the  $\alpha$  path, contacts between  $\text{Pro}^{\text{B28}}\text{-Ala}^{\text{B30}}$  are lost with both  $\text{Gly}^{\text{B'20}}\text{-Gly}^{\text{B'23}}$  (top) and  $\text{Gly}^{\text{A1}}\text{-Val}^{\text{A3}}$  (bottom). The right plots show that along the  $\alpha$  path, nonnative contacts start to form between  $\text{Pro}^{\text{B28}}\text{-Ala}^{\text{B30}}$  and  $\text{Tyr}^{\text{B'16}}$  (top), and between  $\text{Ser}^{\text{B9}}$  and  $\text{Ser}^{\text{B'9}}$ .

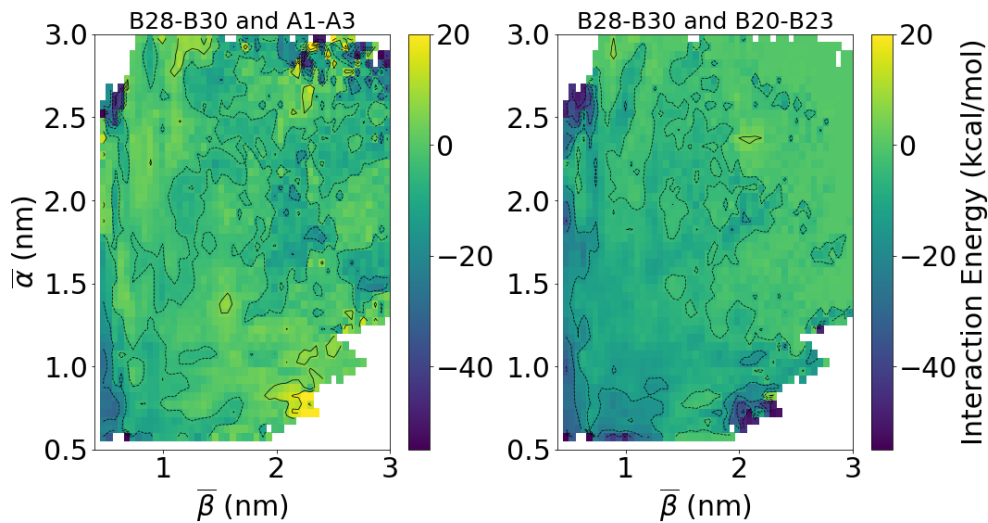


Figure S6: Average total interaction energies of Pro<sup>B28</sup>-Ala<sup>B30</sup> with Gly<sup>A1</sup>-Val<sup>A3</sup> (left) and Gly<sup>B'20</sup>-Gly<sup>B'23</sup> (right). Contour lines shown every 10 kcal/mol. Both of these interactions stabilize the dimer state (lower left corner of both panels).

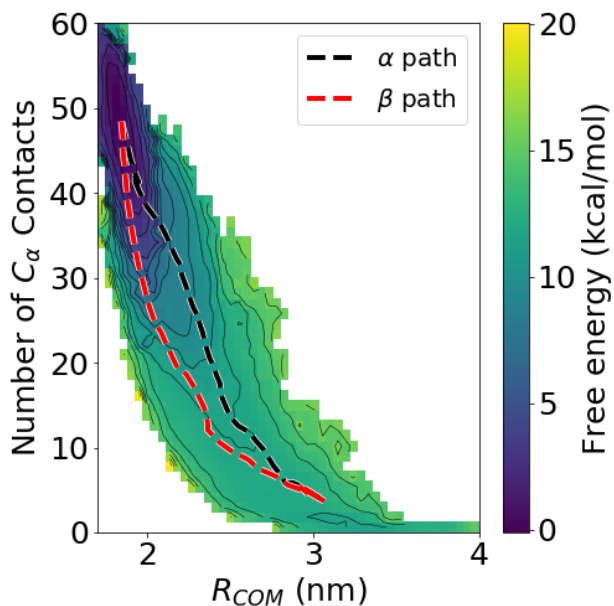


Figure S7: PMF as a function of the center of mass distance between the two monomers ( $R_{COM}$ ) and number of interfacial  $C_\alpha$  contacts (cutoff 7 Å,) the coordinates used by Bagchi and coworkers in refs. 10 and 11.

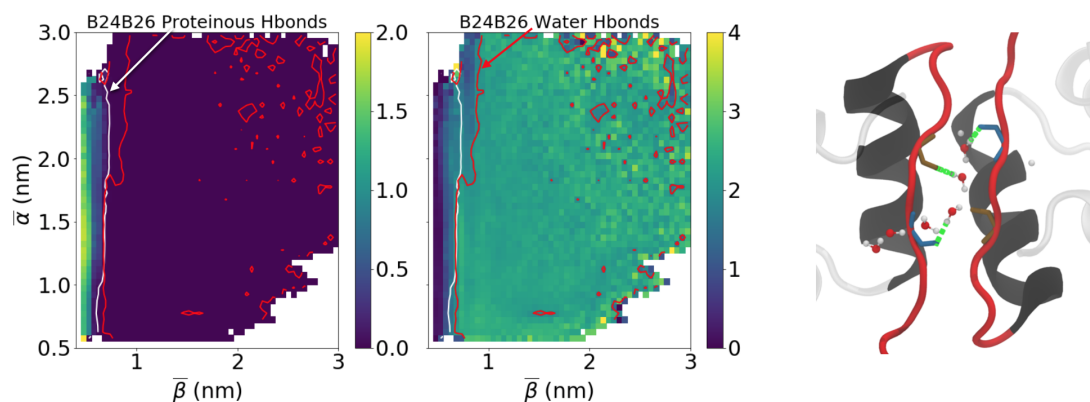


Figure S8: Characterizing interfacial hydrogen bonding. (Left) The average number of protein-protein hydrogen bonds on the beta sheet interface averaged on the PMF, compared to (middle) the average number of hydrogen bonds between those same residues and water. The white contour represents when the protein-protein hydrogen bond character drops to 2% of its initial value, while the red contour shows the point where on average 2 hydrogen bonds have been formed with water on the interface. These contours correlate well in CV space. On the right is a representative structure showing this solvation, with hydrogen bonds between the interfacial residues and water shown in green.

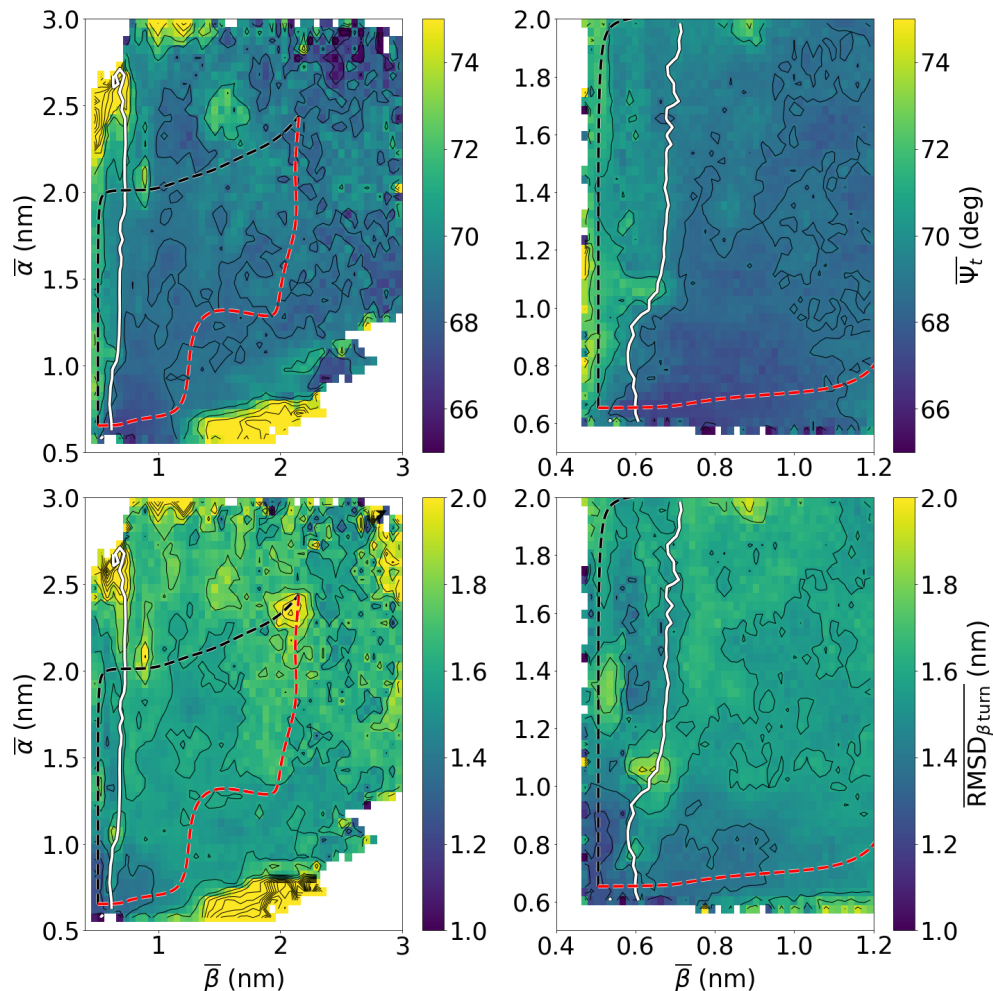


Figure S9: Plots showing the behavior of the interfacial  $\beta$  turn during the dissociation. The average  $\beta$  turn angle (top) and  $\beta$  turn RMSD (bottom) as a function of  $\bar{\alpha}$  and  $\bar{\beta}$ , and zoomed in to the near-dimer regime on the right. On all graphs, the  $\alpha$  and  $\beta$  paths are superimposed, as well as the white contour that signifies the solvation of the  $\beta$  interface. Here, we see the  $\beta$  turn angle increasing along the  $\alpha$  path but not along the  $\beta$  path. Also, the  $\beta$  turn RMSD increases before the  $\beta$  sheets are broken along the  $\alpha$  path, while this only occurs after the  $\beta$  sheets are broken for the  $\beta$  path.

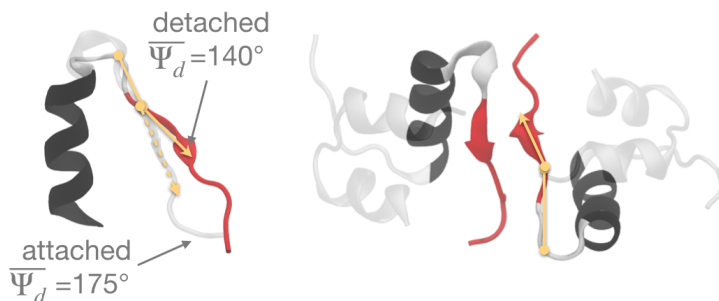


Figure S10: Structures showing detachment. (Left) The detached and attached species overlaid, with the detachment angle explicitly overlaid on top of the structure. (Right) This same detached structure, but now showing the entire dimeric species.



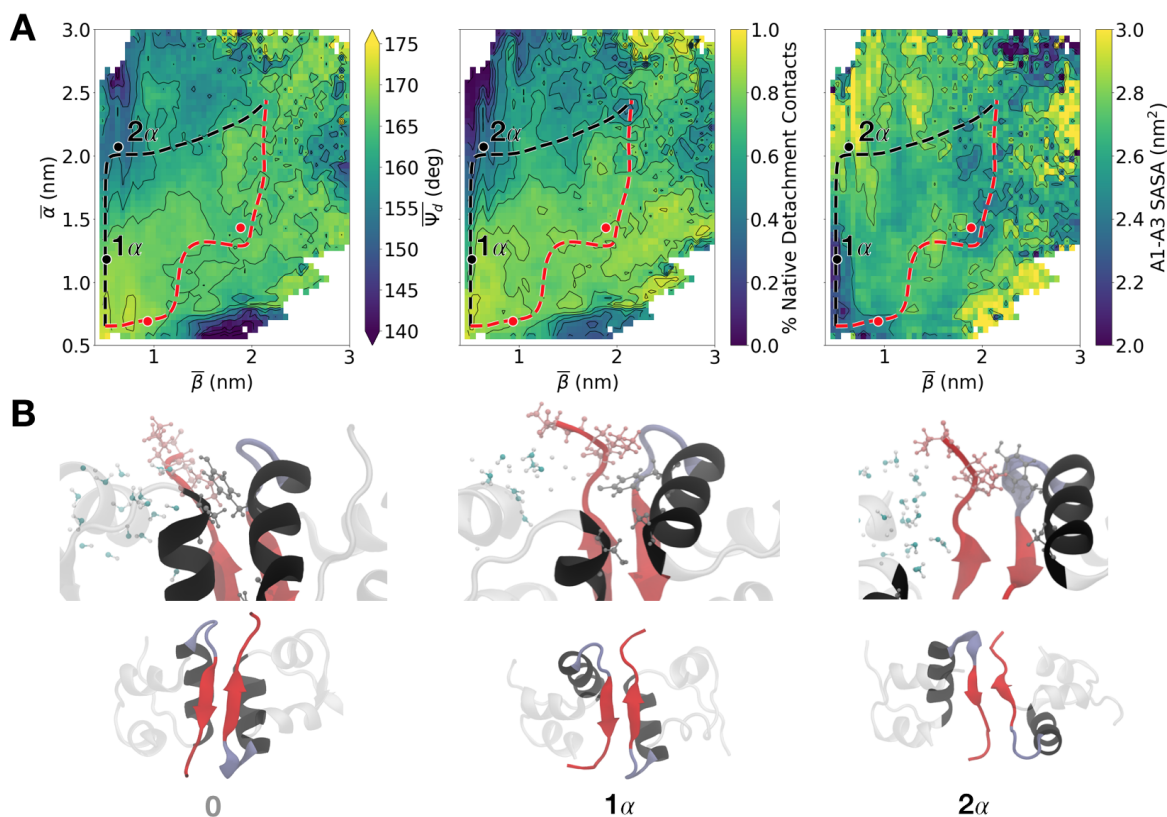


Figure S11: Detachment is correlated with solvation of the N-terminal segment of the A chain. (A) Averages of the detachment angle  $\Psi_d$  (left), percentage of native contacts between Pro<sup>B28</sup>-Ala<sup>B30</sup> and the nearby  $\alpha$  helix of the same monomer (middle), and the SASA for Gly<sup>A1</sup>-Val<sup>A3</sup> of the same monomer. The similarity between the left and middle plots suggests that the detachment angle is an effective measure of the C-terminal segment moving away from the  $\alpha$  helix it is normally tucked against in the native monomeric unit. Furthermore, the similarity to the rightmost plot shows that the solvation of Gly<sup>A1</sup>-Val<sup>A3</sup> is correlated to the large detachment of the B chain's C-terminal segment in the same monomeric unit. (B) Structures showing how the detachment is coupled to the solvation of Gly<sup>A1</sup>-Val<sup>A3</sup>.