processes

MDPI

Article

Machine Learning-Based Two-Dimensional Ultraviolet Spectroscopy for Monitoring Protein Structures and Dynamics

Songnan Jiang, Jiale Jiang, Tong Yan, Huamei Yin, Lu Wang and Jinxiao Zhang *

Guangxi Colleges and Universities Key Laboratory of Surface and Interface Electrochemistry, College of Chemistry and Bioengineering, Guilin University of Technology, Guilin 541006, China
* Correspondence: jxzh@glut.edu.cn

Abstract: Two-dimensional ultraviolet (2DUV) spectroscopy is an emerging spectroscopic technique that offers high resolution and detailed insights into protein structures. However, traditional theoretical calculations of 2DUV spectra for proteins are computationally expensive due to their complex and flexible structures. In this study, we developed a machine learning (ML)-based approach for the rapid and accurate prediction of protein 2DUV spectra. The results demonstrate that, compared to traditional one-dimensional ultraviolet (1DUV) spectroscopy, 2DUV spectroscopy provides higher resolution structural characterization and effectively monitors dynamic processes such as mutations, aggregation, and protein folding. This approach not only offers a cost-effective ML-based solution for predicting 2DUV spectra but also serves as a powerful tool for studying protein structures and dynamics, with potential applications in understanding mechanisms and regulating functions.

Keywords: machine learning; two-dimensional ultraviolet spectroscopy; protein structure; dynamics

check for updates

Academic Editor: Tong Gao

Received: 13 December 2024
Revised: 15 January 2025
Accepted: 20 January 2025
Published: 21 January 2025

Citation: Jiang, S.; Jiang, J.; Yan, T.; Yin, H.; Wang, L.; Zhang, J. Machine Learning-Based Two-Dimensional Ultraviolet Spectroscopy for Monitoring Protein Structures and Dynamics. *Processes* **2025**, *13*, 290. https://doi.org/10.3390/pr13020290

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Proteins are the foundation of life, and their structures are closely related to complex biological functions, disease mechanisms, and drug development [1,2]. Therefore, probing protein structures and tracking their dynamic evolution is crucial for understanding underlying mechanisms and manipulating protein functions [3,4]. Spectroscopic techniques offer unique advantages in investigating protein structures [5,6], and traditional one-dimensional (1D) spectroscopy methods have been widely employed for protein structure determination. X-ray diffraction (XRD) and nuclear magnetic resonance (NMR) provide high-resolution three-dimensional structural information [7–10]. However, XRD requires crystalline samples and is insensitive to planar defects [11], while NMR often requires dissolved samples and is time consuming, costly, and complex [12]. Consequently, XRD and NMR are unsuitable for proteins such as amyloid fibrils that neither crystallize nor dissolve. Vibrational spectroscopy, including infrared and Raman spectroscopy, is effective for probing protein structures based on localized chemical groups, but it fails to detect non-active vibrational modes and lacks the ability to provide global structural information [13]. One-dimensional far-ultraviolet absorption (1DUV) spectroscopy, derived from electronic excitations of peptide backbones under environmental fluctuations, can provide global structural information and monitor dynamic processes [14,15]. However, 1D spectroscopy often suffers from signal congestion, limiting its ability to detect subtle structural changes and monitor dynamic evolution [16,17].

In contrast, two-dimensional (2D) spectroscopy introduces a second spectral dimension to resolve overlapping bands, allowing the characterization of interactions or correlations and tracing subtle intensity changes. Therefore, 2D spectroscopy provides higher resolution and more comprehensive information for studying proteins in biological systems. It has been demonstrated that 2D infrared and Raman spectroscopy have been successfully applied to capture structural changes in proteins [18–20]. Given its success in vibrational spectroscopy, extending 2D spectroscopy to electronic spectroscopy in the ultraviolet (UV) region is both reasonable and promising [20–22]. Two-dimensional ultraviolet absorption (2DUV) spectroscopy provides rich intra- and intermolecular interaction information and is well suited for analyzing global structural features due to its involvement of electronic excitations and coupling under environmental fluctuations [23,24]. Thus, 2DUV spectroscopy holds great potential as a powerful platform for probing subtle protein structures and dynamic processes [25–29].

Despite the promising potential of 2DUV spectroscopy in biological systems, its simulation requires large-scale quantum chemical calculations to model electronic structures and excited states of highly flexible structures and vast conformational spaces, exceeding the capabilities of current computational resources [30,31]. With the rapid advancement of artificial intelligence, machine learning (ML) has been widely applied in chemistry, transitioning from theoretical studies to practical applications in areas such as spectral simulation [32,33], drug development [34,35], catalysis [36,37], and energy prediction [38]. In our previous studies, ML was successfully employed to predict various 1D spectroscopies, including UV, circular dichroism, and infrared spectra [39–41]. These studies demonstrated that ML techniques can generate cost-effective and accurate spectral predictions by capturing complex correlations between protein structures and spectral features, thereby advancing research into protein structure and function.

In this study, we developed an ML protocol to predict the 2DUV spectra of proteins and investigated its applications in probing protein structures and monitoring dynamics. The results demonstrate that, compared to traditional 1DUV spectroscopy, 2DUV spectroscopy provides richer and more precise structural information. It can be used to identify secondary and quaternary structures and effectively track dynamic processes such as mutations, aggregation, and folding. Our study not only establishes a rapid ML-based framework for predicting 2DUV spectra but also introduces a novel approach for identifying subtle protein structures and monitoring their dynamic evolution. Therefore, 2DUV spectroscopy holds significant potential for unraveling underlying mechanisms and manipulating protein functions.

2. Materials and Methods

A divide-and-conquer strategy combined with ML methods was employed to predict the 2DUV spectra of proteins. Proteins are composed of peptide backbones and side-chain amino acid residues, with the 2DUV spectra primarily deriving from two electronic excitations of the peptide backbones: the $n \rightarrow \pi^*$ transition at approximately 220 nm and the $\pi \rightarrow \pi^*$ transition at approximately 190 nm [42]. These electronic excitations are influenced not only by the intrinsic properties of isolated peptide bonds but also by the surrounding environment fluctuations (amino acid residues). To describe the electronic excitations in proteins, the Frenkel exciton model combined with a dipole approximation was adopted [43–46]. This model represents the protein's electronic structure using a Hamiltonian that captures both the excitation energy of peptide bonds and their interactions:

$$\hat{H} = \sum_{ma} \varepsilon_{ma} \hat{B}_{ma}^{\dagger} \hat{B}_{ma} + \sum_{ma,nb}^{m \neq n} J_{ma,nb} \hat{B}_{ma}^{\dagger} \hat{B}_{nb} \quad (1)$$

Here, m and n represent peptide bonds, while a and b correspond to the $n \rightarrow \pi^*$ or $\pi \rightarrow \pi^*$ transitions of the peptide backbones. The Hamiltonian matrix consists of diagonal and off-diagonal elements, where the diagonal term (ε_{ma}) represents the excitation energy of peptide bond m for transition a under environmental fluctuations, while the off-diagonal term ($J_{ma,nb}$) describes the resonant coupling between the excited states of peptide bonds m and n. $\widehat{B}_{ma}^{\dagger}$ and $\widehat{B}_{ma}$ are the creation and annihilation operators for electronic transitions between the ground and excited states of the peptide bonds, respectively.

Under this model, the diagonal term ε_{ma} captures the excitation energy of peptide bonds influenced by their molecular environment, while the off-diagonal term $J_{ma,nb}$ describes the coupling between excited states, which governs energy transfer across the protein structure. The excitation energy ε_{ma} can be expressed as the sum of the intrinsic excitation energy of an isolated peptide $\varepsilon_{0,ma}$ and contributions from the electrostatic interactions with surrounding amino acid residues:

$$\varepsilon_{ma} = \varepsilon_{0,\,ma} + \sum_{l} \frac{1}{4\pi\varepsilon\varepsilon_0} \left(\frac{\boldsymbol{\mu}_{T,\,ma} \cdot \boldsymbol{\mu}_{G,\,l}}{|\boldsymbol{r}_{ml}|^3} - 3 \frac{(\boldsymbol{\mu}_{T,\,ma} \cdot \boldsymbol{r}_{ml})(\boldsymbol{\mu}_{G,\,l} \cdot \boldsymbol{r}_{ml})}{|\boldsymbol{r}_{ml}|^5} \right) \tag{2}$$

In this equation, $\boldsymbol{\mu}_{T,ma}$ and $\boldsymbol{\mu}_{G,l}$ represent the transition dipole moment of peptide bond m and the ground-state dipole moment of amino acid residue l, respectively, and $\boldsymbol{r}_{ml}$ is the distance vector between peptide bond m and residue l. The environmental electrostatic interactions are modeled using dipole interactions, which describe the influence of surrounding residues on the excitation energy of peptide bonds. Similarly, the resonant coupling term $J_{ma,nb}$ is defined as transition dipole interactions between peptide bonds m and n:

$$J_{ma,nb} = \sum_{m,\,n}^{m \neq n} \frac{1}{4\pi\varepsilon\varepsilon_0} \left(\frac{\boldsymbol{\mu}_{T,\,ma} \cdot \boldsymbol{\mu}_{T,\,nb}}{|\boldsymbol{r}_{mn}|^3} - 3 \frac{(\boldsymbol{\mu}_{T,\,ma} \cdot \boldsymbol{r}_{mn})(\boldsymbol{\mu}_{T,\,nb} \cdot \boldsymbol{r}_{mn})}{|\boldsymbol{r}_{mn}|^5} \right) \tag{3}$$

Here, $\boldsymbol{r}_{mn}$ is the distance vector between peptides m and n. The transition dipole moments $\boldsymbol{\mu}_T$ of peptides include both electric dipole moments $\boldsymbol{\mu}_{TE}$ and magnetic dipole moments $\boldsymbol{\mu}_{TM}$, with their interplay being crucial for calculating rotatory strength ($R = |\,\boldsymbol{\mu}_{TM}\,| \cdot |\,\boldsymbol{\mu}_{TE}\,| \cdot cos\theta$).

The computational workflow for predicting 2DUV spectra is illustrated in Figure 1. To calculate the 2DUV spectra of proteins, we can employ a divide-and-conquer strategy to decompose proteins into peptide bonds and residues and simulate the excitation energies ε_0 and the transition dipole moments $\boldsymbol{\mu}_T$ (including $\boldsymbol{\mu}_{TE}$ and $\boldsymbol{\mu}_{TM}$) of the peptides as well as the ground-state dipole moments $\boldsymbol{\mu}_G$ of the residues. By integrating ε_0, $\boldsymbol{\mu}_T$, $\boldsymbol{\mu}_G$, and the corresponding structural information, the exciton Hamiltonian of protein can be constructed using Equations (1)–(3). The 1DUV and 2DUV spectra in the far-UV range (approximately 175–240 nm) can be obtained by diagonalizing the Hamiltonian matrix using the SPECTRON 2.8 program [47].

However, obtaining the parameters ε_0, $\boldsymbol{\mu}_T$, and $\boldsymbol{\mu}_G$ for proteins typically requires computationally expensive theoretical calculations due to their large, flexible molecular structures and vast conformational spaces. Here, a fast and accurate ML approach was employed to predict ε_0, $\boldsymbol{\mu}_T$, and $\boldsymbol{\mu}_G$, as detailed in our prior works [39,40,48]. In the ML training process, a total of 1000 structurally diverse protein structures were first obtained from the RCSB Protein Data Bank (PDB) [49], encompassing a wide range of protein types, including fibrous proteins, globular proteins, keratin, collagen, chaperones, myoglobin, hemoglobin, and denatured proteins, to ensure broad representation and data diversity for machine learning. The 1000 proteins were subsequently decomposed into peptide bonds and amino acid residues, resulting in 50,000 peptide bonds and 200,000 amino acid residues (representing 20 types of residues with 10,000 structures for each type) that were randomly

selected for analysis. The PDB IDs of the 1000 proteins are provided in Table S1. Note that although this dataset captures broad coverage of protein structural classes, it represents only a small fraction of the vast protein universe, which is estimated to include millions of proteins according to databases such as UniProt and the RCSB Protein Data Bank.

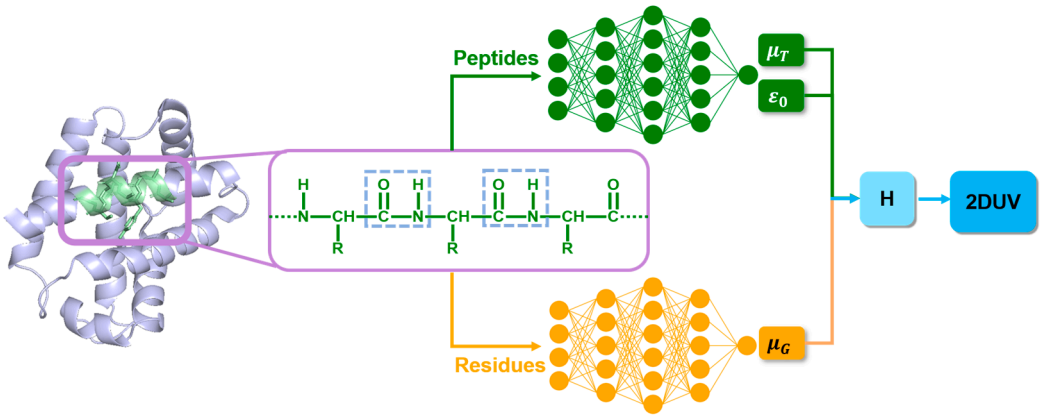


Figure 1. Machine learning protocol for predicting protein 2DUV spectra. Protein structures were obtained from the RCSB Protein Data Bank and decomposed into peptide bonds and amino acid residues. Machine learning models were employed to predict the excitation energies (ε_0) and transition dipole moments ($\boldsymbol{\mu}_T$) of peptide bonds, as well as the ground-state dipole moments of amino acid residues ($\boldsymbol{\mu}_G$). These predictions were integrated to construct the Hamiltonian matrix of the protein system. Finally, the SPECTRON 2.8 software is used to generate the protein's 2DUV spectra.

Afterwards, the divided peptide bonds and residues were converted into molecular descriptors and served as ML inputs for predicting ε_0 and $\boldsymbol{\mu}_T$ of peptides and $\boldsymbol{\mu}_G$ of residues. The molecular descriptors for ε_0, $\boldsymbol{\mu}_T$, and $\boldsymbol{\mu}_G$ were internal coordinates, embedded densities, and converted Cartesian coordinates, respectively. For the ML training reference outputs, density functional theory and time-dependent density functional theory (DFT/TDDFT) calculations were employed: ε_0 and $\boldsymbol{\mu}_T$ of peptides were computed at the PBE0/cc-pVDZ level, and $\boldsymbol{\mu}_G$ of residues was computed at the B3LYP/6-311++G** level. The ML training utilized a deep neural network protocol featuring three hidden layers with 32, 64, and 128 neurons, respectively, along with L2 regularization, for the prediction of ε_0 for peptides and $\boldsymbol{\mu}_G$ for residues. Embedded atomic neural networks (EANNs), incorporating atom-wise embedded density descriptors, were employed for the prediction of $\boldsymbol{\mu}_T$ of peptides with two hidden layers with 30 neurons each. The final optimal ML training models demonstrated high Pearson correlation coefficients (r) and low mean relative errors (MRE), achieving $r > 0.95$ and MRE < 1.5% for both ε_0 and $\boldsymbol{\mu}_T$ of peptides, as well as $r > 0.98$ and MRE < 10% for $\boldsymbol{\mu}_G$ of 20 residues, with the majority of residues exhibiting MRE < 5%. Further details regarding DFT/TDDFT calculations and ML training procedures are available in our previous studies and in the Supplementary Materials.

3. Result and Discussions

3.1. Comparison with Traditional 1DUV Spectroscopy

Traditional 1DUV spectroscopy has been widely utilized for identifying protein structures and can be extended to 2DUV spectroscopy. To evaluate the advantages of 2DUV spectroscopy, we first calculated and analyzed the 1DUV spectra of proteins in the far-UV region. Protein secondary structures, such as α-helices, β-sheets, and random coils, are essential components of protein architecture. Accurate identification of these structures not only provides insight into the backbone organization but also facilitates the design of proteins with specific functional modifications [50,51]. As shown in Figure 2a, distinct secondary structures exhibit unique absorption characteristics in the 1DUV spectra. The

α-helix displays two strong absorption peaks in the range of 44,000 cm^{-1} to 46,000 cm^{-1}, along with a weaker secondary peak at approximately 51,800 cm^{-1} in the higher frequency region. The overall peak shape is highly regular, reflecting the highly ordered structure of α-helix. The β-sheet exhibits a primary absorption peak near 45,800 cm^{-1}, accompanied by a noticeable shoulder, and a broader secondary peak spanning 50,000 cm^{-1} to 52,000 cm^{-1}. The spectrum exhibits a wider signal distribution compared to the α-helix. The random coil shows a major absorption peak at around 46,000 cm^{-1}, with weaker secondary peaks at approximately 50,000 cm^{-1} and 52,000 cm^{-1}, which are related to their disordered and flexible nature. These results demonstrate that 1DUV spectroscopy can identify major secondary structures, offering preliminary insights into protein organization and conformation.

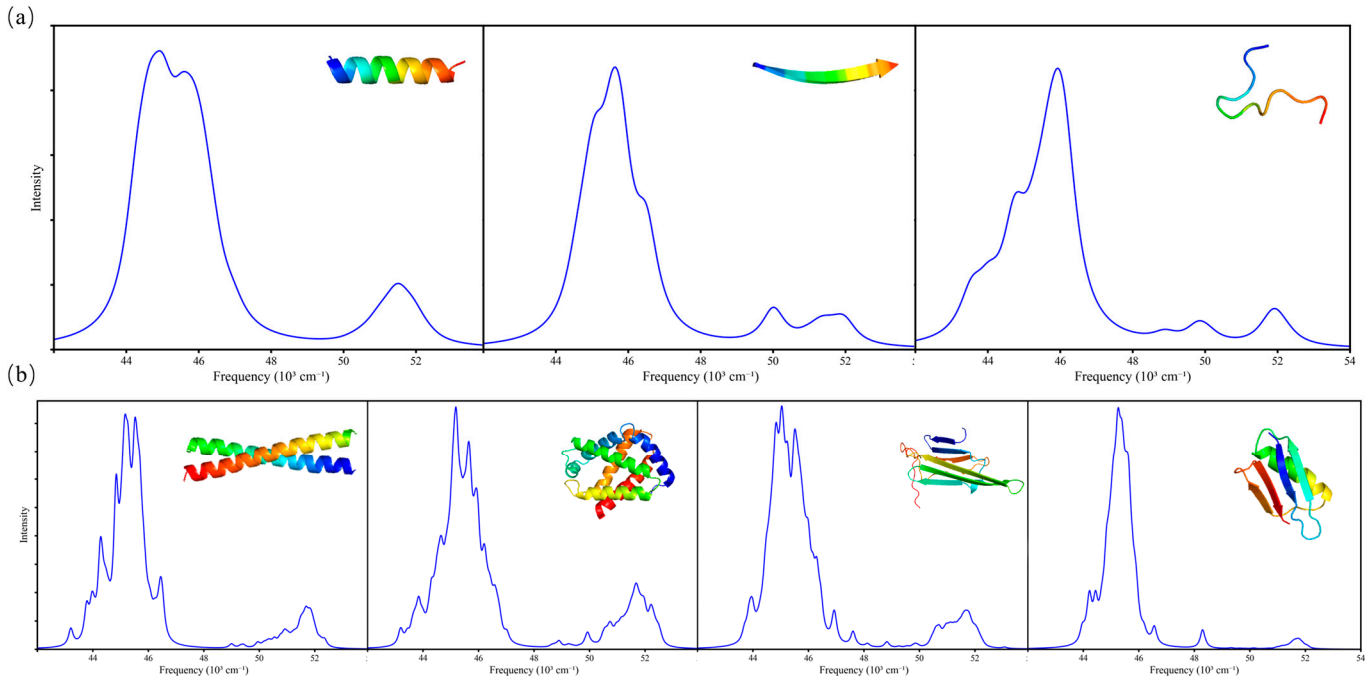


Figure 2. 1DUV spectra of the secondary and quaternary structures of proteins. (**a**) Secondary structures of proteins: α-helix, β-sheet, and random coil. (**b**) Quaternary structures of proteins: two tightly coiled α-helices, short α-helix segments, twisted β-sheets, and an α-helix/β-sheet combination. 1DUV spectra show low resolution, limiting their ability to capture detailed structural features effectively.

The quaternary structure of proteins, which involves the assembly and interaction of multiple subunits, contributes to their specific biological functions, making its study essential for understanding functional mechanisms. Fibrous and globular proteins are abundant in biological systems, with fibrous proteins typically providing structural support and globular proteins performing biochemical functions such as catalysis and signal transduction [52]. We analyzed the 1DUV spectra of four representative fibrous and globular protein fragments: two tightly coiled α-helices (2d3e), short α-helix segments (1hbg), twisted β-sheets (3l1e), and an α-helix/β-sheet combination (2gb1).

As shown in Figure 2b, the 1DUV spectral characteristics of these structures are as follows: The tightly coiled α-helices exhibit a strong primary absorption peak in the range of 44,000 cm^{-1} to 46,000 cm^{-1}, consistent with typical α-helical features, along with a weaker peak around 51,800 cm^{-1}. The signals are strong and clear, although slight irregularities in peak shape arise due to the two-coiled structure. The short-helix segments show a primary absorption peak at 45,000 cm^{-1} and a secondary peak at 45,800 cm^{-1}, along with a weak peak near 51,800 cm^{-1} in the higher frequency region. The signal intensity is less regular compared to the long helices, possibly due to reduced structural regularity. The

twisted β-sheets exhibit a primary absorption peak at around 45,000 cm^{-1}, a secondary peak at around 45,800 cm^{-1}, and a broad weak peak in the 50,000 cm^{-1} to 52,000 cm^{-1} range. These features are consistent with typical β-sheet spectra, but the dispersed and irregular peak shapes reflect conformational variations caused by twisting. The α-helix/β-sheet combination reveals mixed spectral features, with a primary peak at 45,200 cm^{-1} corresponding to the α-helix, a secondary peak at 51,800 cm^{-1} associated with the β-sheet, and a weak peak at 48,200 cm^{-1}, indicating the combined contributions of α-helix and β-sheet structures.

Although 1DUV spectroscopy can provide fundamental structural information of proteins, its resolution and sensitivity are limited in probing complex or subtle structural changes, tracking dynamic evolution, and resolving signal overlap. In contrast, 2DUV spectroscopy spreads the spectra into a two-dimensional space, which will not only provide richer and more detailed structural information but also enables the investigation of intermolecular interactions and dynamic processes of proteins due to its high temporal and spatial resolution. Therefore, its investigation is significant and promising.

3.2. Identify the Secondary and Quaternary Structures of Proteins

To evaluate the potential of 2DUV spectroscopy in characterizing protein structures, we calculated the 2DUV spectra of the same secondary and quaternary structures. As shown in Figure 3a, the 2DUV spectra of different secondary structures exhibit pronounced differences. The α-helix features a prominent positive diagonal peak (red) at approximately 51,500 cm^{-1}, accompanied by two negative sidebands (blue), reflecting a highly symmetric spectral signature. In contrast, the β-sheet shows a relatively broader primary peak in the range of 50,500 cm^{-1} to 51,500 cm^{-1}, along with an additional weaker positive diagonal peak (yellow) near 52,000 cm^{-1}. The number of negative sidebands increases from two to four, indicating a lower degree of symmetry compared to the α-helix. The random coil exhibits a narrow positive diagonal peak around 52,000 cm^{-1}, alongside a weaker positive peak near 50,000 cm^{-1}. The number of negative sidebands increases to six, resulting in a more complex and asymmetric spectral signature. These results highlight the highly ordered nature of α-helices, reflected in their symmetric and simple spectral features, while β-sheets display relatively broader positive peaks and exhibit moderate spectral complexity. In contrast, the disordered structure of random coils results in significantly more complex spectra, characterized by multiple narrow positive diagonal peaks and numerous negative sidebands.

Figure 3b presents the 2DUV spectra of four representative quaternary structures. Tightly coiled α-helices show a strong positive diagonal peak (red) in the range of 51,300 cm^{-1} to 52,000 cm^{-1}, accompanied by two negative sidebands (blue), consistent with the high symmetry characteristic of the tightly coiled structure. Short α-helix segments exhibit a red peak between 51,000 cm^{-1} and 52,000 cm^{-1}, with slightly reduced symmetry. The negative sidebands in this structure appear more complex, likely due to the irregular distribution of short helical fragments. The twisted β-sheet shows a broader prominent positive diagonal peak (red) in the range of 50,600 cm^{-1} to 52,000 cm^{-1}, along with weaker positive peaks (green) near 53,000 cm^{-1} and 48,000 cm^{-1}. This structure features an increased number of negative sidebands, reflecting greater spectral complexity caused by structural distortion. The α-helix/β-sheet composite structure exhibits the most complex spectral features. Random coils serve as connectors between the α-helices and β-sheets, and the resulting spectrum reflects contributions from all three secondary structures. The spectrum includes a prominent positive diagonal peak (red) between 51,100 cm^{-1} and 52,000 cm^{-1}, with a spectral distribution range broader than some asymmetric α-helices but narrower than typical β-sheets. This is likely due to the combined contributions of

α-helices and β-sheets. Additionally, a weaker positive peak around 48,000 cm^{-1} is observed, likely originating from the random coil. The coexistence of multiple secondary structures contributes to the increased spectral complexity. Notably, the positive diagonal peak in the range of 51,100 cm^{-1} to 52,000 cm^{-1} is broader and slightly shifted compared to spectra dominated by random coils, allowing differentiation between these structures.

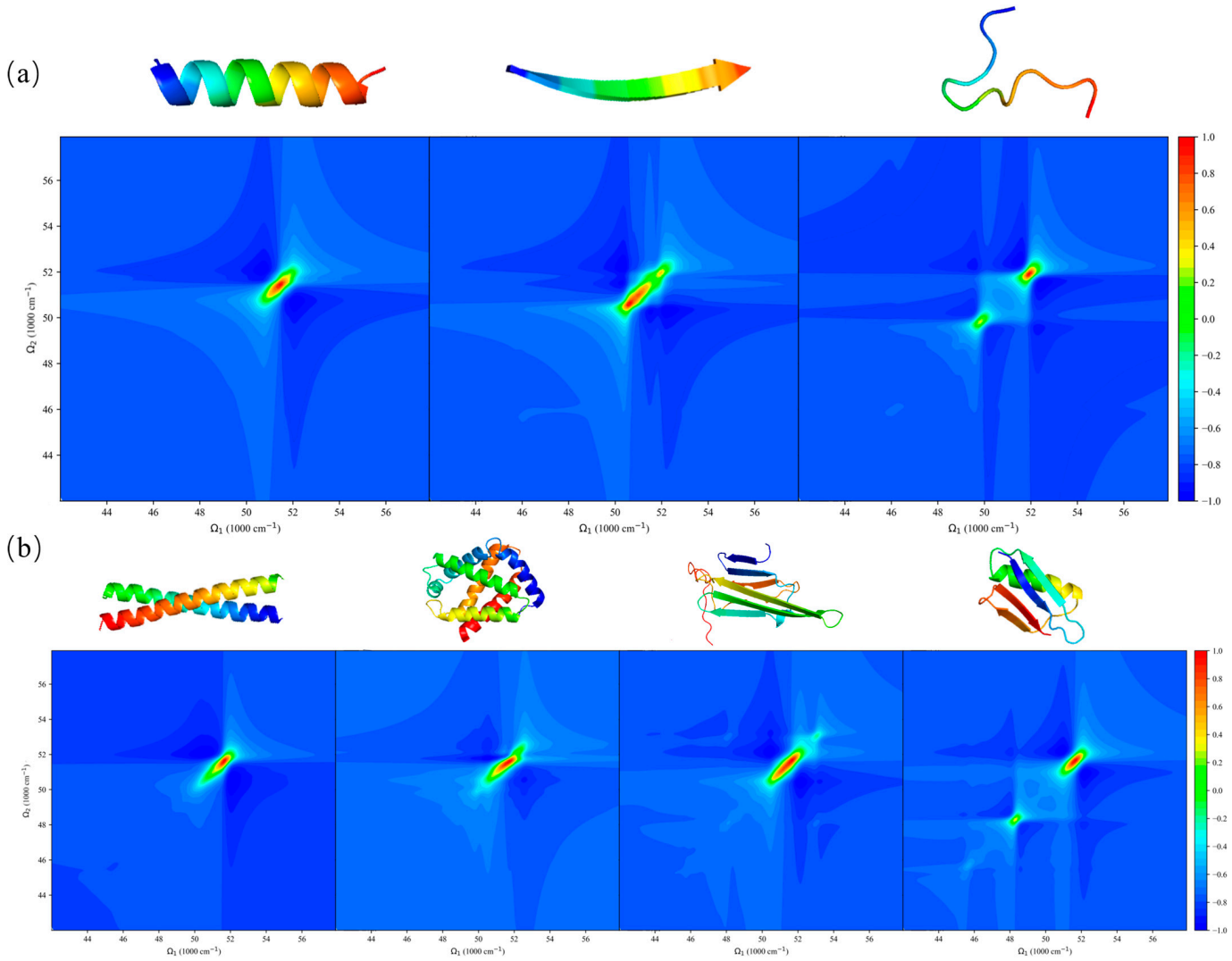


Figure 3. 2DUV spectra of the secondary and quaternary structures of proteins. (**a**) Secondary structures of proteins: α-helix, β-sheet, and random coil. (**b**) Quaternary structures of proteins: two tightly coiled α-helices, short α-helix segments, twisted β-sheets, and an α-helix/β-sheet combination. 2DUV spectra provide high resolution, enabling the identification of more detailed structural features compared to 1DUV spectra.

Therefore, 2DUV spectroscopy provides a novel approach for characterizing protein structures. It enables clear distinction of secondary structures (α-helices, β-sheets, and random coils) and quaternary structures (tightly coiled helices, short helices, twisted β-sheets, and mixed structures) with high resolution. The α-helix is characterized by symmetric features with a prominent positive diagonal peak and two negative sidebands. The β-sheet displays broader primary peaks with increased negative sidebands, while the random coil exhibits a narrow positive diagonal peak with a weaker positive peak and more complex sidebands. For quaternary structures, the irregular distribution of structural elements results in asymmetric spectral features, and the coexistence of multiple secondary structures leads to the integration of their respective spectral signatures. The rich

spectral information and unique features underscore the potential of 2DUV spectroscopy as a powerful tool for analyzing complex protein structures.

3.3. Probing Mutations of Proteins

Protein mutation plays a critical role in the development of certain human diseases and the formation of functional proteins. γS-crystallin, a major structural protein in the human eye lens, is frequently associated with cataracts or lens opacification due to mutations [53]. Wild-type (WT) γS-crystallin is essential for maintaining lens transparency, and its aggregation propensity is closely related to cataract formation. To investigate the effects of mutations on protein structure and aggregation behavior, we analyzed two mutants of WT γS-crystallin: G18V, strongly associated with early-onset cataracts, and G106V, a C-terminal symmetric engineered mutation linked to later-onset cataracts. Although these mutations induce only local structural changes, they significantly impact aggregation propensities in the order of WT < G106V < G18V [54,55].

The 2DUV spectra of WT γS-crystallin and its mutants are presented in Figure 4. Significant differences are observed in the 2DUV spectral features between the native (unmutated) and mutant structures. The 2DUV spectrum of the G106V native structure exhibits a strong positive signal peak (red) between 51,300 cm^{-1} and 52,000 cm^{-1}, along with a weaker positive signal (green) around 50,500 cm^{-1}. For the G106V mutant, the spectrum shows a strong positive signal peak (red) near 51,400 cm^{-1} but with a narrower and more irregular spectral range, indicating mutation-induced conformational changes. The 2DUV spectrum of the G18V native structure features a prominent positive signal peak (red) between 51,000 cm^{-1} and 51,600 cm^{-1}, alongside a weaker positive peak (green) near 51,900 cm^{-1}, forming an irregular elliptical shape. The G18V mutant spectrum, however, shows a more regular positive signal peak at 51,500 cm^{-1} and a new weaker peak at 52,100 cm^{-1} (green). Additionally, two weak negative sidebands (blue) flank the main peak, enhancing the spectral symmetry.

β-sheet amyloid oligomers (Aβ) are one of the key pathogenic factors in neurodegeneration associated with Alzheimer's disease. A detailed understanding of Aβ oligomer structures and their roles in disease pathology is critical for uncovering the molecular basis of Alzheimer's and for developing effective therapeutic and diagnostic strategies [56,57]. However, the study of Aβ oligomers is challenging due to their heterogeneity and instability. To explore their role in neurodegeneration, Sian et al. analyzed the E22G mutant β-sheet amyloid fibrils, where glutamic acid (GLU) is substituted with glutamine (GLN) at position 22 [58]. The 2DUV spectrum of native fibrils reveals a symmetric positive signal peak between 50,800 cm^{-1} and 52,000 cm^{-1}, with a broad spectral range and prominent negative sidebands on either side. In contrast, the E22G mutant fibrils display a slight blue shift of the positive signal peak to approximately 52,000 cm^{-1}, accompanied by a narrower spectral range. These spectral changes suggest that the E22G mutation alters fibril structure, reducing spectral signals and compacting the peak, accompanied by a slight blue shift. These findings demonstrate the ability of 2DUV spectra to sensitively detect subtle local structural changes induced by mutations. Although the spectral signals may not display significant differences between native structures and mutants, they consistently reflect the small-scale structural alterations associated with these mutations. Furthermore, because the 2DUV spectra of both native and mutant structures are derived through identical computational workflows using well-trained ML models, theoretical simulation errors are minimized. This high sensitivity to subtle variations underscores the potential of 2DUV spectroscopy to uncover how localized mutations affect the overall electronic properties and interactions within proteins, offering valuable insights for advancing the early diagnosis and treatment of related diseases.

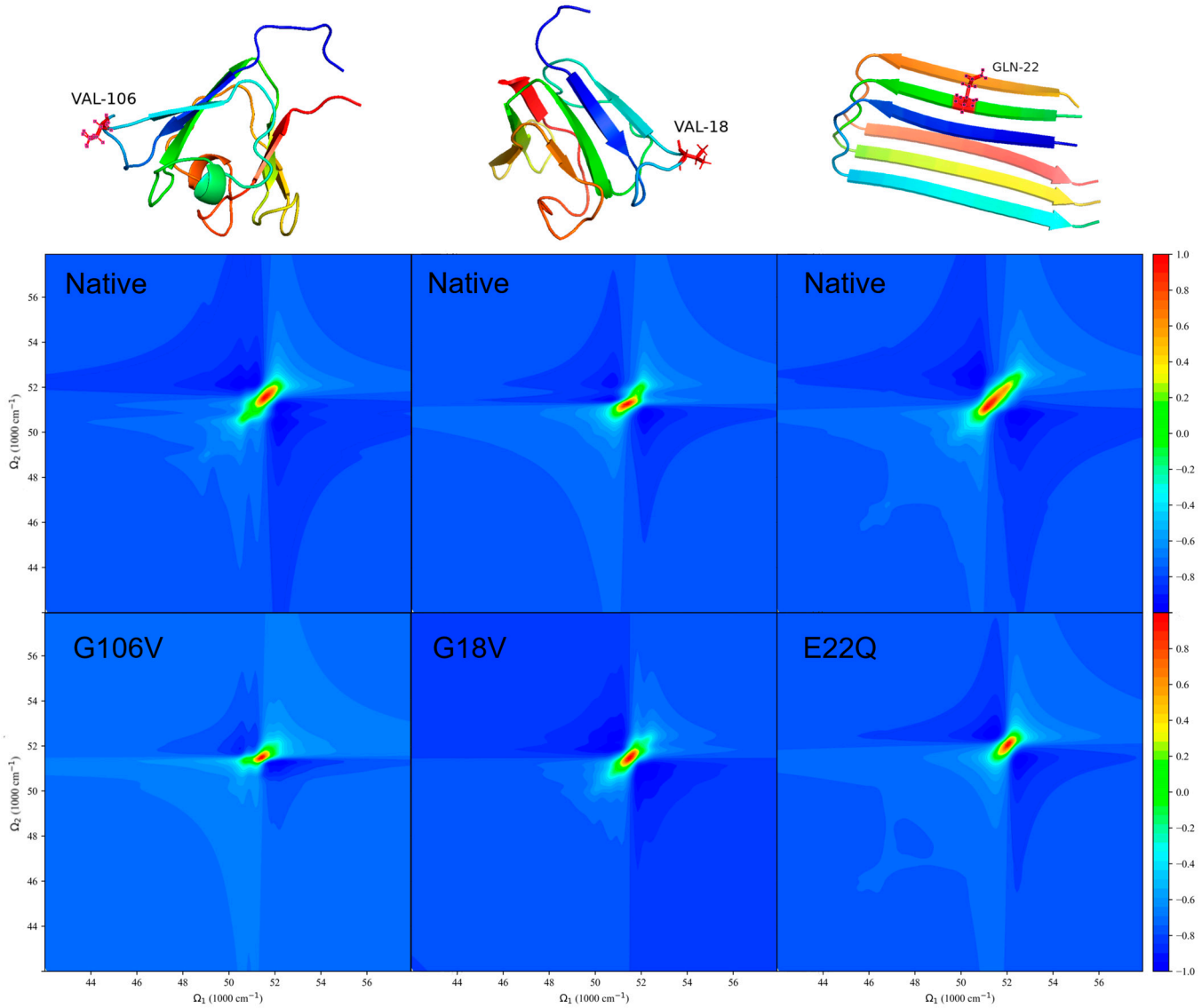


Figure 4. The 2DUV spectra for minoring the mutation of protein. Top: unmutated native structures, bottom: mutants. The G106V and G18V mutations in γS-crystallin affect the C-terminal and N-terminal domains, respectively, with G106V associated with later-onset cataracts and G18V with early-onset cataracts. The E22G mutation refers to the substitution of glutamic acid (GLU) with glutamine (GLN) at position 22 of the β-amyloid fibril sequence.

3.4. Monitoring the Aggregation Process of Proteins

Understanding the structure and growth mechanisms of amyloid fibrils is critical for addressing over 20 diseases associated with protein misfolding. Monitoring the aggregation dynamics of amyloid fibrils is essential to understand their formation mechanisms and to develop targeted treatments for protein misfolding diseases [59]. The aggregation of amyloid proteins involves several intermediate states. To investigate these dynamic changes, we calculated the 2DUV spectra at different aggregation stages, represented by structures with varying numbers of peptide chains.

Figure 5 illustrates the structures of amyloid proteins with one, three, and five peptide chains, representing the early, mid, and fully aggregated stages of amyloid formation. Across all stages, the 2DUV spectra consistently feature a prominent positive diagonal signal peak (red) and two negative sidebands (blue). However, the spectral characteristics undergo significant changes as the number of peptide chains increases. At the initial stage, the positive diagonal peak is concentrated between 51,500 cm^{-1} and 52,100 cm^{-1},

exhibiting high intensity and a narrow range. A weaker green signal peak is observed near 50,500 cm^{-1}, indicating localized structural characteristics. The negative sidebands (blue) are located around 50,700 cm^{-1} and 52,500 cm^{-1}, with broad and diffuse distributions. As aggregation progresses, the positive signal shifts within the range of 51,100–52,000 cm^{-1}, with the main peak merging with adjacent signals to form more continuous spectral features. The merging of the green signal near 50,500 cm^{-1} into the main peak, as well as the narrowing and regularization of the blue sidebands, reflects the increased structural order during peptide chain assembly. In fully aggregated fibrils, the main positive diagonal peak smoothly transitions to a broader range within 51,000–52,300 cm^{-1} with increased intensity, indicating further stabilization of the assembled fibril structure. This smooth transition underscores the cohesive nature of the fibril assembly process, where structural order develops continuously across the fibril.

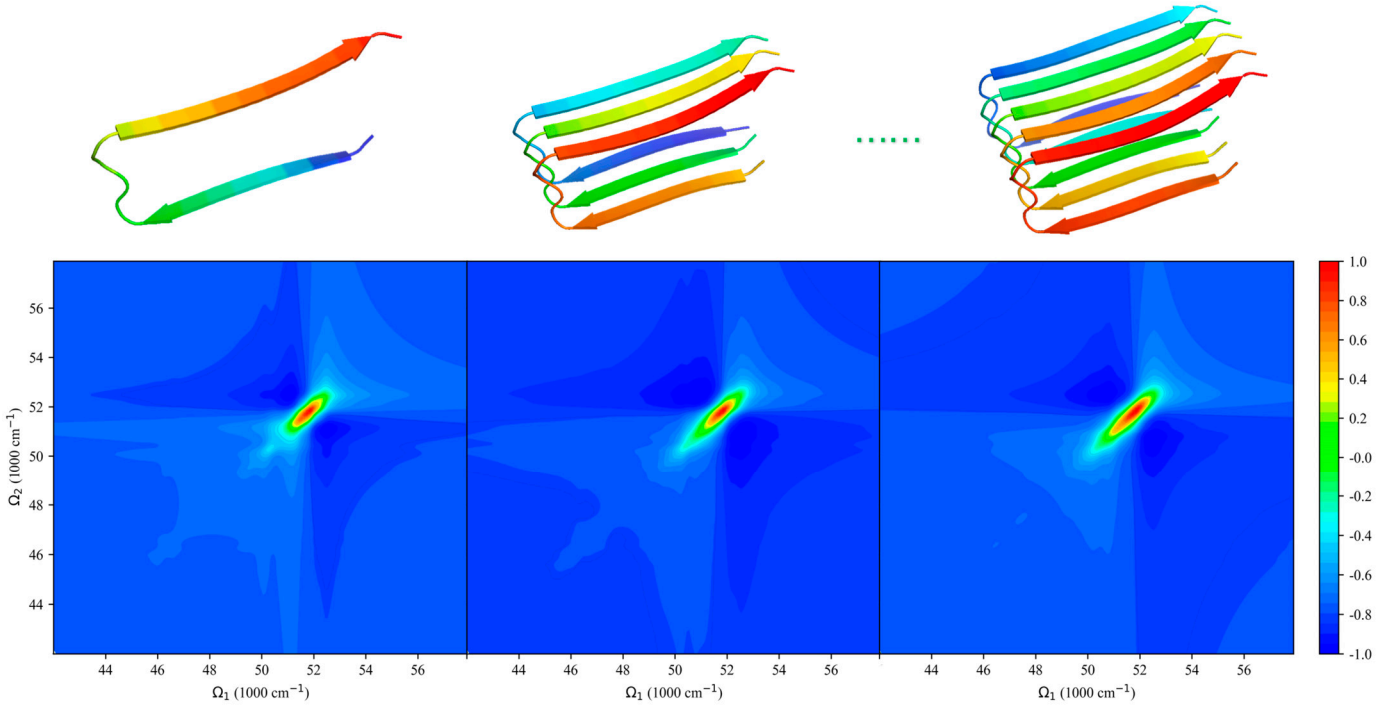


Figure 5. A schematic representation of the growth of a series of Aβ9–40 amyloid fibrils. Fibril formation progresses through three key stages: the early aggregation stage (a single peptide chain), the mid-aggregation stage (three peptide chains), and the fully aggregated stage (five peptide chains).

As the number of peptide chains increases, the 2DUV spectral signals gradually become stronger, broader, and more regular, reflecting the enhanced order and stability of the amyloid fibrils. This indicates that a detailed analysis of 2DUV spectral intensity, range, and peak position will enable the dynamic tracking of protein assembly and fibril growth processes, thus offering valuable insights into understanding the mechanisms of amyloid-associated diseases and developing targeted therapeutic strategies.

3.5. Tracking the Folding Path of Proteins

Protein folding is a crucial process in biological systems, where polypeptide chains adopt unique three-dimensional structures necessary for their functional roles. Tracking the protein folding process is essential for us to uncover the underlying mechanisms and regulate protein functions [60]. Trp-cage, a small polypeptide with a simple and rapid folding process, is widely used as a model for studying protein folding dynamics. In this study, five key folding states of trp-cage were selected, and their 2DUV spectra were predicted to track the dynamic evolution of folding. These states include the initial unfolded

strand (S1), a slightly folded state with random coils (S25), an accelerated folding state with the emergence of α-helices (S50), a continued folding state with increased α-helix content (S75), and the final fully folded cage structure (S100).

As shown in Figure 6, the position and number of positive diagonal signal peaks (red) in the spectra change significantly during the folding process. In the initial unfolded state (S1), the spectrum features three positive diagonal peaks at around 50,000 cm^{-1}, 51,000 cm^{-1}, and 52,000 cm^{-1}, with a weaker signal at around 51,000 cm^{-1}, indicating a broader overall signal distribution. This reflects the loose and dynamic nature of the random structure. In the slightly folded S25 state, the peaks at 51,000 cm^{-1} disappear, merging into the peaks at around 52,000 cm^{-1}, while the 50,000 cm^{-1} peak weakens slightly. This indicates a transition from a large random structure to a partially folded state with random coil content. In the S50 state, the peaks exhibit blue-shifting, converging around 51,000 cm^{-1} with a significantly reduced range, suggesting accelerated folding and the formation of α-helices, leading to a more ordered structure. In the S75 state, the peaks further blue-shift to approximately 51,500 cm^{-1}, with narrower spacing and concentrated signals, reflecting increased α-helix content and enhanced structural stabilization. In the final fully folded cage state (S100), the peaks blue-shift entirely to 52,000 cm^{-1}, with further reduced range and enhanced stability, clearly reflecting the final stable structure dominated by α-helices.

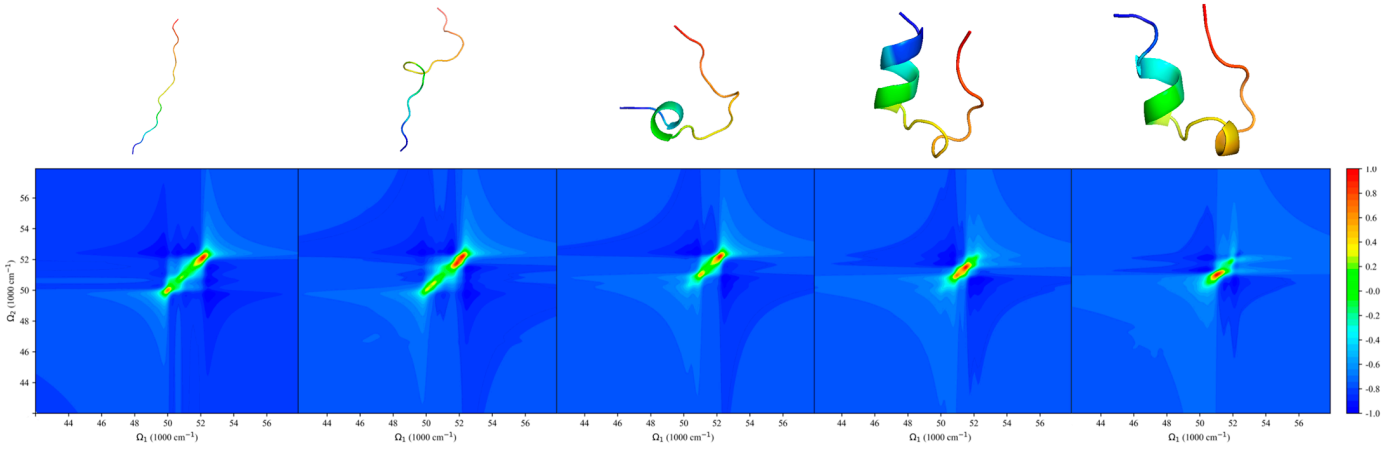


Figure 6. 2DUV spectra of the trp-cage protein along its folding pathway. The folding process of the trp-cage protein includes five key states: the initial unfolded strand (S1), a slightly folded state with random coils (S25), an accelerated folding state with the emergence of α-helices (S50), a continued folding state with increased α-helix content (S75), and the final fully folded cage structure (S100).

Overall, trp-cage transitions from an initial random structure to a stable α-helical-dominated state during folding, and this process can be tracking by significant 2DUV spectral changes. The progressive blue-shifting, reduced peak number, and signal concentration reveal the reduction in random coil content, increased α-helix formation, and overall structural stabilization. These findings demonstrate that 2DUV spectroscopy can serve as a robust tool for capturing dynamic structural evolution during folding, providing valuable insights into protein dynamics and folding mechanisms.

4. Conclusions

In summary, we successfully applied ML methods to predict 2DUV spectra of proteins and explored their applications in structural identification and dynamics tracking. The theoretical results reveal that, compared to traditional 1DUV spectroscopy, 2DUV spectroscopy can probe secondary structures (α-helices, β-sheets, and random coils) and distinguish quaternary structures formed by their combinations with higher resolution. Additionally,

2DUV spectroscopy can track dynamic processes such as mutations, aggregation, and folding, highlighting its ability to analyze protein transitions with high resolution.

The ML-based approach provides an efficient and cost-effective alternative to traditional expensive quantum chemical calculations, making it a practical solution for high-throughput studies in protein research. Its ability to establish robust structure–spectra relationships offers a pathway for systematic investigations of protein function based on spectral data. This capability holds promising potential for advancing our understanding of protein structure–function relationships and uncovering their underlying mechanisms. Additionally, the sensitivity of 2DUV spectroscopy to subtle structural changes and dynamic transitions makes it a valuable tool for studying diseases associated with protein mutations and misfolding, such as cataract formation and Alzheimer's disease. Its ability to monitor assembly and folding pathways with high temporal and spatial resolution further underscores its utility in protein engineering, disease mechanisms, and drug discovery. It is worth noting that while our ML-based 2DUV spectroscopy demonstrates strong performance in common structural identification and dynamic tracking, its application to highly disordered, transient, or complex conformations remains limited due to the constrained diversity of the training dataset relative to the vast number of proteins in existing databases. Looking ahead, the scalability of the ML-based 2DUV prediction model offers significant promise for future applications. By expanding dataset diversity and refining prediction accuracy, this approach could enable the development of more precise structure–spectra relationships and facilitate spectrum-to-structure inversion. Furthermore, the methodology may be extended to other biomolecular systems, such as DNA and RNA, broadening its impact across structural biology and related fields.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pr13020290/s1, Table S1: The PDB IDs of the 1 000 Protein Dataset; Table S2: The ML prediction results for peptide bonds; Table S3: The ML prediction results for 20 residues; Figure S1: 1DUV spectra of 18 proteins. References [21,39–41,43–46,48,52,61] are cited in the Supplementary Materials.

Author Contributions: Investigation, methodology, validation, formal analysis, writing—original draft, S.J.; resources, J.J. and T.Y.; data curation, H.Y. and L.W.; methodology, writing—review and editing, supervision, funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (grant no. 22103019), the Guangxi Natural Science Foundation (grant no. 2021GXNSFBA196024), and the Technology Base and Special Talents Development Foundation of Guangxi Province (grant no. Guike-AD21075005), and the Scientific Research Staring Foundation of Guilin University of Technology (grant no. GUTQDJJ2020127).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Marsh, J.A.; Teichmann, S.A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **2015**, *84*, 551–575. [CrossRef] [PubMed]
2. Mangubat-Medina, A.E.; Martin, S.C.; Hanaya, K.; Ball, Z.T. A vinylogous photocleavage strategy allows direct photocaging of backbone amide structure. *J. Am. Chem. Soc.* **2018**, *140*, 8401–8404. [CrossRef] [PubMed]
3. Pan, X.; Thompson, M.C.; Zhang, Y.; Liu, L.; Fraser, J.S.; Kelly, M.J.S.; Kortemme, T. Expanding the space of protein geometries by computational design of de novo fold families. *Science* **2020**, *369*, 1132–1136. [CrossRef] [PubMed]

4. Chen, C.Y.; Chang, Y.C.; Lin, B.L.; Huang, C.H.; Tsai, M.D. Temperature-resolved Cryo-EM uncovers structural bases of temperature-dependent enzyme functions. *J. Am. Chem. Soc.* **2019**, *141*, 19983–19987. [CrossRef]
5. Hendricks, N.G.; Julian, R.R. Leveraging ultraviolet photodissociation and spectroscopy to investigate peptide and protein three-dimensional structure with mass spectrometry. *Analyst* **2016**, *141*, 4534–4540. [CrossRef]
6. Ianeselli, A.; Orioli, S.; Spagnolli, G.; Faccioli, P.; Cupellini, L.; Jurinovich, S.; Mennucci, B. Atomic detail of protein folding revealed by an Ab initio reappraisal of circular dichroism. *J. Am. Chem. Soc.* **2018**, *140*, 3674–3682. [CrossRef]
7. Kurki, M.; Nesterenko, A.M.; Alsaker, N.E.; Ferreira, T.M.; Kyllonen, S.; Poso, A.; Bartos, P.; Miettinen, M.S. Solid-state NMR validation of OPLS4: Structure of pc-lipid bilayers and its modulation by dehydration. *J. Phys. Chem. B* **2024**, *128*, 4C04719. [CrossRef]
8. Lins, J.; Miloslavina, Y.A.; Avrutina, O.; Theiss, F.; Hofmann, S.; Kolmar, H.; Buntkowsky, G. Enhancing sensitivity of nuclear magnetic resonance in biomolecules: Parahydrogen-induced hyperpolarization in synthetic disulfide-rich miniproteins. *J. Am. Chem. Soc.* **2024**, *146*, 4C11589. [CrossRef]
9. Carpenter, B.P.; Talosig, A.R.; Mulvey, J.T.; Merham, J.G.; Esquivel, J.; Rose, B.; Ogata, A.F.; Fishman, D.A.; Patterson, J.P. Role of molecular modification and protein folding in the nucleation and growth of protein-metal-organic frameworks. *Chem. Mater.* **2022**, *34*, 8336–8344. [CrossRef]
10. Schmuser, L.; Trefz, M.; Roeters, S.J.; Beckner, W.; Pfaendtner, J.; Otzen, D.; Woutersen, S.; Bonn, M.; Schneider, D.; Weidner, T. Membrane structure of aquaporin observed with combined experimental and theoretical sum frequency generation spectroscopy. *Langmuir* **2021**, *37*, 13452–13459. [CrossRef]
11. Tsoukalou, A.; Abdala, P.M.; Stoian, D.; Huang, X.; Willinger, M.G.; Fedorov, A.; Muller, C.R. Structural evolution and dynamics of an In_2O_3 catalyst for CO_2 hydrogenation to methanol: An operando XAS-XRD and in situ TEM study. *J. Am. Chem. Soc.* **2019**, *141*, 13497–13505. [CrossRef] [PubMed]
12. Wang, C.; Zeng, F. Molecular structure characterization of CS_2–NMP extract and residue for malan bituminous coal via solid-state 13C NMR, FTIR, XPS, XRD, and camd techniques. *Energy Fuels* **2020**, *34*, 12142–12157. [CrossRef]
13. Stager, M.A.; Peroza, C.; Villaumie, J.; Bilham, C.; Desmond, C.; Harris, M.; Sambasivan, R.; Rowe, G.; Chen, L.; Tucker, C. Optimizing industrial solid-phase peptide synthesis: Integration of Raman spectroscopy as process analytical technology. *Org. Process Res. Dev.* **2024**, *28*, 4C00432. [CrossRef]
14. Biter, A.B.; Pollet, J.; Chen, W.H.; Strych, U.; Hotez, P.J.; Bottazzi, M.E. A method to probe protein structure from UV absorbance spectra. *Anal. Biochem.* **2019**, *587*, 113450. [CrossRef]
15. Prasad, S.; Mandal, I.; Singh, S.; Paul, A.; Mandal, B.; Venkatramani, R.; Swaminathan, R. Near UV-visible electronic absorption originating from charged amino acids in a monomeric protein. *Chem. Sci.* **2017**, *8*, 5416–5433. [CrossRef]
16. Van, Q.N.; Issaq, H.J.; Jiang, Q.; Li, Q.; Muschik, G.M.; Waybright, T.J.; Lou, H.; Dean, M.; Uitto, J.; Veenstra, T.D. Comparison of 1D and 2D NMR spectroscopy for metabolic profiling. *J. Proteome Res.* **2008**, *7*, 630–639. [CrossRef]
17. Zondlo, N.J. SAR by 1D NMR. *J. Med. Chem.* **2019**, *62*, 9415–9417. [CrossRef]
18. Soenarjo, A.L.; Lan, Z.; Sazanovich, I.V.; Chan, Y.S.; Ringholm, M.; Jha, A.; Klug, D.R. The transition from unfolded to folded g-quadruplex DNA analyzed and interpreted by two-dimensional infrared spectroscopy. *J. Am. Chem. Soc.* **2023**, *145*, 19622–19632. [CrossRef]
19. Lai, Z.; Preketes, N.K.; Jiang, J.; Mukamel, S.; Wang, J. Two-dimensional infrared (2DIR) spectroscopy of the peptide Beta3s folding. *J. Phys. Chem. Lett.* **2013**, *4*, 1913–1917. [CrossRef]
20. Ren, H.; Lai, Z.; Biggs, J.D.; Wang, J.; Mukamel, S. Two-dimensional stimulated resonance Raman spectroscopy study of the Trp-cage peptide folding. *Phys. Chem. Chem. Phys.* **2013**, *15*, 19457–19464. [CrossRef]
21. Zhuang, W.; Hayashi, T.; Mukamel, S. Coherent multidimensional vibrational spectroscopy of biomolecules: Concepts, simulations, and challenges. *Angew. Chem. Int. Ed.* **2009**, *48*, 3750–3781. [CrossRef] [PubMed]
22. Ramos, S.; Thielges, M.C. Site-specific 1D and 2D IR spectroscopy to characterize the conformations and dynamics of protein molecular recognition. *J. Phys. Chem. B* **2019**, *123*, 3551–3566. [CrossRef] [PubMed]
23. Tao, Y.; Wu, Y.; Zhang, L. Advancements of two dimensional correlation spectroscopy in protein researches. *Spectrochim. Acta Part A-Mol. Biomol. Spectrosc.* **2018**, *197*, 185–193. [CrossRef] [PubMed]
24. Silbey, R.J. Principles of nonlinear optical spectroscopy by shaul mukamel (university of rochester). Oxford university press: New york. 1995. XVIII + 543 pp. ISBN 0-19-509278-3. *J. Am. Chem. Soc.* **1996**, *118*, 12872. [CrossRef]
25. Dong, H.; Lewis, N.H.; Oliver, T.A.; Fleming, G.R. Determining the static electronic and vibrational energy correlations via two-dimensional electronic-vibrational spectroscopy. *J. Chem. Phys.* **2015**, *142*, 174201. [CrossRef]
26. Li, Q.; Giussani, A.; Segarra-Marti, J.; Nenov, A.; Rivalta, I.; Voityuk, A.A.; Mukamel, S.; Roca-Sanjuan, D.; Garavelli, M.; Blancafort, L. Multiple decay mechanisms and 2DUV spectroscopic fingerprints of singlet excited solvated adenine-uracil monophosphate. *Chemistry* **2016**, *22*, 7497–7507. [CrossRef]
27. Tumbic, G.W.; Hossan, M.Y.; Thielges, M.C. Protein dynamics by two-dimensional infrared spectroscopy. *Annu. Rev. Anal. Chem.* **2021**, *14*, 299–321. [CrossRef]

28. Jiang, J.; Mukamel, S. Two-dimensional near-ultraviolet spectroscopy of aromatic residues in amyloid fibrils: A first principles study. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2394–2400. [CrossRef]
29. Consani, C.; Aubock, G.; van Mourik, F.; Chergui, M. Ultrafast tryptophan-to-heme electron transfer in myoglobins revealed by UV 2D spectroscopy. *Science* **2013**, *339*, 1586–1589. [CrossRef]
30. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with alphafold. *Nature* **2021**, *596*, 583–589. [CrossRef]
31. Slabinski, L.; Jaroszewski, L.; Rodrigues, A.P.; Rychlewski, L.; Wilson, I.A.; Lesley, S.A.; Godzik, A. The challenge of protein structure determination--lessons from structural genomics. *Protein Sci.* **2007**, *16*, 2472–2482. [CrossRef] [PubMed]
32. Thrall, E.S.; Lee, S.E.; Schrier, J.; Zhao, Y. Machine learning for functional group identification in vibrational spectroscopy: A pedagogical lab for undergraduate chemistry students. *J. Chem. Educ.* **2021**, *98*, 3269–3276. [CrossRef]
33. Oppermann, M.; Spekowius, J.; Bauer, B.; Pfister, R.; Chergui, M.; Helbing, J. Broad-band ultraviolet CD spectroscopy of ultrafast peptide backbone conformational dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 2700–2705. [CrossRef]
34. Zhang, Y.; Chang, K.; Ogunlade, B.; Herndon, L.; Tadesse, L.F.; Kirane, A.R.; Dionne, J.A. From genotype to phenotype: Raman spectroscopy and machine learning for label-free single-cell analysis. *ACS Nano* **2024**, *18*, 18101–18117. [CrossRef] [PubMed]
35. Dixon, K.; Bonon, R.; Ivander, F.; Ale Ebrahim, S.; Namdar, K.; Shayegannia, M.; Khalvati, F.; Kherani, N.P.; Zavodni, A.; Matsuura, N. Using machine learning and silver nanoparticle-based surface-enhanced Raman spectroscopy for classification of cardiovascular disease biomarkers. *ACS Appl. Nano Mater.* **2023**, *6*, 15385–15396. [CrossRef] [PubMed]
36. Ma, S.; Liu, Z.-P. Machine learning for atomic simulation and activity prediction in heterogeneous catalysis: Current status and future. *ACS Catal.* **2020**, *10*, 13213–13226. [CrossRef]
37. Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.-i. Machine learning for catalysis informatics: Recent applications and prospects. *ACS Catal.* **2019**, *10*, 2260–2297. [CrossRef]
38. Grassano, J.S.; Pickering, I.; Roitberg, A.E.; Gonzalez Lebrero, M.C.; Estrin, D.A.; Semelak, J.A. Assessment of embedding schemes in a hybrid machine learning/classical potentials (ML/MM) approach. *J. Chem. Inf. Model.* **2024**, *64*, 4047–4058. [CrossRef]
39. Zhang, J.; Ye, S.; Zhong, K.; Zhang, Y.; Chong, Y.; Zhao, L.; Zhou, H.; Guo, S.; Zhang, G.; Jiang, B.; et al. A machine-learning protocol for ultraviolet protein-backbone absorption spectroscopy under environmental fluctuations. *J. Phys. Chem. B* **2021**, *125*, 6171–6178. [CrossRef]
40. Zhao, L.; Zhang, J.; Zhang, Y.; Ye, S.; Zhang, G.; Chen, X.; Jiang, B.; Jiang, J. Accurate machine learning prediction of protein circular dichroism spectra with embedded density descriptors. *JACS Au* **2021**, *1*, 2377–2384. [CrossRef]
41. Ye, S.; Zhong, K.; Huang, Y.; Zhang, G.; Sun, C.; Jiang, J. Artificial intelligence-based amide-II infrared spectroscopy simulation for monitoring protein hydrogen bonding dynamics. *J. Am. Chem. Soc.* **2024**, *146*, 2663–2672. [CrossRef] [PubMed]
42. Reed, J.; Kinzel, V. Near- and far-ultraviolet circular dichroism of the catalytic subunit of adenosine cyclic 5′-monophosphate dependent protein kinase. *Biochemistry* **1984**, *23*, 1357–1362. [CrossRef] [PubMed]
43. Kasha, M.; Rawls, H.R.; Ashraf El-Bayoumi, M. The exciton model in molecular spectroscopy. *Pure Appl. Chem.* **1965**, *11*, 371–392.
44. Abramavicius, D.; Palmieri, B.; Mukamel, S. Extracting single and two-exciton couplings in photosynthetic complexes by coherent two-dimensional electronic spectra. *J. Chem. Phys.* **2008**, *357*, 79–84. [CrossRef]
45. Frenkel, J. On the transformation of light into heat in solids. I. *Phys. Rev.* **1931**, *37*, 17–44. [CrossRef]
46. Zhang, Y.; Luo, Y.; Zhang, Y.; Yu, Y.J.; Kuang, Y.M.; Zhang, L.; Meng, Q.S.; Luo, Y.; Yang, J.L.; Dong, Z.C.; et al. Visualizing coherent intermolecular dipole-dipole coupling in real space. *Nature* **2016**, *531*, 623–627. [CrossRef]
47. Abramavicius, D.; Palmieri, B.; Voronine, D.V.; Sanda, F.; Mukamel, S. Coherent multidimensional optical spectroscopy of excitons in molecular aggregates; quasiparticle versus supermolecule perspectives. *Chem. Rev.* **2009**, *109*, 2350–2408. [CrossRef]
48. Zhang, Y.; Ye, S.; Zhang, J.; Hu, C.; Jiang, J.; Jiang, B. Efficient and accurate simulations of vibrational and electronic spectra with symmetry-preserving neural network models for tensorial properties. *J. Phys. Chem. B* **2020**, *124*, 7284–7290. [CrossRef]
49. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
50. Kocincova, L.; Jaresova, M.; Byska, J.; Parulek, J.; Hauser, H.; Kozlikova, B. Comparative visualization of protein secondary structures. *BMC Bioinform.* **2017**, *18*, 23. [CrossRef]
51. Chakrabarti, P. On the pathway of the formation of secondary structures in proteins. *Proteins* **2023**, *93*, 396–399. [CrossRef] [PubMed]
52. Abramavicius, D.; Jiang, J.; Bulheller, B.M.; Hirst, J.D.; Mukamel, S. Simulation study of chiral two-dimensional ultraviolet spectroscopy of the protein backbone. *J. Am. Chem. Soc.* **2010**, *132*, 7769–7775. [CrossRef] [PubMed]
53. Sun, H.; Ma, Z.; Li, Y.; Liu, B.; Li, Z.; Ding, X.; Gao, Y.; Ma, W.; Tang, X.; Li, X.; et al. Gamma-S crystallin gene (CRYGS) mutation causes dominant progressive cortical cataract in humans. *J. Med. Genet.* **2005**, *42*, 706–710. [CrossRef] [PubMed]
54. Jiang, J.; Golchert, K.J.; Kingsley, C.N.; Brubaker, W.D.; Martin, R.W.; Mukamel, S. Exploring the aggregation propensity of Gammas-crystallin protein variants using two-dimensional spectroscopic tools. *J. Phys. Chem. B* **2013**, *117*, 14294–14301. [CrossRef] [PubMed]

55. Ma, Z.; Piszczek, G.; Wingfield, P.T.; Sergeev, Y.V.; Hejtmancik, J.F. The G18V CRYGS mutation associated with human cataracts increases Gammas-crystallin sensitivity to thermal and chemical stress. *Biochemistry* **2009**, *48*, 7334–7341. [CrossRef]
56. Kreutzer, A.G.; Guaglianone, G.; Yoo, S.; Parrocha, C.M.T.; Ruttenberg, S.M.; Malonis, R.J.; Tong, K.; Lin, Y.F.; Nguyen, J.T.; Howitz, W.J.; et al. Probing differences among Abeta oligomers with two triangular trimers derived from Abeta. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, E2219216120. [CrossRef]
57. Lam, A.R.; Jiang, J.; Mukamel, S. Distinguishing amyloid fibril structures in Alzheimer's disease (AD) by two-dimensional ultraviolet (2DUV) spectroscopy. *Biochemistry* **2011**, *50*, 9809–9816. [CrossRef]
58. Sian, A.K.; Frears, E.R.; El-Agnaf, O.M.A.; Patel, B.P.; Manca, M.F.; Siligardi, G.; Hussain, R.; Austen, B.M. Oligomerization of β-amyloid of the Alzheimer's and the Dutch-cerebral-haemorrhage types. *Biochem. J.* **2000**, *349*, 299–308. [CrossRef]
59. Jiang, J.; Mukamel, S. Probing amyloid fibril growth by two-dimensional near-ultraviolet spectroscopy. *J. Phys. Chem. B* **2011**, *115*, 6321–6328. [CrossRef]
60. Jiang, J.; Lai, Z.; Wang, J.; Mukamel, S. Signatures of the protein folding pathway in two-dimensional ultraviolet spectroscopy. *J. Phys. Chem. Lett.* **2014**, *5*, 1341–1346. [CrossRef]
61. Besley, N.A.; Hirst, J.D. Theoretical studies toward quantitative protein circular dichroism calculations. *J. Am. Chem. Soc.* **1999**, *121*, 9636–9644. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.