Full Paper

Israel Journal of Chemistry

DOI: 10.1002/ijch.201300141

Exploring the Protein Folding Dynamics of Beta3s with Two-Dimensional Ultraviolet (2DUV) Spectroscopy

Zaizhi Lai,[a] Jun Jiang,[b, c] Shaul Mukamel,[d] and Jin Wang*[a]

Abstract: Ultraviolet (UV) spectra of proteins originate from electronic excitations of their backbone chromophore and aromatic side chains and provide a sensitive probe of the secondary structure. Recently developed femtosecond lasers allow multidimensional spectroscopy to be extended into the UV regime. Two-dimensional UV (2DUV) techniques, with short pulses, provide a promising tool to study the structures and dynamics of proteins. We combined 2DUV spectroscopy and molecular dynamics generated free energy profiles to simulate the protein electronic transitions and UV photon echo signals to monitor the protein folding process of the small protein Beta3s. Near-ultraviolet (NUV) and far-ultraviolet (FUV) signals illustrate the variation of the 2D correlation plots when the protein evolves along the underlying free energy landscape. Chiral polarization configurations of the NUV and FUV pulses are sensitive to protein structural evolution. This work provides a protocol for applying multidimensional UV spectroscopy to study protein folding.

Keywords: amino acids · laser spectroscopy · molecular dynamics · protein folding · UV/Vis spectroscopy

1 Introduction

Theoretical and experimental studies have provided considerable insights into the protein folding process. However, monitoring protein folding dynamics is still challenging. Experiments on the kinetics and thermodynamics of folding do not usually provide atomic-level structural changes, and thus, the microscopic dynamics characterizing the folding process. Computer simulations performed at various levels of complexity, from simple lattice models to all-atom models, fill in some of the gaps in our knowledge of protein folding. However, most folding processes of interest occur on timescales (microsecond to second) that are inaccessible to standard all-atom molecular dynamics (MD) simulations. Faster folders may be used to overcome this limitation. In recent years, many fast-folding proteins have been characterized to fill this need. With modern supercomputers, long simulation times with fast-folding protein may provide the first direct insight into the mechanism of protein folding. According to the current view, protein folding is envisioned to proceed along a moderately rough energy landscape,[1–4] the major features of which are the local minima and overall downhill slope toward the native state. An important notion of this theory is the funnel-like shape of the landscape on which the protein folds.

Optical spectroscopy is a powerful tool to probe the structural details and transformation of biological molecules. Experimental techniques, including NMR,[5] linear optical, Raman, and fluorescence spectroscopy;[6] X-ray scattering;[7] and Laue diffraction,[8] have been widely used to detect proteins structures. However, experimental methods often lack the time resolution for monitoring the whole folding dynamics and structural information at high resolution to observe ultrafast protein processes, and they result in indirect information about the structures along the folding pathway on the free energy surface of proteins. Recent simulation advances in multidimensional correlation spectroscopy,[9–11] which provides much higher time resolution, has emerged as a new probe to investigate the mechanism of protein folding. The technique employs sequences of laser pulses to probe the electronic or vibrational degrees of freedom and detects correlated

[a] Z. Lai, J. Wang
Department of Chemistry and Physics
University of New York at Stony Brook
Stony Brook, NY 11794 (USA)
Fax: (+1) 631-632-7960
e-mail: jin.wang.1@stonybrook.edu

[b] J. Jiang
School of Chemistry and Materials Science
University of Science and Technology of China
Hefei, Anhui 230026 (P. R. China)

[c] J. Jiang
Guizhou Provincial Key Laboratory of Computational Nano-Material Science
Guizhou Normal College, Guiyang, 550018 (P. R. China)

[d] S. Mukamel
Department of Chemistry, University of California
Irvine, CA 92697 (USA)

events during controlled time intervals. The resulting multipoint correlation functions contain detailed structural and dynamics information. Studies so far have given some success in the IR,[12–14] near-IR, and visible spectral regions, however, extension to the UV domain is in its infancy. New, intense, and stable femtosecond lasers with high repetition rates allow multidimensional spectroscopy to be extended into the UV range. Two-dimensional UV (2DUV), with short pulses and high-quality polarization control, provides a more refined resolution for the structural and dynamic properties of a protein. The 2DUV signals require the computation of the system Hamiltonian, including the environment at the quantum mechanics (QM) level, which is extremely expensive. Traditionally, some empirical methods, such as dipole approximation[26] and the map method,[10,27] have been applied to reduce computational cost, yet these methods require empirically fitted parameters. The EHEF (excitation Hamiltonian with electrostatic fluctuations) algorithm,[11] free from empirical parameters, provides an efficient approach to calculate the accurate system Hamiltonian involving the environment at the QM level. Herein, we report the first-principles simulation of 2DUV spectra of protein folding based on EHEF algorithms. Additionally, we applied the chirality-induced pulses technique to probe the structural changes of a protein. Signals with carefully designed chirality-induced polarization configurations have high resolution for describing protein structural evolution.

We studied two types of electronic transition induced by the localized chromophores present in the protein structures. The $n\pi^*$ and $\pi\pi^*$ electronic transitions of the peptide bonds of the proteins lead to spectra in the far-ultraviolet (FUV) region of 190–250 nm. The coupling of this type of transition is very sensitive to the formation or changes of the secondary structure of proteins.[15] Another type of the transition involves aromatic side chains, which lead to near-ultraviolet (NUV) spectra with wavelengths longer than 250 nm. In the three natural aromatic amino acids phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp), the NUV chromophores can be defined as benzene, phenol, and indole, respectively. Each of these chromophores has four valence electronic excitation states, ^1L_b, ^1L_a, ^1B_a, and ^1B_b, and transitions from the ground state to these excitation states induce near-ultraviolet (NUV) spectroscopy ($\lambda > 250$ nm).[11] Particularly, the transitions from the ground state to the two lower excitation states ^1L_b and ^1L_a provide the main contributions to the NUV signals.

The simulation of multidimensional correlation spectra coupled with MD calculations can provide additional insights into the protein folding processes and underlying energy landscape. Herein, the 2DUV spectra and MD simulations were combined to study the folding behavior of Beta3s, which is a de novo mini-protein. Beta3s includes 20 amino acids (Thr1-Trp2-Ile3-Gln4-Asn5-Gly6-Ser7-Thr8-Lys9-Trp10-Tyr11-Gln12-Asn13-Gly14-Ser15-Thr16-Lys17-Ile18-Tyr19-Thr20), and its solution conformation has been studied by NMR spectroscopy.[16] Data of nuclear Overhauser enhancement (NOE) and chemical shifts show that these 20 amino acids can fold into a single structured compact form, the three-stranded antiparallel β-sheet conformation with turns at Gly14-Ser15 and Gly6-Ser7, in equilibrium solution.[16] Also, the protein folds cooperatively and spontaneously into a Beta3s in the range of 4–14 μs, which is by far one of the fastest folding proteins known. Another interesting feature is that this protein contains four aromatic amino acids, which are Trp2, Trp10, Tyr11, and Tyr19. The transitions of these aromatic residues in the NUV regime reveal more detailed information on the folding process.

2 Methods

2.1 MD Simulations

All MD simulations and part of the analysis of the trajectories were performed with the CHARMM c35b5 package by using a parameter set TOPH19/PARAM19.[17] Herein, the folding simulations were performed by MD simulation with an implicit model for the solvent. This approximation is justified by recent explicit solvent MD simulations in which the folding behavior of Betanova, which is a synthetic three-stranded antiparallel β-sheet mini-protein with Asn-Gly as turn residues, has been studied, and it has been shown that the solvent does not influence significantly the results of the folding process with explicit solvents.[18] The covalent bond length, including hydrogen atoms were fixed with the SHAKE algorithm, and the nonbonded interaction list was extended to 7.5 Å. A constant temperature, 330 K, was maintained by assigning atom velocities from a Gaussian distribution for the different trajectories. From the native structure, 20 trajectories were simulated, and the simulation time for each trajectory was set to 500 ns. The total 10 μs folding simulation time provided efficient data to build up the free energy landscape upon which we can choose typical structures on the pathway of protein folding. We used the root-mean-square deviation (RMSD) and the fraction of native contacts of C_α atoms (Q) as the reaction coordinates to build up the free energy landscape. The native contacts were derived from CSU software,[19] and a contact was defined as "formed" when its pair distance fell within a range of $r_{native} \pm 2$ Å, in which r_{native} was the distance of that contact in the native structure between two atoms during the folding process. Along the folding dominant pathway from the unfolded state to the folded state, we chose 100 points on the free energy landscape. The distance between any two adjacent points was moderate, so that these points could cover the whole pathway. Around each point on the free energy landscape, there were many subconformation fluctuations. Therefore, we used the cluster method to choose 1 typical structure and 100 sub-

conformations around the typical structure for each point. After choosing the structures, we could calculate the exciton Hamiltonian and obtain the multidimensional UV spectroscopy signals.

2.2 Exciton Hamiltonian

The electronic transition of chromophores, including the peptide unit, benzene, phenol, and indole, can be modeled by the Frenkel exciton Hamiltonian[9,20,21] with the Heitler–London approximation [Eq. (1)]:

$$\hat{H} = \sum_{ma} \varepsilon_{ma} \hat{B}^{+}_{ma} \hat{B}_{ma} + \sum_{ma,nb}^{m \neq n} J_{ma,nb} B^{I}_{ma} \hat{B}_{nb} \tag{1}$$

in which $m(n)$ means $m(n)$th chromophore; a and b represent the electronic transition of the chromophores, including the nπ* transition, the ππ* transition, and the four electronic transitions of aromatic groups, as mentioned above; ε_{ma} is the excited-state energy; $\hat{B}^{+}_{ma}$is the creation Pauli operator; $\hat{B}_{ma}$is the corresponding annihilation operator; and $J_{ma,nb}$ is the resonant coupling between the a electronic transition of the mth chromophore and the b electronic transition of the nth chromophore.

The excited-state energy, $\varepsilon_{\mathrm{me}}$, can be calculated from Equation (2):

$$\varepsilon_{ma} = \varepsilon^{(0)}_{ma,Iso} + \sum_{k}^{k \neq m} \frac{1}{4\pi\varepsilon\varepsilon_0} \iint dr_m dr_k \frac{[\rho^{ee}_{ma}(r_k) - \rho^{gg}_{m}(r_m)]\rho^{gg}_{k}(r_k)}{|r_m - r_k|} \tag{2}$$

in which the first term, $\varepsilon^{(0)}_{ma,Iso}$, is the excited-state energy of the isolated chromophore, while the second term represents inter- and intramolecular electrostatic correlations; k is the index number of the atomic or molecular sites around the excited chromophore; $\rho^{ee}_{ma}(r_m)$ and $\rho^{gg}_{m}(r_m)$ represent the charge density of the a excited and ground states of the isolated chromophore; and $\rho^{gg}_{k}(r_k)$ denotes the charge density of the ground state of the surrounding atoms. Herein, we set the dielectric constant, ε, to one. The resonant coupling of any two different electronic transitions is given by Equation (3):

$$J_{ma,nb} = \frac{1}{4\pi\varepsilon\varepsilon_0} \iint dr_m dr_k \frac{\rho^{eg}_{ma}(r_m)\rho^{ge}_{nb}(r_n)}{|r_m - r_n|} \tag{3}$$

in which $\rho^{eg}_{ma}(r_m)$ and $\rho^{ge}_{nb}(r_n)$ are the transition charge densities.

The first term, $\varepsilon^{(0)}_{ma,Iso}$, in Equation (2), the excited-state energy of the isolated chromophore, can be calculated by the complete-active space self-consistent field method with a self-consistent reaction field (CASSCF/SCRF) and multi-configurational second-order perturbation theory (CASPT2), which can be performed by the MOLCAS program package.[22] The correlated interactions between the chromophore and local electronic environment, described as the second term in Equation (2), and the coupling between the electronic transitions of chromophores in Equation (3) can be calculated by the EHEF algorithm[11] at a high ab initio level. After fixing all of the parameters, the global electronic excitation of the protein is computed by diagonalizing the exciton Hamiltonian, and thus, we can perform the calculation of a two-dimensional photo echo signal.

2.3 2D Signal

The detailed calculation protocol of the 2D signals is given in ref. [9]. Briefly, the 2DUV signal is generated by three impulsive coherent short laser pulses with wave vectors $\mathbf{k_1}$, $\mathbf{k_2}$, and $\mathbf{k_3}$. The laser pulses have the same frequency and are resonant with the electronic transitions of the chromophores in the protein. One can detect the coherent signal field along the phase-matching direction: $\mathbf{k_4} = -\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}$ after the pulses interact with the protein system. The signal $\mathbf{k_4}$ is recorded as a function of three delay times: $\mathbf{t_1}$ between pulses $\mathbf{k_1}$ and $\mathbf{k_2}$; $\mathbf{t_2}$ between pulses $\mathbf{k_2}$ and $\mathbf{k_3}$; and $\mathbf{t_3}$ between pulses $\mathbf{k_3}$ and $\mathbf{k_4}$. We set $\mathbf{t_2}$ to zero. The 2D signals were obtained by two-dimensional Fourier transform: $\mathbf{t_1} \rightarrow \Omega_1$ and $\mathbf{t_3} \rightarrow \mathbf{\Omega_3}$. Calculations were performed for the nonchiral (*xxxx*) and chirality-induced (*xxxy*) polarization configurations; *xxxx*(*xxxy*) means the $\mathbf{k_1}$ laser pulse with polarizations along the $x(y)$ axis, while the other three are all x polarized. For the FUV signals, four Gaussian laser pulses are centered at 52000 cm^{-1} (192 nm) with a full-width at half-maximum (FWHM) bandwidth of 3754 cm^{-1} (14 nm). Also, for the NUV signals, lasers were centered at 37000 cm^{-1} (270 nm) with a FWHM bandwidth of 3000 cm^{-1} (22 nm). All calculated 2DUV spectra were normalized and were plotted from −1 to 1 to enhance weak signal features.

3. Results and Discussion

3.1 Free Energy Landscape

Figure 1 shows the simulated free energy landscape of Beta3s. The free energy was plotted as a function of the RMSD and Q and obtained by $F = -\log(P)$, in which P was the statistic population obtained from all 10 μs MD simulated data. Along the folding pathway, we chose five points, the 1st (P1), 25th (P25), 50th (P50), 75th (P75), and 100th point (P100), to represent the spectra of the FUV and NUV. P1 represents the unfolded state of Beta3s, P100 means the folded state of Beta3s, and P25, P50, and P100 are the medial states during the folding process, which occurs from P1 to P100 on the underlying folding free energy landscape. Figure 1 also shows the typical structures of these five points. P1 has an extended

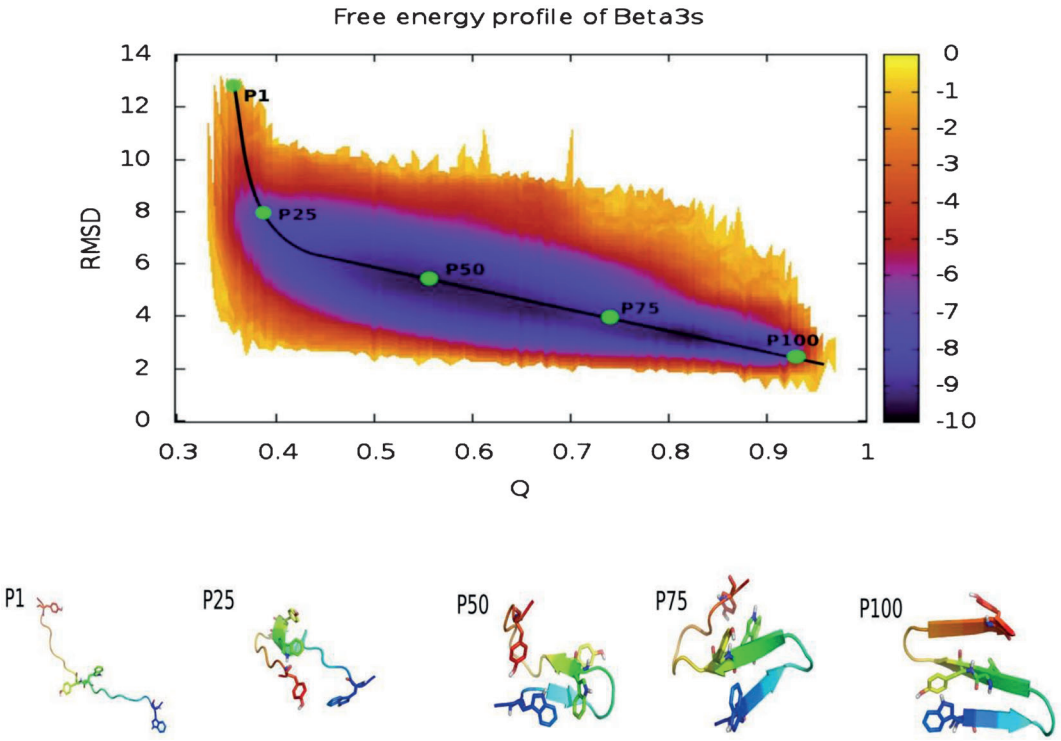


Figure 1. Free energy profile of Beta3s folding. Five typical structures on the folding pathway were marked as P1, P25, P50, P75, and P100, and the corresponding structures are shown in the bottom of the figure.

strand structure and P25 has a random coil structure, yet residues 8–13 form a very short α helix. When the protein arrives at P50, two strands with some β-sheet structure begin to form and P75 includes two β-sheets in the structure with another strand of random coil. P100 is the folded state, which consists of three antiparallel β sheets.

3.2 NUV Spectra

The NUV signals are induced by electronic transitions of the side chains in the aromatic amino acids, including Phe, Tyr, and Trp. For each aromatic amino acid, there are four transitions: ^1L_b, ^1L_a, ^1B_a, and ^1B_b. Generally, the transitions of Trp are stronger than those of Tyr and then Phe, and the ^1L_a transition of Trp is stronger than that of ^1L_b. In Beta3s, there are four aromatic amino acids: Trp2, Trp10, Tyr11, and Tyr19. In the folded structure of Beta3s, Trp10 and Tyr11 are in the middle strand, while Trp2 and Tyr19 are in the other two strands. The NUV linear absorption (LA) spectra are displayed in Figure 2a–e. They all have a dominant peak at around 38400 cm^{-1} (260 nm) and some shoulders located in the region of 35700–37000 cm^{-1}. The dominant peak is induced by the ^1L_a transition of Trp2 and Trp10, while the shoulders are caused by the ^1L_b transitions of Tyr and Trp. In P100, the shoulders are less pronounced, which implies that the coupling of the aromatic amino acids tends to cancel out the signals induced by the ^1L_b transitions in the folded structure. The nonchiral (*xxxx*) signals in Figure 2g–j are also dominated by the ^1L_a transitions. They all show a negative diagonal peak with the center at 38400 cm^{-1} and two positive side bands. The diagonal peaks are broad along the diagonal direction, and narrow in the antidiagonal direction.

The circular dichroism (CD) spectra and chirality-induced (CI) 2DNUV signals (*xxxy*) are much more sensitive to structural variation. Figure 3a has two peaks: a negative peak at 38000 cm^{-1} (263 nm) originating from the ^1L_a transitions of Trp2 and Trp10, and a broad positive peak at around 35700 cm^{-1} (280 nm) induced by the ^1L_b transitions of Trp and Tyr. The refined structure of this positive peak shows that the signal induced by the ^1L_b transition of Trp (35700 cm^{-1}) is slightly stronger than that induced by the ^1L_b transition of Tyr (36500 cm^{-1}). Figure 3f has one negative peak, the center of which is at 38000 cm^{-1}, and a positive peak does not appear in the diagonal line, since the intensity of the ^1L_a transition is much stronger than that of the ^1L_b transition. The side bands located near the diagonal peaks are weak, which implies that the coupling between the aromatic amino acids is not strong due to the extended strand structure in P1. Although Trp10 and Tyr11 are adjacent in the protein sequence of Beta3s, the weak side bands in Figure 3f imply that they do not have strong interactions. Their coupling may be reduced by the opposite orientations of the side chains. As the protein evolves from P1 to P25 on

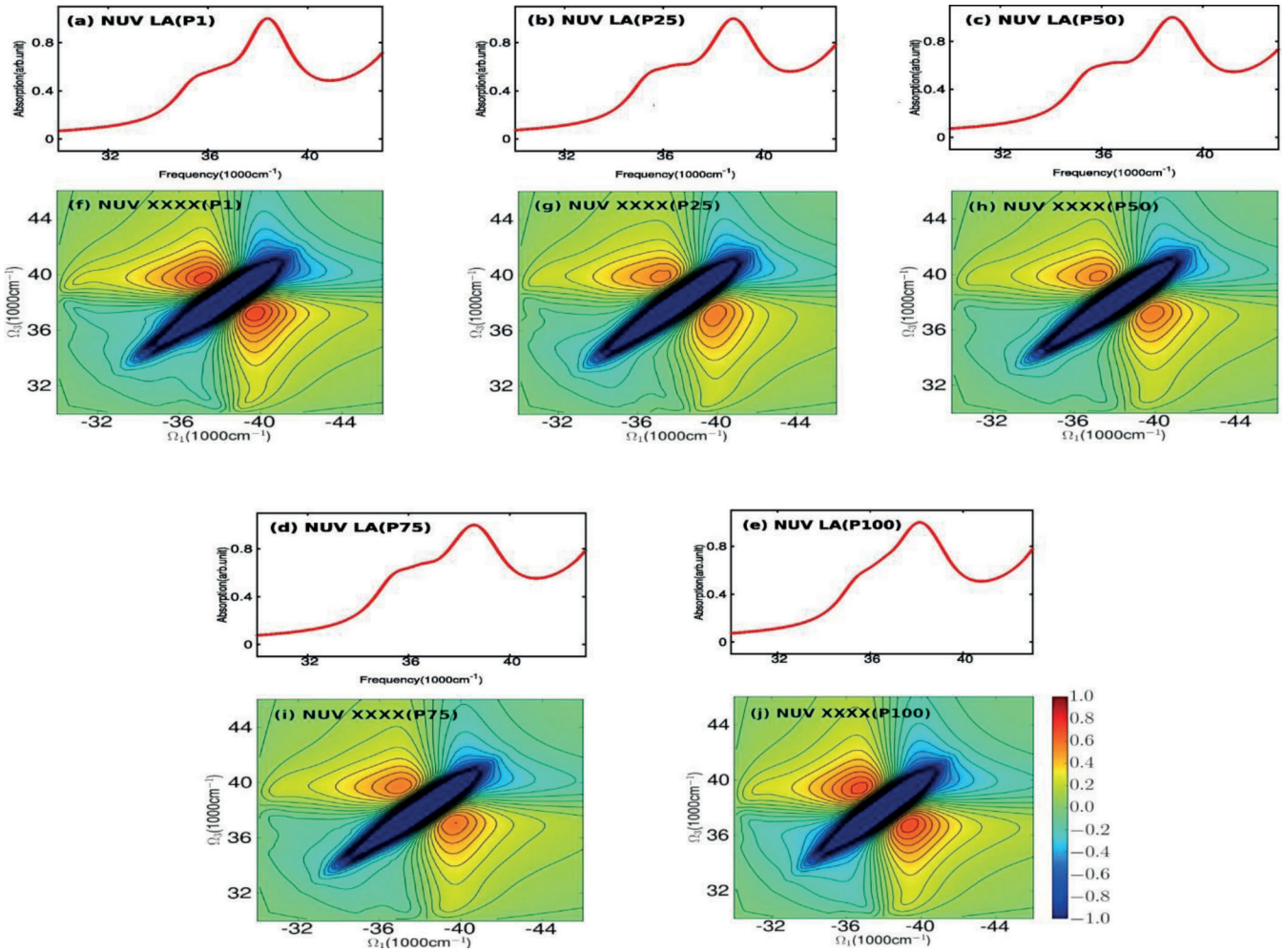


Figure 2. NUV LA and nonchiral 2DNUV spectra. LA spectra of P1 (a), P25 (b), P50 (c), P75 (d), and P100 (e). Nonchiral (*xxxx*) 2DNUV spectra of P1 (f), P25 (g), P50 (h), P75 (i), and P100 (j).

the free energy landscape, the CI 2DNUV spectrum of P25 displays two peaks, the negative peak at 36000 cm^{-1} and the positive peak at 39000 cm^{-1}, as shown in Figure 3g, and the corresponding CD spectrum (Figure 3b) shows the negative peak at around 36000 cm^{-1} and a positive peak at around 39000 cm^{-1}, as shown in Figure 3b. Also, the intensity of the CD spectrum at 260–265 nm is very weak. In P25, the couplings among the chromophores are increased due to packing of the strand. However, they are weak because the protein is still in the loosened structure. For the CI 2DNUV spectrum, these weak couplings cause the cancellation of the dominant peak at 38000 cm^{-1} and split it into two opposite peaks through the μ–μ coupling mechanism.[23] It is hard to determine the interactions precisely due to the random coil structures.

When the protein folds to P50, the CI 2DNUV signals show one dominant peak on the diagonal line again. The center of the peak around 38000 cm^{-1} originates mainly from the ^1L_a transition of Trp. This change may be caused by the structural properties of P50, in which, although two strands have not developed the complete antiparallel β sheets, the preliminary β-sheet structures begin to appear. When the protein folds from the random coil to the more ordered structure, the side chain of the residue Trp is exposed stably to the environment again and induces the main signals of the diagonal peak. However, due to the compact nature of the structure, coupling between the electronic transitions suppresses the length of the peak on the diagonal line. Also, the main peak blueshifts from 38100 (262 nm) to 39200 cm^{-1} (255 nm), as shown in Figure 3c. Another interesting feature of the CI 2DNUV signals is the side bands. The off-diagonal band at $\Omega_1 =$ 38500 cm^{-1}, $\Omega_3 = 36500$ cm^{-1} implies coupling between the ^1L_a transition of Trp and the ^1L_b transition of Tyr. Due to the orientations of their side chains, the coupling between Trp10 and Tyr11 is very weak, as shown in Figure 3f. In addition, the distance between Trp2 and Tyr19 is large. Therefore, the coupling between Trp2 and Tyr11 or Trp10 and Tyr19 dominates the off-diagonal bands.

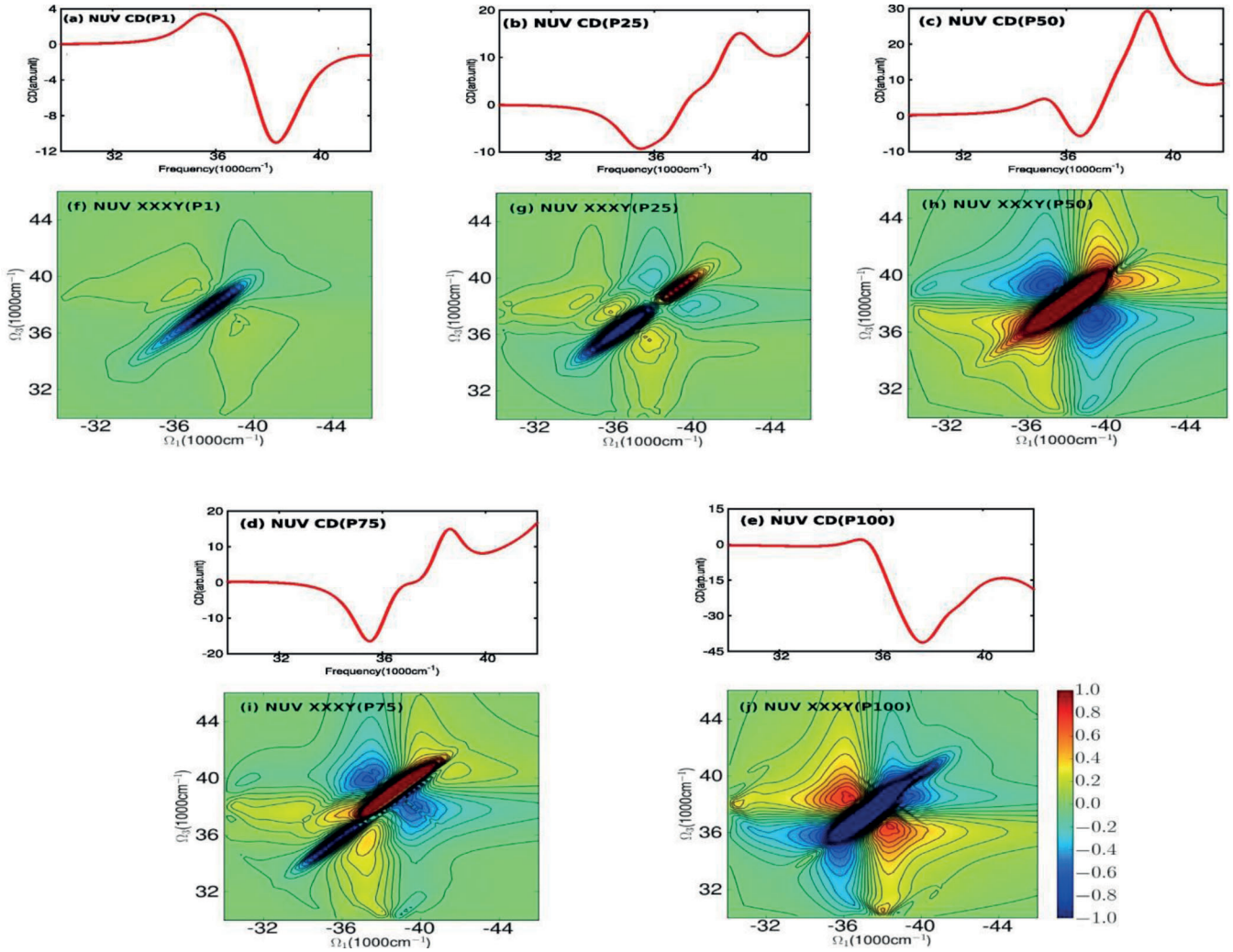


Figure 3. NUV CD and CI 2DNUV spectra. NUV CD spectra of P1 (a), P25 (b), P50 (c), P75 (d), and P100 (e). CI (*xxxy*) 2DNUV signals of P1 (f), P25 (g), P50 (h), P75 (i), and P100 (j) along the dominant folding pathway.

Since the Trp and Tyr residues that contribute to the off-diagonal band at $\Omega_1 = 38500\ cm^{-1}$, $\Omega_3 = 36500\ cm^{-1}$ on the different sheets in the protein, the stable off-diagonal signal may illustrate the formation of the two sheets during the protein folding of Beta3s. This is a promising property for an experimental design of 2DUV, which is not available from one-dimensional spectra.

When the protein evolves from P50 to P75, the initial two β sheets develop to complete β sheets, while the third strand retains the random coil structure. The diagonal peak of the CI 2DNUV signal separates into two peaks. One peak is still at around $38000\ cm^{-1}$ because of the ^1L_a transition of Trp, and the other peak is located at around $35500\ cm^{-1}$ from the ^1L_b transition of Trp. From the CD spectrum of P50, we find that there is a weak peak at around $35500\ cm^{-1}$ (281 nm), yet, compared with the peak at around $38700\ cm^{-1}$ (256 nm), this peak is so weak that it is hidden in the CI 2DNUV spectrum of P50. As the β strands grow in P75, the protein becomes more organized and the orientations of the aromatic side chains become more rigid and identical. Increased intensity of the ^1L_b transition can be observed when other aromatic rings juxtapose with the Trp side chain[23] in P75. In the lower triangle of the CI 2DNUV spectrum of P75, the off-diagonal band ($\Omega_1 = 36500\ cm^{-1}$, $\Omega_3 = 35500\ cm^{-1}$) is induced by coupling between the ^1L_b transition of Trp and the ^1L_b transition of Tyr. Also, the other off-diagonal band at around $\Omega_1 = 38500\ cm^{-1}$, $\Omega_3 = 37500\ cm^{-1}$ may indicate the interaction between the ^1L_a transition of Trp and the ^1L_b transition of Tyr.

The CI 2DNUV spectrum of the folded state (P100) has a butterfly pattern. The ^1L_a transition of Trp induces the diagonal negative signal at $38000\ cm^{-1}$, which is similar to the diagonal peak of P1. Since the P1 structure is an extended strand and P100 has three organized antiparallel β sheets, the dominant diagonal peaks in the CI 2DNUV spectra of P1 and P100 indicate that the ^1L_a transition of Trp is very intensive in the ordered structures.

Moreover, the broad diagonal negative band from 35000 to 39000 cm^{-1} suggests that the ^1L_b transition of Tyr and Trp may also contribute mainly to the diagonal peak signals. Additionally, the off-diagonal band around $\Omega_1 = 38500\ cm^{-1}$, $\Omega_3 = 36500\ cm^{-1}$ is intensive in P100, implying strong coupling between the ^1L_a transition of Trp and ^1L_b transition of Tyr, which can also be observed in P50 and P75. The stable signal from this coupling carries information about the formation of the β sheets. In addition, the cross-peak at around $\Omega_1 = 35300\ cm^{-1}$, $\Omega_3 = 36300\ cm^{-1}$ that first appears in P100 is excited by coupling between the ^1L_b transition of Trp and Tyr; this indicates the strong interaction between these two residues due to the ordered arrangement of the aromatic side chain in the folded state of Beta3s. The CD spectrum of P100 shows a strong peak near 37700 cm^{-1} (265 nm), induced by the ^1L_a transition of Trp, and a weak peak near 35000 cm^{-1} (285 nm), induced by the ^1L_b transition of Trp; this suggests that the peaks at around 35000 and 37700 cm^{-1} can be good measurements for ordered structures. Therefore, the 2D spectra, especially the CI spectra, provide a good tool for investigating the protein folding process.

3.3 FUV Spectra

The FUV spectra originate from the peptide bond, which primarily absorbs UV light at around 200 nm.[11] The LA spectra are shown in Figure 4a–e. They are very similar for P1 to P100 and all have a single band at around 195 nm (51300 cm^{-1}). LA signals can be mapped into the diagonal line ($\Omega_3 = -\Omega_1$) of the nonchiral 2DUV spectrum (*xxxx*), as shown in Figure 4f–j. From P1 to P100, all nonchiral 2DUV spectra have strong diagonal peaks and relatively weak cross-peaks distributed symmetrically around the diagonal peaks. Comparing nonchiral 2DUV spectra of P1 and P100, we find that the diagonal negative peak of the β sheet is slightly broader than that of the strand. Also, the lower cross-peak of the folded state is stronger than that of the unfolded state. Along the protein folding path, the conformation of Beta3s changes from the extended strand to a sheet structure. The LA and 2D nonchiral spectra indicate that the signals are not sensitive to protein secondary structure motifs. Also, the diagonal peaks are stronger than the cross-peaks, indicating that the electrostatic interaction between the different

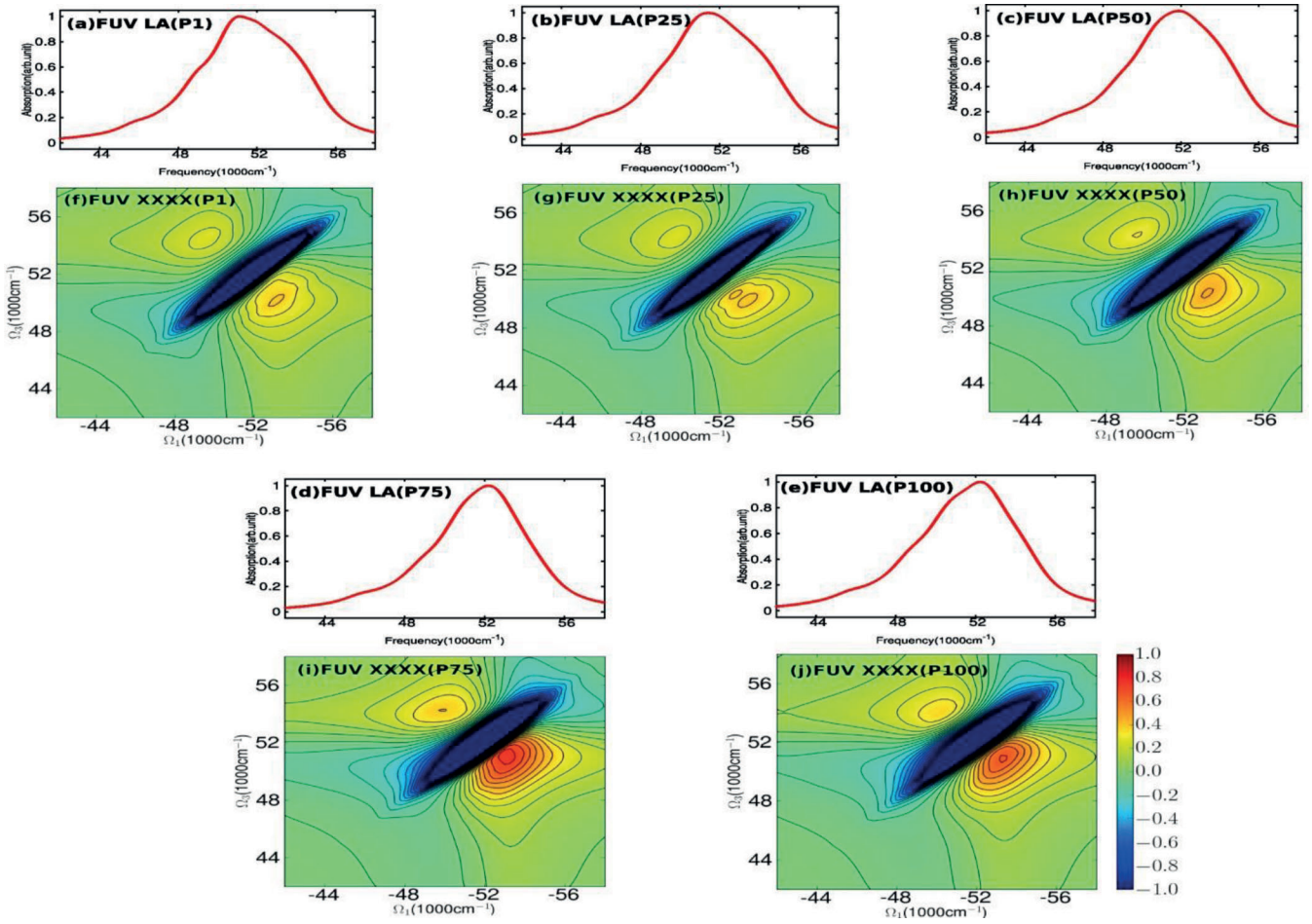


Figure 4. FUV LA and nonchiral 2DFUV spectra. FUV LA spectra of points P1 (a), P25 (b), P50 (c), P75 (d), and P100 (e). Nonchiral (*xxxx*) 2DFUV signals of P1 (f), P25 (g), P50 (h), P75 (i), and P100 (j).

amides is weaker than the amide transition energy. The *xxxx* spectra of P75 and P100, however, show that when beta3s folds to the more ordered structure, the intensity of the cross-peak increases, which implies that the interaction between different amides is stronger.

The CD spectra, shown in Figure 5a–e, are sensitive to the structural changes of protein folding. In Figure 5a, there is a negative band at 180 nm (55500 cm^{-1}) and a positive band at 196 nm (51000 cm^{-1}), reflecting the structure of β strands in the protein.[24,25] These two bands also appear in the CD spectra of P25 to P100, implying the β structure in the whole folding process, yet the amplitudes of these two bands decrease when the protein folds into the final structure. It is well known that the folded structure of Beta3s includes two type II β turns[16] and three antiparallel β sheets. The formation of the type II β-turns can be reflected in the blue side (190 nm) of the CD spectra. The structures of P1, P25, and P50 lack type II β-turns, and there is no peak at around 190 nm (52600 cm^{-1}). When the protein evolves into P75, the blue side of Figure 5d shows that a negative shoulder at 190 nm (52600 cm^{-1}) begins to emerge and finally develops into the negative peak in Figure 5e, indicating the formation the type II β turn in the folding process.[28] The formation of the β sheets is also reflected in the CD spectra. From P50 to P100, the positive band near 46500 cm^{-1} (215 nm) gradually changes into the negative band, implying a decrease in the random coil and increase in the β-sheet structure. This becomes significant in P100, which agrees well with experimental results.[25,29]

The CI 2DUV spectra, as shown in Figure 5f–j, carry rich information about the structural evolution of protein folding. The diagonal section of the strand spectrum, as shown in Figure 5f, has a positive peak, 1a, at 50500 cm^{-1} and a negative peak, 1b, at 54000 cm^{-1}. Strong cross-peaks labeled 1c, 1d, 1e, and 1f show certain asymmetri-

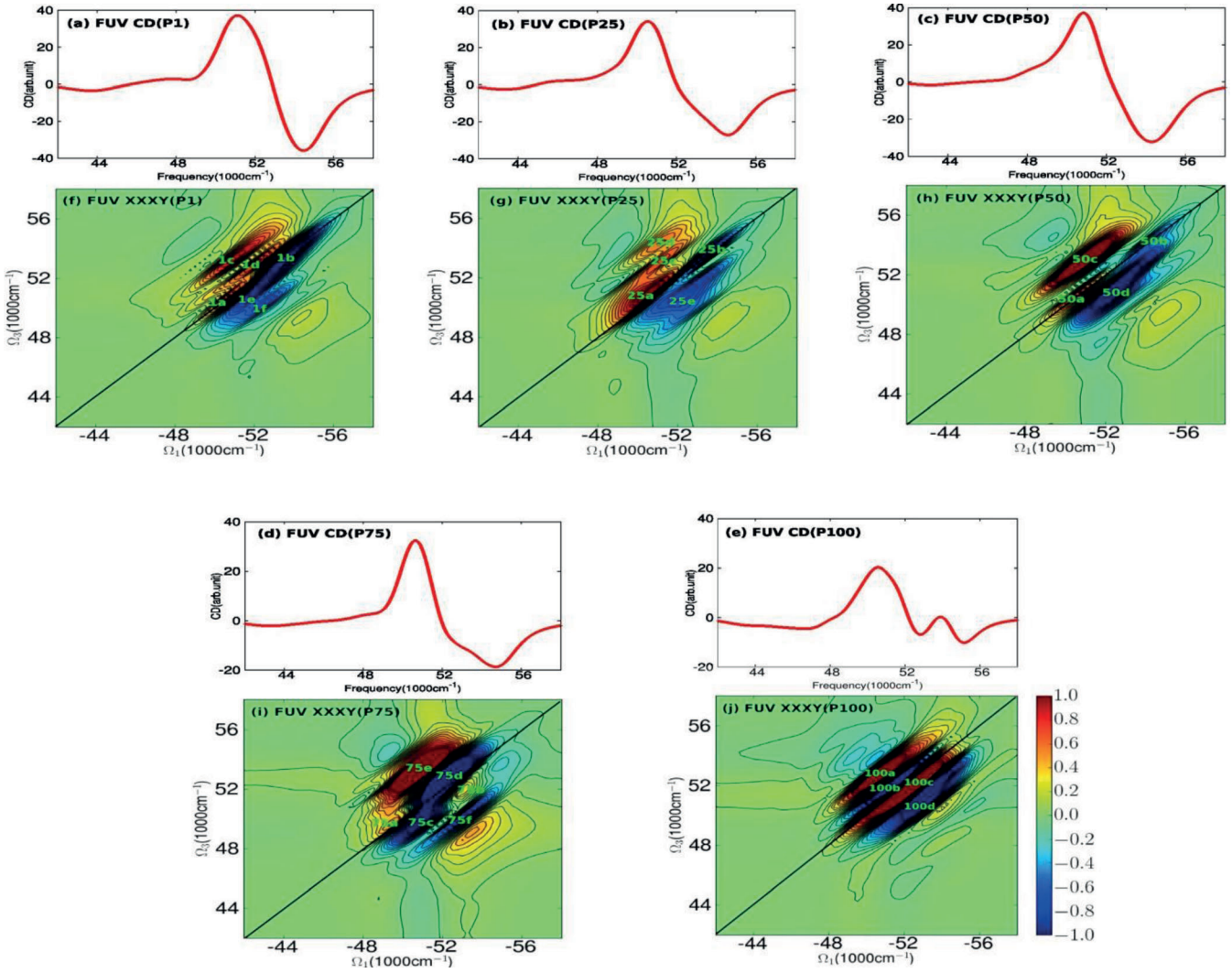


Figure 5. FUV CD and CI 2DFUV spectra. FUV CD spectra of P1 (a), P25 (b), P50 (c), P75 (d), and P100 (e). CI (*xxxy*) 2DFUV signals of P1 (f), P25 (g), P50 (h), P75 (i), and P100 (j).

cal distribution patterns along the diagonal peaks, although peaks 1e and 1f connect on one side and peaks 1c and 1d separate on the other side, indicating that the structure of P1 is mainly of strand structure (symmetry) and also mixes with some random coil structure (no symmetry). The diagonal positive peak 1a at 50500 cm^{-1} is mainly induced by the separate $\pi\pi^*$ of the whole peptide backbone, while the cross-peaks 1c, 1d, 1e, and 1f indicate coupling between $\pi\pi^*$ and ^1B_a, $\pi\pi^*$, and ^1B_b. This pattern of the signal from the strand agrees well with our previous study.[11] When the protein reaches P25, the negative diagonal peak 1b of the strand at 54000 cm^{-1} shifts a little lower to peak 25b, and a strong positive peak, 25a, develops at 50500 cm^{-1}. The asymmetric distribution at P1 disappears at P25 due to the increase of the random coil. Peaks 1e and 1f in Figure 5f merge into peak 25e in Figure 5g. Peaks 25a and 25e become broader in the antidiagonal direction due to the appearance of the turn structure and compactness in the random coil.[11] As the protein evolves into P50, peaks 25d and 25c in P25 merge into one strong side band, 50c, the center of which is $\Omega_1 =$ 51000 cm^{-1}, $\Omega_3 = 53000$ cm^{-1}, indicating strong interaction between the $\pi\pi^*$ transition of the peptide bond and the ^1B_a, ^1B_b transition of the aromatic side chains. The diagonal peak 50a becomes weak and separates from the side peak 50c, and the lower triangle peak 50d becomes stronger and more compact than peak 25e; this indicates that more ordered structures tend to cancel out the signals of the diagonal peaks and strength of the signals of the side bands. This feature becomes clearer in the CI 2DUV signal of P75, in which two β sheets already form. The two diagonal peaks of P75 are so weak that we can only find the four side peaks 75a, 75b, 75c, and 75d beside the diagonal line. From P50 to P75, another very interesting feature of the CI 2DUV signal is that the pattern of the signal in P75 becomes more symmetrical than that in P50, especially peaks 75a, 75b, 75c, and 75d are symmetrically located beside the diagonal line. These two features are very clear in the CI 2DUV signal of the folded state, as shown in Figure 5j, in which there are four strong, compact side peaks, 100a, 100b, 100c, and 100d, located symmetrically beside the diagonal line and there are no diagonal peaks. Interestingly, the positive and negative signs of the side peaks also show a symmetrical distribution beside the diagonal line. These symmetrical properties indicate the formation of the three antiparallel β sheets. The CI 2DUV spectra show many advantages as a powerful tool to detect the protein folding process.

4 Conclusions

Lately, multidimensional time-resolved ultrafast spectroscopy has provided a powerful tool in investigating protein folding. Herein we presented a study of the protein folding of Beta3s by combining multidimensional UV spectroscopy. We applied a new technique of computing the system Hamiltonian at the QM level and CI pulses to calculate the 2DUV spectra of the structures, which were chose from the pathway of the underlying free energy landscape. Atomistic MD simulations using the CHARMM force field were applied to sample the conformations and build up the underlying free energy surface. Coherent 1D spectra, including LA and CD, and 2D spectra, including chiral and nonchiral, in the NUV and FUV region were simulated to investigate the folding conformational evolution on the underlying free energy landscape. In both the NUV and FUV regions, the patterns of LA and 2D nonchiral spectra were similar for the structural evolution of folding, implying that these signals were insensitive to structural changes. In contrast, the CD and 2D chiral signals in the NUV and FUV regions were very sensitive to structural changes, reflecting the folding process on the pathway. When the protein first collapsed from the random coil state to an intermediate state, the appearance of the stable off-diagonal peak of the CI 2DNUV spectra induced by the aromatic residues on the strands implied the formation of a native-like turn and orientation of two strands. Further folding progress yielded a well-formed β-sheet structure on the two strands with good interstrand contacts, which made the pattern of the CI 2DFUV spectra more symmetrical. Finally, the symmetrical pattern of the CI 2DFUV spectra indicated the formation of the whole antiparallel β sheets. By studying the variation of the spectra, we could uncover the local and global protein folding process. This is a promising property for an experimental design of 2DUV and hard to obtain by other spectroscopic methods. All of this shows that 2DUV spectroscopy is a powerful tool for investigating protein folding with many advantages. Coherent 2D experiments in the UV region are now possible, but require state of the art technology.[30,31] By directly connecting 2D spectroscopy experiments with molecular simulations, we can uncover the underlying mechanisms of folding and see how the folding process actually occurs in time by monitoring the folding pathways. Our work provides a general approach to connect 2DUV experiments with MD simulations by theoretical calculations.

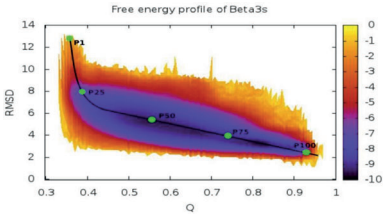
Acknowledgments

We thank the National Science Foundation for support. S. M. gratefully acknowledges the support of the National Institutes of Health through grant number GM-59230, and the National Science Foundation through Grant No. CHE-1058791. J. J. thanks the support of the National Science Foundation of China (91221104) and the Construction Project for Guizhou Provincial Key Laboratories (ZJ[2011]4007).

References

[1] Y. Levy, P. G. Wolynes, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 511–516.
[2] B. A. Shoemaker, J. J. Portman, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8868–8873.
[3] G. A. Papoian, P. G. Wolynes, *Biopolymers* **2003**, *68*, 333–349.
[4] J. Wang, G. M. Verkhivker, *Phys. Rev. Lett.* **2003**, *90*, 188101.
[5] J. Balbach, *J. Am. Chem. Soc.* **2000**, *122*, 5887.
[6] X. Chen, P. Li, J. Holtz, Z. Chi, V. Pajcini, A. Asher, *J. Am. Chem. Soc.* **1996**, *118*, 9705.
[7] S. Arai, M. Hirai, *Biophys. J.* **1999**, *76*, 2192–2197.
[8] T. Hori, H. Moriyama, J. Kawaguchi, Y. Hayashi-Iwasaki, T. Oshima, N. Tanaka, *Protein Eng.* **2000**, *13*, 527–533.
[9] D. Abramavicius, B. Palmieri, D. V. Voronine, F. Sanda, S. Mukamel, *Chem. Rev.* **2009**, *109*, 2305–2408.
[10] W. Zhuang, T. Hayashi, S. Mukamel, *Angew. Chem. Int. Ed.* **2009**, *48*, 3750–3781.
[11] J. Jiang, S. Mukamel, *Phys. Chem. Chem. Phys.* **2011**, *13*, 2394–2400.
[12] C. N. J. Marai, S. Mukamel, J. Wang, *PMC Biophys.* **2010**, *3*, 8.
[13] Z. Z. Lai, N. K. Preketes, S. Mukamel, J. Wang, *J. Phys. Chem. B* **2013**, *117*, 46614669.
[14] Z. Z. Lai, N. K. Preketes, J. Jiang, S. Mukamel, J. Wang, *J. Phys. Chem. Lett.* **2013**, *4*, 19131917.
[15] D. Abramavicius, J. Jiang, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Am. Chem. Soc.* **2010**, *132*, 7769–7775.
[16] E. De Alba, J. Santoro, M. Rico, M. A. Jimenez, *Protein Sci.* **1999**, *8*, 854–865.
[17] B. R. Brooks, R. E. Bruccileri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187–217.
[18] B. D. Bursulaya, C. L. Brooks, *J. Am. Chem. Soc.* **1999**, *121*, 9947–9951.
[19] V. Sobolev, R. C. Wade, G. Vriend, M. Edelman, *Proteins Struct. Funct. Genet.* **1996**, *25*, 120–129.
[20] Y. Frenkel, *Phys. Rev.* **1931**, *37*, 17–44.
[21] S. Schenkl, F. van Mourik, G. van der Zwan, S. Haacke, M. Chergui, *Science* **2005**, *309*, 917–920.
[22] G. Karlstrom, R. Lindh, P. Malmquist, B. Roos, U. Ryde, V. Veryazov, P. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady, L. Seijo, *Comput. Mater. Sci.* **2003**, *28*, 222–239.
[23] E. H. Strickland, *Crit. Rev. Biochem.* **1974**, *2*, 113–175.
[24] N. Greenfield, G. D. Fasman, *Biochemistry* **1969**, *8*, 4108–4116.
[25] E. K. Koepf, H. M. Petrass, M. Sudol, J. W. Kelly, *Protein Sci.* **1999**, *8*, 841–853.
[26] Y. Luo, P. Norman, H. Argen, *J. Chem. Phys.* **1998**, *109*, 3589–3595.
[27] Y. S. Lin, J. M. Shorb, P. Mukherjee, M. T. Zanni, J. T. Skinner, *J. Phys. Chem. B* **2009**, *113*, 592–602.
[28] S. Brahms, J. Brahms, G. Spach, A. Brack, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3208–3212.
[29] B. Ranjbar, P. Gill, *Chem. Biol. Drug Des.* **2009**, *74*, 101–120.
[30] B.-A. West, J.-M. Womick, A.-M. Moran, *J. Phys. Chem. A* **2011**, *115*, 8630–8637.
[31] C. Consani, G. Aubock, F. van Mourik, M. Chergui, *Science* **2013**, *339*, 1586–1589.

Received: December 20, 2014
Accepted: April 5, 2014
Published online: ■■ ▮▮, 0000



*Z. Lai, J. Jiang, S. Mukamel, J. Wang**

■■–■■

Exploring the Protein Folding Dynamics of Beta3s with Two-Dimensional Ultraviolet (2DUV) Spectroscopy