Two-Dimensional Ultraviolet (2DUV) Spectroscopic Tools for Identifying Fibrillation Propensity of Protein Residue Sequences

ARTICLE *in* ANGEWANDTE CHEMIE INTERNATIONAL EDITION · DECEMBER 2010

Impact Factor: 11.26 · DOI: 10.1002/anie.201005093 · Source: PubMed

2 AUTHORS, INCLUDING:

Shaul Mukamel
University of California, Irvine
852 PUBLICATIONS **23,715** CITATIONS

Published in final edited form as:
Angew Chem Int Ed Engl. 2010 December 10; 49(50): 9666–9669. doi:10.1002/anie.201005093.

Two Dimensional Ultraviolet (2DUV) Spectroscopic Tools for Identifying Fibrillation Propensity of Protein Residue Sequences**

Jun Jiang and **Shaul Mukamel**
Dr. J. Jiang, Prof. Dr. S. Mukamel., Chemistry Department, University of California, Irvine, California, USA., Fax: (+1) 949-824-8571

Jun Jiang: jiangj1@uci.edu; Shaul Mukamel: smukamel@uci.edu

More than 20 neurodegenerative diseases[1–3] are associated with the formation and deposition of amyloid fibrils of misfolded proteins.[4–8] Various amyloid-forming proteins across diverse systems are known to share common structures and fibril formation kinetics.[9,10] This suggests the existence of common molecular mechanisms for different amyloid diseases. A recent article argued that "The most common mechanism by which proteins aggregate consists of the incorporation of relatively short sequence segments into β-sheet-like assemblies".[11] It has been conjectured that the fibrillation propensities of proteins depend strongly on their sequences.[11–13] Many theoretical tools have been proposed to describe and characterize this sequence-dependence.[11–16] Their success in predicting the fibrillation propensity of various protein sequences should help understand why protein form fibrils.

The study of amyloid fibrils would greatly benefit from adequate tools that correlate the fibrillation propensity of proteins with physical or chemical properties accessible by both theory and experiment. However, the parameters or factors currently available in the theoretical tools for evaluating the fibrillation propensity are not accessible experimentally. For instance, the TANGO[14], Waltz[11], and Zyggregator[15] computational tools provide different "aggregation scores", which are not connected to physical or chemical properties accessible by experiment. The 3D profile method[13] and the PASTA algorithm[16] are based on the energy function of protein sequences. As shown in a recent work of Eisenberg *et. al.*,[12] the ability of proteins to form fibrils can be effectively predicted by their free energies. Again, it is extremely hard to measure free energies of separated protein segments, and the required discrimination of energies within several kcal/mol is not possible by current experiments. Moreover, the above criteria normally rely on detailed structural information, which is not available for most aggregates due to the lack of suitable probes with atomic resolution.[4,6,7]

Coherent ultrafast two dimensional (2D) optical spectroscopy is a novel tool for studying the structure and dynamical properties of bio-molecules.[17–21] These techniques have been successfully employed for studying fibril structure and aggregation kinetics and mechanisms.[18,22–26] Recent progress in laser technology opens up new windows of observation[27] by extending these techniques into the ultraviolet (UV).[28–31] 2DUV may

**This research was supported by the National Institutes of Health (Grand GM059230 and GM091364), and the National Science Foundation (Grant CHE-0745892). We thank Prof. Yi Luo for helpful discussions.

Correspondence to: Shaul Mukamel, smukamel@uci.edu.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

use coherent ultrafast (20 fs) pulses for the characterization of bio-molecules.[32] It provides an effective way for probing the geometric and electronic structures, and the inter- and intra-molecular interactions of amyloid fibrils.[24]

It was reported recently[12] that the fibrillation propensity of proteins decreases as the Rosetta free energy[33] is increased. This is computed by the software RosettaDesign[34,35] as a potential energy function, which depends on the following ingredients: (i) the Lennard–Jones potential with van der Walls radii and well depth for atoms, (ii) the solvation energy computed using the Lazaridis–Karplus implicit solvation model, (iii) the side-chain-main-chain hydrogen bonding term, (iv) the torsion potentials estimated from PDB statistics, (v) a unique reference value for each amino acid type and (vi) electrostatic interactions between residues. The Rosetta energy depends strongly on through space interactions among residues. The relationships between such interactions and the fibrillation propensity could thus be used to predict the fibril formation. Here we use advanced theoretical tools developed for 2DUV spectroscopy of protein backbone transitions, to study various protein residue sequences with known fibrillation propensity. Simulations are carried out using a QM/MM (Quantum Mechanics/Molecular Mechanics) approach (details are described in the Supporting Information).[32,36] We identify direct spectroscopic signatures of the Rosetta free energies. Combined with the established connection between the fibrillation propensity and the Rosetta energy this study demonstrates that 2DUV signals may be used to identify protein residue sequences capable of forming amyloid-like fibrils.

In a joint experimental and theoretical investigation, Eisenberg *et. al.*[12] had shown that the fibrillation propensity of some protein segments is relatively insensitive to the residue composition but is strongly sequence-dependent. Short protein segments with 4–10 residues tend to form amyloid-like fibrils provided their Rosetta energies are below −23 kcal/mol. If the sequence of a fibrillizing segment is shuffled to obtain a higher energy, it loses its tendency to form fibrils.

The Rosetta free energy consists primarily of three terms: the intrinsic free energy of the protein backbones and amino acids, solvation energy, and interaction energy between residues. The former two only depend on the residue composition and are not sensitive to the sequence. Interactions among residues vary with their couplings and should depend on sequence shuffling. We argue that the variation of the Rosetta energy associated with the shuffle of protein sequences should be induced by the change of interactions among residues. These interactions in turn affect the excitonic energy splitting of protein electronic transitions. In an earlier simulation of protein backbone UV spectra,[36] we found that residue interactions split the distribution of protein transition energies into several peaks. The two strong transitions at 48000 cm^{-1} (~210nm) and 53000 cm^{-1} (~190nm) result from the splitting of the $\pi \rightarrow \pi^*$ backbone transitions. Thus, the transition frequency patterns of proteins reflect the strength of the residue-residue interactions (RRI), which affect the Rosetta energy.

Eisenberg *et. al.* have performed a systematic study of the fibrillation propensity of the Bovine pancreatic ribonuclease A (RNase A) protein and other 13 short protein segments[12], for which the theoretical prediction of the Rosetta energy/fibrillation propensity relation has been validated by experiment. We have therefore used these proteins, to explore 2DUV characteristics associated with the Rosetta energy and fibrillation propensity. We have labelled the eight RNase A segments used in [12] as R1-F, R2-F, R3-F, R4-F, R5-F, R6-F, R7-NF, and R8-NF. F (NF) denotes fibril forming (non-forming) sequences with Rosetta free energy lower (higher) than −23 kcal/mol. The distributions of the backbone transition frequencies and UV spectra were calculated by using 500 molecular dynamic (MD) snapshots. The distributions of the $\pi \rightarrow \pi^*$ backbone transition frequencies

(which dominate the UV signals of proteins[36]) are displayed in the far UV region (46000~56000 cm^{-1}) for six RNase A segments in Fig. 1. The four left (F) panels of Fig. 1 contain four or five peaks. The two right (NF) panels contain fewer (two or three) peaks. This trend survives the differences in residue compositions. We have also calculated the same distributions for three segments of the wild RNase A and their shuffled sequences (Fig. S1 in the Support Information). The shuffled sequences of R5-NF, R6-NF, and R8-NF are found to contain fewer and narrower peaks than the corresponding F sequences.

Ordinary one dimensional ultraviolet (1DUV) spectroscopy of proteins provides information about the distribution of electronic transition frequencies of singly-excited states which depend on residue-residue interactions (RRI).[36] UV resonant Raman spectra show characteristic vibrational features associated with the fibrillation.[37,38] We have first simulated the 1DUV spectra (linear absorption and circular dichroism) of several proteins, and looked for the fibrillation propensity/1DUV relations. We found that the differences of spectral broadening induced by the change of RRI are not easily observable in these highly congested spectra (Fig. S2 and S3 in the Support Information).

2DUV photon echo signals depend on both singly and doubly excited electronic states, and thus contain richer spectral features associated with the geometric and electronic structures of proteins than 1DUV. 2DUV signals probe doubly-excited states, and additional 2DUV peaks imply that excitations of the protein segments have more diverse transition energies. 2DUV spectra are very sensitive to variations in protein backbone transitions, which in turn depend on RRI.

The fibrillation propensity/transition frequency relation was examined for both non-chiral (xxxx) and chirality-induced (xxxy) pulse polarization configurations. We found the 2DUV xxxx spectra of RNase A segments to be insensitive to the fibrillation propensity, as described in the Support Information (Fig. S4 and S5). The chiral 2DUV xxxy signals in contrast are much richer and show a high sensitivity to the protein geometry and fibrillation propensity. This is illustrated by the simulated spectra of RNase A segments R1-F, R2-F, R3-F, R4-F, R7-NF, and R8-NF shown in Fig. 2 (A). The F spectra show more elaborate peak patterns, than the NF. The spectra of segments R7-NF and R8-NF contain only four peaks, while the other F segments have at least six. Simulated 2DUV xxxy spectra of R5, R6, and R8 with the F and NF shuffled sequences are shown in Fig. 2 (B). We denote the energy difference between the shuffled sequences with lower and higher energy by ΔE. Again, the spectra of the F group contain more peaks than of the NF. In the NF segments, the signals are mostly distributed close to diagonal line. Many additional small crosspeaks are observed in the xxxy spectra of F segments. Since crosspeaks are induced by couplings between chromophores, this also suggests the strong RRI in the F group.

The fibril formation propensity threshold of −23 kcal/mol for the Rosetta energy had also been tested and validated in thirteen wild protein segments. These segments were further shuffled to get sequences with the lowest and highest Rosetta energies. [12] We have simulated the 2DUV spectra of twelve of these protein segments (Labelled P1 to P12) whose crystal structures are available in the protein data bank. The shuffled sequences with the lowest and highest Rosetta energies are labelled as P*xx*-L and P*xx*-H. Simulated spectra of P*xx*-L and P*xx*-H (in Figs. S6, S7 and S8 in the Support Information) help validate the correlation between the Rosetta energy and 2DUV spectral features. We found that an increase of Rosetta energy upon sequence shuffling is always accompanied by a decrease of the number of 2DUV xxxy peaks. These reflect the reduction of RRI which is correlated with the ability to form fibrils. These spectroscopic features are consistent with the transition diversity illustrated in Figs. S6, S7 and S8 in the Support Information, where in all cases the

protein sequence with the lowest Rosetta energy shows much broader distributions with a larger number of peaks than the shuffled sequence with the highest Rosetta energy.

The differences of 2DUV spectra induced by sequence shuffling reflect variations of RRI. We have subtracted the 2DUV xxxy spectra of the sequences with lower Rosetta energies (and fibrils observed) from those with higher energies for three RNase A segments (R5, R6, R8) and eight other protein segments (P1 to P8). The resulting difference spectra displayed in Fig. 2 (C) show different patterns for various protein segments. However, we found a notable common feature in all cases: a positive (red) diagonal peak at around 54000~58000 cm^{-1} which is sometimes accompanied by two negative (blue) side bands. This feature is also common in the 2DUV xxxy spectra of a series of $A\beta_{1–40}$ amyloid fibrils of various size. [24] One example of A $\beta_{1–40}$ amyloid fibrils is shown at the bottom right panel of Fig. 2 (C). In all cases shuffling the protein sequence to decrease the Rosetta energy and enhance the ability to form amyloid-like fibrils, results in additional positive diagonal signals at 54000~58000 cm^{-1} in the 2DUV xxxy spectra.

In summary, the present computational QM/MM investigation of the fibrillation propensity of protein segments reveals a Rosetta energy/fibrillation propensity relation in proteins. We found that protein segments with high fibrillation propensity and lower Rosetta energy have more diversified transition energy distributions than those with low fibrillation propensity and higher Rosetta energy. Decrease of the Rosetta energy implies increasing RRI. Fluctuations of the QM Hamiltonians of the system are then amplified, increasing the splittings of the electronic transition energies, and broadening their distribution.

UV spectroscopy of proteins contains information about RRI and transition diversification, which is related to the fibrillation propensity. Due to high spectral congestion, 1DUV techniques such as absorption and CD do not provide clear signatures of the fibrillation propensity. 2DUV signals are much richer and better resolved. Simulated 2DUV spectra of six RNase A segments and twelve other protein segments and their shuffled sequences show that 2DUV xxxy signals are very sensitive to variations of RRI. We further demonstrated that chiral 2DUV (xxxy) spectra of protein segments with high fibrillation propensity contain additional peaks than those with low fibrillation propensity. 2DUV xxxy signals directly represent the electronic transition diversity of proteins, and provide a convenient tool for identifying the fibrillation propensity of protein sequences.

2DUV difference spectra can help understand the variations of inter-molecular interactions in proteins. 2DUV xxxy spectroscopy differences of shuffled protein sequences share a common spectral feature: a positive diagonal peak at around 54000~58000 cm^{-1}, which also dominates the signals of typical amyloid $A\beta_{1–40}$ fibrils. In order to form fibrils, protein sequences with high Rosetta energy should be shuffled. An additional positive diagonal peak at around 54000~58000 cm^{-1} in their 2DUV xxxy spectra can serve as a indicator of the ability of protein sequences to form amyloid-like fibrils.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Tycko R. Curr. Opin. Struct. Biol. 2004; 14:96. [PubMed: 15102455]
2. Caughey B, Lansbury PT. Annu. Rev. Neurosci. 2003; 26:267. [PubMed: 12704221]
3. Lester-Coll NE, Riviera EJ, Soscia SJ, Doiron K, Wands JR, de la Monte SM. J Alzheimers Dis. 2006; 9:13. [PubMed: 16627931]

4. Rishton GM. Nat. Chem. Biol. 2008; 4:159. [PubMed: 18277973]
5. Feng BY, Toyama BH, Wille H, Colby DW, Collins SR, May BCH, Prusiner SB, Weissman J, Shoichet BK. Nat. Chem. Biol. 2008; 4:197. [PubMed: 18223646]
6. Chiti F, Dobson CM. Nat. Chem. Biol. 2009; 5:15. [PubMed: 19088715]
7. Fawzi NL, Yap EH, Okabe Y, Kohlstedt KL, Brown SP, Head-Gordon T. Acc. Chem. Res. 2008; 41:1037. [PubMed: 18646868]
8. Shankar GM, Li S, Mehta TH, G.-M A, Shepardson NE, Smith I, Brett FM, Farrell MA, Rowan MJ, Lemere CA, Regan CM, Walsh DM, Sabatini BL, Selkoe DJ. Nat. Med. 2008; 14:837. [PubMed: 18568035]
9. Sunde M, Blake CCF. Q. Rev. Biophys. 1998; 31:1. [PubMed: 9717197]
10. Dobson CM. Nature. 2003; 426:884. [PubMed: 14685248]
11. Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JWH, Rousseau F. Nat. Meth. 2010; 7:237.
12. Goldschmidta L, Tenga PK, Riekb R, Eisenberga D. Proc. Natl. Acad. Sci. USA. 2010; 107:3487. [PubMed: 20133726]
13. Thompson MJ. Proc Natl Acad Sci USA. 2006; 103:4074. [PubMed: 16537487]
14. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Nat Biotechnol. 2004; 22:1302. [PubMed: 15361882]
15. Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M. J Mol Biol. 2008; 380:425. [PubMed: 18514226]
16. Trovato A, Chiti F, Maritan A, Seno F. PLoS Comput Biol. 2006; 2
17. Chung HS, Ganim Z, Jones KC, Tokmakoff A. Proc. Natl. Acad. Sci. USA. 2007; 104:14237. [PubMed: 17551015]
18. Kim YS, Liu L, Axelsen PH, Hochstrasser RM. Proc. Natl. Acad. Sci. USA. 2008; 105:7720. [PubMed: 18499799]
19. Zhuang W, Hayashi T, Mukamel S. Agnew Chem. Int. Ed. 2009; 48:3750.
20. Mukamel S, Tanimura Y, Hamm P. Accounts of Chemical Research. 2009; 42:1207. [PubMed: 19754111]
21. Shim SH, Gupta R, Ling YL, Strasfeld DB, Raleigh DP, Zanni MT. Proc. Natl. Acad. Sci. USA. 2009; 106:6614. [PubMed: 19346479]
22. Strasfeld DB, Ling YL, Shim SH, Zanni MT. J. Am. Chem. Soc. 2008; 130:6698. [PubMed: 18459774]
23. Zhuang W, Abramavicius D, Voronine DV, Mukamel S. Proc. Natl. Acad. Sci. USA. 2007; 104:14233. [PubMed: 17675411]
24. Jiang J, Abramavicius D, Falvo C, Bulheller BM, Hirst JD, Mukamel S. J. Phys. Chem. B. 2010 In press.
25. Zanni MT, Ge NH, Kim YS, Hochstrasser RM. Proc. Nat. Acad. Sci. 2001; 98:11265. [PubMed: 11562493]
26. Zhuang W, Sgourakis NG, Li ZY, Garcia AE, Mukamel S. Proc. Nat. Acad. Sci. 2010; 107:15687. [PubMed: 20798063]
27. Corkum, P.; Silvestri, SD.; Nelson, KA.; Riedle, E.; Schoenlein, RW. Ultrafast Phenomena XVI. Springer; 2009.
28. Oskouei AA, Bram O, Cannizzo A, van Mourik F, Tortschanoff A, Chergui M. J. Mol. Liq. 2008; 141:118.
29. Oskouei AA, Bram O, Cannizzo A, van Mourik F, Tortschanoff A, Chergui M. Chem. Phys. 2008; 350:104.
30. Tseng C, Matsika S, Weinacht TC. Opt. Express. 2009; 17:18788. [PubMed: 20372612]
31. Beutler M, Ghotbi M, Noack F, Brida D, Manzoni C, Cerullo G. Opt. Lett. 2009; 34:710. [PubMed: 19282907]
32. Abramavicius D, Jiang J, Bulheller BM, Hirst JD, Mukamel S. J. Am. Chem. Soc. 2010; 132
33. Kuhlman B, Baker D. Proc. Natl. Acad. Sci. USA. 2000; 97:10383. [PubMed: 10984534]
34. Liu BKY. Nucleic Acids Research. 2006; 34:W235. [PubMed: 16845000]

35. Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J. Biochemistry. 2010; 49:2987. [PubMed: 20235548]
36. Jiang J, Abramavicius D, Bulheller BM, Hirst JD, Mukamel S. J. Phys. Chem. B. 2010; 114
37. Mikhonin A, Bykov S, Myshakina N, Asher SA. J. Phys. Chem. B. 2006; 110:1928. [PubMed: 16471764]
38. Lednev IK, Shashilov VA, Xu M. Current Science. 2009; 97:180.

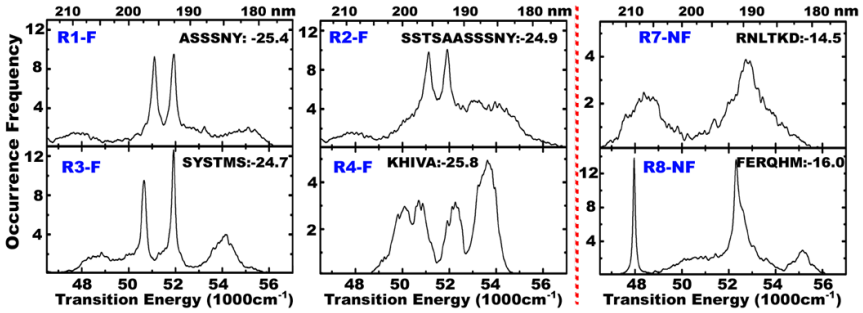

Figure 1.

Histograms of the $\pi \rightarrow \pi^*$ transition energies of eight RNase A segments as indicated. The residue sequences and Rosetta energies (in kcal/mol) are given in each panel.

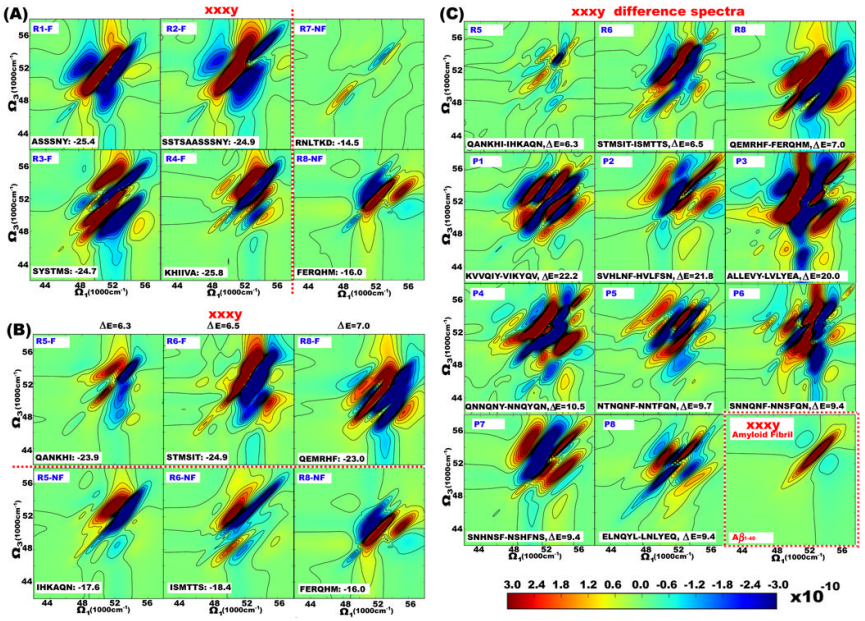

Figure 2.

(A): Simulated chiral 2DUV xxxy spectra for the same systems of Fig. 1. (B): Simulated chiral 2DUV xxxy pectra for the RNase A segments R5-F, R6-F, R8-F, (top) and the corresponding shuffled sequences R5-NF, R6-NF, R8-NF, (bottom). (C): 2DUV xxxy difference spectra (lower energy – higher energy) between the shuffled sequences with lower energies (and fibril forming) and higher energies for three RNase A segments (R5, R6, R8) and eight other protein segments (P1 to P8). The residue sequences and energy differences (in kcal/mol) are given at the bottom of each spectra graph. Bottom right panel: 2DUV xxxy spectroscopy of a typical amyloid fibril $A\beta_{1-40}$. 2DUV magnitude scale bar is plotted at the right bottom.