

Supplementary Materials:

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

S1. Computational Details

Proteins are composed of peptide bonds and amino acid residues. The far-ultraviolet absorption spectra of proteins predominantly arise from the electronic excitations of their peptide backbones. These excitations are highly sensitive to surrounding environmental fluctuations, particularly the amino acid residues. A Frenkel exciton model based on a divide-and-conquer strategy can provide an effective approach for describing the electronic excitations [1–3]:

$$\hat{H} = \sum_{ma} \varepsilon_{ma} \hat{B}_{ma}^\dagger \hat{B}_{ma} + \sum_{ma,nb}^{m \neq n} J_{ma,nb} \hat{B}_{ma}^\dagger \hat{B}_{nb} \quad (S1)$$

Here, m and n represent peptide bonds, while a and b correspond to 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. \hat{B}_{ma}^\dagger and \hat{B}_{ma} are the creation and annihilation operators for electronic transitions between the ground and excited states of the peptide bonds, respectively. The excitation energy of the peptide bond, ε_0 , can be expressed as the sum of the excitation energy of an isolated peptide bond, $\varepsilon_{0,ma}$, and the electrostatic interaction energy with environment fluctuations, primarily the amino acid residues [4,5]:

$$\varepsilon_{ma} = \varepsilon_{0,ma} + \sum_l \frac{1}{4\pi\epsilon\epsilon_0} \iint d\mathbf{r}_m d\mathbf{r}_l \left(\frac{[\rho_{T,ma}(\mathbf{r}_m) - \rho_{G,ma}(\mathbf{r}_m)] \rho_{G,l}(\mathbf{r}_l)}{|\mathbf{r}_m - \mathbf{r}_l|} \right) \quad (S2)$$

Here, $\rho_{T,ma}$ and $\rho_{G,ma}$ represent the transition state and ground-state charge densities of the peptide bond, respectively, while $\rho_{G,l}$ represents the ground-state charge density of the amino acid residue. l denotes the amino acid residue. The spatial coordinates are denoted by \mathbf{r} , with T and G referring to the transition and ground states. The electronic resonance coupling between excited states is given by [6]:

$$J_{ma,nb} = \frac{1}{4\pi\epsilon\epsilon_0} \iint d\mathbf{r}_m d\mathbf{r}_n \frac{\rho_{T,ma}(\mathbf{r}_m) \rho_{T,nb}(\mathbf{r}_n)}{|\mathbf{r}_m - \mathbf{r}_n|} \quad (S3)$$

Equations (S2) and (S3) reveal that simulating peptide bonds under environmental fluctuations and dynamic structural evolution involves computationally expensive excited-state calculations and two-electron integrals. Addressing the high structural flexibility and vast conformational spaces of proteins remains a significant computational bottleneck.

To tackle these challenges, approximate methods based on divide-and-conquer strategies are commonly used. The dipole approximation, a generalized divide-and-conquer method, is an effective approach for simplifying electronic interactions [7-9]. Under this approximation, the excitation energy $\varepsilon_{0,ma}$ of a peptide bond is represented as:

$$\varepsilon_{ma} = \varepsilon_{0,ma} + \sum_l \frac{1}{4\pi\epsilon\epsilon_0} \left(\frac{(\boldsymbol{\mu}_{T,ma} \cdot \boldsymbol{\mu}_{G,l})}{|\mathbf{r}_{ml}|^3} - 3 \frac{(\boldsymbol{\mu}_{T,ma} \cdot \mathbf{r}_{ml})(\boldsymbol{\mu}_{G,l} \cdot \mathbf{r}_{ml})}{|\mathbf{r}_{ml}|^5} \right) \quad (S4)$$

The coupling between excited states is similarly expressed as:

$$J_{ma,nb} = \sum_{m,n}^{m \neq n} \frac{1}{4\pi\epsilon\epsilon_0} \left(\frac{(\boldsymbol{\mu}_{T,ma} \cdot \boldsymbol{\mu}_{T,nb})}{|\mathbf{r}_{mn}|^3} - 3 \frac{(\boldsymbol{\mu}_{T,ma} \cdot \mathbf{r}_{mn})(\boldsymbol{\mu}_{T,nb} \cdot \mathbf{r}_{mn})}{|\mathbf{r}_{mn}|^5} \right) \quad (S5)$$

Here, $\boldsymbol{\mu}_{T,ma}$ and $\boldsymbol{\mu}_{G,l}$ represent the transition dipole moment of the peptide bond and the ground-state dipole moment of the surrounding amino acid residue, respectively. 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 ($R = |\boldsymbol{\mu}_{TM}| \cdot |\boldsymbol{\mu}_{TE}| \cdot \cos \theta$).

By integrating the geometric structure of a protein, the excitation energies $\epsilon_{0,ma}$, transition dipole moments $\mu_{T,ma}$, and ground-state dipole moments $\mu_{G,l}$, the exciton Hamiltonian can be constructed using Equations (S4) and (S5). Diagonalizing the Hamiltonian allows for calculating the one-dimensional and two-dimensional ultraviolet absorption spectra (1DUV and 2DUV), which is implemented in the SPECTRON package [10].

With the rapid development of artificial intelligence (AI), machine learning (ML) has become a powerful tool for tackling complex chemical problems, including spectral simulation, drug discovery, catalysis, and energy prediction. Here, ML was employed to predict $\epsilon_{0,ma}$, $\mu_{T,ma}$, and $\mu_{G,l}$. The training dataset was derived from 1000 structurally diverse proteins obtained from the RCSB Protein Data Bank (PDB), as shown in Table S1. These proteins were decomposed into peptide bonds and amino acid residues, which were subsequently converted into molecular descriptors to serve as inputs for machine learning (ML) models. To enhance diversity, statistical robustness, and to mitigate potential sampling biases, peptide bonds and residues were evenly and randomly sampled from those derived from the decomposition. Specifically, approximately equal numbers of peptide bonds and residues were randomly selected from each protein, thus ensuring balanced representation across the dataset. Each peptide bond was represented as an N-methylacetamide (NMA) molecule. The molecular descriptors used included internal coordinates (bond lengths, angles, and dihedral angles), embedded density descriptors (evaluating density-like properties as squared linear combinations of Gaussian-type orbitals), and Cartesian coordinates (standardized to a common orientation and center).

A deep neural network protocol was employed for training ϵ_0 for peptides and μ_G for residues. This protocol consisted of three hidden layers containing 32, 64, and 128 neurons, respectively, and incorporated L2 regularization to mitigate overfitting. The rectified linear unit (ReLU) activation function was applied in each hidden layer to prevent gradient vanishing and to minimize the impact of noise. The Adam optimizer was utilized for dynamically adjusting the learning rate. For the prediction of μ_T for peptides, embedded atomic neural networks (EANNs) were employed, incorporating atom-wise embedded density descriptors. A total of 36 descriptors were used to represent the local environment of each atom in the NMA molecule, and each EANN consisted of 2 hidden layers with 30 neurons in each layer. To further enhance model robustness, early stopping was implemented, halting training if the validation loss increased consecutively for six epochs. Additionally, the Levenberg–Marquardt algorithm was applied to optimize the training process. For the ML training reference outputs, density functional theory (DFT) and time-dependent density functional theory (TDDFT) calculations were employed: ϵ_0 and μ_T of peptides were computed at the PBE0/cc-pVDZ level, and μ_T of residues was computed at the B3LYP/6-311++G** level.

To evaluate the accuracy and reliability of the ML model, we employed two key metrics: the Pearson correlation coefficient (r) and the mean relative error (MRE). The Pearson correlation coefficient, ranging from -1 to 1 , measures the linear relationship between the ML predicted and DFT/TDDFT calculated reference values, where values close to 1 indicate a strong positive correlation, reflecting the model’s ability to accurately capture the relationship between molecular descriptors and target properties. The MRE quantifies the average relative deviation between the predicted and reference values, with lower MRE values typically indicating higher prediction accuracy. As shown in Tables S2 and S3, the model achieved $r > 0.95$ and $\text{MRE} < 1.5\%$ for ϵ_0 and μ_T of peptides, as well as $r > 0.98$ and $\text{MRE} < 10\%$ for μ_G of residues, with the majority of residues exhibiting $\text{MRE} < 5\%$. These results demonstrate the ML model’s high precision and robustness in handling complex protein systems. Further details regarding dataset preparation, descriptor calculations, and ML protocols are provided in our previous studies [7,9,11].

Table S1. The PDB IDs of the 1000 protein dataset. The dataset comprised 1000 proteins downloaded from the RCSB Protein Data Bank (PDB), selected to ensure broad representation and diversity. These proteins were utilized to extract 50 000 peptide bonds and 200 000 amino acid residues, representing all 20 residue types, with 10 000 structures sampled for each type. The selected proteins encompassed a wide range of categories, including fibrous proteins, globular proteins, keratin, collagen, chaperones, myoglobin, hemoglobin, and denatured proteins, providing a comprehensive dataset for machine learning model development and testing.

1A00	1A01	1A0U	1A3O	1A4F	1A6G	1A6M	1ABY	1AH6	1AH8	1AJ9	1AMX
1ANB	1AOX	1B0B	1B86	1B9Q	1BBB	1BF8	1BIJ	1BKV	1BUW	1BUY	1BVC
1C40	1CBL	1CG5	1CG8	1CH4	1CK7	1CLG	1CMY	1CN4	1CO9	1COH	1CPZ
1DG4	1DGF	1DGH	1DKE	1DKG	1DKX	1DKY	1DLW	1DM1	1DXU	1DY2	1DZI
1ECD	1EER	1EZU	1F4J	1FAW	1FCS	1FDH	1FHJ	1FM1	1FSZ	1FUJ	1G08
1G0A	1G3J	1GCV	1GJN	1GR3	1GVL	1GXD	1GZX	1H1X	1HAB	1HBA	1HBH
1HBS	1HCO	1HGA	1HGB	1HGC	1HK7	1HX1	1HYL	1I6Z	1I7X	1IBE	1IRD
1IWH	1J14	1J3Z	1J52	1J7W	1J7Y	1JBK	1JJ9	1JWN	1JY7	1JZK	1JZL
1JZM	1K0V	1K0Y	1K9O	1KD2	1KHY	1KIU	1KKE	1KR7	1LFL	1LFQ	1LFT
1LFV	1LI1	1M3D	1M9P	1MBA	1MBD	1MBN	1MBO	1MBS	1MGN	1MKO	1MOH
1MWB	1MYH	1MYI	1MYM	1MYZ	1MZ0	1N9X	1NEJ	1NIH	1NPF	1NPG	1NQP
1NWI	1NWN	1O1I	1O1K	1O1N	1O91	1P9H	1PBX	1PMB	1Q5L	1Q7D	1QI8
1QPW	1QQW	1QUN	1QVR	1QXD	1R1X	1R1Y	1ROC	1RPS	1RTX	1RVW	1S5Y
1S69	1S6A	1SB6	1SDK	1SDL	1SHR	1SI4	1SLU	1SPG	1SS8	1SWM	1T08
1T60	1T7S	1THB	1U5M	1U7S	1U97	1UIW	1UMK	1US7	1USU	1UVY	1V4U
1V4W	1V4X	1V8X	1V9Q	1W09	1W0A	1W0B	1WG3	1WVP	1WXV	1X46	1X9F
1XUC	1XXT	1XYE	1XZ2	1XZY	1Y01	1Y09	1Y4P	1Y5J	1Y8H	1Y8I	1YCA
1YDZ	1YEO	1YEQ	1YGF	1YHU	1YIE	1YJP	1YKT	1YMB	1YOU	1YVQ	1YVT
1YZI	1Z2G	1Z8U	1ZAV	1ZE3	1ZTQ	1ZWH	2A3G	2AA1	2AKP	2AV0	2B7H
2BPR	2BRC	2BRE	2BW9	2BWH	2C0K	2CG9	2CGE	2D1N	2D2M	2D3E	2D5X
2D5Z	2D60	2D6C	2DHB	2DN1	2DN2	2DN3	2DXM	2E2D	2E2Y	2E3O	2E3R
2EKU	2EVP	2F6A	2FAM	2FRF	2FRJ	2FSE	2FXS	2G0S	2G12	2GTL	2H35
2H8D	2H8F	2HBC	2HBD	2HBF	2HBS	2HCO	2HHB	2HHD	2HHE	2HP8	2HUE
2HZ1	2IDC	2IN4	2IW2	2IWS	2JHO	2KHO	2LKV	2LLL	2LLP	2LM1	2LWP
2LYJ	2LYK	2LYL	2LYP	2LYQ	2LYR	2LYS	2M0M	2M6Z	2M8S	2MB5	2MGO
2MIQ	2MZE	2MZI	2N8R	2NB0	2ND2	2ND3	2ND5	2NRL	2NSR	2NX0	2O5L
2O5Q	2O5S	2OHB	2OJ5	2OKN	2PEI	2PEO	2PEQ	2PGH	2QIF	2QLS	2QSP
2QSS	2QU0	2R1H	2R80	2R9Y	2RAO	2SEB	2UUR	2V1E	2V1F	2V1I	2V1K
2V53	2V7Y	2VLY	2VW5	2W6V	2W6W	2W72	2XD6	2XI6	2XIF	2XIL	2XJ6
2XKI	2XX4	2YRS	2Z44	2Z46	2Z6S	2Z6T	2Z85	2Z9Y	2Z9Z	2ZLV	2ZLW
2ZLX	2ZSP	2ZSS	2ZSY	3A0G	3A2G	3A59	3AEH	3AK5	3AQ5	3ASE	3ASW
3B75	3BJ1	3BWU	3C11	3CIU	3D17	3D1K	3D7O	3DHR	3DLL	3DPO	3DPQ
3DUT	3EDA	3EJH	3ELM	3EOK	3EU1	3FH9	3FP8	3FS4	3FZH	3FZK	3GKV
3GLN	3GOU	3GQG	3GQP	3GYS	3H0X	3H3T	3HC9	3HF4	3HQV	3IA3	3IC0
3IC2	3IUC	3K8B	3KEK	3LDL	3LDN	3LDO	3LDP	3LDQ	3LJZ	3LQD	3LR7
3LW2	3M0B	3M38	3M3B	3MBA	3MJP	3MJU	3MVF	3N3F	3NL7	3NML	3O2X
3ODQ	3OGB	3OVU	3PEL	3PI8	3PI9	3QJE	3QL1	3QZL	3QZM	3QZN	3QZO
3RIK	3RJR	3RTL	3RUR	3S48	3S5C	3S5H	3S5K	3SDH	3SZK	3TFB	3TNU
3TVC	3UHI	3UT2	3V03	3V2V	3VFE	3VM5	3VM9	3VND	3VNW	3VQK	3VQL
3VQM	3W6L	3WFT	3WHM	3WI8	3WTG	3WV1	3WVL	3WYO	3ZHC	3ZHD	3ZHK
3ZHL	4A7B	4AIX	4AIZ	4AJ0	4AU2	4B2T	4B9Q	4BB2	4BJ3	4BKL	4BNR
4C0N	4C44	4CTD	4CUD	4CUE	4CUF	4D0E	4D2U	4D8N	4DC5	4DF3	4DOU
4EO5	4EZN	4EZO	4EZP	4EZR	4EZW	4EZX	4F01	4F4O	4F68	4FC3	4FCT
4FCW	4FVL	4FWZ	4GR7	4H32	4HRR	4HRT	4HSE	4HWC	4I0C	4I0Y	4I1E

4I2S	4I37	4I3N	4I96	4IJ2	4JA7	4JA9	4JB0	4JB2	4JSD	4JSO	4K07
4K5Q	4K6G	4K6H	4K6K	4KJT	4L2A	4L2D	4LJ6	4LJA	4M4B	4M56	4M8U
4MA7	4MBN	4MKF	4MKG	4MKH	4MPB	4MPR	4MQK	4MTH	4N79	4N7P	4N8W
4NI0	4NSM	4NWE	4NWH	4O4T	4O4Z	4OF9	4OJ0	4OOD	4OW4	4PNJ	4QBY
4R1E	4RMB	4RRP	4RX9	4TQL	4TYU	4U3H	4U5T	4U8U	4UOS	4UOT	4UOX
4UOY	4URG	4URQ	4URS	4UZV	4W68	4W70	4W81	4W94	4WJG	4WUY	4XIF
4XS0	4Y00	4YU3	4YU4	4Z3V	4ZLY	4ZRY	5AKS	5AO6	5AQG	5AQI	5AQO
5AQT	5AUY	5AZQ	5B5O	5B85	5BOY	5BX0	5CE5	5CJB	5CMV	5CN5	5CNC
5CTD	5CTI	5CVA	5CVB	5D0Q	5D5R	5E83	5E84	5E85	5EII	5EIV	5F2R
5FFO	5FQD	5FWL	5FWP	5FXP	5GHU	5GW4	5GW5	5HCL	5HLY	5HQ3	5HY8
5IAT	5IAX	5IKS	5ILM	5ILP	5ILR	5J3P	5J3S	5J3Z	5JG9	5JHI	5JI4
5JOM	5KA0	5KER	5KI0	5KKK	5KRR	5KSI	5KSJ	5KVN	5KWX	5KWZ	5KX0
5KX1	5KX2	5M4G	5M4J	5M4L	5M9M	5MBY	5MC1	5MU0	5MV3	5MZU	5N30
5N4H	5NAX	5NI1	5NIR	5NJX	5NRO	5NX3	5O4P	5OBU	5OCX	5OFO	5OMP
5OMY	5OPW	5OPX	5OU8	5OU9	5OWI	5OWJ	5PKC	5Q5Z	5QEH	5R4J	5SV3
5SV7	5SXD	5THP	5TU7	5TU8	5TU9	5U2L	5U2U	5UCB	5UCU	5UE2	5UE5
5UEA	5UEK	5URC	5UT7	5UT9	5UWK	5UYX	5V4M	5V4N	5VPN	5VQP	5VY8
5VY9	5VZN	5VZO	5VZP	5VZQ	5W0S	5WOG	5X2R	5X2S	5XKV	5Y45	5YAN
5YCE	5YP8	5YPB	5YUP	5YZF	5Z5O	5ZBA	5ZHB	5ZUI	5ZYK	5ZZF	5ZZG
5ZZT	5ZZY	6A06	6A0H	6A0V	6A0Y	6A19	6A1W	6A23	6A2U	6A32	6A39
6A3C	6AHF	6AIT	6ASY	6AXB	6BB5	6BIN	6BJR	6BNR	6BWU	6CD2	6CF0
6CQG	6CQV	6D45	6D6S	6DDK	6DFM	6DJU	6DL9	6DTC	6E14	6E15	6E0F
6E0G	6E2J	6E7G	6E7H	6EC0	6ED3	6EOF	6F0Y	6F17	6F25	6FQF	6FSE
6FZW	6G5A	6G5T	6GCQ	6GZD	6H2P	6H2Q	6HAL	6HBI	6HBW	6HG7	6HV2
6IHX	6II1	6IWK	6J0A	6J81	6JBX	6JP1	6M8F	6MV0	6N02	6N8V	6N8Z
6NBC	6NBD	6ND8	6NDH	6O5V	6O69	6OG3	6QFF	6QFH	6QH9	6QI8	6REU
6S0F	6TSZ	6UUUV	6VGK	6W75	6XV4	6Y6W	7ABP	7ACN	7AHL	7AME	7API
7BNA	7CA2	7CCP	7CEI	7CEL	7CGT	7DFR	7FAB	7FD1	7GAT	7GCH	7HSC
7HVP	7ICD	7ICE	7ICF	7ICN	7ICO	7ICQ	7ICR	7ICV	7INS	7KME	7LPR
7LYZ	7LZM	7MHT	7MSF	7NN9	7NSE	7PAZ	7PCK	7PTD	7R1R	7REQ	7RSA
7RXN	7STD	7TIM	7TLN	7WGA	7XIM	7YAS	7ZNF	821P	830C	8A3H	8AAT
8ABP	8ACN	8ADH	8AME	8API	8AT1	8ATC	8BNA	8CA2	8CAT	8CGT	8CHO
8CPA	8CPP	8DFR	8DRH	8EST	8FAB	8GCH	8GEP	8GPB	8GSS	8HVP	8I1B
8ICA	8ICZ	8JDW	8KME	8LDH	8LPR	8LYZ	8MHT	8MSI	8NSE	8OHM	8PAZ
8PCH	8PRK	8PRN	8PSH	8PTI	8RAT	8RNT	8RSA	8RUC	8RXN	8TFV	8TIM
8TLI	8TLN	8XIA	8XIM	9ABP	9AME	9ANT	9ATC	9CA2	9CGT	9DNA	9EST
9GAA	9GAC	9GAF	9GPB	9GSS	9HVP	9ICA	9ICC	9ICE	9ICH	9ICJ	9ICK
9ICM	9ICO	9ICQ	9ICS	9ICU	9ICV	9ICY	9ILB	9INS	9JDW	9LDB	9LDT
9LPR	9LYZ	9MHT	9MSI	9NSE	9PAI	9PAP	9PCY	9PTI	9RAT	9RNT	9RSA
9RUB	9WGA	9XIA	9XIM								

Table S2. The ML prediction results for peptide bonds.

Name	Metrics	$n \rightarrow \pi^*$	$\pi \rightarrow \pi^*$
ε_0	r	0.9616	0.9512
	MRE/%	0.363	0.252
μ_{TE}	r	0.9973	0.9590
	MRE	0.191	1.499
μ_{TM}	r	0.9973	0.9729
	MRE	0.062	0.646

Table S3. The ML prediction results for 20 residues.

Residues	GLY	ALA	LEU	ILE	VAL	PRO	PHE	MET	TRP	SER
r	0.9963	0.9974	0.9939	0.9966	0.9984	0.9986	0.9945	0.9932	0.9954	0.9982
MRE/%	4.095	4.0265	6.0746	4.3792	3.4479	3.0722	6.4502	7.1671	6.1629	3.9615
Residues	GLN	THR	CYS	ASN	TYR	ASP	GLU	LYS	ARG	HIS
r	0.9942	0.9852	0.9948	0.9979	0.9962	0.997	0.9964	0.9997	0.9996	0.9978
MRE/%	6.4224	9.3273	6.5005	3.8518	5.1823	4.9678	4.8288	1.6229	1.6891	3.8792

S2. Predicted One-Dimensional Ultraviolet (1DUV) Spectra of 18 Proteins

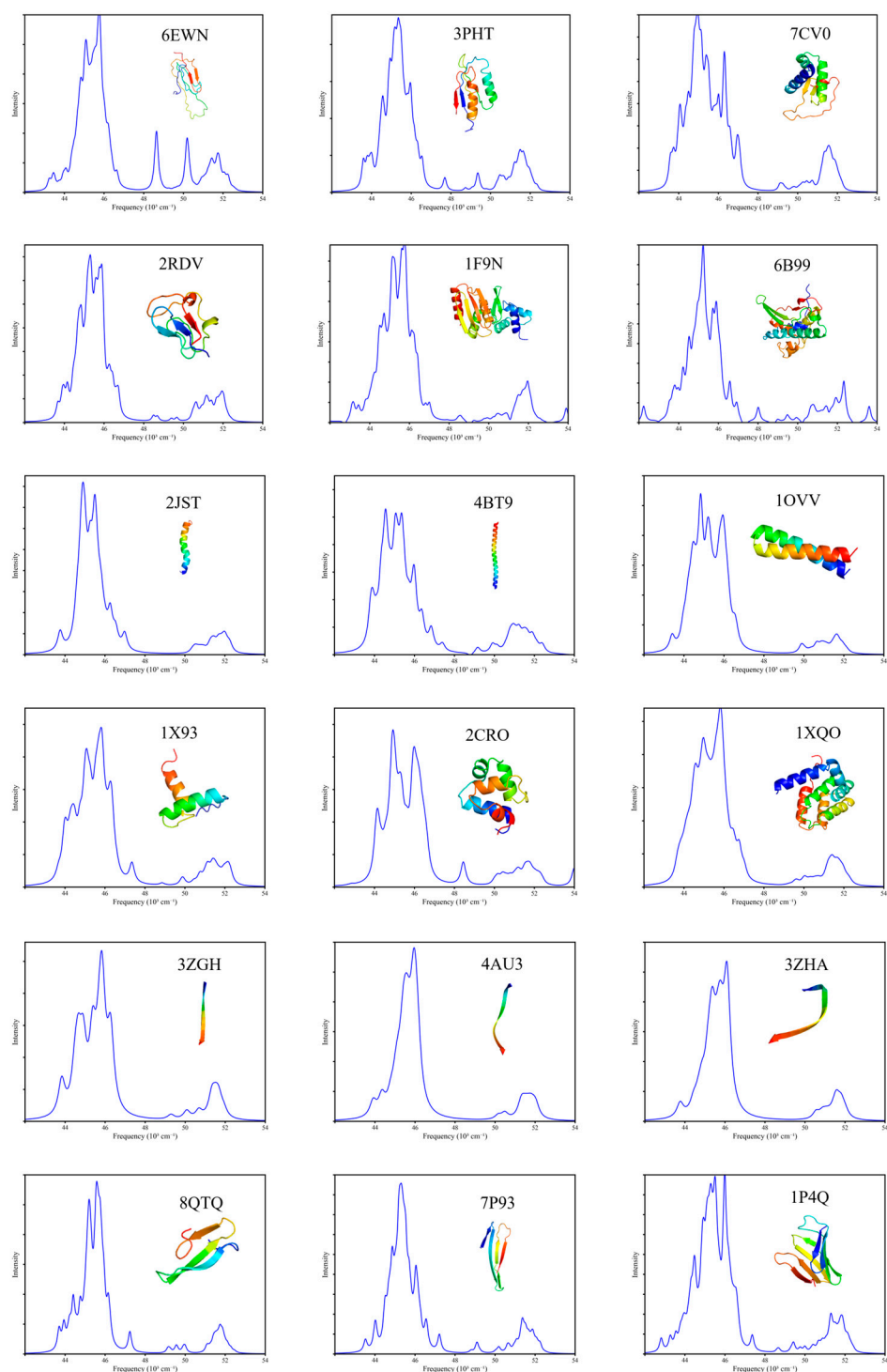


Figure S1. 1DUV spectra of the 18 proteins.

References

1. Abramavicius, D.; Palmieri, B.; Mukamel, S. Extracting single and two-exciton couplings in photosynthetic complexes by coherent two-dimensional electronic spectra. *Journal of Chemical Physics* **2008**, *357*, 79-84.
2. Frenkel, J. On the transformation of light into heat in solids. I. *Physical Review* **1931**, *37*, 17-44.
3. Abramavicius, D.; Jiang, J.; Bulheller, B.M.; Hirst, J.D.; Mukamel, S. Simulation study of chiral two-dimensional ultraviolet spectroscopy of the protein backbone. *Journal of the American Chemical Society* **2010**, *132*, 7769-7775.
4. Kasha, M.; Rawls, H.R.; Ashraf El-Bayoumi, M. The exciton model in molecular spectroscopy. *Pure and Applied Chemistry* **1965**, *11*, 371-392.
5. 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.
6. Besley, N.A.; Hirst, J.D. Theoretical studies toward quantitative protein circular dichroism calculations. *Journal of the American Chemical Society* **1999**, *121*, 9636-9644.
7. 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.
8. 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. *Journal of the American Chemical Society* **2024**, *146*, 2663-2672.
9. 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. *The Journal of Physical Chemistry B* **2021**, *125*, 6171-6178.
10. Zhuang, W.; Hayashi, T.; Mukamel, S. Coherent multidimensional vibrational spectroscopy of biomolecules: Concepts, simulations, and challenges. *Angewandte Chemie international Edition* **2009**, *48*, 3750-3781.
11. 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. *The Journal of Physical Chemistry B* **2020**, *124*, 7284-7290.