Supporting Information for

Computational IR Spectroscopy of Insulin Dimer Structure and Conformational Heterogeneity

Chi-Jui Feng[1], Anton Sinitskiy[2], Vijay Pande[2], and Andrei Tokmakoff[1,*]

[1]*Department of Chemistry, James Franck Institute and Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, United States*

[2]*Department of Bioengineering, Stanford University, Stanford, CA 94305, United States*

Contents

* Author to whom correspondence should be addressed. Electronic mail: tokmakoff@uchicago.edu. Telephone: (773) 834-7696.

Characterization of Structural Collective Variables for Markov States

To quantitatively analyze native contacts in the dimer state, the number of contacts were calculated by defining atom pairs of interest and using a rational switching function in Eqn. (1) to determine if the pairs were in contact.

$$s\left(r_{ij}\right)=\begin{cases}1, \forall r_{ij} \le d_0 \\ \left(1-\left(\frac{r_{ij}}{r_0}\right)^n\right) \Big/ \left(1-\left(\frac{r_{ij}}{r_0}\right)^m\right), r_{ij} > d_0\end{cases} \quad (1)$$

To calculate n_α and n_β, the number of α contacts and β contacts at the dimer interface for a particular structure, the average over $s(r_{ij})$ is calculated for all pairs of C_α atoms summarized in Table S1. Pairs of Cα atoms for calculating number of contacts.[1] M1 and M2 represents the monomer 1 and 2 in the dimer structure. for comparison with the result in ref. 1. To compare with results from Bagchi and co-workers, we also computed the number of inter-monomer contacts using every single possible pair of C_α atoms between monomer 1 (M1) and monomer 2 (M2), resulting in 51 × 51 = 2601 pairs.[2] The value of r_0 used in the rational switching function was set to 7 Å to match Bagchi and co-workers calculations, as well as setting n = 6 and m = 12.[3] For each Markov state consisting of 100 structures, the average number of the contacts was computed over the contacts of those 100 structures within the same Markov state with equal weight. All of the contacts were computed using the open-source, community-developed PLUMED library,[4] version 2.5.2.[5]

α contact	β contact
B9 (M1) – B13 (M2)	B24 (M1) – B24 (M2)
B9 (M1) – B16 (M2)	B24 (M1) – B25 (M2)
B12 (M1) – B16 (M2)	B24 (M1) – B26 (M2)
B13 (M1) – B9 (M2)	B25 (M1) – B24 (M2)
B13 (M1) – B13 (M2)	B25 (M1) – B25 (M2)
B16 (M1) – B9 (M2)	B25 (M1) – B26 (M2)
B16 (M1)– B12 (M2)	B26 (M1) – B24 (M2)
	B26 (M1) – B25 (M2)
	B26 (M1)– B26 (M2)

Table S1. Pairs of C_α atoms for calculating number of contacts.[1] M1 and M2 represents the monomer 1 and 2 in the dimer structure.

The number of hydrogen bonds (n_{HB}) along the β-strand were also analyzed to illustrate the local structural changes of the residues involved in the β-sheet (Figure S1). The presence of a hydrogen bond is determined using geometric criteria, such that the distance between the O on C=O and H on N–H is less than or equal to 3.5 Å, and the $\mathrm{H\cdots C=O}$ angle, $\angle_{\mathrm{HOC}}$, is greater or equal to 150°. The calculation was done using the GROMACS utility. The average number of hydrogen bonds of each Markov state, $\langle n_{HB} \rangle$, was computed over 100 structures within the same Markov state with equal weight.

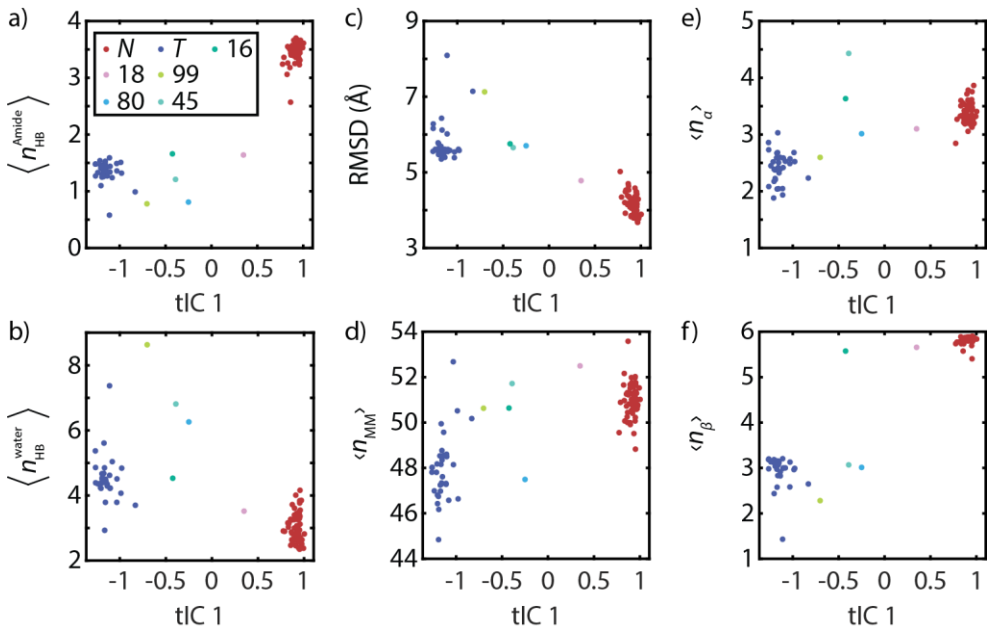

Figure S1. Average contact variables of Markov states along tIC1, including (a) average number of amide hydrogen bonds from B23 to B26 $\langle n_{HB}^{amide} \rangle$, (b) average number of water-amide hydrogen bonds from B23 to B26 $\langle n_{HB}^{water} \rangle$, (c) average RMSD of the heavy atoms with respect to the crystal structure, (d) average number of inter-monomer contacts $\langle n_{MM} \rangle$, (e) average number of α-contact $\langle n_\alpha \rangle$, and (f) number of β-contact $\langle n_\beta \rangle$. The average was computed over the structures within the same Markov state with equal weight.

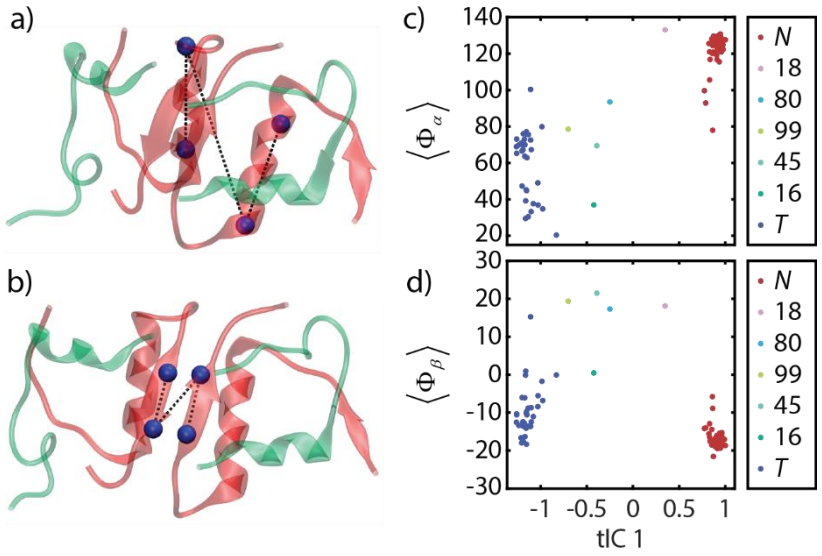

Figure S2. Structural characterization of the MSM including pseudo-dihedral angles of α-helices Φ_α and β-sheet Φ_β. (a) Illustration of the α pseudo-dihedral angle. The blue spheres represent the centers of mass (COMs) used for defining the pseudo dihedral angle. (b) Illustration of the β pseudo-dihedral angle. (c–d) Distribution of pseudo-dihedral angles along tIC1.

To illustrate relevant global changes of different dimer structures, we also defined a set of pseudo dihedral angles shown in Figure S2, including the α pseudo-dihedral angle Φ_α, and the β pseudo-dihedral angle Φ_β. These dihedral angles were defined using four centers of mass (COM). The α pseudo-dihedral angle was computed using COMs of the backbone atoms in B9–B11 of M1, the backbone atoms in B17–B19 of M1, the backbone atoms in B17–B19 of M2, and the backbone atoms in B9–B11 of M2. Similarly, β-pseudo dihedral angle was calculated using COMs of the backbone atoms of B24 backbone atoms in M1, B26 backbone atoms in M1, B26 backbone atoms in M2, and B24 backbone atoms in M2.

Comparison of MSM Structures to Dimer Dissociation Free Energy Surface

To investigate the relationship of native and twisted MSM structures with dimer dissociation and association mechanisms, we correlated collective variables characterizing the Markov states with recently calculated free energy surfaces along those CVs.[1] The overlap of these two characterizations of insulin dimer are presented in Figure S3, illustrating a close correspondence of the MSM native state and the global dimer free energy minimum. The native dimer state, however, does not show a clean mapping to a distinct free energy basin along all CVs, notably along the helix pseudo-dihedral coordinate.

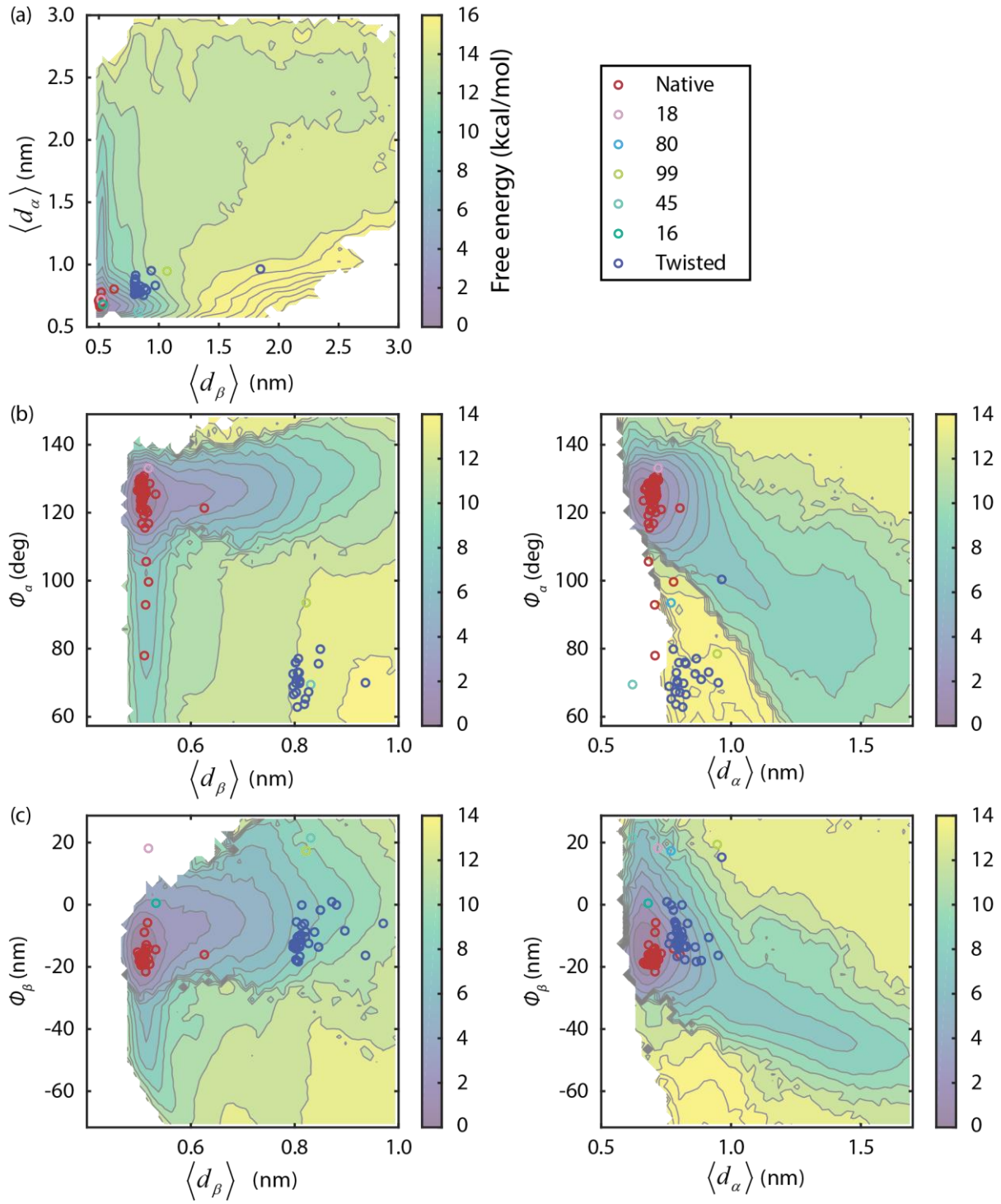

Figure S3. Projection of dimer MSM onto potentials of mean force (PMFs) generated from sampling of dimer dissociation.[1] (a) PMF as a function of average distance of α and β contacts. (b) PMF as function of α pseudo-dihedral angle, average β distance (left), and average α distance (right). (c) PMF as function of β pseudo-dihedral angle, average β distance (left), and average α distance (right).

Twelve-State Lumping of Dimer MSM

In addition to identifying hub or intermediate states, the top 20 tICs and eigenvectors of the transition matrix indicate that slower kinetics are associated with subgroups of the native and twisted states which correspond to clustered groups in our network plot. As a result, we investigated a reduction of the full MSM state space using a Robust Perron Cluster Cluster Analysis (PCCA+) of the first 20 eigenvectors of the transition matrix.[6] We concluded that both structural, spectral, and kinetic variations can be usefully lumped into 12 states outlined in Figure S4, including 4 states which lie within the native state (N_0, N_1, N_2, N_3), three in the twisted state (T_1, T_2, T_3), and 5 single Markov states from the 100-state MSM. The corresponding network plot for the 12-state lumping is shown in Figure S5.

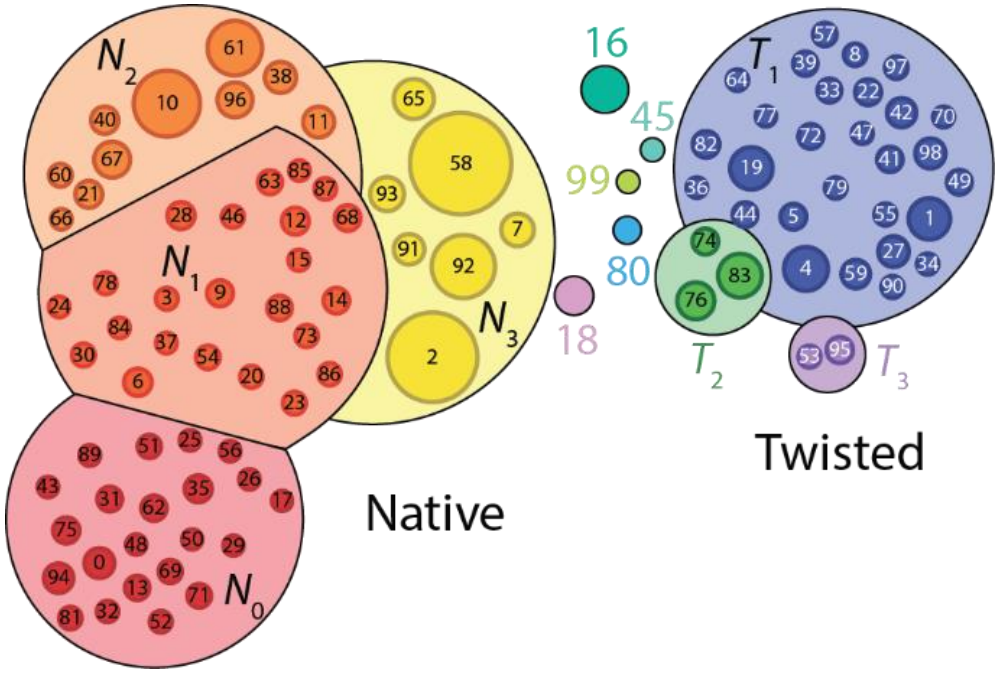

Figure S4. Markov State Model network plot of insulin dimer with state indices for all 100 Markov states and 12 coarse-grained states.

The N_1 state has 16% of the population and with a backbone that closely resembles the crystal structure and well-folded secondary structural elements, whereas the N_2 states (21%) have the A1 helix unfolded in one insulin monomer, and are a relatively isolated block of the transition matrix. N_3 are the most populated (29%) and these are states that are typically traversed in converting between native and twisted states. Almost all N_3 conformations retain the intermolecular α and β contacts of the crystal structure, but have one or both A1 helices unfolded in both monomers, with considerable conformational disorder for the termini of all chains. The T_1 state represents most of the twisted state population (20%) and most configurations are well-folded low disorder structures, whereas the T_2 and T_3 states are kinetically separated and generally have more disordered A chains. T_3 retains the twisted dimer structures but the configuration of the A chain differs from T_1 and is disordered.

The network plot readily identifies the intermediate states 80 and 45 as on-pathway intermediates in the conversion between native and twisted forms, with most of the flux passing between N_3 and T_1. The states 99, 16, 18, and T_3 are observed to be off-pathway intermediates or kinetic traps. Since states 80 and 45 have water intercalated between β-strands on each monomer, we conclude that the primary mechanism of dimer twisting involves water disrupting the specific interactions of the β-sheet without significant disruption to the hydrophobic core. The remaining weak contacts between hydrophobic sidechains of the B chain helix provide the orientational flexibility to reconfigure the β-strand sidechains and contacts in its new configuration.

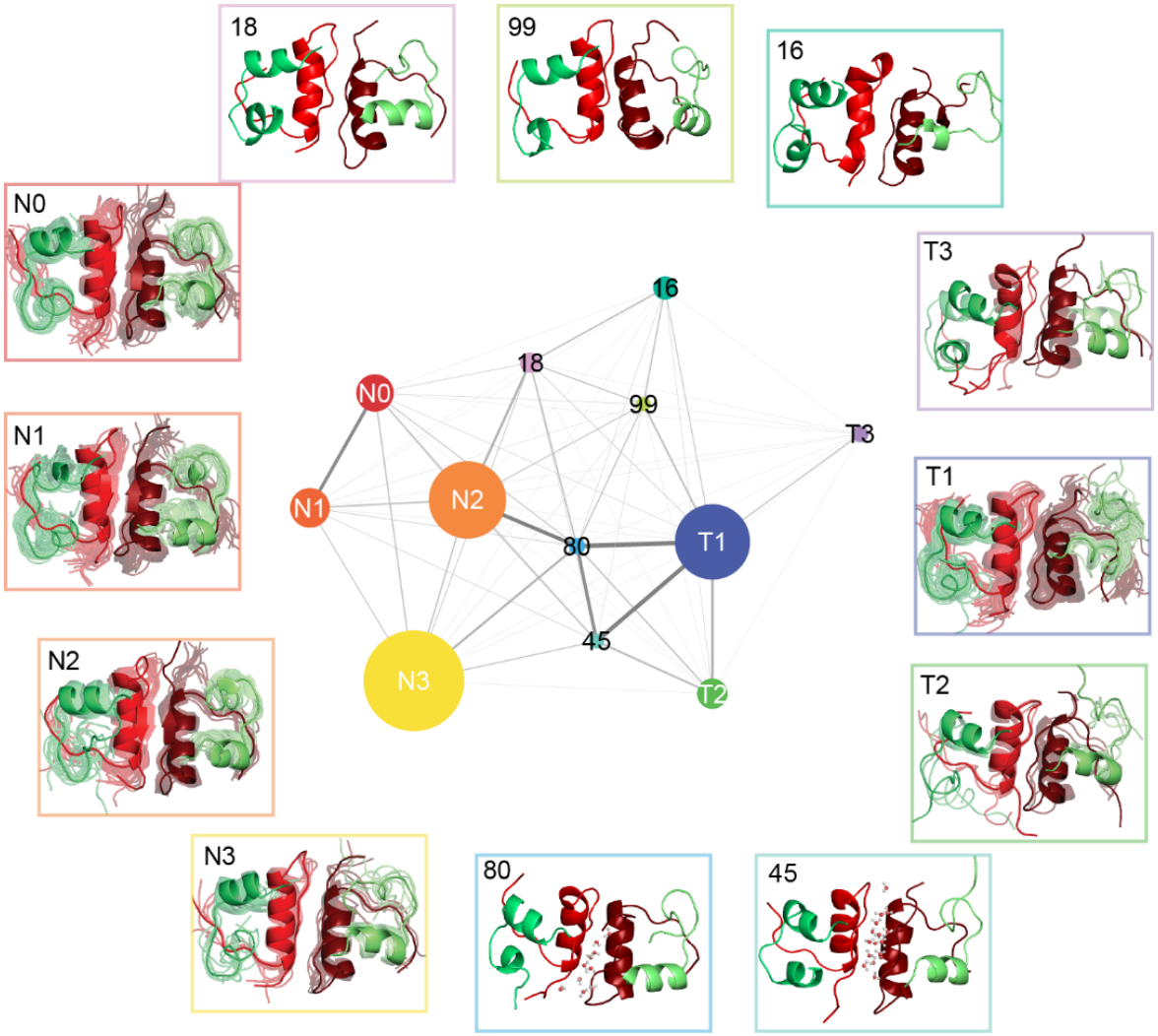

Figure S5. Network plot of reduced 12-state model. Medoids of each reduced state are shown on the side.

Vibrational Exciton Hamiltonian and Spectroscopic Maps

To facilitate understandings of the amide I IR spectra and correlate structures to spectra using spectroscopic maps described in the main manuscript, the vibrational exciton Hamiltonian can be written as

$$\hat{H}_{\mathrm{AmI}} = \sum_{n=1}^{N} \omega_n \,|n\rangle\langle n| + \sum_{n=1,m\neq n}^{N} J_{mn} \,|m\rangle\langle n| \tag{2}$$

The indices m and n indicate the site of vibration excitation. $|n\rangle$ refers to a excitation of site n with excitation energy ω_n whereas all the other sites remain in the vibrational ground state, and J_{mn} is the coupling between the sites n and m.

The 4P map is used to predict the amide I vibrational site energies (i.e. frequencies). The site frequency of an amide group in an MD frame is estimated based on the electrostatic potential evaluated at the position of the C, O, N, H atoms in the amide group:

$$\omega(t) = \omega_0 + \sum_{i=1}^{N_{\mathrm{sites}}} C_i \Phi_i(t) \tag{3}$$

In Eqn. (3), the site index i loops through the C, O, N, H atoms, and the associated coefficient C_i is given by the map. The map used throughout the study is 4PN-150,[7] which is optimized against the isotope-edited NuG2b protein. The map parameters are summarized in the following Table S2. Note that when proline residue is present, the electrostatics of the C_D atom will be evaluated instead of a normal H atom since its position corresponds to the H atom in normal amide bonds. Even though the parametrization of 4PN-150 map does not account for proline units, the only proline isotope-label in the present study is located at B28, which may influence the accuracy of the frequency prediction. However, all the other site-specific isotope labels should still be reasonable for interpretation purpose.

4PN-150 map was parametrized under MD simulations of CHARMM27 FF with SPC/E water model, but the coefficients were determined using CHARMM charges with modified glycine charges and TIP3P water model charges. However, the MD simulations of constructing the Markov State Models and trajectories for spectral simulations were performed based on AMBER99sb-ildn FF with TIP3P water. To account for differences in the electrostatics, we used CHARMM charges with modified glycine charges instead of AMBER charges for evaluating electrostatics when mapping site frequencies. Since the CHARMM charges do not have NHE/NH2 group that is present in the AMBER FF, the NHE/NH2 charges are adapted into the electrostatic calculations.

In addition to the frequency calculation of each site, the off-diagonal vibrational coupling J between the two amide oscillators is obtained with coupling maps. There are two primary types of vibrational coupling, through-bond coupling and through-space coupling. Through-bond coupling between adjacent sites is generated by the DFT-based nearest-neighbor coupling map, and through-space coupling is computed by a transition charge coupling map.[8] Note that our calculations do not account for C=O stretch from the terminal COOH group and sidechains.

	ω_0 (cm^{-1})	C_C (cm^{-1}.e/E_H)	C_O (cm^{-1}.e/E_H)	C_N (cm^{-1}.e/E_H)	C_H (cm^{-1}.e/E_H)	$\omega_{0,PRO}$ (cm^{-1})
4PN-150	1746.4	1121.7	-1571.7	-150	600	1730.0

Table S2. Map parameters used in this study.[7]

Additional Figures on Calculated Site Frequencies and Vibrational Couplings

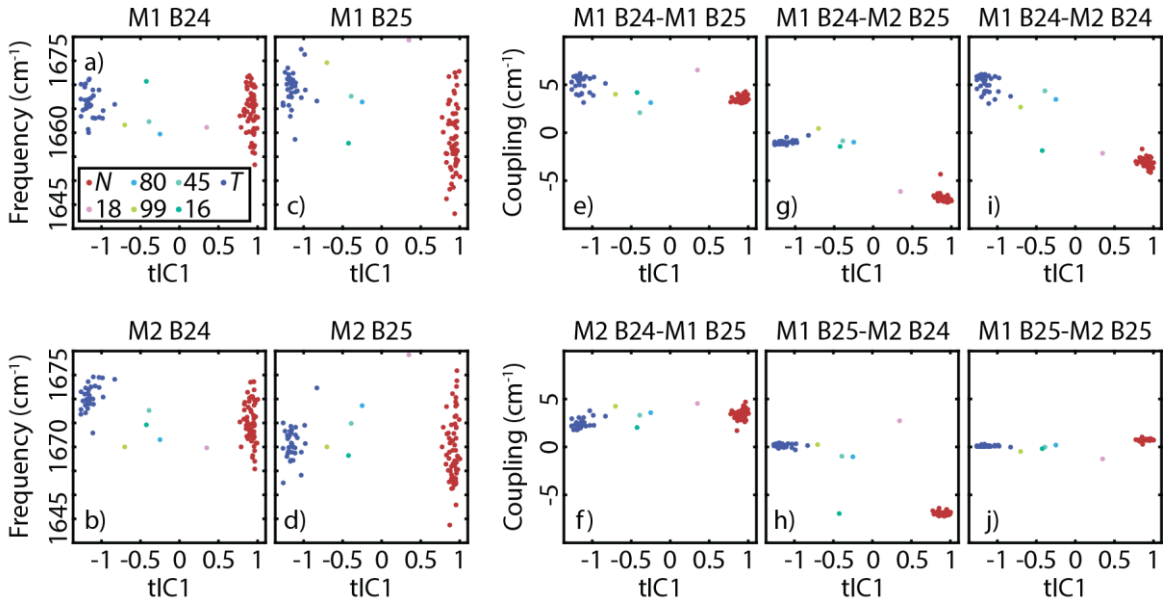

Figure S6. Average amide I parameters of the dimer MSMs. (a–d) Amide I frequency (cm^{-1}) along tIC1 of (a) monomer 1 (M1) B24F, (b) monomer 2 (M2) B24F, (c) M1 B25F, and (d) M2 B25F. (e–j) Excitonic coupling (cm^{-1}) along tIC1 between (e) M1 B24F and M1 B25F, (f) M2 B24F and M2 B25F, (g) M1 B24F and M2 B25F, (h) M1 B25F and M2 B24F, (i) M1 B24F and M2 B24F, and (j) M1 B25F and M2 B25F.

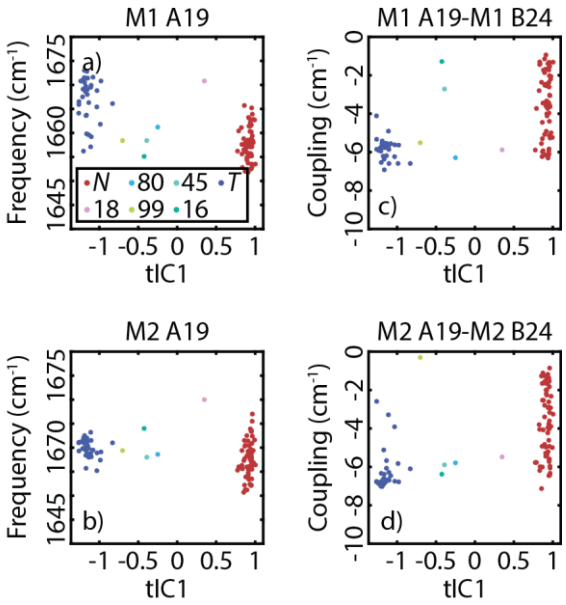

Figure S7. Average amide I parameters of the dimer MSMs. (a–b) Amide I frequency (cm^{-1}) along tIC1 of (a) M1 A29Y, and (b) M2 A19Y. (c–d) Excitonic coupling (cm^{-1}) along tIC1 between (c) M1 A19Y and M1 B24F, and (d) M2 A19Y and M2 B24F.

Supporting References

1. 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**, 5571-87.
2. Banerjee, P.; Mondal, S.; Bagchi, B., Effect of ethanol on insulin dimer dissociation. *J Chem Phys* **2019,** *150* (8), 084902.
3. 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* (11), 114902.
4. Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T. M. F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A., Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* **2019,** *16* (8), 670-673.
5. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers for an old bird. *Comput Phys Commun* **2014,** *185* (2), 604-613.
6. Deuflhard, P.; Weber, M., Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications* **2005,** *398*, 161-184.
7. Reppert, M.; Tokmakoff, A., Communication: Quantitative multi-site frequency maps for amide I vibrational spectroscopy. *J Chem Phys* **2015,** *143* (6), 061102.
8. la Cour Jansen, T.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J., Modeling the amide I bands of small peptides. *J Chem Phys* **2006,** *125* (4), 44312.