

# Computational IR Spectroscopy of Insulin Dimer Structure and Conformational Heterogeneity

Chi-Jui Feng<sup>1</sup>, Anton Sinititskiy<sup>2</sup>, Vijay Pande<sup>2</sup>, and Andrei Tokmakoff<sup>1,\*</sup>

<sup>1</sup>*Department of Chemistry, James Franck Institute and Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, United States*

<sup>2</sup>*Department of Bioengineering, Stanford University, Stanford, CA 94305, United States*

## Abstract

We have investigated the structure and conformational dynamics of insulin dimer using a Markov state model (MSM) built from extensive unbiased atomistic MD simulations, and performed infrared spectral simulations of the insulin MSM to describe how structural variation within the dimer can be experimentally resolved. Our model reveals two significant conformations to the dimer: a dominant native state consistent with other experimental structures of the dimer, and a twisted state with a structure that appears to reflect a  $\sim 55^\circ$  clockwise rotation of the native dimer interface. The twisted state primarily influences the contacts involving the C-terminus of insulin's B chain, shifting the registry of its intermolecular hydrogen bonds and reorganizing its sidechain packing. The MSM kinetics predict that these configurations exchange on a  $14 \mu\text{s}$  timescale, largely passing through two Markov states with a solvated dimer interface. Computational amide I spectroscopy of site-specifically  $^{13}\text{C}^{18}\text{O}$  labeled amides indicates that the native and twisted conformation can be distinguished through a series of single and dual labels involving the B24F, B25F, and B26Y residues. Additional structural heterogeneity and disorder is observed within the native and twisted states, and amide I spectroscopy can also be used to gain insight into this variation. This study will provide important interpretive tools for IR spectroscopic investigations of insulin structure, and transient IR kinetics experiments studying the conformational dynamics of insulin dimer.

---

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

## Introduction

Protein structure-function relationships have been rethought over the past two decades to account for the functional role of conformational disorder, which appears in intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs).<sup>1</sup> Proteins with structural heterogeneity and conformational disorder contain a variety of thermally accessible conformers undergoing rapid structural fluctuations or activated interconversion kinetics between free-energy basins on a complex free energy landscape.<sup>2-4</sup> Such IDPs and proteins with IDRs have been observed to be involved in many biological processes including regulation, signaling, and coupled-folding and binding to functional partners.<sup>1,5</sup>

Structural characterization of IDPs and proteins with IDRs requires an ensemble description, which creates numerous experimental challenges. Ensemble-structure determination is naturally an ill-posed problem in which degrees of freedom in relevant conformational states far exceed limited number of measurements and information content of experiments.<sup>6</sup> Also, fast conformational fluctuations and activated conformational dynamics cannot be decoupled,<sup>7</sup> meaning that conformational characterization inherently requires experimental probes with both high structural and temporal resolutions. Traditional structural tools are often limited by their intrinsic time resolution, which prohibits one from accessing conformational fluctuations and interconversion of conformers with time scales spanning many decades from picoseconds to microseconds.<sup>8-11</sup> NMR spectroscopy that measures chemical shifts and J couplings is limited by the coalescence time scale of ms such that faster conformational dynamics are averaged whereas relaxation experiments provide indirect structural information.<sup>12</sup> Optical spectroscopies do carry the advantage of femtosecond time-scales for their light-matter interaction, but in most cases have little structural information content. Advances are being made with IR and 2D IR spectroscopies which probe structure sensitive molecular vibrations with fs-ps time resolution, and thereby have the capability of structural characterization on a peptide or protein structure that is essentially static.<sup>13-23</sup>

Despite intensive experimental advances to investigating protein structures, all experimental methods face challenges of structural interpretation, which inevitably require structure-based models. Molecular dynamics (MD) simulation offers atomistic descriptions of protein structures and motions. Also, recent efforts of force field (FF) developments have improved the ability to study disordered proteins.<sup>24</sup> Clustering methods and kinetic frameworks have been developed to analyze conformational dynamics such as time-structure independent component analysis (tICA)<sup>25-28</sup> and Markov State Models (MSMs),<sup>29-30</sup> which provides a natural structural basis for computing experimental properties and studying interconversion kinetics. Computational advances have laid foundations of rationalizing experimental evidence, predicting experimental outcomes, and even helping design suitable experiments.

Direct quantitative comparison of atomistic protein structures and IR experiments is now possible using computational amide I spectroscopy.<sup>31</sup> Amide I IR spectroscopy, which probes the C=O stretching vibration of the polypeptide backbone in the 1600–1700 cm<sup>-1</sup> frequency range, can be used to interrogate local hydrogen bonding contacts to the carbonyl oxygen and structure-sensitive couplings between different amide carbonyls. This method can be used to predict traditional IR absorption and 2D IR spectra for simulated conformational distributions drawn from MD trajectories or MSMs, providing a unique route to investigate structure-spectrum correlations. Specifically, amide I vibrational frequencies can be predicted to high accuracy using spectroscopic

“maps” that relate frequency to the local electrostatic potential or electric field from MD force fields at specific amide carbonyl sites.<sup>32-39</sup> Similarly, maps for vibrational coupling between different amide I vibrations are used to calculate the interaction of multiple backbone amide groups.<sup>40-43</sup> These maps have reached the point of predicting amide I spectroscopic observables to a high level of accuracy with  $2\text{ cm}^{-1}$  frequency uncertainty and provided a direct way to structurally interpret experimental IR spectra.<sup>38</sup> This approach has additional power when interpreting site-specific isotope-edited IR spectroscopy using computational spectroscopy with MD simulations and MSMs. This approach has been used for a number of peptides and proteins, such as spectroscopic investigation of various secondary structures,<sup>44-46</sup> conformational characterization of TrpZip2,<sup>16</sup> disordered peptides and amyloid fibrils,<sup>22,47</sup> structural disorder of NTL9<sup>17</sup> and CD3 $\zeta$  transmembrane domain,<sup>48</sup> and investigating permeation mechanism of potassium ion in KcsA.<sup>20,49-50</sup>

In this study, we investigate the structural heterogeneity of insulin homodimer using a MSM built on MD simulations, and show how this structural variation can be interrogated using amide I IR spectroscopy of site-specifically isotope-labelled amide carbonyls. Insulin dimer dissociation is a necessary step prior to binding to insulin receptor, which is regarded as a coupled unfolding and unbinding process studied both computationally and experimentally.<sup>51-55</sup> Insulin monomer is a 51-residue peptide composed of two disulfide bonded chains with 21 residues on chain A and 30 residues on chain B. It has three  $\alpha$ -helical segments, a  $\beta$ -turn located from B20Gly to B23Gly, and the B-chain C-terminus ranging from B24Phe–B30Thr is known to be disordered with the extent depending on mutations and solution environment.<sup>56-63</sup> This intrinsically disordered region folds and binds into a well-defined inter-monomer  $\beta$ -sheet in the native dimer structure stabilized by hydrogen bonding and sidechain packing of the aromatic triplet of B24Phe, B25Phe, and B26Tyr.<sup>51, 63-65</sup> This B-chain C-terminus is also involved in the insulin receptor binding, exhibiting detachment of the B-chain C-terminus, a significant dihedral rotation on B24Phe, and hinging motion upon recognition with the insulin receptor,<sup>66-69</sup> which may share similar conformational transitions as in the dimerization.<sup>54</sup>

Insulin dimer has also become a useful model system for investigating coupled folding and binding dynamics. Computational insulin dimerization studies have focused almost entirely on the conformational characterization in the monomeric state, with the current viewpoint that the dimer structure resembles published crystal structures. However, recent simulation study on mutant dimer, showed that mutation of B24Phe to Gly resulted in additional dimer conformations including strongly interacting dimer and weakly interacting dimer, which involves conformational change between B10His and B13Glu, and increased solvation of the dimer interface.<sup>70</sup> A detailed description of the dimer conformational distribution still requires investigation, in particular accounting for changes of solvation environment such as pH, ionic strength, and temperature that can mediate dimer conformational changes and in turn dynamics of dimer dissociation and association.<sup>71-72</sup>

Here, we present a computational study to characterize equilibrium conformational ensemble of insulin dimer in aqueous solution using MSMs built off extensive MD simulations, and computational amide I spectroscopy to predict how conformational substates can be observed in experiments. The MSM of insulin dimer provides an all-atom, high-resolution structural basis with associated interconversion rates to compare to IR structural and kinetics measurements. In addition to structures that are similar to previous experimental results, the MSM reveals a second major conformational state, the twisted dimer state with register-shifted  $\beta$ -strands, that has not

been observed experimentally and computationally. Conformational exchange between twisted state and native state is predicted to occur on a 10-20  $\mu$ s time scale. With the help of computational amide I spectroscopy, we proposed site-specific isotope labels that can effectively distinguish these two major conformational states, as well as intermediate configurations visited as they interconvert. This study forms a computational basis for experimental investigation of insulin dimer structures and kinetic experiments that can resolve conformational transition between these states.

## Methods

### Molecular Dynamics Simulations

MD sampling was initiated from the native crystal structure of wild-type human insulin dimer (PDB: 3W7Y) using the AMBER99sb-ildn force field (FF) and TIP3P water model.<sup>73-74</sup> To match the preferred low pH conditions for infrared spectroscopy that increase solubility, titratable side chains including HIS, GLU, and N-terminal  $\text{NH}_3^+$  were protonated. The COOH group was replaced by CONH<sub>2</sub> group because the AMBER FF does not have parameters for COOH. The protonated insulin dimer was solvated in a cubic water box, having 9000 water molecules and additional Na<sup>+</sup> and Cl<sup>-</sup> ions to represent the ionic strength of 0.15 M.

Potential energy minimization was performed to ensure a reasonable starting structure for further temperature equilibration. To equilibrate temperature, the system was then gradually heated to 310 K in the NVT ensemble for 20 ps. Subsequent density equilibration was performed in the NPT ensemble at 310 K and 1 atm for 100 ps. Production runs were performed in the NPT ensemble at 310 K and 1 atm using OpenMM on Folding@home.<sup>75-76</sup> Several rounds of MD simulations were performed, with starting structures reseeded from previous rounds to accumulate more statistics on rare conformational transitions and to explore new configurations faster than a direct single MD simulation. The aggregate sampling of insulin dimer consisted of 409 MD trajectories with the total sampling of 1.71 ms.

### Construction of Dimer Markov State Model

The Markov State Model (MSM) of insulin dimer was constructed using MSMBuilder.<sup>77</sup> For clustering structures from MD sampling into conformational states, the collective variables (CVs) were inter-atomic distances between 102 C $\alpha$  atoms in the dimer, resulting in  $102 \times 101/2 = 5151$  C $\alpha$ -C $\alpha$  pairs. To ensure robustness of the results to outliers in the data, the quantile range was used instead of standard deviations. The distances were scaled with RobustScaler module to remove the medians and to the quantile range between the 25th quartile and the 75<sup>th</sup> quartile. Centering and scaling were performed independently on each pair. After scaling, time-structure independent components analysis (tICA) was applied on the time series data of all C $\alpha$ -C $\alpha$  pairs with the lag time of 125 ns.<sup>27-28</sup> The first 20 independent components (tICs) from tICA were selected as a subspace for subsequent k-medoids clustering, resulting in  $k = 100$  states.<sup>78-79</sup> These states were used as the basis for building the MSM with the lag time of 150 ns, with structures in each state drawn from the original MD sampling.

### Visualization of the MSM Network

To visualize the MSM, network graphs were generated using Gephi 0.9.2,<sup>80</sup> and the corresponding network layout was produced using the ForceAtlas algorithm.<sup>80</sup> Markov states are

treated as nodes with the radius proportional to its equilibrium population. Each node is connected with edges whose thicknesses reflect the sum of forward and backward transition probabilities between nodes. The ForceAtlas algorithm treats this network as a coupled spring-mass system, in which the spring constants correspond to the sum of the transition probabilities. Repulsive forces are added between each node to avoid spatial overlap of the nodes. The algorithm minimized the overall energy of the system by rearranging the layout of the network such that states stay in proximity to each other when they interconvert rapidly.

### Simulations of Amide I Spectra

Amide I spectral simulations were performed using a mixed quantum-classical model that builds on atomistic structures drawn from classical MD simulations.<sup>31</sup> Simulations of explicitly solvated structures for all 100 configurations within each of the 100 Markov states were performed using GROMACS 4.6.7,<sup>81</sup> using the AMBER99sb-ildn force field and TIP3P water.<sup>73, 82</sup> For each configuration, the solvent and ions were equilibrated around the position-restrained peptide for 100 ps at 300 K using the Berendsen thermostat.<sup>83</sup> Subsequent 1 ns production runs on the unrestrained protein were performed using the Nosé–Hoover thermostat under NVT conditions with a 1 fs integration step, and 20 fs/frame sampling rate for spectral simulations.<sup>84-85</sup> The final spectra for each Markov state was obtained by averaging the calculated spectra over all 100 initial configurations.

IR vibrational spectra were calculated from the Fourier transform of a transition dipole time-correlation function obtained from a mixed quantum-classical model implemented in the freely available `g_amide` and `g_spec` programs.<sup>86-87</sup> The model treats the amide I vibrations as a set of coupled oscillators assigned to each amide group of the backbone.<sup>31, 88</sup> Collective electrostatic variables are used to translate, or “map”, a series of instantaneous structures along a trajectory onto a time-dependent Hamiltonian and transition dipole moment for the amide I vibrations. The amide oscillators, or “sites”, are identified by atomic positions of the CONH peptide backbone linkages. The vibrational frequency of each site is generated from a 4-site potential map (4P) which evaluates the electrostatic potential at the C, O, N and H positions and maps it to a vibrational frequency. The 4P map used in this study, 4PN-150, has a frequency prediction accuracy of  $\sigma=2.25\text{ cm}^{-1}$ .<sup>38</sup> In addition to the frequency of each site, vibrational coupling between amide I oscillators is obtained using coupling maps for mechanical through-bond coupling and electrostatic through-space coupling.<sup>42</sup> Additional details are provided in the SI.

Amide I transition dipole correlation functions were calculated using a dynamic wavefunction propagation method,<sup>89</sup> using a Trotter expansion to reduce computation time.<sup>90-91</sup> The window time for calculations was set to 11 ps, equivalent to  $3\text{ cm}^{-1}$  frequency resolution. The model includes a 1.0 ps vibrational lifetime for amide I modes, to match experiments.<sup>88, 91</sup> The isotope frequency shift of  $^{13}\text{C}^{18}\text{O}$  isotope labels relative to  $^{12}\text{C}^{16}\text{O}$  is set to  $-65\text{ cm}^{-1}$ .<sup>92-93</sup>

## Results

### Insulin Dimer MSM

We performed 1.7 ms of aggregated unbiased equilibrium MD sampling of insulin dimer and constructed an all-atom, 100-state MSM using time-structure independent components analysis (tICA)<sup>25, 27-28</sup> and k-medoids clustering. tICA variationally combines coordinates to maximize their auto-correlation time such that the resulting independent components (tICs) encode structural and kinetic information ordered by decreasing implied time scale.<sup>27-28</sup> K-medoids clustering is chosen to construct the MSM to identify statistically well-sampled states and to obtain a real-structure medoid instead of an average structure for the cluster.

The MSM contains dimer structures that are mostly compact, with a negligibly small number of configurations that are fully dissociated or loosely bound structures with non-specific contacts. Monomer conformations in the medoid structures are mostly within the range of folds observed in experimental structures, and individual Markov states vary in the degree of the structural disorder. Structural variation between dimers primarily reflects non-native contacts at the dimer interface between monomers, conformational variation, and unfolded segments. The time scales for exchange between states vary from a few  $\mu$ s to tens of  $\mu$ s. A summary of structural variables describing all 100 states is provided in the SI.

An overview of the MSM is presented as a network plot in Fig. 1a. Each Markov state is represented as a node (circle) whose radius is proportional to its equilibrium population in the MSM. Edges connecting these nodes correspond to pathways of state interconversions, with line thickness proportional to the sum of interconversion rates between pairs of states. The network layout is optimized such that states in proximity show fast state interconversion, giving a coordinate-free visualization of the equilibrium population and kinetics of the MSM.

The nodes of the network diagram in Fig. 1a are color-mapped to tIC1 values, illustrating that tIC1, the slowest kinetic process of the system, describes shifts in population between two large groupings of states. Color coding the network plot by the state's average RMSD of heavy atoms relative to the dimer crystal structure (Fig. 1b), we observe that it is correlated with tIC1 (the correlation coefficient  $\rho = -0.64$ ). The RMSD values vary from 3.7 Å to 8.1 Å across all 100 states, but the distribution is bimodally separated by states in the lower left with low RMSD configurations closest to x-ray dimer structures with values of 4–5 Å, and high RMSD values of >5.5 Å in the top right. This indicates that tIC1 is related to a significant conformational change in the dimer. We also calculated correlations of tIC1 to several collective variables (CVs), and found strong correlations (>0.8) to torsion angles, distances, and hydrogen bonds involving the B chains at the dimer interface (see Table 1 and SI).

We used tIC1 values to group these two dominant clusters in order to understand the conformational changes that it describes. These coarse-grained states are shown as dashed circles in Fig. 1, and with the structural differences illustrated in Fig. 2. These exhibit twisted motion of the two monomers at the dimer interface, which leads to a disruption of native  $\beta$ -sheet contacts and a reconfiguration of sidechains at the dimer interface. Structural changes along tIC1 are described by several CVs used in previous simulation studies,<sup>53-54</sup> which are summarized in Table 1.

|  | Native ( <i>N</i> ) | Twisted ( <i>T</i> ) | $\rho$     |
|--|---------------------|----------------------|------------|
| $p$ (%)  | 66                  | 27                   |            |
| RMSD ( $\text{\AA}$ )  | 4.4 (0.3)           | 5.6 (0.3)            | -0.64      |
| $\langle \Phi_\alpha \rangle$ ( $^\circ$ )                   | 115.2 (15.3)        | 58.8 (14.7)          | 0.94       |
| $\langle n_{\text{HB}}^{\text{Amide}} \rangle$               | 3.4 (0.2)           | 1.4 (0.1)            | 0.96       |
| $\langle n_{\text{HB}}^{\text{water}} \rangle$               | 3.0 (0.3)           | 4.2 (0.5)            | -0.71      |
| $\langle n_{\text{MM}} \rangle$                              | 50.45 (0.65)        | 48.31 (1.84)         | 0.79       |
| $\langle n_\alpha \rangle$                                   | 3.28 (0.23)         | 2.47 (0.24)          | 0.69       |
| $\langle n_\beta \rangle$                                    | 5.79 (0.06)         | 2.93 (0.25)          | 0.97       |
| B22R $\phi/\psi$ ( $^\circ$ )                                | -99.2/4.6           | -83.2/130.2          | -0.82/0.95 |
| B23G $\phi/\psi$ ( $^\circ$ )                                | 59.1/120.5          | -65.0/-66.8          | 0.95/0.94  |
| $\langle d_\alpha \rangle$ ( $\text{\AA}$ )                  | 7.1 (0.33)          | 8.0 (0.42)           | -0.80      |
| $\langle d_\beta \rangle$ ( $\text{\AA}$ )                   | 5.1 (0.09)          | 8.5 (0.69)           | -0.83      |
| $\langle n_{\text{HB}}^{\text{A19}\cdots\text{B24}} \rangle$ | 0.60 (0.44)         | 0.80 (0.38)          | -0.51      |

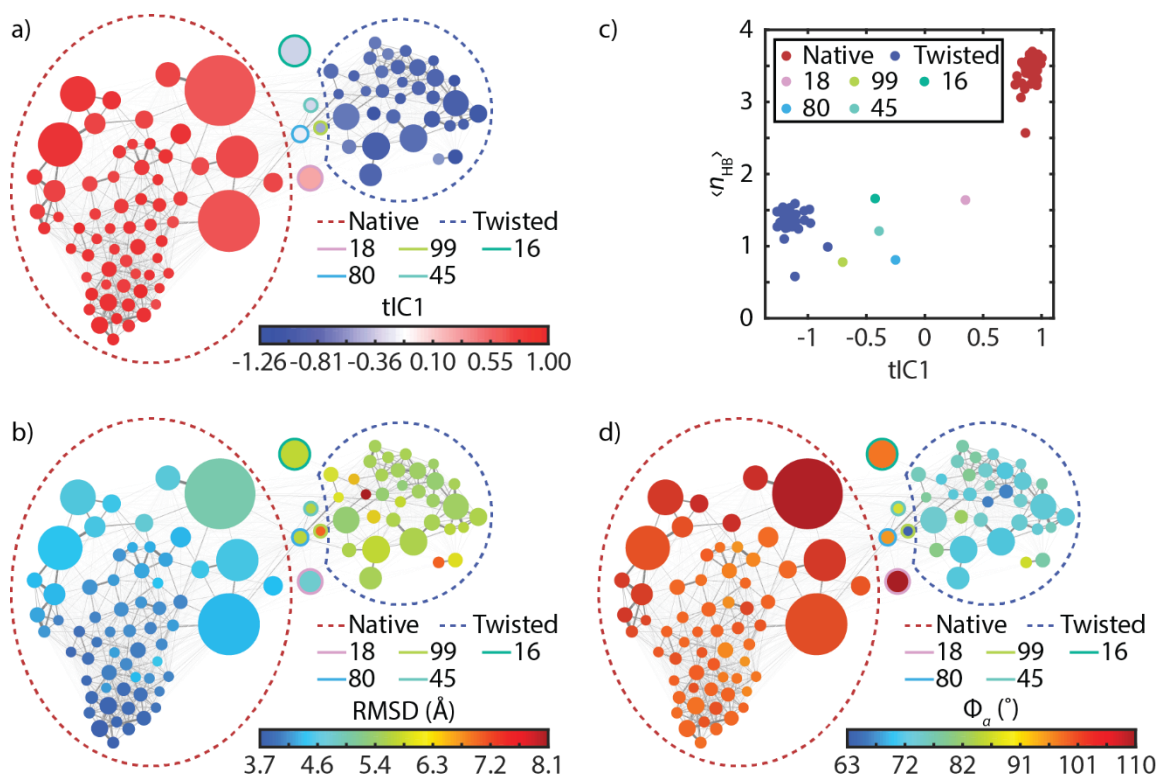
**Table 1.** Physical properties of the native (*N*) and twisted (*T*) dimer structures from the MSM, and their correlation coefficient to tIC1 ( $\rho$ ): population percentage  $p$ , RMSD of heavy atoms with respect to the crystal structure, average pseudo-dihedral angle between the B-chain helices  $\langle \Phi_\alpha \rangle$ , average number of amide HBs between inter-monomer  $\beta$  residues including B23G, B24F, B25F, B26Y  $\langle n_{\text{HB}}^{\text{Amide}} \rangle$ , average number of water-amide HBs of these  $\beta$  residues  $\langle n_{\text{HB}}^{\text{water}} \rangle$ , average number of inter-monomer contacts  $\langle n_{\text{MM}} \rangle$ ,  $\alpha$  contacts  $\langle n_\alpha \rangle$ , and  $\beta$  contacts  $\langle n_\beta \rangle$ , backbone torsion angles  $\phi/\psi$  for B22R and B23G, average distance of  $\alpha$  contacts  $\langle d_\alpha \rangle$  and  $\beta$  contacts  $\langle d_\beta \rangle$ , and average number of HBs between A19 amide unit and B24 amide unit  $\langle n_{\text{HB}}^{\text{A19}\cdots\text{B24}} \rangle$ . Bracket average indicates the average over all structures within the same Markov state whereas the bar average refers to the weighted average over states based on their equilibrium population.

The dominant coarse-grained state, the native dimer (*N*), has the largest population of 66%, an average RMSD value of 4.4  $\text{\AA}$ , and tIC1 values from 1.0 to 0.77. Configurations within this state are similar to the crystal structure, exhibiting an intact inter-monomer  $\beta$ -sheet with an average of 3.4 inter-monomer hydrogen bonds (HBs)  $\langle n_{\text{HB}}^{\text{Amide}} \rangle$  of  $\beta$  residues including B23G, B24F, B25F and B26Y. The native state also has two pairs of intermolecular sidechain contacts between B24F and B26F (Fig. 2), as well as contacts between the two B25F sidechains and between the B16Y and B26Y sidechains, all of which contribute significantly to stabilizing the dimer.<sup>94</sup> Most of the conformational variation within the native state arises from disorder away from the dimer interface in the N-termini of the B chain and in the fold of the N-terminal helix of the A chain.

The other dominant coarse-grained state, the twisted dimer (*T*), has a population of 27%, a larger RMSD of 5.8  $\text{\AA}$ , and tIC1 values from -0.83 to -1.26. Structures within this state appear as if one twisted the monomers of the native state clockwise (Fig. 2) around the inter-monomer axis,



resulting in several conformational changes at the dimer interface. The  $\beta$ -sheet residues of the B-chain remain aligned but have a registry shift relative to the native structure by one amide unit and a decrease of 2 HBs between strands ( $\langle n_{\text{HB}}^{\text{Amide}} \rangle = 1.4$ ). These strands appear wrapped around the dimer rather than the flat sheet observed in the native state. The unusual B20G–B23G turn of the native structure adopts a new configuration with canonical torsion angles. The native  $\beta$ -strand contacts between F and Y residues are replaced with non-native contacts between the two B24F sidechains. Most prominently the twisting motion is seen through the change in the relative orientation of the two B-chain helices, which can be quantified through an  $\alpha$ -helix pseudo-dihedral angle  $\langle \Phi_\alpha \rangle$ .<sup>54</sup> The helices rotate relative to each other by  $-55^\circ$  on average from  $115^\circ$  to  $59^\circ$  between  $N$  and  $T$  states. Fig. 1d shows a network plot color-coded by the pseudo-dihedral angle, illustrating how this conformation change identifies the coarse-grained states.

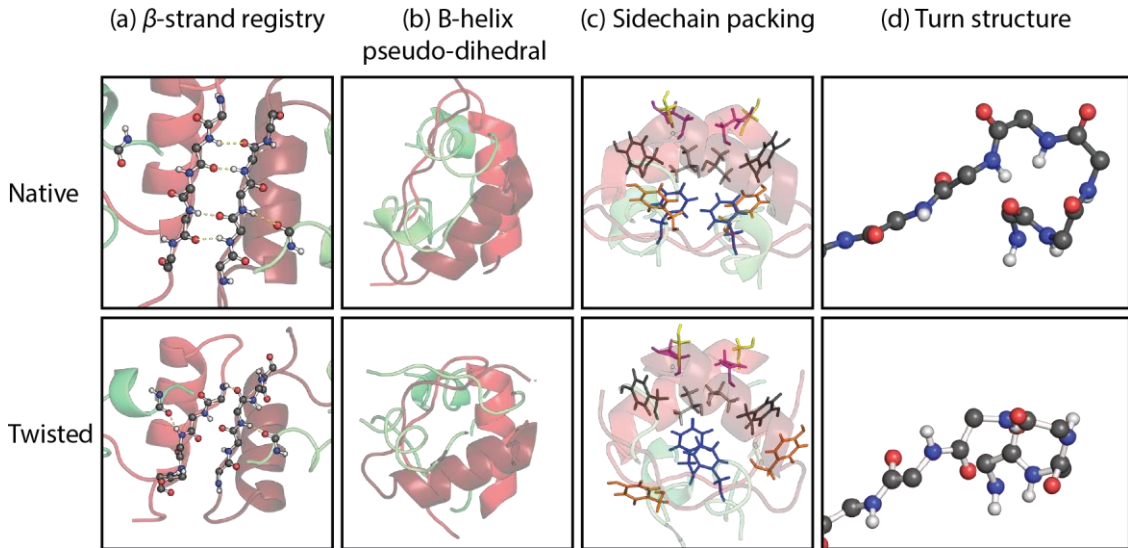


**Figure 1.** (a) Network representation of the dimer MSM with nodes color-coded by tIC1 values accounting for the slowest global process in the transition matrix. Thickness and color of the edges connecting nodes are proportional to the interconversion probabilities. Colored dashed circles identify the native and twisted states identified from a coarse-grained 3-state k-medoids clustering along tIC1. A full assignment of nodes to specific states is found in the SI. (b) Network plot color-coded by heavy atom RMSD with respect to crystal structure (PDB: 3W7Y). (c) Correlation of tIC1 with the average number of  $\beta$ -sheet HBs for the Markov state ( $\rho = 0.96$ ). (d) Network plot color-coded by  $\alpha$  pseudo-dihedral angles ( $\rho = 0.94$ ).

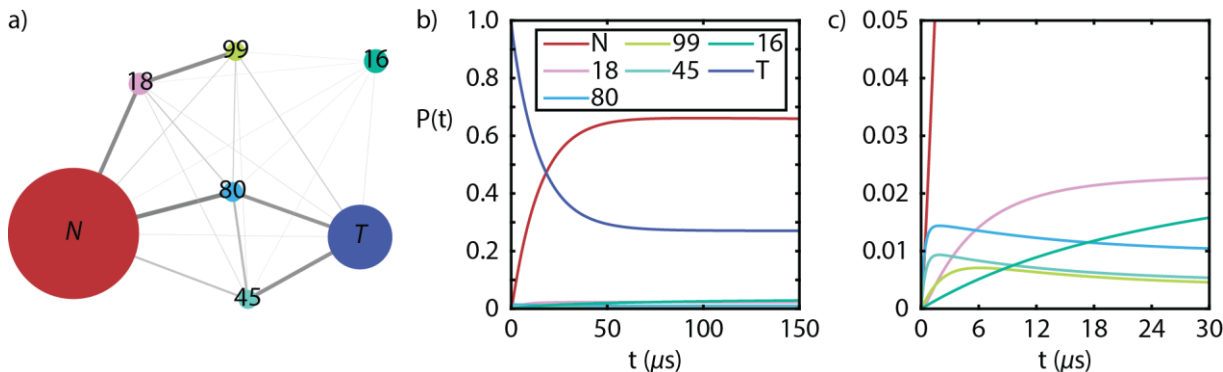
We quantified changes to the average number of inter-monomer contacts involving the B-chain  $\alpha$ -helix ( $n_\alpha$ ), the  $\beta$ -sheet residues ( $n_\beta$ ), and all residues including non-native contacts ( $n_{\text{MM}}$ ),



as described in the SI. This showed that there was only a slight decrease of the number of  $\alpha$  contacts from the native state to the twisted state, whereas the number of  $\beta$  contacts decreases by  $\sim 3$  contacts, accounting for majority of loss of inter-monomer contacts (Table 1). This observation indicates that the conformational change along tIC1 perturbs mostly on the local structure along  $\beta$ -sheet residues and sidechain packing while minimally disrupting other contacts.



**Figure 2.** Structural differences between the native,  $\langle \text{tIC1} \rangle \sim 0.8$ , and twisted states,  $\langle \text{tIC1} \rangle \sim -1.1$ , illustrated with MSM states 0 and 4 medoid structures. The columns illustrate the shift in backbone hydrogen bond registry from B24 to B26, the rotation of the B1 helix pseudo-dihedral angle from  $103^\circ$  to  $74^\circ$ , the changes in sidechain contacts at the dimer interface Blue: B24F, Orange: B26Y; Gray: B16Y; Light gray: B12V; Magenta: B13E; Yellow: B9S, and the change in turn structure for the B19C-B23G residues.

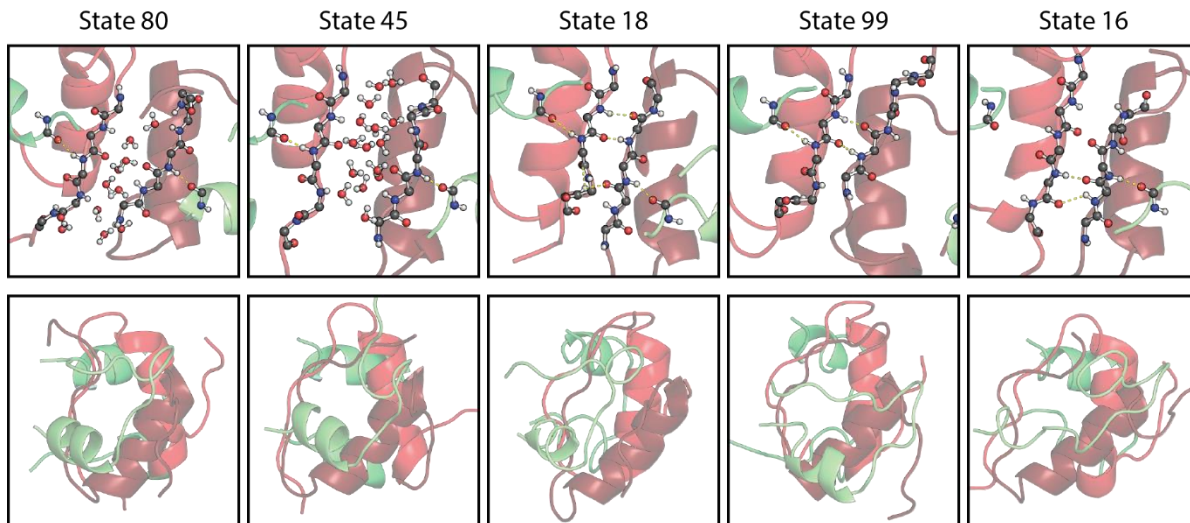


**Figure 3.** Seven-state lumping of native and twisted states with five intermediates. (a) Network plot for the new states and transition matrix. (b) Calculated equilibration kinetics tracking the exchange between native and twisted states when the population is initially in the twisted state.

## Kinetics

The remaining five states (16, 18, 45, 80, and 99) with intermediate tIC1 values are structurally diverse and account for 7% of the population (Fig. 4). Analysis of the top 20 tICs indicates that 11 tICs describe slow kinetics involving the transfer of population between one of these five intermediate Markov states and the rest, suggesting that these intermediate states are kinetic traps. From Fig. 1, we also infer that these intermediate states may play an important role in the transitions between the native and twisted states. To further investigate these intermediate states, we reduced the full MSM state space by lumping native and twisted states together and reduced the transition matrix for the resulting seven states using the method of Hummer and Szabo.<sup>95</sup> The network plot for this seven-state lumping is shown in Fig. 3a. One can see that states 80 and 45 act as on-pathway intermediates for the conversion of native and twisted forms, whereas 18, 99, and 16, which are identified in tICs 2–7, appear to be off-pathway, kinetic traps in this exchange process.

With this seven-state lumping we also investigated the kinetics of dimer twisting with the reduced rate matrix. Fig. 3b shows the time dependent population changes in the 7 states when the system is initiated entirely in the twisted state. Native and twisted populations exchange with a time constant of 14  $\mu$ s. Further reduction of the MSM to 3 states in which 80 and 45 are lumped as intermediates leads to a small increase in the observed kinetics to 21  $\mu$ s for the exchange between *N* and *T*. In Fig. 3c we also see that population in states 80 and 45 rise with a time-scale of 460 ns and then re-equilibrate with a 21  $\mu$ s decay as expected of on-path intermediates, whereas states 18 and 16 simply rise slowly to their equilibrium value. State 99's behavior lies between the other indicating that it also plays a non-negligible role in the exchange of native and twisted forms.



**Figure 4.** Structure of intermediate Markov states, showing a representative frame of each state with backbone atoms from A19Y and B23G–B27T, and a rotated side view illustrating the helix dihedral.

Representative structure for the intermediate states are illustrated in Fig. 4. States 80 and 45 have water molecules that penetrate the two  $\beta$ -strands such that all of the native HBs present in the  $\beta$ -sheet are replaced by water molecules. This suggests that the primary mechanism of dimer twisting involves water disrupting the specific interactions of the  $\beta$ -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  $\beta$ -strand sidechains and contacts in its new configuration.

A closer look at the makeup of the native and twisted states reveals that both contain conformational substates, which correspond to clustered groups in our network plot. Based on 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,<sup>96</sup> we identify four conformational substates which have native dimer contacts, but differ in their fold away from the dimer interface. For instance, two low RMSD native clusters mirror the crystal structure of insulin ( $N_0$ ,  $N_1$ ), whereas the another varies by the unfolding of the A1 helix of one insulin monomer ( $N_2$ ). The third native substate ( $N_3$ ) retains the intermolecular  $\alpha$  and  $\beta$  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 kinetics reveal that most conversion between native and twisted passes through the  $N_3$  state. PCCA+ also reveals that the twisted state is primarily one block of well-folded configurations ( $T_1$ ), with two minor substates ( $T_2$  and  $T_3$ ) that retain the twisted dimer interface, but vary in the structure and disorder of the A chains and chain termini. These states are discussed further below.

## Computational spectroscopy

### Spectral calculations for all 100 states and spectral trends.

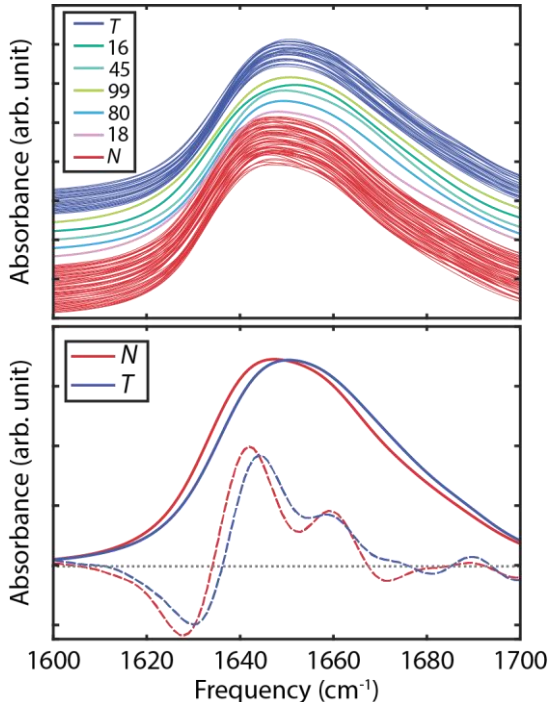
The MSM predicts the presence of two dominant conformational states for insulin dimer, one of which corresponds to the well-known dimer structure, in addition to an observed rate constant of 14  $\mu$ s for the interconversion between native and twisted forms. The prediction of these large-scale conformational changes that have not been previously observed, perhaps due to the  $\mu$ s interconversion timescale, raises the question of how such conformational changes could be observed experimentally. For this purpose, we investigated how local and global conformational changes of insulin dimer could be characterized with amide I infrared spectroscopy.

The amide I vibrational frequency shifts in proportion to the local electric field experienced by the carbonyl, and different carbonyl oscillators can couple to one another by through-bond (mechanical) and through-space (dipole-dipole) couplings.<sup>31</sup> The patterns of characteristic CO frequency shifts and couplings for different secondary structures gives rise to characteristic frequencies and band shapes for  $\alpha$ -helices and  $\beta$ -sheets. Most importantly, it is now possible to computationally model the protein amide I spectrum on the basis of atomistic structures drawn from MD simulation with quantitative accuracy. This tool has been used to characterize and refine conformational ensembles in peptides and small proteins.<sup>16-17, 19, 88, 91, 97</sup>

Amide I spectroscopy can be performed in different manners when combined with site-specific isotope labeling strategies. In the absence of labels, the relatively small frequency variations among the different CO vibrations ( $\sigma \sim 10$  cm<sup>-1</sup>) and similar coupling strength ( $V \sim 0-10$  cm<sup>-1</sup>) means that vibrations spectrally overlap. This leads to broad absorption bands which are insensitive to local structural variation but can be used to quantify secondary structure content. We refer to these as unlabeled (UL) spectra. Alternatively, an isotope-labeled carbonyl—here <sup>13</sup>C<sup>18</sup>O—can be used to shift the vibrational frequency well outside the band ( $\approx -60$  cm<sup>-1</sup>), which

both spectrally isolates and vibrationally decouples it from other amides in order to identify site-specific contacts.<sup>98-99</sup> In addition to the single-label experiments, dual-label experiments which insert pair of specific isotope labeled carbonyls selected to interact strongly when in close proximity and alignment, are particularly effective for characterizing hydrogen bonding contacts between two residues of the main chain.

With computational IR spectroscopy, one can compute a spectrum from structures as an interpretive tool, and also predict which isotope labels will be most informative for revealing specific changes in conformation, solvation environment, and hydrogen-bond contacts. We used this strategy to computationally study the IR spectra associated with all 100 Markov states to identify isotope labeling strategies for investigations of structural heterogeneity in insulin dimer.



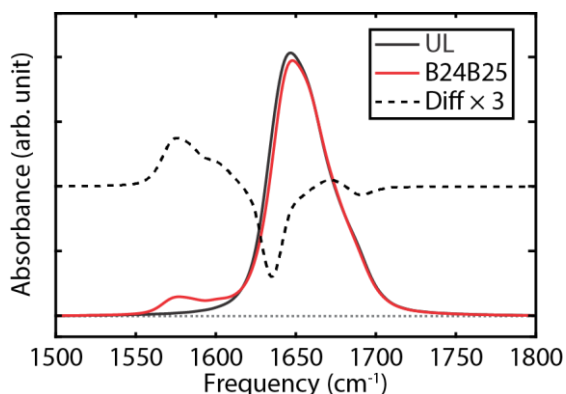
**Figure 5.** (a) Simulated UL FTIR spectra for all 100 Markov states ordered by increasing tIC 1 from top (twisted) to bottom (native). Spectra are vertically displaced for presentation purposes, and colored by their assignment to seven coarse-grained states (b) Comparison of population-weighted average IR spectra (solid line) and second-derivative spectra (dashed line) for the native and twisted states.

To begin we calculated the UL amide I absorption spectrum for all 100 Markov states. These spectra are shown in Fig. 5a, ordered by the state’s tIC1 value. We observe that all spectra are featureless asymmetric absorption bands, similar to experimental IR absorption spectra for the dimer,<sup>51-52</sup> but with little variation in the lineshape between states. Although the lineshape variations are nearly imperceptible within the *N* and *T* MSM states, the tIC1 value is found to correlate well with the frequency of the absorption maximum ( $\rho = -0.89$ ). Population-weighted average spectra over the *N* and *T* states reveals a predicted 3 cm<sup>-1</sup> band shift between states, from

$\langle \omega_N \rangle \sim 1650 \text{ cm}^{-1}$  to  $\langle \omega_T \rangle \sim 1647 \text{ cm}^{-1}$  (Fig. 5b). The asymmetry of the native spectrum can be explained in terms of the two transitions expected from anti-parallel  $\beta$ -sheets,<sup>44, 51-52</sup> including a weak  $\nu_{\parallel}$  vibrational transition at  $1680 \text{ cm}^{-1}$  and a stronger  $\nu_{\perp}$  band at  $1635 \text{ cm}^{-1}$ . Second derivative spectra only marginally improve the spectral differences, so although there are predictable differences, it appears difficult to distinguish  $N$  and  $T$  configurations from UL spectra in practice.

Since the  $N$  and  $T$  states vary most with the change of  $\beta$ -strand hydrogen bonds and other contacts at the dimer interface, changes in IR spectra between these states are more likely to be observed in spectra from isotope labels placed to interrogate these contacts.<sup>98-99</sup> To identify the most promising candidates, we performed calculations of isotope-edited IR spectra for all 100 states of the MSM starting with single site-specific labels for all 49 amide linkages of the monomer peptide backbone. These single labels shift in frequency depending on the local electric field experienced by the amide carbonyl, but is qualitatively best understood as being sensitive to the number and strength of hydrogen bonds to the carbonyl oxygen. Note that a single isotope label in this homodimer will result in two labels that can couple with one another depending on their proximity. The resulting spectra were analyzed individually or averaged by MSM population over all  $N$  and  $T$  coarse-grained states. Additionally, we performed calculations on 16 additional dual labels selected to isolate particular intra- and intermolecular contacts between the two amide groups.

To illustrate isotope labeling IR spectroscopy, Fig. 6 shows simulated IR spectra of UL insulin (black curve) and B24B25 double-labeled insulin for one MSM state with a native configuration (red curve). Upon  $^{13}\text{C}^{18}\text{O}$  isotopic substitution on both B24 and B25 amide units on the  $\beta$ -sheet, there are additional isotope-edited features appearing between  $1550\text{--}1620 \text{ cm}^{-1}$ . To highlight vibrational features associated with labeling, we calculate the difference spectrum between B24B25-labeled insulin and UL insulin (dashed curve). From this isotope-labelled difference spectrum, one can see the positive absorption change with the peak frequency of  $1576 \text{ cm}^{-1}$  corresponding to the labeled amide I vibrations, as well as negative features at  $1635$  and  $1691 \text{ cm}^{-1}$  that arise from the loss of those unlabeled residues.



**Figure 6.** Simulated FTIR spectra of the native state 0 of the MSM. Simulated UL FTIR spectrum (black), B24B25 labeled FTIR spectrum (red) and difference spectrum,  $\Delta A$ , between labeled spectrum and unlabeled spectrum (black dashed). Difference spectrum has been vertically displaced for presentation purpose.

Examples of the calculated isotope difference spectra for residues showing the largest frequency shifts between *N* and *T* states ( $>5\text{ cm}^{-1}$ ) are shown in Fig. 7a. Overall 18% of the labels provided a spectrally resolvable distinction between *N* and *T* states. The largest spectral differences are observed for the B24, B25, and B26 single labels that form the  $\beta$ -sheet in the native state, as well as double-labels that include a label on one of these sites. Residues at the N-terminus of the A chain also report on a significant conformational change in the A2 helix relative to the B chain between *N* and *T* states, and the A4 and B11 labels report on a change in amide hydrogen bonding strength within the A1 and B helices, respectively. In all cases the frequency shifts observed are less than the linewidth, indicating that quantifying *N* and *T* populations in a mixture will be challenging with only one label; however, the pattern of spectral variation among multiple labels can be used for a more accurate determination.

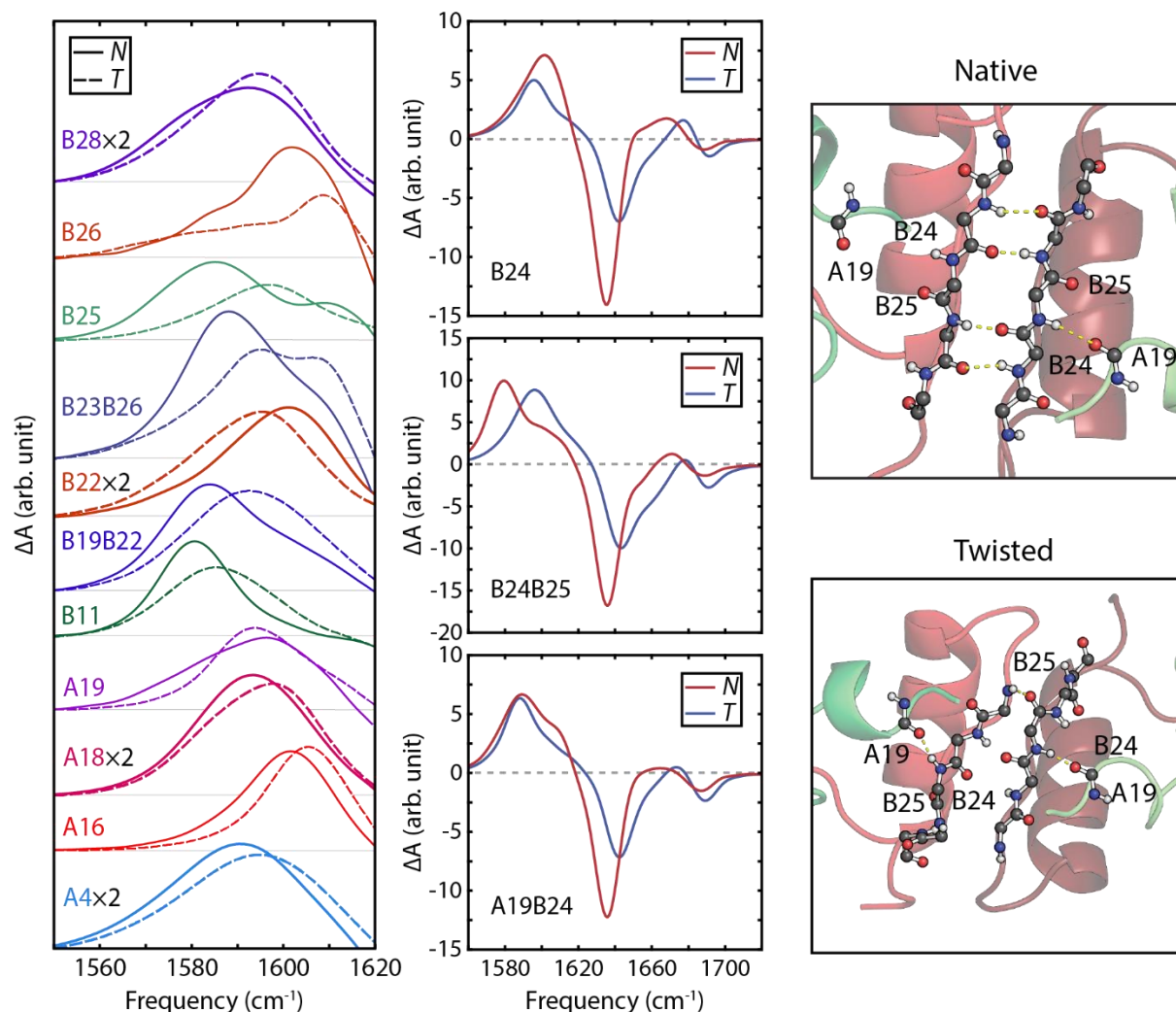
Looking closer at labels involving the dimer  $\beta$ -sheet residues, we now focus on the B24 single label, B24B25 dual label, and A19B24 dual label. The B24 and B24B25 labels are expected to probe intermolecular hydrogen-bond contacts within the  $\beta$ -sheet, whereas the A19B24 dual label should be sensitive to the intramolecular hydrogen bond between the A19 and B24 amide units in the native dimer, as illustrated in Fig. 7. The resulting isotope-labeled difference spectra for *N* and *T* states are shown in Fig. 7.

Overall, all three difference spectra exhibit common features, including an increase of absorption due to the isotope labelled residues between  $1560\text{--}1620\text{ cm}^{-1}$ , and the loss of intensity from the unlabeled band in the frequency range above  $1620\text{ cm}^{-1}$ . Labeled difference spectra of the native states (blue curves) share common loss features at  $1635\text{ cm}^{-1}$  and  $\sim 1690\text{ cm}^{-1}$  corresponding to the  $\nu_{\perp}$  and  $\nu_{\parallel}$  modes of the  $\beta$ -sheet, respectively. For each label, the  $\nu_{\perp}$  loss peak for the *T* state is suppressed by about half from the *N* state and blue-shifted to  $1642\text{ cm}^{-1}$ , and the  $\nu_{\perp}$  loss peak blue-shifts to  $1693\text{ cm}^{-1}$ . We observe this pattern also in the A19, B25, and B23B26 label spectra.

Structurally, the B24-isotope label is sensitive to the register shift of the two  $\beta$ -strands between *N* and *T* configurations, partially due to a decrease in hydrogen bonding, but also because of the through-space coupling between B24 labels on each monomer changes significantly ( $-3\text{ cm}^{-1}$  to  $5\text{ cm}^{-1}$ ). The calculated B24 difference spectra predicts that the *N* state has an asymmetric isotope-labeled amide I band peaked at  $1602\text{ cm}^{-1}$ , whereas the *T* state exhibits a frequency downshift to  $1596\text{ cm}^{-1}$ , a decrease in intensity and change in spectral lineshape.

The B24B25 dual-label of each monomer introduces four labels total into the dimer, which effectively isolates the  $\beta$ -sheet of the native state. In Fig. 7, the B24B25 difference spectrum of the *N* state has a peak frequency at  $1580\text{ cm}^{-1}$  and a shoulder at  $1605\text{ cm}^{-1}$ , which we attribute to the isotope-edited  $\nu_{\perp}$  mode and  $\nu_{\parallel}$   $\beta$ -sheet modes.<sup>44, 100</sup> The *T* state, in contrast, shows a symmetric labeled band with a peak frequency of  $1596\text{ cm}^{-1}$ .





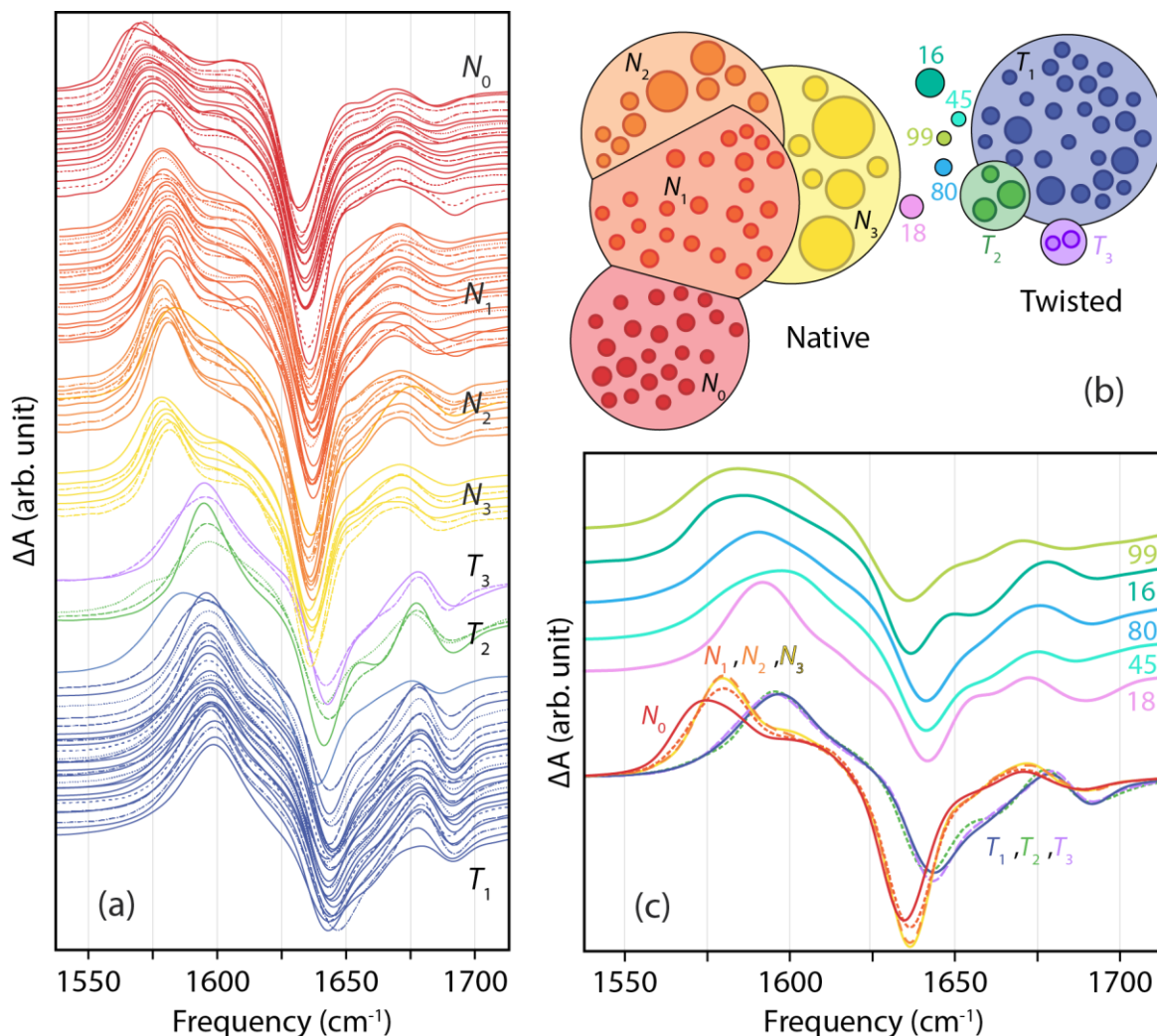
**Figure 7.** Left: Simulated isotope labeled IR difference spectra for several labels illustrating patterns of frequency shifts between the native and twisted states. Center: Simulated isotope difference spectra for B24, B24B25, and A19B24 labels including gain and loss features for *N* and *T* states. Right: Representative structures of both *N* and *T* states indicating structural differences in the A19, B24, and B25 carbonyls.

The A19B24 dual label probes the intramolecular H-bond contact between the A19 and B24 amide units away from dimer interface. The labeled difference spectrum of *N* shows a peak at 1589 cm<sup>-1</sup> and a pronounced shoulder at ~1610 cm<sup>-1</sup>, whereas the *T* state is observed to have a more symmetric peak with about the same peak frequency. This reflects the changes in the number of H-bonds between the B24 N–H and A19 C=O, from an average of 0.86 for the *T* state to 0.59 for the *N* state.

These calculations establish that there are labeling strategies available to distinguish *N* and *T* configurations, however, with some of the labels investigated we also found patterns of spectral variation within with *N* and *T* states with slight variation in spectral lineshape corresponding to clusters within our network plot. In Fig. 8, we illustrate these shifts with the B24B25 dual label and compare how spectra for all native and twisted MSM states maps onto coarse-grained substates



that were lumped with assistance of PCCA+. Although the  $N$  and  $T$  states share the general features described above, a closer look within 7  $N$  and  $T$  coarse-grained substates reveals that the individual MSM states also have a different frequency, linewidth and lineshape to their label transition and loss features (Fig. 8a). For instance, we find the subset of states  $N_0$ , which correspond to the kinetically clustered lowest RMSD states of the MSM, have the lowest frequency labelled bands ( $<1575\text{ cm}^{-1}$ ) compared to the  $N_1$ ,  $N_2$ , and  $N_3$  substates ( $>1575\text{ cm}^{-1}$ ). This is illustrated by the averaged spectra in Fig. 8c. This comparison also reveals that there is very little spectral variation for the B24B25 label within the individual or coarse-grained twisted states. Only one  $T$  state (77) is clearly distinguishable by its low frequency resonance. Overall 16% of the labels calculated showed such spectral variation within substates.



**Figure 8.** Spectral variation of the B24B25 label difference spectra among the 100 Markov states. (a) Individual spectra for native and twisted states ordered by peak transition frequency within the coarse-grained states obtained by PCCA+. (b) Corresponding color-coded native and twisted substates and intermediate states in 12-state coarse graining. (c) Comparison of population-weighted spectra for the four native and three twisted substates and the spectra of intermediate states.

Fig. 8c also compares the  $N$  and  $T$  spectra with those calculated for the five intermediate states. We observe that the intermediate states have significant spectral variation and are distinct from the native and twisted states. The spectra for these intermediate states do not share clear similarities with either  $N$  or  $T$  substates, and are much broader and featureless. Only state 18 has a clearly identifiable sharp resonance in its spectrum. These observations suggest that it may also be possible to distinguish the populations of intermediate states in IR kinetics measurements of structural interconversion of insulin dimer.

## Discussion and Conclusions

Our investigation of the structural variation of insulin dimer using extensive all-atom simulations, Markov state modeling, and computational amide I spectroscopy predicts the presence of two dominant conformations to insulin dimer, and illustrate how they can be experimentally resolved through isotope-edited IR spectroscopy. The native and twisted dimer conformations primarily differ by the change of contacts at the interface between the two bound monomers, and in the backbone hydrogen bonding and sidechain packing of the B-chain  $\beta$ -strand residues that form the intermolecular  $\beta$  sheet in the native configuration. The MSM kinetics indicate that the exchange of native and twisted populations occurs on a 21  $\mu$ s time-scale.

The twisted dimer conformation has, to our knowledge, not been experimentally observed. This is not surprising, given that it is a higher energy state than the native form and the predicted exchange kinetics are rapid; however, if present, confirming the presence of a twisted structure could have several consequences. On a fundamental level its presence should be considered in several factors varying from its influence on biological processes to interpretation of NMR experiments. It would influence our understanding of the role of many common B chain mutants found in insulin medications on the monomer-dimer equilibrium, and other principles for drug design. More generally, it provides evidence of the structural rearrangements and dynamical processes that can occur at protein-protein binding interfaces of complexes that are thought to bind in a unique site-specific manner.

From a dynamics perspective, it remains unknown what role a twisted dimer could play in the dimer dissociation and association processes, perhaps as an intermediate. Recent simulations of insulin dimer dissociation free energies described a broad distribution of possible energetically favorable dissociation pathways, bounded by two limiting cases: (1) a sequential process of disrupting B-chain  $\alpha$ -helix contacts prior to B-chain  $\beta$  contacts (the  $\alpha$  path), or (2)  $\beta$  contacts prior to  $\alpha$  contacts ( $\beta$  path).<sup>54</sup> To investigate connections between MSM intermediate states and on-path structures during the course of dissociation in that study, we projected the dimer MSM onto collective variables (CVs) describing the dissociation (see Figure S3). As one might expect, the well-solvated  $\beta$  strands of states 45 and 80 lie along the  $\beta$  path when observing CVs involving  $\beta$  contacts, but the  $\alpha$  pseudo-dihedral rotation lies closer to the  $\alpha$  path in a high free energy region rarely visited in the sampling of dimer dissociation. It is possible that the dimer MSM contains structures not sampled for the biased sampling of dissociation free energy landscapes,<sup>53-54</sup> but it is also possible that variations in the side-chain protonation state or force field contribute to this discrepancy. On the other hand, projecting the dimer MSM onto the free energy landscape constructed by Bagchi and coworkers showed that the native state and the twisted state lie in the

same free energy basin (Table 1 and Figs. 3–4 in ref. 53). As a result, the role of the twisted dimer in insulin dimer dissociation remains unclear at this time.

The simulation conditions we used were selected with IR spectroscopy in mind, since protonated sidechains (low pH) improves insulin solubility, destabilizes insulin hexamer, and reduces IR background absorptions from asymmetric  $\text{COO}^-$  vibrations in the same region as the labels. It is possible that these conditions favored the stabilization of the twisted configuration; however, a clear rationale is not apparent. The protonatable sidechains are away from the dimer binding interface, except for the case of B21E which may influence the conformation of the B19C–B23G turn.

To search for the presence of the twisted state, we find that IR spectroscopy targeting the B24, B25, and B26 residues in single and pairwise  $^{13}\text{C}^{18}\text{O}$  isotope labelling provides the best strategy for spectroscopically distinguishing the *N* and *T* configurations. Temperature and pH dependent studies could be used to influence the equilibrium between *N* and *T* states. Such IR spectra could also be used to track the exchange kinetics between *N* and *T* states when used as the probe of a temperature-jump experiment. Separately, we did investigate the variation of computed UV circular dichroism spectra<sup>101</sup> between *N* and *T* states, and found no significant change in the spectral shape, but a decrease in the magnitude of the molar ellipticity in the *T* state.

While a few labels result in large spectral differences between *N* and *T* states, most of the expected spectral changes for any particular label are predicted to show a small up-shift or down-shift in vibrational frequency, often less than the linewidth of the transition. Therefore, a robust strategy for studying insulin dimer is best performed with multiple labels whose pattern of spectral peaks can act as a type of “bar code” to identify the presence of the twisted dimer state. Indeed, we believe that the calculated isotope-labeled difference spectra for all 100 Markov states form a unique basis set for structural ensemble refinement from experiments using maximum entropy or Bayesian refinement tools.<sup>19, 91</sup> Finally, we note that the amide I spectral simulation tools presented here are equally applicable to 2D IR spectroscopy, which has improved capabilities for resolving isotope peak positions and distributions of spectra encoding structural variation. These observations set the stage for IR experimental studies to study insulin dimer structure and the dissociation/association equilibrium between the dimer(s) and monomer, and the kinetics of the coupled dimer conformational change and the dimer dissociation processes.

## Acknowledgments

This work was supported by the National Institutes of Health (R01-GM118774), and made use of resources provided by the University of Chicago Research Computing Center. C.-J. F. and A.T. thank Adam Antoszewski, Bodhi Vani, and Aaron Dinner at the University of Chicago for fruitful discussions on structural descriptions of the dimer MSM.

## References:

1. Tompa, P., Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **2012**, *37* (12), 509-16.
2. Papoian, G. A., Proteins with weakly funneled energy landscapes challenge the classical structure-function paradigm. *Proc Natl Acad Sci U S A* **2008**, *105* (38), 14237-8.
3. Burger, V.; Gurry, T.; Stultz, C., Intrinsically Disordered Proteins: Where Computation Meets Experiment. *Polymers* **2014**, *6* (10), 2684-2719.
4. Flock, T.; Weatheritt, R. J.; Latysheva, N. S.; Babu, M. M., Controlling entropy to tune the functions of intrinsically disordered regions. *Curr Opin Struct Biol* **2014**, *26*, 62-72.
5. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M., Classification of intrinsically disordered regions and proteins. *Chem Rev* **2014**, *114* (13), 6589-631.
6. Rieping, W.; Habeck, M.; Nilges, M., Inferential structure determination. *Science* **2005**, *309* (5732), 303-6.
7. Buchenberg, S.; Schaudinnus, N.; Stock, G., Hierarchical Biomolecular Dynamics: Picosecond Hydrogen Bonding Regulates Microsecond Conformational Transitions. *J Chem Theory Comput* **2015**, *11* (3), 1330-6.
8. Fleming, G. R.; Wolyne, P. G., Chemical Dynamics in Solution. *Physics Today* **1990**, *43* (5), 36-43.
9. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007**, *450* (7172), 964-72.
10. Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M., Principles of protein structural ensemble determination. *Curr Opin Struct Biol* **2017**, *42*, 106-116.
11. Markwick, P. R.; Malliavin, T.; Nilges, M., Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol* **2008**, *4* (9), e1000168.
12. Bryant, R. G., The NMR time scale. *Journal of Chemical Education* **1983**, *60* (11), 933.
13. Hamm, P.; Zanni, M., *Concepts and Methods of 2D Infrared Spectroscopy*. Cambridge University Press: 2011.
14. Baiz, C. R.; Reppert, M.; Tokmakoff, A., An Introduction to Protein 2D IR Spectroscopy. *Ultrafast Infrared Vibrational Spectroscopy* **2013**, 361-403.
15. Woutersen, S.; Hamm, P., Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy. *Journal of Physical Chemistry B* **2000**, *104* (47), 11316-11320.
16. Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L.; Knoester, J., Melting of a beta-hairpin peptide using isotope-edited 2D IR spectroscopy and simulations. *J Phys Chem B* **2010**, *114* (34), 10913-24.
17. Baiz, C. R.; Tokmakoff, A., Structural disorder of folded proteins: isotope-edited 2D IR spectroscopy and Markov state modeling. *Biophys J* **2015**, *108* (7), 1747-1757.
18. Feng, Y.; Huang, J.; Kim, S.; Shim, J. H.; MacKerell, A. D., Jr.; Ge, N. H., Structure of Penta-Alanine Investigated by Two-Dimensional Infrared Spectroscopy and Molecular Dynamics Simulation. *J Phys Chem B* **2016**, *120* (24), 5325-39.
19. Reppert, M.; Roy, A. R.; Tempkin, J. O.; Dinner, A. R.; Tokmakoff, A., Refining Disordered Peptide Ensembles with Computational Amide I Spectroscopy: Application to Elastin-Like Peptides. *J Phys Chem B* **2016**, *120* (44), 11395-11404.
20. Kratochvil, H. T.; Carr, J. K.; Matulef, K.; Annen, A. W.; Li, H.; Maj, M.; Ostmeier, J.; Serrano, A. L.; Raghuraman, H.; Moran, S. D.; Skinner, J. L.; Perozo, E.; Roux, B.; Valiyaveetil, F. I.; Zanni, M. T., Instantaneous ion configurations in the K<sup>+</sup> ion channel selectivity filter revealed by 2D IR spectroscopy. *Science* **2016**, *353* (6303), 1040-1044.
21. Ghosh, A.; Ostrander, J. S.; Zanni, M. T., Watching Proteins Wiggle: Mapping Structures with Two-Dimensional Infrared Spectroscopy. *Chem Rev* **2017**, *117* (16), 10726-10759.
22. Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P. N.; Chiu, C. C.; Cao, P.; Raleigh, D. P.; de Pablo, J. J.; Nowick, J. S.; Zanni, M. T., Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient beta-sheet. *Proc Natl Acad Sci U S A* **2013**, *110* (48), 19285-90.
23. Lomont, J. P.; Ostrander, J. S.; Ho, J. J.; Petti, M. K.; Zanni, M. T., Not All beta-Sheets Are the Same: Amyloid Infrared Spectra, Transition Dipole Strengths, and Couplings Investigated by 2D IR Spectroscopy. *J Phys Chem B* **2017**, *121* (38), 8935-8945.
24. Huang, J.; MacKerell, A. D., Jr., Force field development and simulations of intrinsically disordered proteins. *Curr Opin Struct Biol* **2018**, *48*, 40-48.
25. Molgedey, L.; Schuster, H. G., Separation of a mixture of independent signals using time delayed correlations. *Phys Rev Lett* **1994**, *72* (23), 3634-3637.

26. Naritomi, Y.; Fuchigami, S., Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J Chem Phys* **2011**, *134* (6), 065101.
27. Schwantes, C. R.; Pande, V. S., Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J Chem Theory Comput* **2013**, *9* (4), 2000-2009.
28. Perez-Hernandez, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noe, F., Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* **2013**, *139* (1), 015102.
29. Husic, B. E.; Pande, V. S., Markov State Models: From an Art to a Science. *J Am Chem Soc* **2018**, *140* (7), 2386-2396.
30. Noe, F.; Rosta, E., Markov Models of Molecular Kinetics. *J Chem Phys* **2019**, *151* (19), 190401.
31. Reppert, M.; Tokmakoff, A., Computational Amide I 2D IR Spectroscopy as a Probe of Protein Structure and Dynamics. *Annu Rev Phys Chem* **2016**, *67*, 359-86.
32. Bouř, P.; Keiderling, T. A., Empirical modeling of the peptide amide I band IR intensity in water solution. *The Journal of Chemical Physics* **2003**, *119* (21), 11253-11262.
33. Ham, S.; Kim, J.-H.; Lee, H.; Cho, M., Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA–nD<sub>2</sub>O complexes. *The Journal of Chemical Physics* **2003**, *118* (8), 3491-3498.
34. Hayashi, T.; Zhuang, W.; Mukamel, S., Electrostatic DFT map for the complete vibrational amide band of NMA. *J Phys Chem A* **2005**, *109* (43), 9747-59.
35. la Cour Jansen, T.; Knoester, J., A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy. *J Chem Phys* **2006**, *124* (4), 044502.
36. Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L., Development and validation of transferable amide I vibrational frequency maps for peptides. *J Phys Chem B* **2011**, *115* (13), 3713-24.
37. Reppert, M.; Tokmakoff, A., Electrostatic frequency shifts in amide I vibrational spectra: direct parameterization against experiment. *J Chem Phys* **2013**, *138* (13), 134116.
38. Reppert, M.; Tokmakoff, A., Communication: Quantitative multi-site frequency maps for amide I vibrational spectroscopy. *J Chem Phys* **2015**, *143* (6), 061102.
39. Torii, H., Amide I Vibrational Properties Affected by Hydrogen Bonding Out-of-Plane of the Peptide Group. *J Phys Chem Lett* **2015**, *6* (4), 727-33.
40. Torii, H.; Tasumi, M., Ab initio molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. *Journal of Raman Spectroscopy* **1998**, *29* (1), 81-86.
41. Ham, S.; Cha, S.; Choi, J.-H.; Cho, M., Amide I modes of tripeptides: Hessian matrix reconstruction and isotope effects. *The Journal of Chemical Physics* **2003**, *119* (3), 1451-1461.
42. 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.
43. Hayashi, T.; Mukamel, S., Vibrational-exciton couplings for the amide I, II, III, and A modes of peptides. *J Phys Chem B* **2007**, *111* (37), 11032-46.
44. Cheatum, C. M.; Tokmakoff, A.; Knoester, J., Signatures of beta-sheet secondary structures in linear and two-dimensional infrared spectroscopy. *J Chem Phys* **2004**, *120* (17), 8201-15.
45. Sengupta, N.; Maekawa, H.; Zhuang, W.; Toniolo, C.; Mukamel, S.; Tobias, D. J.; Ge, N. H., Sensitivity of 2D IR spectra to peptide helicity: a concerted experimental and simulation study of an octapeptide. *J Phys Chem B* **2009**, *113* (35), 12037-49.
46. Woys, A. M.; Almeida, A. M.; Wang, L.; Chiu, C. C.; McGovern, M.; de Pablo, J. J.; Skinner, J. L.; Gellman, S. H.; Zanni, M. T., Parallel beta-sheet vibrational couplings revealed by 2D IR spectroscopy of an isotopically labeled macrocycle: quantitative benchmark for the interpretation of amyloid and protein infrared spectra. *J Am Chem Soc* **2012**, *134* (46), 19118-28.
47. Wang, L.; Middleton, C. T.; Singh, S.; Reddy, A. S.; Woys, A. M.; Strasfeld, D. B.; Marek, P.; Raleigh, D. P.; de Pablo, J. J.; Zanni, M. T.; Skinner, J. L., 2DIR spectroscopy of human amylin fibrils reflects stable beta-sheet structure. *J Am Chem Soc* **2011**, *133* (40), 16062-71.
48. Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T., Structural disorder of the CD3zeta transmembrane domain studied with 2D IR spectroscopy and molecular dynamics simulations. *J Phys Chem B* **2006**, *110* (48), 24740-9.
49. Stevenson, P.; Gotz, C.; Baiz, C. R.; Akerboom, J.; Tokmakoff, A.; Vaziri, A., Visualizing KcsA conformational changes upon ion binding by infrared spectroscopy and atomistic modeling. *J Phys Chem B* **2015**, *119* (18), 5824-31.

50. Kratochvil, H. T.; Maj, M.; Matulef, K.; Annen, A. W.; Ostmeyer, J.; Perozo, E.; Roux, B.; Valiyaveetil, F. I.; Zanni, M. T., Probing the Effects of Gating on the Ion Occupancy of the K(+) Channel Selectivity Filter Using Two-Dimensional Infrared Spectroscopy. *J Am Chem Soc* **2017**, *139* (26), 8837-8845.
51. Ganim, Z.; Jones, K. C.; Tokmakoff, A., Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy. *Phys Chem Chem Phys* **2010**, *12* (14), 3579-88.
52. Zhang, X. X.; Jones, K. C.; Fitzpatrick, A.; Peng, C. S.; Feng, C. J.; Baiz, C. R.; Tokmakoff, A., Studying Protein-Protein Binding through T-Jump Induced Dissociation: Transient 2D IR Spectroscopy of Insulin Dimer. *J Phys Chem B* **2016**, *120* (23), 5134-45.
53. 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.
54. 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.
55. Desmond, J. L.; Koner, D.; Meuwly, M., Probing the Differential Dynamics of the Monomeric and Dimeric Insulin from Amide-I IR Spectroscopy. *J Phys Chem B* **2019**, *123* (30), 6588-6598.
56. Hua, Q. X.; Shoelson, S. E.; Kochoyan, M.; Weiss, M. A., Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **1991**, *354* (6350), 238-41.
57. Ludvigsen, S.; Roy, M.; Thogersen, H.; Kaarsholm, N. C., High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr-->His, as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **1994**, *33* (26), 7998-8006.
58. Olsen, H. B.; Ludvigsen, S.; Kaarsholm, N. C., Solution structure of an engineered insulin monomer at neutral pH. *Biochemistry* **1996**, *35* (27), 8836-45.
59. Keller, D.; Clausen, R.; Josefsen, K.; Led, J. J., Flexibility and bioactivity of insulin: an NMR investigation of the solution structure and folding of an unusually flexible human insulin mutant with increased biological activity. *Biochemistry* **2001**, *40* (35), 10732-40.
60. Ludvigsen, S.; Olsen, H. B.; Kaarsholm, N. C., A structural switch in a mutant insulin exposes key residues for receptor binding. *J Mol Biol* **1998**, *279* (1), 1-7.
61. Kosinova, L.; Veverka, V.; Novotna, P.; Collinsova, M.; Urbanova, M.; Moody, N. R.; Turkenburg, J. P.; Jiracek, J.; Brzozowski, A. M.; Zakova, L., Insight into the structural and biological relevance of the T/R transition of the N-terminus of the B-chain in human insulin. *Biochemistry* **2014**, *53* (21), 3392-402.
62. Bocian, W.; Sitkowski, J.; Bednarek, E.; Tarnowska, A.; Kawecki, R.; Kozerski, L., Structure of human insulin monomer in water/acetonitrile solution. *J Biomol NMR* **2008**, *40* (1), 55-64.
63. Zoete, V.; Meuwly, M.; Karplus, M., A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer. *J Mol Biol* **2004**, *342* (3), 913-29.
64. Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. M.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D.; et al., The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Philos Trans R Soc Lond B Biol Sci* **1988**, *319* (1195), 369-456.
65. Jørgensen, A. M. M.; Kristensen, S. M.; Led, J. J.; Balschmidt, P., Three-dimensional solution structure of an insulin dimer. *Journal of Molecular Biology* **1992**, *227* (4), 1146-1163.
66. Menting, J. G.; Yang, Y.; Chan, S. J.; Phillips, N. B.; Smith, B. J.; Whittaker, J.; Wickramasinghe, N. P.; Whittaker, L. J.; Pandeyarajan, V.; Wan, Z. L.; Yadav, S. P.; Carroll, J. M.; Strokes, N.; Roberts, C. T., Jr.; Ismail-Beigi, F.; Milewski, W.; Steiner, D. F.; Chauhan, V. S.; Ward, C. W.; Weiss, M. A.; Lawrence, M. C., Protective hinge in insulin opens to enable its receptor engagement. *Proc Natl Acad Sci U S A* **2014**, *111* (33), E3395-404.
67. Croll, T. I.; Smith, B. J.; Margetts, M. B.; Whittaker, J.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C., Higher-Resolution Structure of the Human Insulin Receptor Ectodomain: Multi-Modal Inclusion of the Insert Domain. *Structure* **2016**, *24* (3), 469-76.
68. Gutmann, T.; Kim, K. H.; Grzybek, M.; Walz, T.; Coskun, U., Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. *J Cell Biol* **2018**, *217* (5), 1643-1649.
69. Weis, F.; Menting, J. G.; Margetts, M. B.; Chan, S. J.; Xu, Y.; Tennagels, N.; Wohlfart, P.; Langer, T.; Muller, C. W.; Dreyer, M. K.; Lawrence, M. C., The signalling conformation of the insulin receptor ectodomain. *Nat Commun* **2018**, *9* (1), 4420.
70. Raghunathan, S.; El Hage, K.; Desmond, J. L.; Zhang, L.; Meuwly, M., The Role of Water in the Stability of Wild-type and Mutant Insulin Dimers. *J Phys Chem B* **2018**, *122* (28), 7038-7048.
71. Wicky, B. I. M.; Shammas, S. L.; Clarke, J., Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. *Proc Natl Acad Sci U S A* **2017**, *114* (37), 9882-9887.



72. Boreikaite, V.; Wicky, B. I. M.; Watt, I. N.; Clarke, J.; Walker, J. E., Extrinsic conditions influence the self-association and structure of IF1, the regulatory protein of mitochondrial ATP synthase. *Proc Natl Acad Sci U S A* **2019**, *116* (21), 10354-10359.
73. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-8.
74. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79* (2), 926-935.
75. Shirts, M.; Pande, V. S., COMPUTING: Screen Savers of the World Unite! *Science* **2000**, *290* (5498), 1903-4.
76. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **2017**, *13* (7), e1005659.
77. Harrigan, M. P.; Sultan, M. M.; Hernandez, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; McGibbon, R. T.; Pande, V. S., MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys J* **2017**, *112* (1), 10-15.
78. Keller, B.; Daura, X.; van Gunsteren, W. F., Comparing geometric and kinetic cluster algorithms for molecular simulation data. *J Chem Phys* **2010**, *132* (7), 074110.
79. Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S., MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J Chem Theory Comput* **2011**, *7* (10), 3412-3419.
80. Bastian, M.; Heymann, S.; Jacomy, M. In *Gephi: an open source software for exploring and manipulating networks*, Third international AAAI conference on weblogs and social media, 2009.
81. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845-54.
82. Boonstra, S.; Onck, P. R.; Giessen, E., CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J Phys Chem B* **2016**, *120* (15), 3692-8.
83. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984**, *81* (8), 3684-3690.
84. Nosé, S., A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *Journal of Chemical Physics* **1984**, *81* (1), 511-519.
85. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys* **1985**, *31* (3), 1695-1697.
86. Reppert, M., g\_amide. **2017**.
87. Reppert, M.; Feng, C.-J., g\_spec. **2017**.
88. Feng, C. J.; Tokmakoff, A., The dynamics of peptide-water interactions in dialanine: An ultrafast amide I 2D IR and computational spectroscopy study. *J Chem Phys* **2017**, *147* (8), 085101.
89. Torii, H., Effects of intermolecular vibrational coupling and liquid dynamics on the polarized Raman and two-dimensional infrared spectral profiles of liquid N,N-dimethylformamide analyzed with a time-domain computational method. *J Phys Chem A* **2006**, *110* (14), 4822-32.
90. Liang, C.; Jansen, T. L., An Efficient N(3)-Scaling Propagation Scheme for Simulating Two-Dimensional Infrared and Visible Spectra. *J Chem Theory Comput* **2012**, *8* (5), 1706-13.
91. Feng, C. J.; Dhayalan, B.; Tokmakoff, A., Refinement of Peptide Conformational Ensembles by 2D IR Spectroscopy: Application to Ala-Ala-Ala. *Biophys J* **2018**, *114* (12), 2820-2832.
92. Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T., Site-specific examination of secondary structure and orientation determination in membrane proteins: The peptidic<sup>13</sup>C-<sup>18</sup>O group as a novel infrared probe. *Biopolymers* **2001**, *59* (6), 396-401.
93. Decatur, S. M., Elucidation of residue-level structure and dynamics of polypeptides via isotope-edited infrared spectroscopy. *Acc Chem Res* **2006**, *39* (3), 169-75.
94. Zoete, V.; Meuwly, M.; Karplus, M., Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition. *Proteins* **2005**, *61* (1), 79-93.
95. Hummer, G.; Szabo, A., Optimal Dimensionality Reduction of Multistate Kinetic and Markov-State Models. *J Phys Chem B* **2015**, *119* (29), 9029-37.
96. Deuffhard, P.; Weber, M., Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications* **2005**, *398*, 161-184.



97. Baiz, C. R.; Lin, Y. S.; Peng, C. S.; Beauchamp, K. A.; Voelz, V. A.; Pande, V. S.; Tokmakoff, A., A molecular interpretation of 2D IR protein folding experiments with Markov state models. *Biophys J* **2014**, *106* (6), 1359-70.
98. Dong, J.; Wan, Z. L.; Chu, Y. C.; Nakagawa, S. N.; Katsoyannis, P. G.; Weiss, M. A.; Carey, P. R., Isotope-edited Raman spectroscopy of proteins: a general strategy to probe individual peptide bonds with application to insulin. *J Am Chem Soc* **2001**, *123* (32), 7919-20.
99. Dhayalan, B.; Fitzpatrick, A.; Mandal, K.; Whittaker, J.; Weiss, M. A.; Tokmakoff, A.; Kent, S. B., Efficient Total Chemical Synthesis of (13) C=(18) O Isotopomers of Human Insulin for Isotope-Edited FTIR. *Chembiochem* **2016**, *17* (5), 415-20.
100. Miyazawa, T.; Blout, E. R., The Infrared Spectra of Polypeptides in Various Conformations: Amide I and II Bands1. *Journal of the American Chemical Society* **1961**, *83* (3), 712-719.
101. Mavridis, L.; Janes, R. W., PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics* **2017**, *33* (1), 56-63.