THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Exploring the Aggregation Propensity of γS-Crystallin Protein Variants Using Two-Dimensional Spectroscopic Tools

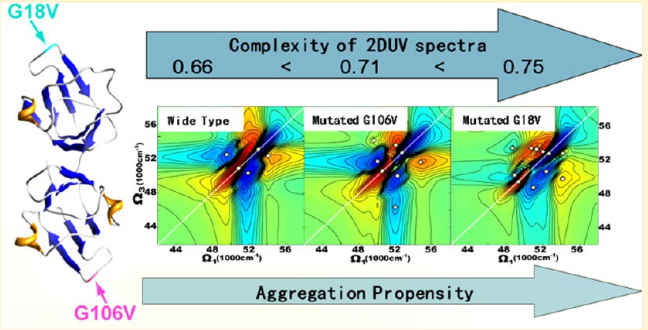
Jun Jiang,*[,†,‡] Kory J. Golchert,[‡] Carolyn N. Kingsley,[‡] William D. Brubaker,[§] Rachel W. Martin,*[,‡,§] and Shaul Mukamel*[,‡]

[†]Department of Chemical Physics, University of Science and Technology of China, Hefei, China

[‡]Chemistry Department and [§]Department of Molecular Biology and Biochemistry, University of California Irvine, Irvine, California, United States

Supporting Information

ABSTRACT: The formation of amyloid fibrils is associated with many serious diseases as well as diverse biological functions. Despite the importance of these aggregates, predicting the aggregation propensity of a particular sequence is a major challenge. We report a joint 2D nuclear magnetic resonance (NMR) and ultraviolet (2DUV) study of fibrillization in the wild-type and two aggregation-prone mutants of the eye lens protein γS-crystallin. Simulations show that the complexity of 2DUV signals as measured by their "approximate entropy" is a good indicator for the conformational entropy and in turn is strongly correlated with its aggregation propensity. These findings are in agreement with high-resolution NMR experiments and are corroborated for amyloid fibrils. The 2DUV technique is complementary to high-resolution structural methods and has the potential to make the evaluation of the aggregation propensity for protein variant propensity of protein structure more accessible to both theory and experiment. The approximate entropy of experimental 2DUV signals can be used for fast screening, enabling identification of variants with high fibrillization propensity for the much more time-consuming NMR structural studies, potentially expediting the characterization of protein variants associated with cataract and other protein aggregation diseases.



1. INTRODUCTION

Understanding the contributions of protein sequence and structure to fibril formation is critical for revealing why some of these protein aggregates are implicated in human diseases[1] and others play functional roles in organisms ranging from bacteria to humans.[2] Common structures and fibril formation kinetics have been observed in proteins with diverse sequences, functions, and native secondary structures,[3,4] suggesting the existence of a common aggregation mechanism, although specific side-chain packing was found to be critical for the formation of fibrils in several short peptides.[5] There is surprising diversity among the structures solved so far.[6−8] It has been conjectured that the fibrillization propensity of the protein depends strongly on its Rosetta free energy,[9] which in turn relies on the conformational entropy, which varies with its residue composition and sequence. Recently, several computational algorithms were developed to evaluate the protein aggregation propensity of particular amino acid motifs,[10−12] and their success demonstrated the relationship between the protein sequence space and aggregation ability. Despite the predictive power of the Rosetta free energy and the other sequence-based algorithms, their practical applicability for evaluating peptide sequences is limited by the fact that they are not based on an experimental observable. These findings underscore the importance of not only solving more structures but also developing methods for rapid measurement of conformational entropy and identification of peptide sequences with high fibrillization propensity.

Here, we report a joint 2D nuclear magnetic resonance (NMR) and two-dimensional ultraviolet (2DUV) study of protein aggregation propensity and its connection to structure. NMR has emerged as an important method for elucidating the molecular-scale organization of large fibrillar protein aggregates,[6,13−15] but even with recent advances, each structure determination requires a significant effort. Coherent ultrafast two-dimensional (2D) correlation spectroscopy is widely used in the infrared and visible regimes for investigating the structure and dynamics in biomolecules.[16,17] Extension to the UV is possible with existing laser technology. 2DUV signals are uniquely capable of producing a multidimensional projection of electronic couplings in proteins, revealing fine detailed interactions among residues. Correlated plots of 1D NMR and UV spectra have been used[18] to identify various chemical

Received: August 9, 2013
Revised: October 16, 2013

   dx.doi.org/10.1021/jp408000k | J. Phys. Chem. B XXXX, XXX, XXX−XXX

species in a mixture, illustrating the complementarity of the methods.

We focus on the wild-type (WT) and two aggregation-prone variants of γS-crystallin, a major structural protein of the human eye lens. Crystallin aggregation causes cataract or opacification of the lens. The G18V variant leads to hereditary, early onset cataract formation,[19] while the engineered variant G106V has an analogous mutation in a symmetry-related position in the C-terminal domain. The WT has a predominantly β-sheet structure consisting of two double Greek key domains, where the residues G18 and G106 are located on surface-exposed loops in the N-terminal and C-terminal domains, respectively (see the Supporting Information). Even though the two variants show only minor changes in the local sequence and structure, they have dramatic aggregation propensity differences.[20] The aggregation propensity varies as WT < G106V < G18V. The disease-related G18V is more prone to aggregation than G106V even though its folded state is more stable (unfolding at 66 °C compared to 59 °C for G106V and 72 °C for WT). Dynamic light scattering measurements indicate that G18V is not monomeric over the entire temperature range studied (down to 16 °C), while G106V and WT begin to aggregate at 37 and 49 °C, respectively.[20]

2. METHODS

A. Experimental Spectroscopic Measurements. Uniformly $^{13}C^{15}N$-labeled G106V-γS was expressed and purified similarly to WT and G18V as previously described.[21] The purified G106V protein was dialyzed into 10 mM sodium phosphate, pH 6.9, 0.05% NaN_3. The addition of 50 mM NaCl and 1 mM DTT was made before concentrating the protein. The final NMR sample conditions were as follows: 2 mM G106V γS-crystallin in 10% D_2O, 20 mM TMSP/90% 10 mM sodium phosphate, 50 mM NaCl, 1 mM DTT, pH 6.9, 0.05% NaN_3. NMR experiments were performed at 22 °C on a Varian UnityINOVA system operating at 800 MHz and equipped with a HCN 5 mm triaxis PFG triple-resonance probe. Decoupling of ^{13}C and ^{15}N nuclei was performed by WURST[22] and GARP[23] sequences, respectively. ^1H shifts were referenced to trimethylsilyl propanoic acid (TMSP), and ^{13}C and ^{15}N were referenced indirectly to TMSP. NMR data were processed using NMRPipe[24] and analyzed using Sparky.[25]

Linear absorption (LA) spectra were recorded using a NanoDrop 2000 (Wilmington, DE) spectrophotometer. FUV absorbance measurements are shown between 190 and 260 nm for WT, G18V, and G106V at a concentration of 0.2 mg/mL and near-UV (NUV) absorbance between 260 and 345 nm at a concentration of 1.0 mg/mL. CD data were taken on a Jasco (Easton, MD) J-810 spectrometer. FUV signals were measured between 180 and 260 nm for WT, G18V, and G106V at concentrations of 77, 51, and 89 μg/mL, respectively. Near-UV signals were measured between 260 and 345 nm for each protein at a concentration of 8.1, 6.2, and 14.6 mg/mL. Fluorescence data were recorded using a NanoDrop 3000 (Wilmington, DE) fluorospectrometer.

B. Simulation Protocol. Simulations of 2DUV spectra were carried out using a QM/MM (quantum mechanics/molecular mechanics) computational protocol based on the exciton Hamiltonian with electrostatic fluctuations (EHEF) algorithm.[26,27] MM simulations of proteins in water were carried out with the CHARMM22 force field[28] for the protein and the TIP3P model for water[29] using the software package NAMD.[30] Electronic excitations of isolated amide and aromatic chromophores were treated at a high ab initio level, the complete active space self-consistent field (CASSCF).[31] The surrounding protein amino acids and the solvent were treated by density function theory (DFT).[32] EHEF also provides a convenient interface for QM and MM calculations. The utilization of atomic frames in EHEF enables us to compute Coulomb interactions in all molecular dynamics (MD) snapshots, taking into account environmental fluctuations at the QM level yet avoiding repeated QM calculations. The Frenkel exciton Hamiltonian[33,34] was constructed using the method described in refs 26 and 27. Effective fluctuating QM Hamiltonians were constructed using the matrix method in the DichroCalc core. Nonlinear optical signals of molecular systems and their complexes relevant for the resonant electronic spectroscopy have been calculated with the SPECTRON code.[35]

The present work focuses on two UV regions, (I) the FUV from 190 to 250 nm, as shown in Figure 1A, in which signals

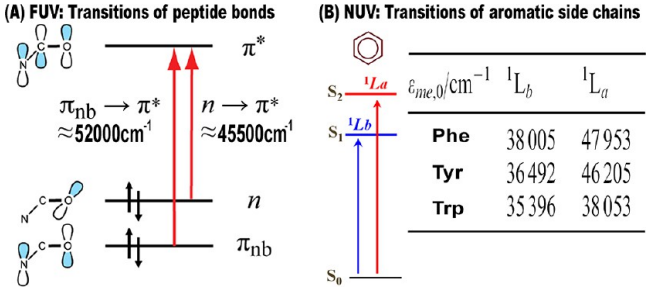


Figure 1. Electronic transitions of protein structural elements that induced FUV (A) and NUV (B) signals.

are induced by protein backbone amide n → π* (~220 nm) and π → π* (~190 nm) electronic transitions and (II) the NUV from 250 to 300 nm, as shown in Figure 1B, where ^1L_b and ^1L_a transitions of aromatic side chains dominate. The full inhomogeneous UV spectra are obtained by averaging over 1000 MD snapshots for LA, circular dichroism (CD), and 2D spectra. LA spectra were normalized to 1, and CD and 2D spectra were given in absolute values.

The 2DUV photon echo signal is generated by four impulsive Gaussian femtosecond laser pulses with central frequency 52000 cm^{-1} (190 nm for FUV signals) or 37000 cm^{-1} (270 nm for NUV signals) and full width at half-maximum (fwhm) of 3754 cm^{-1}, which span the absorption band in the FUV and NUV regions. The pulses (in chronological order) have wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$, and $\mathbf{k}_4$, with $\mathbf{k}_4 = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$. The absorption change of the $\mathbf{k}_4$ beam is recorded as a function of the three consecutive delay times, t_1, t_2, and t_3. 2D signals are calculated by two-dimensional Fourier transform $t_1 \rightarrow \Omega_1$ and $t_3 \rightarrow \Omega_3$ with t_2 varying. Dynamic 2D photon echo snapshots are captured with exciton energy relaxations during time delay t_2, and the static signal is recorded with $t_2 = 0$. When two excitons generated by electronic peptide backbone transitions approach the same amide group, they scatter due to Pauli exclusion, generating the signals. We have used the nonlinear exciton equations (NEE) approach, and the scattering matrix is built as described in ref 35. This method avoids the diagonalization of doubly excited states and can be used in large proteins with hundreds of units. Calculations were performed for both the ordinary nonchiral (*xxxx*) and chirality-induced (*xxxy*) polarization configurations. In the later, the $\mathbf{k}_1$

dx.doi.org/10.1021/jp408000k | J. Phys. Chem. B XXXX, XXX, XXX−XXX

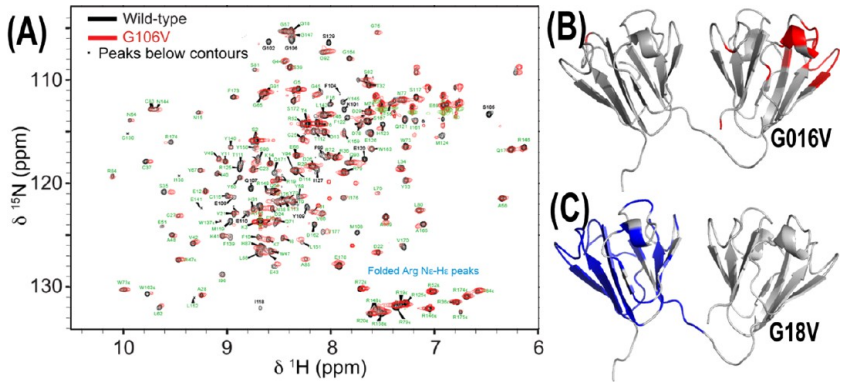


Figure 2. (A) 2D $^{1}H-^{15}N$ HSQC spectra of G106V (red) overlaid with WT (black). The chemical shift assignments (assigned for WT) are labeled to illustrate the extent of chemical shift differences. Residues that do not shift between the WT and mutant protein are labeled in green on each overlay. In the case of G106V, most of residues that shift are located near the mutation site. (B) Residues (red) in G106V for which cross peaks are shifted or absent relative to the WT spectrum. (C) Residues (blue) that have a significant change (>0.1 ppm) in chemical shift in G18V, relative to WT. The shifted residues are plotted on the homology model of WT human gS-crystallin.

Table 1. Structural Data (Packing Density of the Residues 106 and 18 and Average of All Residues) and the Number of Signal Peaks and ApEn of 2DFUV and 2DNUV Signal Data of WT, G106V, and G18V crystallins[a]

	packing density			2DFUV peaks		2DFUV ApEn		2DNUV peaks		2DNUV ApEn	
	R18	R106	all	*xxxx*	*xxxy*	*xxxx*	*xxxy*	*xxxx*	*xxxy*	*xxxx*	*xxxy*
WT	11.32	9.72	12.6	3	6	0.42	0.66	3	7	0.40	0.54
G106V	10.8	8.3	12.5	3	8	0.41	0.71	3	8	0.41	0.58
G18V	10.5	9.3	12.4	3	11	0.42	0.75	3	16	0.41	0.60

[a]ApEn is computed along the scanning line shown in Figure 4A. Data are based on 1000 MD snapshots of each variant.

pulse is polarized along the *y* axis, and the other three pulses are *x*- polarized. 2DUV spectra depend on molecular interactions among peptides and aromatic amino acids and thereby reflect the structural distribution of proteins. Because chiral polarization configurations of the UV pulses are sensitive to the structural chirality, rotation angle, and bulk size of the protein samples, those chiral 2DUV signals could detect long-range interactions between amide groups and always carry more detailed structural information than the nonchiral ones.

2D spectra are plotted using a nonlinear scale that reveals both the strong and weak signals

$$\operatorname{arcsinh}(c\mathbf{S}) = \ln(c\mathbf{S} + \sqrt{1 + c^2\mathbf{S}^2}) \quad (1)$$

The signal **S** is multiplied by a scale factor *c* to make it close to 1, so that weak amplitudes are amplified; for $c\mathbf{S} < 1$, the scale is linear, $\operatorname{arcsinh}(c\mathbf{S}) \approx c\mathbf{S}$, and for larger $c\mathbf{S}$, it becomes logarithmic, $\operatorname{arcsinh}(c\mathbf{S}) \approx (\mathbf{S}|\mathbf{S}|^{-1}) \ln(2|c\mathbf{S}|)$.

3. RESULTS AND DISCUSSION

A. NMR of γS-Crystallin Variants. NMR spectra provide a direct window into the structural and conformational changes associated with protein mutations. Our previous NMR study has shown that for the cataract-related variant G18V, chemical shift changes are widespread throughout the N-terminal domain.[21] Figure 2 shows 2D $^{1}H-^{15}N$ heteronuclear single-quantum (HSQC) spectra of G106V (red) overlaid with WT (black). The spectra show excellent chemical shift dispersion, indicating that each residue has a unique chemical environment, that is, both the WT and variant are folded. If the protein were unfolded, the dispersion would be much weaker because all residues would be exposed to solvent and residues of the same type would have similar chemical shifts. This is also consistent with the CD and UV fluorescence measurements. However, the number of peaks that are shifted, indicating subtle local structural or conformational differences, is dramatically different for G106V than the previous measurements for G18V. Here, G106V is compared with WT; residues labeled in green are identifiable in both spectra, while labels in black correspond to peaks in the WT spectrum that do not have overlaid counterparts in G106V. These include F99, E100, K101, G102, F104, S105, Q107, M108, Y109, E110, I118, E120, I127, and S129. Although broadening or small shifts may be seen elsewhere, large chemical shift changes in G106V are thus localized to the region around the mutated residue. The situation is quite different for G18V; not only do large chemical shift changes occur for almost the entire N-terminal domain, but many residues display two separate sets of peaks at different chemical shifts (assigned separately from 3D data), indicating the presence of two conformations that are slowly interconverting on the NMR time scale.

B. Structures of γS-Crystallin Variants. A homology model of human γS-crystallin was constructed based on the solution-state NMR structure of murine γS-crystallin.[36] MD trajectories for the equilibrium state of each variant generated with MM simulations indicate only subtle secondary structure changes induced by the mutations. In both G18V and G106V, a local opening of the loop around the mutated residue is observed (see the Supporting Information). The decrease in compactness in both variants can be quantified by the computed packing densities (the number of amino acids within a sphere of 9 Å radius of the C-α atom of a residue[37]) of residues 18 and 106 listed in Table 1. Residues 18 in G18V and 106 in G106V have smaller packing densities compared to the corresponding residues in the WT protein. The local conformational entropy (which is inversely related to the packing density[37]) around the mutated residues is therefore

dx.doi.org/10.1021/jp408000k | J. Phys. Chem. B XXXX, XXX, XXX−XXX

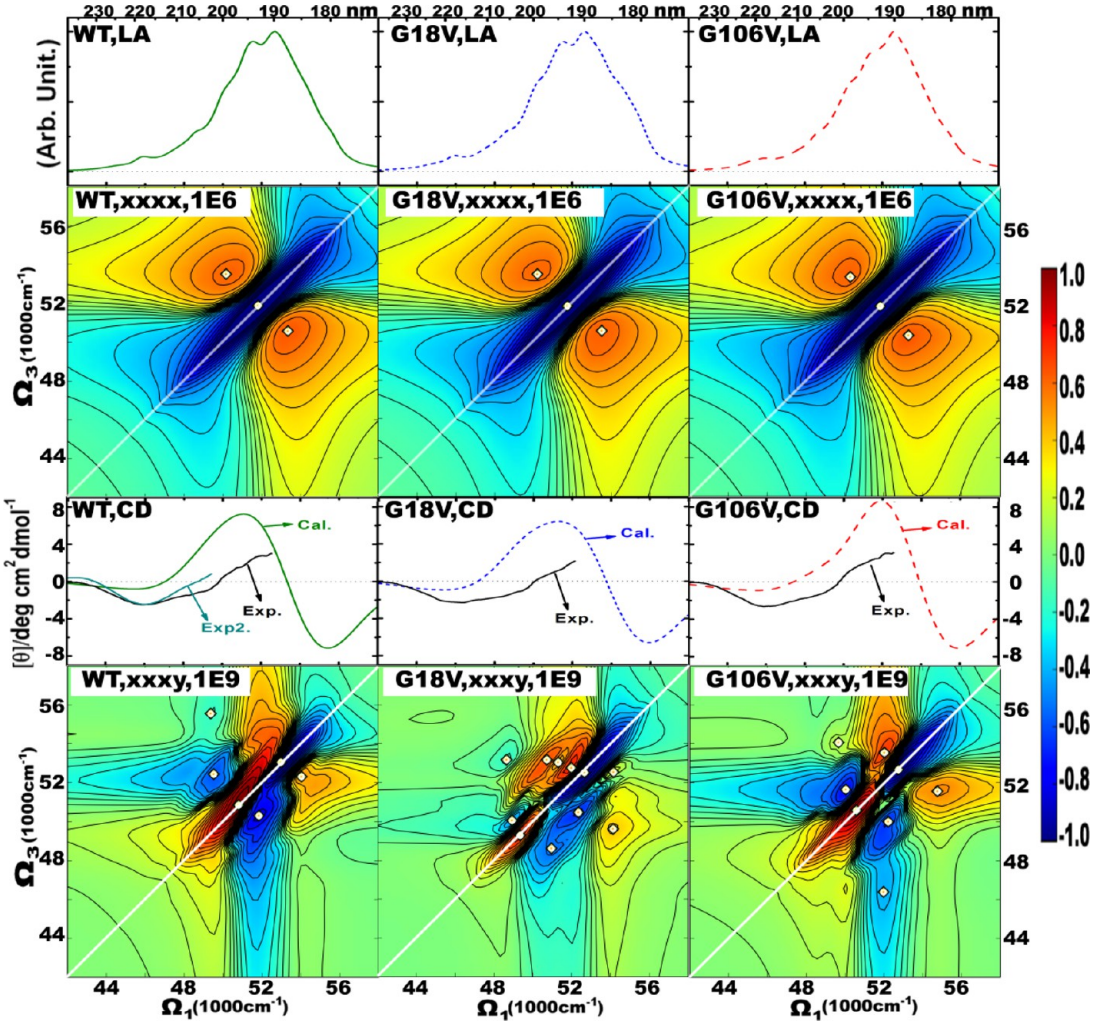


Figure 3. Left column: Spectra of the WT γS-crystallin. From top to bottom: LA, nonchiral (*xxxx*) 2DFUV signal, CD, chiral (*xxxy*) 2DFUV signal. All signals are averaged over 1000 MD snapshots. The black and dark cyan lines in the CD represent present (exp1) and earlier (exp2[19]) experimentally measured spectra, respectively. 2DFUV signals are displayed on a nonlinear scale that interpolates between logarithmic for small values and linear for large values[27] (see the Support Information); the scale factor *c* makes the signal close to one and is given on the top of each panel. A white diagonal line is drawn to highlight the symmetry of the signals, and peaks are marked by white square dots. Middle and right columns: Same as the left column but for the G18V and G106V variants, respectively.

higher than that in WT. This is consistent with the increase of conformational complexity due to mutations found in NMR measurements. Thus, the large aggregation propensity differences of the three variants despite the very small differences in the overall secondary and tertiary structures may be attributed to conformational entropy variations.

C. FUV Absorption and CD Spectra. UV spectra simulations were carried out using a QM/MM computational protocol.[27] We have focused on two spectral regions, (i) the FUV (190−250 nm), where signals are induced by protein backbone electronic transitions and (ii) the NUV (250−300 nm) of aromatic side chains. For the conventional one-dimensional (1D) spectra in the FUV, LA and CD of the three proteins are displayed in the first and third rows of Figure 3. All LA spectra are dominated by a strong ∼190 nm peak. The experimental CD spectra contain typical CD features of a β-sheet, one negative peak between 205 and 240 nm and one positive peak between 185 and 205 nm. These are well-reproduced by the simulations. The LA and CD spectra of the three variants are virtually indistinguishable because their overall secondary structures are almost identical.

D. 2DUV Spectra. We now turn to 2DUV photon echo signals, which can probe subtle differences in protein secondary and tertiary structures that are not resolved by the 1D spectra.[27,38] Signals from randomly oriented ensembles were calculated for two pulse-polarization configurations, nonchiral *xxxx* (all laser pulses have parallel polarization along the *x* axis) and chirality-induced *xxxy* (the last pulse is *y*-axis-polarized). The 2DFUV *xxxx* spectra shown in Figure 3 are dominated by the negative diagonal ∼52000 cm^{-1} peak accompanied by two positive sidebands. Like the LA, the three *xxxx* spectra are very similar. In contrast, the chirality-induced *xxxy* signals displayed at the bottom row of Figure 3 are very sensitive to small changes in structure and flexibility. The three proteins exhibit a similar pattern along the diagonal, a positive peak at ∼50000 cm^{-1} and a negative one at ∼53000 cm^{-1}, together with weaker sidebands. The cross peaks carry rich information about the couplings between chromophores and reveal fine details characteristic of each of the three variants. In the vicinity of (Ω_1 = 50000 cm^{-1}, Ω_3 = 53000 cm^{-1}), WT has no cross peak, G18V shows two cross peaks near the sideband of the negative diagonal peak, and G106V exhibits one weak cross peak in that

dx.doi.org/10.1021/jp408000k | J. Phys. Chem. B XXXX, XXX, XXX−XXX

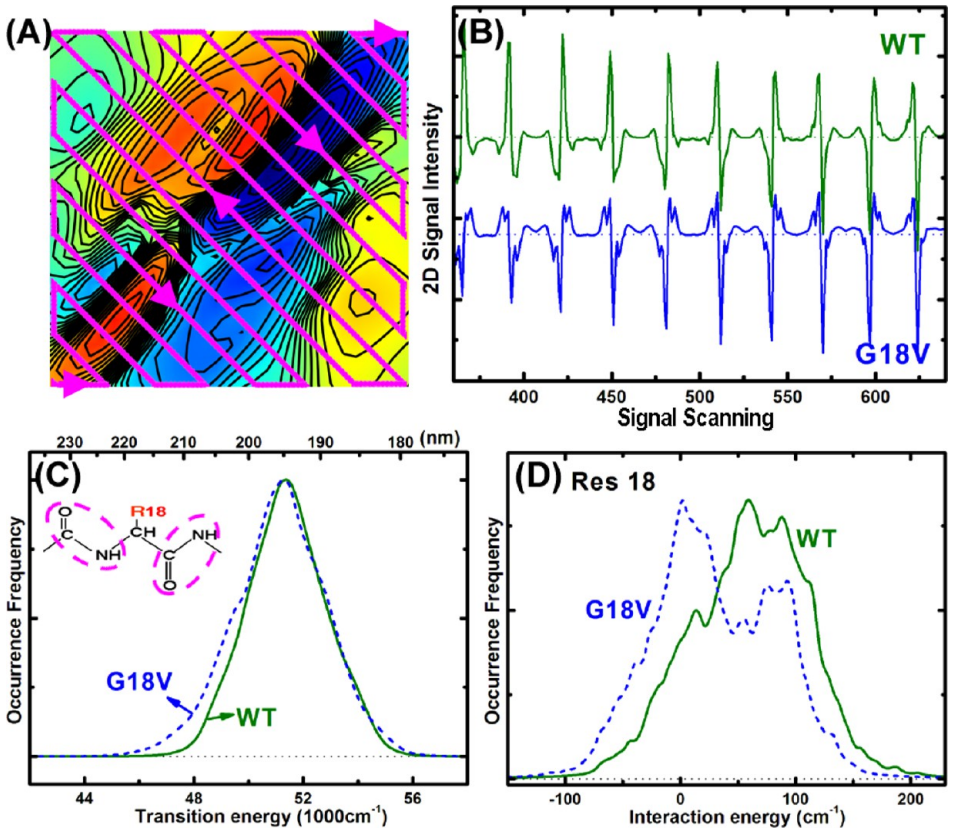


Figure 4. (A) Signal scanning line in the 2DFUV *xxxy* of G18V. (B) 2DFUV *xxxy* signal evolution curves of WT and G18V along the scanning line. (C) Transition energy histograms of two backbone amides linked to residue 18 in WT and G18V in 1000 MD snapshots. (D) Histograms of interactions of residue 18 with the protein backbone in WT and G18V obtained by 1000 MD snapshots.

region. The peaks are marked by white square dots, and their number is listed in Table 1. As we proceed from WT through G106V to G18V, the 2DFUV *xxxy* spectra contain more peaks and become more complicated.

Spectral complexity is a measure of the structural richness of a dynamical system. In Figure 4A, we display a scanning line perpendicular to the diagonal of the 2D contour map, starting from the bottom left (lower energy) to the upper right (higher energy) corner. The projections of the 2DFUV *xxxy* signals of WT and G18V along this line are depicted in Figure 4B. G18V has more peaks than WT, consistent with the NMR data that show more than one conformation for G18V but only one for WT.[21] Preliminary results also indicate that G106V has a single stable conformation at room temperature. We describe the complexity of these projections by the approximate entropy (ApEn),[39] which has been widely used in a variety of contexts to quantify subtle irregularities in biomedical data. For example, the ApEn of electroencephalogram data has been successfully used to diagnose Alzheimer's disease.[40] Table 1 shows that ApEn varies in the order of G18V > G106V > WT for *xxxy* but is almost constant for *xxxx*. This is consistent with the number of 2DFUV peaks; however, ApEn provides a better measure of signal complexity.

The complexity of the 2DUV signals results from chromophore couplings and molecular interactions, which shift and split the electronic transitions. The distributions of the backbone transition frequencies of the two amides linked to residue 18 in WT and G18V calculated using 1000 MD protein configurations are displayed in Figure 4C. The broader distribution for G18V compared to WT is consistent with the more complex coupling pattern. The calculated distributions of energies of residue 18 due to interactions with the protein backbone are displayed in Figure 4D. Again, the G18V distribution has more pronounced structure than WT.

The same conclusion may be reached by examining the 2DNUV spectra of the aromatic residue transitions. γS-Crystallin contains 25 aromatic side chains (9 Phe, 12 Tyr, and 4 Trp), distributed throughout the protein. The computed spectra are displayed in Figure 5. The corresponding numbers of peaks and ApEn values are given in Table. 1. As in the 2DFUV, the 2DNUV chiral (*xxxy*) spectra of G18V have a larger number of peaks, more complicated patterns, and higher ApEn values, maintaining the order G18V > G106V > WT.

E. Aggregation Propensity of γS-Crystallin Variants. The function of many biological systems depends on the formation of ordered lower-entropy structures.[42] The loss of complexity that accompanies amyloid fibril formation was reported in Parkinson's and Alzheimer's diseases.[40,43] We expect a close relationship between the conformational entropy and the aggregation propensity of a protein sequence. A good indicator of aggregation ability recently found by Eisenberg et al. is the Rosetta free energy,[9] which is however hard to access experimentally. Our previous study[38] demonstrated that protein segments with lower Rosetta free energy and higher fibrillization propensity always have more diversified intramolecular interactions and complicated 2DUV spectral patterns.

By examining the conformational entropy, diversity of chromophore couplings and interactions, and ApEn of 2D signals, we conclude that the tendency of the three γS-

dx.doi.org/10.1021/jp408000k | J. Phys. Chem. B XXXX, XXX, XXX−XXX

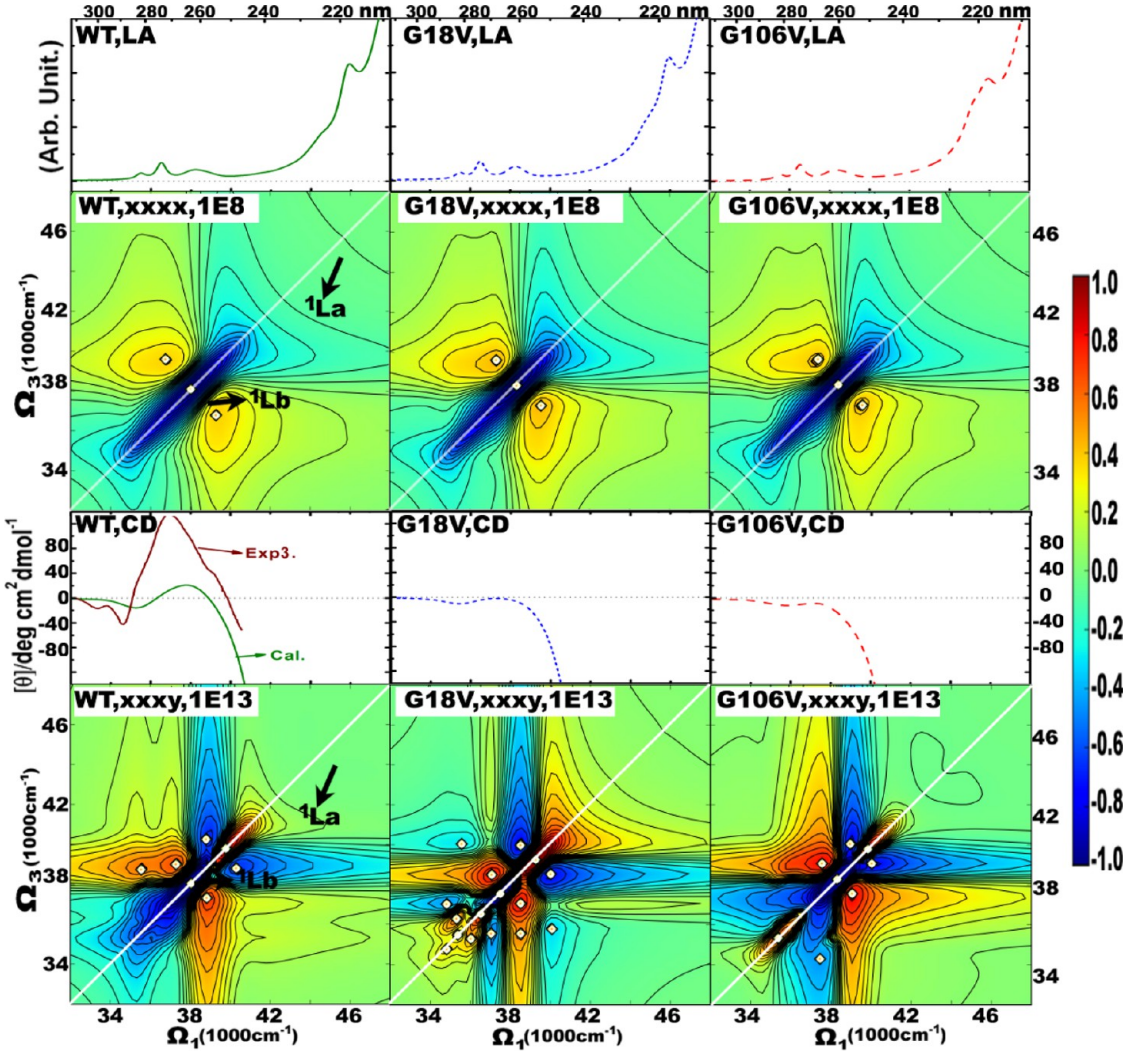


Figure 5. The same as Figure 3, except for the NUV 1D and 2D spectra. The brown curve in the CD represent the reported experimental spectra (exp3).[41]

crystallins to aggregate varies in the order G18V > G106V > WT, consistent with previous experimental measurements.

The present study demonstrates that 2DUV spectra contain characteristic signatures for the WT, G18V, and G106V variants that are consistent with the available experimental data.[20] A strong correlation between 2DUV spectra and experimental biophysical and NMR data is established.

4. SUMMARY AND CONCLUSIONS

The present combined NMR and 2DUV investigation of γS-crystallin variants reveals a close relationship between conformational complexity and aggregation propensity. We found that G18V and G106V mutations only cause minor structural changes to WT γS-crystallin but effectively increase the conformational entropy and complexity of electronic transitions. 2DUV signals capture the changes induced by mutations. Studies on the γS-crystallin variants demonstrate a general rule that the complexity of the 2DUV signals is a good indicator of the ability of the protein to aggregate. It also gives a general implication that the underlying mechanism of amyloid fibril aggregation is the minimization of the system entropy. ApEn of the 2DUV signals used to quantify the complexity of the proteins thus provides a direct measure of the protein aggregation ability, which would promote the quantitative assessments of environmental conditions in aggregation studies. Because aggregation lowers the conformational entropy, we expect that the change of 2DUV spectral complexity can be used to monitor the aggregation kinetics.

Our findings and protocol are not limited to γS-crystallin and should apply to other fibrillization-prone proteins. To illustrate that, we have carried out extensive calculations on three β-amyloid variants involving E22 mutations (see the Supporting Information) and demonstrated that the order of their 2DFUV and 2DNUV spectral complexities is the same as the aggregation propensity order revealed by experiments.[44] We further investigated 12 model protein segments,[38] shuffling the sequence to lower the Rosetta free energy. We found that this resulted in higher 2DUV complexity and hence aggregation propensity (see the Supporting Information). The Rosetta free energy is an established indicator of amyloid fibril formation propensity;[9] however, it is only accessible by theoretical calculations, whereas 2DUV complexity can be measured experimentally. Because the formation and deposition of amyloid fibrils are associated with more than 20 serious diseases and the mechanisms involved are not well understood, predicting the aggregation propensity of proteins is a timely

dx.doi.org/10.1021/jp408000k | J. Phys. Chem. B XXXX, XXX, XXX−XXX

issue of broad scientific interest. The study of amyloid fibrils will greatly benefit from adequate tools that correlate the aggregation ability of proteins with physical or chemical properties accessible by both theory and experiment.

These results suggest a new protocol for choosing protein variants for detailed structural studies. Even though the collection of NMR spectra presented here is straightforward, taking approximately 20 min of experimental time using a standard pulse sequence, it does require preparation of a relatively large amount of ^{15}N-labeled sample, and obtaining the resonance assignments needed to make sense of the chemical shift changes is much more time-consuming and expensive. The three-dimensional sequences needed to make assignments of residues along the protein backbone require ^{13}C−^{15}N labeling and several days of experiment time, even before the constraints needed to calculate the structure are obtained. 2DUV, in contrast, can provide a fast screening method to determine which protein variants have interesting changes in structure or conformation. The complementary information provided by both techniques allows one to identify protein structural details. The 2DUV technique may thus be employed for experimental determination of which protein variants make worthy targets and are selected for more time-consuming NMR structural studies.

■ ASSOCIATED CONTENT

Supporting Information

The calculation of the approximate entropy (ApEn), structure analysis, transition population analysis, 2DFUV and 2DNUV signals of β-amyloid variants, and ApEn of 2DUV correlated to the Rosetta energy. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: jiangj1@ustc.edu.cn (J.J.).
*E-mail: rwmartin@uci.edu (R.W.M.).
*E-mail: smukamel@uci.edu (S.M.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Institutes of Health (Grants GM059230, GM091364, and EY021514), National Science Foundation (Grant CHE-1058791), and National Natural Science Foundation of China (Grant 91221104). We thank Dr. Alfredo Freites and Prof. Doug Tobias for providing MD trajectories and Prof. Yi Luo and Dr. Yu Zhang for helpful discussions.

■ REFERENCES

(1) Lansbury, P. T.; Lashuel, H. A. A century-old debate on protein aggregation and neurodegeneration enters the clinic. *Nature* **2006**, *443*, 774−779.
(2) Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W. Functional amyloid — from bacteria to humans. *Trends Biochem. Sci.* **2007**, *32*, 217−224.
(3) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884−890.
(4) Shim, S.-H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6614−6619.
(5) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **2007**, *447*, 453−457.
(6) Heise, H. Solid-state NMR spectroscopy of amyloid proteins. *ChemBioChem* **2008**, *9*, 179−189.
(7) Tycko, R. Solid state NMR studies of amyloid fibril structure. *Annu. Rev. Phys. Chem.* **2011**, *62*, 279−299.
(8) Greenwald, J.; Riek, R. Biology of amyloid: Structure, function, and regulation. *Structure* **2010**, *18*, 1244−1260.
(9) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 3487−3492.
(10) Maurer-Stroh, S.; Debulpaep, M.; Kuemmerer, N.; de la Paz, M. L.; Martins, I. C.; Reumers, J.; Morris, K. L.; Copland, A.; Serpell, L.; Serrano, L. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Nat. Methods* **2010**, *7*, 237−242.
(11) Thompson, M. J.; Sievers, S. A.; Karanicolas, J.; Ivanova, M. I.; Baker, D.; Eisenberg, D. The 3d profile methods for identifying fibril-forming segments of proteins. *Proc. Nati. Acad. Sci. U.S.A.* **2006**, *103*, 4074−4078.
(12) Trovato, A.; Chiti, F.; Maritan, A.; Seno, F. Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Comput. Biol.* **2006**, *2*, e170.
(13) Paravastu, A. K.; Leapman, R. D.; Yau, W.-M.; Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18349−18354.
(14) Wasmer, C.; Lange, A.; van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. Amyloid fibrils of the HET-s(218−289) prion form a β solenoid with a triangular hydrophobic core. *Science* **2008**, *319*, 1523−1526.
(15) Helmus, J. J.; Surewicz, K.; Nadaud, P. S.; Surewicz, W.; Jaroniec, C. P. Molecular conformation and dynamics of the Y145Stop variant of human prion protein in amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6284−6289.
(16) Brixner, T.; Stenger, J.; Vaswani, H. M.; Cho, M.; Blankenship, R. E.; Fleming, G. R. Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature* **2005**, *434*, 625−628.
(17) Mukamel, S.; Abramavicius, D.; Yang, L.; Zhuang, W.; Schweigert, I. V.; Voronine, D. Coherent multidimensional optical probes for electron correlations and exciton dynamics: From NMR to X-rays. *Acc. Chem. Res.* **2009**, *42*, 553−562.
(18) Koeppe, B.; Tolstoy, P. M.; Guo, J.; Nibbering, E. T. J.; Elsaesser, T. Two-dimensional UV−vis/NMR correlation spectroscopy: A heterospectral signal assignment of hydrogen-bonded complexes. *J. Phys. Chem. Lett.* **2011**, *2*, 1106−1110.
(19) Ma, Z.; Piszczek, G.; Wingfield, P. T.; Sergeev, Y. V.; Hejtmancik, J. F. The G18V CRYGS mutation associated with human cataracts increases γS-crystallin sensitivity to thermal and chemical stress. *Biochemistry* **2009**, *48*, 7334−7341.
(20) Brubaker, W. D.; Freites, J. A.; Golchert, K. J.; Shapiro, R. A.; Morikis, V.; Tobias, D. J.; Martin, R. W. Separating instability from aggregation propensity in γS-crystallin variants. *Biophys. J.* **2011**, *100*, 498−506.
(21) Brubaker, W. D.; Martin, R. W. ^{1}H, ^{13}C, and ^{15}N assignments of wild-type human γs-crystallin and its cataract-related variant γs-g18v. *Biomol. NMR Assignments* **2011**, DOI: 10.1007/s12104-011-9326-1.
(22) Kupce, E.; Freeman, R. Adiabatic pulses for wide-band inversion and broad-band decoupling. *J. Magn. Reson., Ser. A* **1995**, *115*, 273−276.
(23) Shaka, A.; Barker, P. B.; Freeman, R. Computer-optimized decoupling scheme for wideband applications and low level operation. *J. Magn. Reson.* **1985**, *64*, 547−552.
(24) Delaglio, F.; Grezesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.
(25) Goddard, T. D.; Kneller, D. G. *Sparky* 3; University of California: San Francisco, CA, 2011.

(26) Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. Ultraviolet spectroscopy of protein backbone transitions in aqueous solution: Combined QM and MM simulations. *J. Phys. Chem. B* **2010**, *114*, 8270−8277.
(27) 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.
(28) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(30) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(31) Yamamoto, N.; Vreven, T.; Robb, M.; Frisch, M.; Schlegel, H. A direct derivative mc-scf procedure. *Chem. Phys. Lett.* **1996**, *250*, 373−378.
(32) Becke, A. D. Density-functional thermochemistry. 3. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(33) Frenkel, Y. On the transformation of light into heat in solids. I. *J. Phys. Rev.* **1931**, *37*, 17−44.
(34) Abramavicius, D.; Palmieri, B.; Mukamel, S. Extracting single and two-exciton couplings in photosynthetic complexes by coherent two-dimensional electronic spectra. *Chem. Phys.* **2009**, *357*, 79−84.
(35) Abramavicius, D.; Palmieri, B.; Voronine, D. V.; Šanda, F.; Mukamel, S. Coherent multidimensional optical spectroscopy of excitons in molecular aggregates; quasiparticle versus supermolecule perspectives. *Chem. Rev.* **2009**, *109*, 2350−2408.
(36) Wu, Z.; Delaglio, F.; Wyatt, K.; Wistow, G.; Bax, A. Solution structure of γs-crystallin by molecular fragment replacement NMR. *Protein Sci.* **2005**, *14*, 3101−3114.
(37) Liao, H.; Yeh, W.; Chiang, D.; Jernigan, R.; Lustig, B. Protein sequence entropy is closely related to packing density and hydrophobicity. *Protein Eng., Des. Sel.* **2005**, *18*, 59−64.
(38) Jiang, J.; Mukamel, S. Two dimensional ultraviolet (2DUV) spectroscopic tools for identifying fibrillation propensity of protein residue sequences. *Agnew. Chem., Int. Ed.* **2010**, *49*, 9666−9669.
(39) Pincus, S. M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2297−2301.
(40) Escudero, J.; Abasolo, D.; Hornero, R.; Espino, P.; Lopez, M. Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. *Physiol. Meas.* **2006**, *27*, 1091−1106.
(41) Mandal, K.; Chakrabarti, B.; Thomson, J.; Siezen, R. J. Structure and stability of γ-Crystallins. *J. Biol. Chem.* **1987**, *262*, 8096−8102.
(42) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448*, 325−329.
(43) Vaillancourt, D. E.; Newell, K. M. The dynamics of resting and postural tremor in Parkinson's disease. *Clin. Neurophysiol.* **2000**, *111*, 2046−2056.
(44) Betts, V.; Leissring, M. A.; Dolios, G.; Wang, R.; Selkoe, D. J.; Walsh, D. M. Aggregation and catabolism of disease-associated intra-Aβ mutations: Reduced proteolysis of Aβa21g by neprilysin. *Neurobiol. Dis.* **2008**, *31*, 442−450.