Article

pubs.acs.org/Biomac

Fibrillation Mechanism of a Model Intrinsically Disordered Protein Revealed by 2D Correlation Deep UV Resonance Raman Spectroscopy

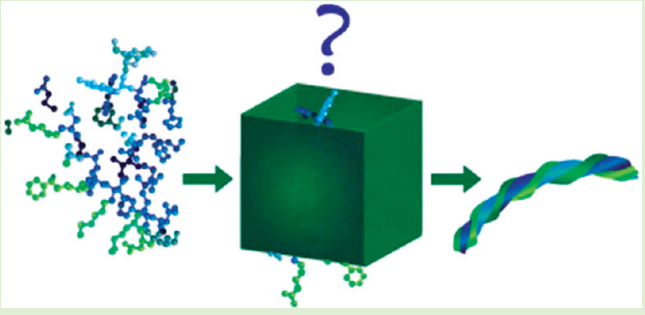
Vitali Sikirzhytski,[†] Natalya I. Topilina,[†] Gaius A. Takor,[†] Seiichiro Higashiya,[†] John T. Welch,[†] Vladimir N. Uversky,[‡,§] and Igor K. Lednev*[,†]

[†]Department of Chemistry, University at Albany, SUNY, 1400 Washington Avenue, Albany, New York 12222, United States

[‡]Department of Molecular Medicine and Byrd Alzheimer's Research Institute, University of South Florida, Tampa, Florida 33612, United States

[§]Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

ABSTRACT: Understanding of numerous biological functions of intrinsically disordered proteins (IDPs) is of significant interest to modern life science research. A large variety of serious debilitating diseases are associated with the malfunction of IDPs including neurodegenerative disorders and systemic amyloidosis. Here we report on the molecular mechanism of amyloid fibrillation of a model IDP (YE8) using 2D correlation deep UV resonance Raman spectroscopy. YE8 is a genetically engineered polypeptide, which is completely unordered at neutral pH yet exhibits all properties of a fibrillogenic protein at low pH. The very first step of the fibrillation process involves structural rearrangements of YE8 at the global structure level without the detectable appearance of secondary structural elements. The formation of β-sheet species follows the global structural changes and proceeds via the simultaneous formation of turns and β-strands. The kinetic mechanism revealed is an important new contribution to understanding of the general fibrillation mechanism proposed for IDP.



INTRODUCTION

Using the large de novo, genetically engineered polypeptide $GH_6[(GA)_3GY(GA)_3GE]_8GAH_6$ (YE8) as a model of an intrinsically disordered protein, it was possible to demonstrate that global structural changes absent the appearance of secondary structural features can be central to fibrillation processes. Intrinsically disordered proteins (IDPs), which lack stable secondary or unique tertiary structure under physiological conditions,[1−4] play a crucial role in a large variety of human diseases ranging from neurodegenerative disorders to systemic amyloidosis[5−8] as well as in a variety of other diseases.[9] Transitions of α-synuclein, amyloid β peptide, tau-protein, prion protein, huntingtin protein with polyQ expansion, islet amyloid polypeptide, or atrial natriuretic factor, among others, from soluble, natively unfolded forms into insoluble plaques consisting of β-sheet-rich amyloid fibrils are associated with pathological evidence of Parkinson's disease, Alzheimer's disease, spongiform encephalopathies, Huntington's disease, type II diabetes, or atrial amyloidosis.[5−8] Whereas the precise molecular mechanisms of amyloid fibrillation remain elusive,[10−16] the process typically begins with the formation of a monomeric amyloidogenic conformation, followed by nucleation (formation of specific aggregation-prone oligomers) and propagation to form proto-fibrils and proto-filaments, and finally mature fibrils.[17−22] Protein fibrillation is highly dependent on the initial protein structure (folded or unfolded), amino acid sequence, concentration, pH, or environmental conditions.[22−27] After reviewing data for the amyloidogeneses of more than 20 IDPs, both related and unrelated to human disease, we found that in contrast with compact globular proteins that require partial unfolding prior to the subsequent structural rearrangements and aggregation eventually leading to the amyloid fibril formation, partial folding is an obligatory prerequisite for the initiation of IDP amyloidogenesis.[7,8,28] Our work addresses one of the key issues in the amyloid fibrillation of IDP where structural rearrangements are not constrained by the initial conformation, which is the transformation of a polypeptide at global and secondary structural levels in the early stages of fibrillation. Recently, it was demonstrated that the de novo, genetically engineered polypeptide $GH_6[(GA)_3GY(GA)_3GE]_8GAH_6$ (YE8) is intrinsically disordered and exhibits all of the properties of a typical fibrillogenic protein. YE8 is mostly disordered at neutral pH yet forms amyloid-like fibrils at low pH, thereby providing an excellent model system for the study of the mechanism of IDP fibrillation.[14,29−31]

Received: February 5, 2012
Revised: March 28, 2012
Published: April 19, 2012

Deep UV resonance Raman (DUVRR) spectroscopy combined with advanced statistics is a powerful tool for structural characterization of proteins in all stages of fibrillation.[29,32−39] Two-dimensional correlation spectroscopy (2DCoS) is one of the most powerful and versatile analytical methods for the analysis of various types of dynamic Raman data.[34,37,40] Together, 2DCoS and DUVRR spectroscopy were used to probe the kinetics of YE8 fibrillation. The sequence of events involved global structural rearrangements, followed by the development of a fibrillar β-sheet where the turns and β-strands formed simultaneously. The kinetic scheme established provides new insights into the fibrillation mechanism of an IDP.

EXPERIMENTAL SECTION

Materials and Methods. *Analysis of the Intrinsic Disorder Propensity of YE8.* Predictions of intrinsic disorder in YE8 were performed using a set of predictors from the PONDR family, PONDR VLXT, VL3, VSL2 predictors, a PONDR-FIT metapredictor and SPINE-D predictor. Access to PONDR VLXT and VL3 was provided by Molecular Kinetics, Inc. (http://www.pondr.com). PONDR (Predictor Of Natural Disordered Regions) is a set of neural network predictors of disordered regions on the basis of local amino acid composition, flexibility, hydropathy, and other factors. These predictors classify each residue within a sequence as either ordered or disordered. PONDR VLXT integrates three feed-forward neural networks: the Variously characterized Long, version 1 (VL1) predictor, which predicts nonterminal residues,[41] and the X-ray characterized N- and C- terminal predictors (XT), which predict terminal residues.[42] Output for the VL1 predictor starts and ends 11 amino acids from the termini. The XT predictors output provides predictions up to 14 amino acids from their respective ends. A simple average is taken for the overlapping predictions, and a sliding window of 9 amino acids is used to smooth the prediction values along the length of the sequence. Unsmoothed prediction values from the XT predictors are used for the first and last four sequence positions.

PONDR VL3 predictor is a feed-forward neural network that was trained on regions of 152 long regions of disorder that were characterized by various methods. The set of ordered proteins consisted of 290 PDB-Select-25 chains having no disordered residues. This predictor is based on 20 attributes (18 amino acid frequencies, average flexibility, and sequence complexity) in an input window of length 41. The raw predictions are averaged over an output window of length 31 to obtain the final prediction for a given position. The putative boundaries between order and disorder were corrected using the order/disorder boundary predictor. The closest maximum prediction from the boundary predictor (above 0.8) became the new boundary between the ordered and disordered regions.[43]

PONDR VSL2 predictor was trained on variously characterized, short and long disordered regions. Here two specialized predictors were first built and optimized for short (≤30 residues) and long disordered regions (>30 residues), respectively. A meta-predictor was then trained to integrate the specialized predictors into the final predictor model.[44] PONDR-FIT is a meta-predictor[45] that combines six individual predictors, which are PONDR VL-XT,[41] VSL2,[44] VL3,[43] FoldIndex,[46] IUPred,[47] and TopIDP.[48] This meta-predictor is moderately more accurate than each of the component predictors. In addition to PONDR family members, disorder propensity of YE8 was further evaluated by SPINE-D predictor,[49] which is a single neural-network-based technique that makes a three-state prediction

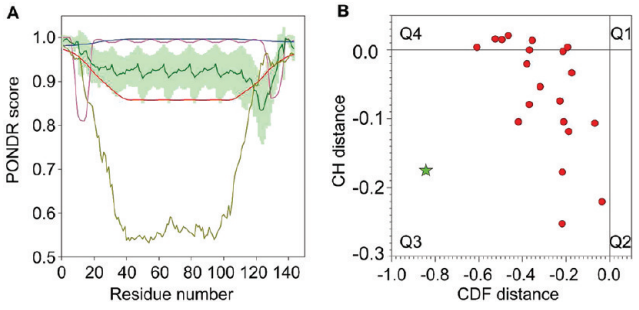


Figure 1. (A) Intrinsic disorder propensity of YE8 as evaluated by PONDR VLXT (pink line), VL3 (red line), VSL2 (blue line), PONDR-FIT (green line), and SPINE-D (dark yellow line). Light-green shadows represent standard errors of disorder prediction by PONDR-FIT. Note: all curves are located above the 0.5 threshold. (B) CH-CDF plot for a number of amyloidogenic proteins associated with human neurodegenerative diseases (red circles) and the artificial IDP YE8 (green star). Note: all proteins are located within the quadrants Q3 and Q4 that correspond to mostly disordered proteins.

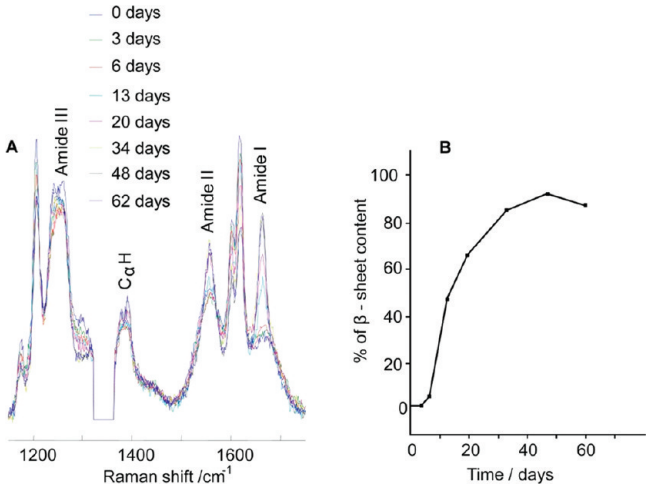


Figure 2. Kinetic study of YE8 polypeptide fibrillation. (A) Experimental DUVRR spectra of 44 μM YE8 polypeptide incubated for various times. Amide I mode consists of carbonyl C═O stretching, with smaller contribution from C−N stretching and N−H bending modes.[71,72] Both amide II and amide III bands involve significant C−N stretching, N−H stretching, and C−C stretching. The CαH bending vibrational mode involves Cα-H symmetric bending and C−Cα stretching.[72] (B) Kinetics of YE8 folding in terms of the amount of β-sheet formed for the sample with concentration of 44 μM.

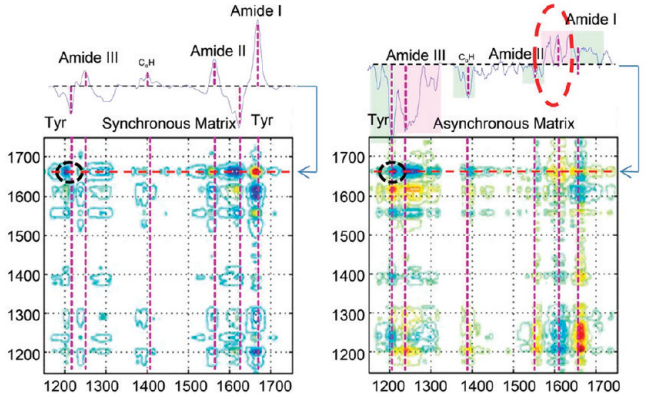


Figure 3. Synchronous (left panel) and asynchronous (right panel) 2D correlation maps of Raman spectra recorded during YE8 polypeptide fibrillation. Spectral profiles are cross sections of 2D correlation maps at $\nu 2$ = 1664 cm^{-1} (red horizontal line). Pink vertical lines and labels indicate the position of amide I, amide III, amide III, CαH, and tyrosine Raman bands (colored ovals). Shading was used to distinguish Raman bands of tyrosine and protein back band on the cross-section of asynchronous 2D correlation.

dx.doi.org/10.1021/bm300193f | Biomacromolecules 2012, 13, 1503−1509

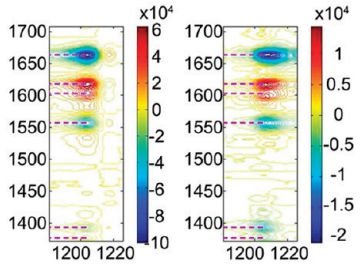


Figure 4. Synchronous (left panel) and asynchronous (right panel) 2D correlation maps of the tyrosine ν7a region of YE8 polypeptide DUVRR spectra. The same sign of corresponding synchronous and asynchronous correlation features indicates that the change in CαH (1393 cm^{-1}), amide II (1558 cm^{-1}), and amide I (1664 cm^{-1}) bands is delayed relative to the change in tyrosine ν7a Raman band.

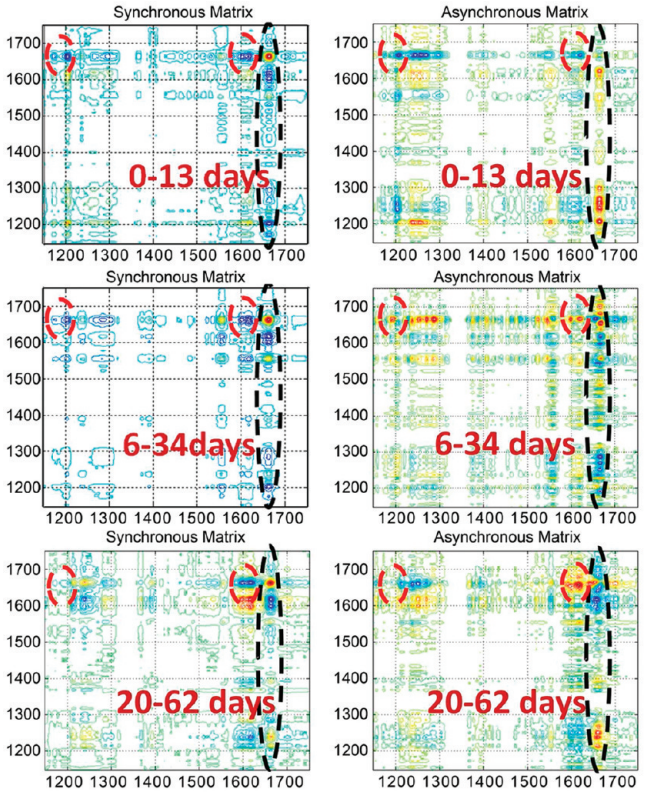


Figure 5. Moving window 2DCoS results in identical behavior of all tyrosine Raman bands with respect to the correlation with amide I band and indicates a complex kinetic mechanism of global and secondary structural rearrangements. Synchronous (left panels) and asynchronous (right panels) 2D correlation maps of YE8 polypeptide DUVRR spectra. Tyrosine peaks at 1175, 1207, 1602, and 1620 cm^{-1} (colored ovals) change first in the earlier stages of fibrillation, whereas β-sheet (amide I) development is accelerated during the fibril growth phase.

(ordered residues and disordered residues in short and long disordered regions) first and reduces it into a two-state prediction afterward.[49]

CH-CDF Analysis of YE8 and Amyloidogenic Proteins Involved in the Neurodegenerative Diseases. In this analysis, the coordinates of each spot are calculated as a distance of the corresponding protein in the CH-plot (charge-hydropathy plot)[3,50] from the boundary (Y-coordinate) and an average distance of the respective cumulative distribution function (CDF) curve from the CDF boundary (X-coordinate).[50−52]

Preparation and Characterization of the YE8 Polypeptide. The design, synthesis, and folding of the polypeptide YE8 consisting of 8

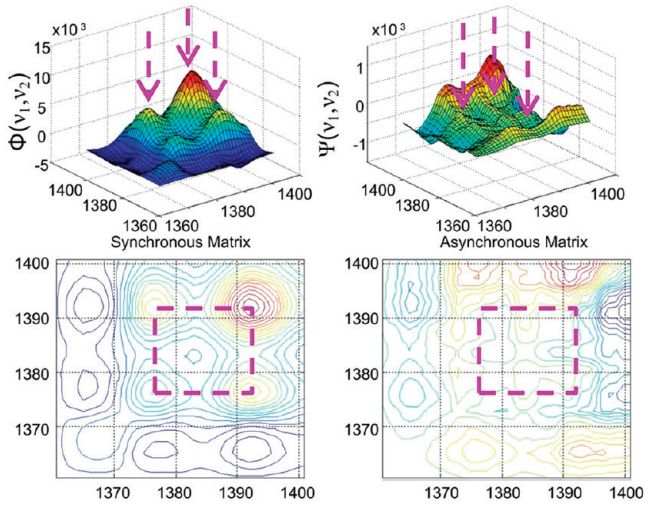
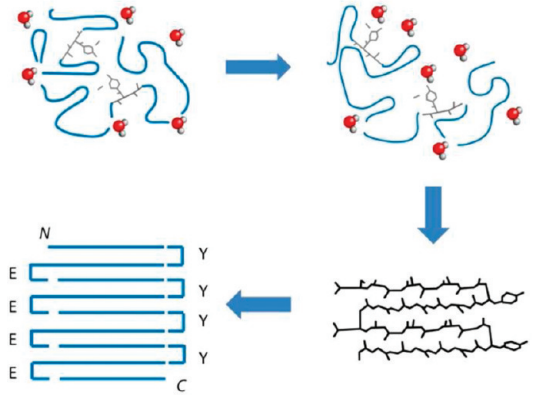


Figure 6. Synchronous (left panel) and asynchronous (right panel) 2D correlation maps of the CαH region of YE8 polypeptide DUVRR spectra. The absence of an asynchronous cross-peak at (1377, 1392 cm^{-1}) is an indicative of simultaneous appearance of 1377 and 1392 cm^{-1} CαH Raman peaks.

Scheme 1. Fibrillation Mechanism of Initially Unordered Polypeptide YE8[a]



[a]First, global structural rearrangements result in change in tyrosine local environment, followed by simultaneous formation of turns and β-strands. Further aggregation leads to the growth of proto-filaments, proto-fibrils, and finally mature fibrils.

repeats of a 16 amino acid monomer have been described elsewhere.[14,29,30] GAGAGA repeats form β-strands with the alternating turn groups tyrosine (Y) and glutamic acid (E) decorating the edges of the antiparallel β-sheet. After expression, purification, dialysis against doubly distilled water, and centrifugation at 15 000g for 45 min, deep UV Raman and CD spectroscopy were used to verify disordered structure of the polypeptide at neutral pH. The fibrillation of a 44 μM solution YE8 polypeptide at pH 3.5 and room temperature was monitored by deep UV Raman spectroscopy[14] with a home-built Raman instrument with deep UV excitation capability that has been described elsewhere.[53] A 197 nm laser beam (~1 mW) was focused into a spinning Suprasil NMR tube containing 150 μL of solution where the scattered radiation was collected in a backscattering geometry. GRAMS/AI 7.01 software was used for Raman data treatment.

Raman Spectra Preprocessing. Raman spectra were calibrated using Teflon as an external standard. Base line correction was performed using multipoint linear approximation (GRAMS/AI 7.01). The data set was preprocessed with a third-order Savitzky−Golay

Scheme 2. Modified Fibrillation Mechanism of Initially Unordered Polypeptide

Initial ensemble of conformations → altered ensemble of conformations (global structural changes) →

formation of β-sheet (including the nucleation) → fibrillar structure

Scheme 3. General Fibrillation Mechanism of Initially Unordered Polypeptide

Intrinsically unordered protein or polypeptide → partially folded conformation → fibrillar structure

smoothing (window width of 15) to reduce the noise.[54] Spectral intensities around a buffer line at ~1350 cm^{-1} were forced to zero.

2DCoS Analysis. 2DCoS[55−58] allows for decomposing and monitoring the evolution of overlapping bands in the Raman spectra of proteins into individual components representing various secondary structure elements or aromatic amino acid residues. Correlation among spectral regions permits the correct assignment of vibrational bands to protein and peptide structural motifs. 2DCoS can provide insight into the sequence of protein conformational changes, thus making it possible to elucidate the nature of structural transitions and the kinetic reaction mechanism. Synchronous $\Phi(\nu_1, \nu_2)$ and asynchronous $\Psi(\nu_1, \nu_2)$ 2D correlation maps represent the real and imaginary parts, respectively, of the cross-correlation function calculated using the following equation

$$\Phi(\nu_1, \nu_2) + i\Psi(\nu_1, \nu_2) = \frac{1}{\pi(T_{min} - T_{max})} \int_0^\infty \tilde{Y}_{\nu 1}(w) \cdot \tilde{Y}^*_{\nu 2}(w)\, dw \quad (1)$$

The term $\tilde{Y}_{\nu 1}(w)$ is the forward Fourier transform of the observed spectral changes at a spectral variable $\nu 1$ with respect to the external variable T (time in our case). The term $\tilde{Y}_{\nu 2}(w)$ is the conjugate of the Fourier transform of the observed spectral changes at a spectral variable ν_2. The relationship between variables ν_1 and ν_2 can be described as: if $\Phi(\nu_1, \nu_2) \bullet \Psi(\nu_1, \nu_2) > 0$, then the spectral changes at ν_1 occur before that at ν_2; if $\Phi(\nu_1, \nu_2) \bullet \Psi(\nu_1, \nu_2) < 0$, then the spectral changes order is the reverse.

Results and Discussion. *Analysis of the Intrinsic Disorder Propensity of YE8.* YE8, a β-sheet forming polypeptide consisting of 16 repeats,[29,30] was designed to have a set of identical, weakly interacting, very simple six amino acid β-strands GAGAGA, which are similar to the β-strand motif found in *Bombyx mori* silk and in other de novo designed polypeptides.[59,60] Previous studies showed that YE8 was mostly disordered at neutral pH yet formed amyloid fibrils at acidic pH.[29,30] Because besides disorder-promoting residues A, G, and E this artificial polypeptide contains a noticeable number of order-promoting residues Y and H, we decided to evaluate the intrinsic propensity of YE8 for being disordered by several publicly available disorder predictors, such as members of the PONDR family VLXT, VL3, VSL2, a meta-predictor PONDR-FIT, and a newly developed rather accurate SPINE-D predictor. The choice of these predictors is determined by their characteristic features. For example, PONDR VLXT is very sensitive to local sequence peculiarities and therefore can be used for identifying functionally important sites within the disordered regions.[61,62] PONDR VL3 is better for proteins that are experimentally known to be 100% disordered or possess long disordered regions, whereas PONDR VSL2 is statistically better for proteins containing both structured and disordered regions. A meta-predictor PONDR-FIT is statistically not different from PONDR VL3 for fully disordered and fully structured proteins and slightly better than PONDR VSL2 when both structured and disordered regions are present. Finally, SPINE-D is comparable to a meta-predictor in predicting disordered residues in long disordered regions and superior in short disordered regions.

Figure 1A represents the results of this analysis and shows that for all predictors tested the corresponding disorder score curves are located well above the 0.5 threshold supporting the predominantly disordered nature of this model protein. At the next step, we compared the disorder predisposition of YE8 artificial protein with the disorderedness of proteins known to be involved in various human neurodegenerative diseases, such as Aβ; tau protein; prion protein; α-, β-, and γ-synucleins; huntingtin; DRPLA protein (atrophin-1); androgen receptor; ataxin-1; ataxin-2; ataxin-3; P/Q-type calcium channel α1A subunit; ataxin-7; TATA-box-binding protein; ABri; glial fibrillary acidic protein; mitochondrial DNA polymerase γ; DNA excision repair protein ERCC-6; and survival motor neuron protein, which previously were shown to contain a significant amount of intrinsic disorder.[7,8] To this end, the CH-CDF plot analysis was utilized. This method is based on the methodological difference between the two binary disorder predictors (i.e., predictors that evaluate the predisposition of a given protein to be ordered or disordered as a whole), CH-plot[3,50] and CDF.[50−52] In fact, the CH-plot is a linear classifier that takes into account only two parameters of the particular sequence (charge and hydropathy), whereas CDF analysis is dependent on the output of the PONDR predictor, a nonlinear classifier, which was trained to distinguish order and disorder based on a significantly larger feature space. According to these methodological differences, CH-plot analysis, being applied to the relatively short proteins, is predisposed to discriminate proteins with substantial amount of extended disorder (extended IDPs) from proteins with compact conformations (collapsed IDPs and ordered globular proteins). PONDR-based CDF analysis may discriminate all disordered conformations, including collapsed disordered species, from ordered globular proteins. Therefore, this discrepancy in the disorder prediction by CDF and CH-plot provides a computational tool to discriminate proteins with extended disorder from collapsed IDPs. Positive and negative Y values in corresponding plot correspond to proteins predicted by CH-plot analysis to be extended or compact, respectively. On the contrary, positive and negative X values are attributed to proteins predicted by the CDF analysis to be ordered or intrinsically disordered, respectively. Therefore, the resultant quadrants of CDF-CH phase space correspond to the following expectations: Q1, proteins predicted to be extended by CH-plots but ordered by CDFs; Q2, ordered proteins; Q3, proteins predicted to be disordered by CDFs but compact by CH-plots (i.e., putative collapsed IDPs); and Q4, proteins predicted to be disordered by both methods.

Figure 1B shows that the proteins analyzed here known to be involved in various human neurodegenerative diseases are expected to be mostly disordered. Furthermore, the majority of these proteins are grouped within the Q3 quadrant, suggesting that Aβ, prion protein, α- and γ-synucleins, huntingtin, androgen receptor, ataxin-1, ataxin-3, P/Q-type calcium channel α1A subunit, TATA-box-binding protein, ABri, mitochondrial DNA polymerase γ, DNA excision repair protein ERCC-6, and survival motor neuron protein are expected to belong to the subclass of the collapsed disordered proteins. Tau protein, β-synuclein, DRPLA protein (atrophin-1), ataxin-2, ataxin-7, and glial fibrillary acidic protein, being located within the Q4 quadrant, are expected to belong to the subclass of extended IDPs. Because these six proteins are located close to the quadrant boundary, they are also expected to have more collapsed structure than typical random coil-like polypeptide.

Figure 1B illustrates also that although YE8 is located in the Q3 quadrant it is rather distant from the other amyloidogenic proteins, suggesting that this polypeptide should possess rather specific properties, where the relative high compaction degree is combined with the high disorder degree. In other words, according to this analysis YE8 is expected to be rather collapsed but be mostly devoid of the developed secondary structure. These data agree well with the

experimental characterization of YE8H, which revealed that at neutral pH this protein is highly disordered and with the results of the secondary structure prediction by a number of computational tools available at the ExPASy Proteomics Tool portal (http://expasy.org/tools/, data not shown). Computational tools used at this stage include CFSSP (http://www.biogem.org/tool/chou-fasman/), GOR4 (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), JRED3 (http://www.compbio.dundee.ac.uk/www-jpred/), JUFO (http://www.meilerlab.org/index.php/servers), PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/), and PredicProtein (http://www.predictprotein.org/). Furthermore, according to the JPRED3 analysis (http://www.compbio.dundee.ac.uk/www-jpred/), YE8 is expected to have 48 buried residues with less than 25% solvent accessibility, suggesting its relatively compact state.

Analysis of the YE8 Fibrillation by DUVRR Spectroscopy. The Raman spectra of the YE8 polypeptide (Figure 2A) are dominated by pronounced tyrosine and amide bands, which report on the polypeptide global and secondary structures, respectively.[53,63−67] Intense narrow amide I and amide II Raman bands are indicative of fibrillar β-sheet secondary structure.[31,65] The percentages of β-sheet content estimated by multivariate curve resolution-alternating least-squares[68] in the Raman spectra of YE8 polypeptide at various time points are shown in Figure 2B. Amide III and C(C_αH) bending mode are also sensitive to the amide backbone conformation.[67,69,70] In particular, C(C_αH) bending band (~1370−1400 cm^{-1}) is weak for peptides in the helical form and strong in β-sheet (~1395−1400 cm^{-1}) and random coil (~1380−1385 cm^{-1}).[37,66]

Amide I and amide II bands increase with YE8 polypeptide incubation time, indicating the transformation of unordered structure into amyloid β-sheet form. The C(C_αH) band slightly increases and splits into 1377 and 1392 cm^{-1} subbands with time. The C(C_αH) band (1392 cm^{-1}) is characteristic of a β-sheet, whereas lower band (1377 cm^{-1}) is typical of unordered structure.[66] The appearance of 1377 cm^{-1} Raman band was assigned to turn formation and was confirmed by 2DCoS (see below). All tyrosine peaks (Figure 2A) decrease during the fibrillation, reflecting the change in tyrosine local environment.

Analysis of the YE8 Fibrillation by the 2D Correlation DUVRR. A two-component approach describes the simplest model of initially unordered polypeptide fibrillation, a direct transition from unordered to amyloid-like β-sheet structure. The application of 2DCoS revealed the higher levels of complexity of the process and established a sequential order of events.

Figure 3 shows synchronous (left panel) and asynchronous (right panel) 2D correlation maps calculated using DUVRR spectra of 44 μM YE8 fibrillation. Figure 4 shows enlarged synchronous and asynchronous 2D-Raman correlation maps for the tyrosine ν_{7a} (1208 cm^{-1}) spectral region. Here the noise level is at least two orders of magnitude lower than the level of the signal. Synchronous map demonstrates peaks centered at (1208, 1377 cm^{-1}), (1208, 1393 cm^{-1}), (1208, 1558 cm^{-1}), and (1208, 1664 cm^{-1}), indicating an increase in the intensity of C_αH, amide II, and amide I Raman bands while tyrosine ν_{7a} is decreasing (Figure 2). Corresponding correlation features on the asynchronous map are also negative. According to Noda's rules, the same sign of corresponding synchronous and asynchronous correlation features indicate that the change in amide Raman bands is delayed relative to the change in tyrosine ν_{7a} Raman band.[53] Consequently, the very first step of the fibrillation process involves global structural rearrangements of YE8 polypeptide resulting in a change in the tyrosine local environment. The formation of β-sheet is preceded by the global structural rearrangements. It is worth noting that the clear picture of the early stages of the fibrillation process can be provided only by Moving Window 2D correlation analysis.

Moving Window 2D Correlation Spectroscopy (Figure 5) was used to resolve the complex behavior of correlation features between tyrosine and amide Raman bands (Figure 3). In particular, while 1208 cm^{-1} tyrosine Raman band shows negative correlation feature with amide I band on the asynchronous map, 1602 and 1620 cm^{-1} tyrosine Raman bands exhibit positive correlation with amide I band. It was hypothesized that this correlation was indicative of changes in tyrosine local environment and that YE8 structural rearrangements correlate differently in different stages of fibrillation. To test this hypothesis, we utilized the moving window approach. Figure 5 demonstrates 2D correlation maps calculated for three characteristic stages of fibrillation including initiation lag phase (0−13 days), elongation or fibril growth phase (6−34 days), and plateau phase (20−62 days). This simple approach eliminates the ambiguity in the asynchronous map interpretation. All tyrosine Raman bands show the same correlation with amide I band for a single window, although the behavior varies from window to window. Specifically, β-sheet (amide I) development is delayed in the earlier stages of fibrillation and then accelerates relative to the tyrosine environment change reported by Raman peaks at 1175, 1207, 1602, and 1620 cm^{-1}.

As a whole, 2D correlation analysis of the selected data subsets reinforced the conclusions based on the entire Raman data set, providing deeper insight into the mechanism of YE8 fibrillation. Synchronous and asynchronous maps obtained for individual time windows are quite noisy, reflecting the small number of spectra used. Figure 5 (upper panels) demonstrates that tyrosine (1208 cm^{-1}) behavior changes in earlier stages of the process, before secondary structure modification indicated by amide III (1225−1300 cm^{-1}). The amide III region is resolved on asynchronous map presenting characteristic features at 1237, 1257, 1284, and 1300 cm^{-1}. Changes in the amide I and amide II spectral regions are also delayed relative to tyrosine peaks changes at 1605 and 1620 cm^{-1}.

Synchronous and asynchronous maps calculated using Raman spectra acquired during the elongation phase of fibrillation (6−34 days) demonstrate the reverse order of events: perturbation of the secondary structure precedes changes in the tyrosine environment. The inverted order is indicative of YE8 polypeptide transition into fibrils, rapid formation of β-sheet secondary structure, and relatively slower changes in tyrosine local environment. Final stages of the YE8 fibrillation are characterized by noticeable changes at 1237 (β-sheet), 1605, and 1620 cm^{-1} (tyrosine) as well as at the amide I and amide II regions (Figure 5, bottom panels).

The complicated initial 2DCoS maps can be explained by the formation of intermediates. However, Joint diagonalization Jade,[73] second-order blind identification (SOBI) in the Fourier space data,[74] and second-order nonstationary source separation (SEONS),[75] which have been successfully used to extract pure spectra of lysozyme, partially unfolded intermediate, and nucleus,[37] are not able to separate nucleus and fibrillar signatures. These results are consistent with our previous studies that showed high similarity of DUVRR spectroscopic signatures of the completely fibrillated protein and precursor nucleus.[37]

Simultaneous Formation of Turns and β-Strands. Two-dimensional correlation analysis of C_αH spectral region provided further insight into YE8 polypeptide fibrillation process (Figure 6). It is worth noting that the local map helps to show spectral changes within the CαH region (Figure 6) and avoid interference from relatively larger spectral changes in the intense amide peaks (Figure 3). The wide C_αH band centered at 1385 cm^{-1} characteristic for unordered structure splits into two narrower 1377 and 1392 cm^{-1} Raman bands. The positive (1377, 1392 cm^{-1}) synchronous cross peak reflects the increase in these Raman bands with incubation time. The lack of a corresponding peak on the asynchronous 2D correlation map suggests that these two peaks experience completely synchronous changes. The appearance of a narrow Raman band between 1370 and 1400 cm^{-1}, a spectral region typically associated with unordered conformations (wide peak at 1385 cm^{-1}), is a quite remarkable. In contrast, the β-sheet conformation is characterized by relatively narrow C_αH peaks at ~1395−1400 cm^{-1}. α-Helix has a very low Raman cross-section and does not contribute to the observed changes.[32,35] Consequently, the 1377 cm^{-1} Raman band can be assigned tentatively to the turns required for formation of the β-sheet structure.[39] This assignment is also consistent with the Raman signature of turns obtained for YEHK (ref 39, figure 2),[39] a polypeptide similar in sequence to YE8 that has a wide C_αH peak centered near 1375 cm^{-1}. The 1360−1410 cm^{-1} spectral region in the Raman spectrum of completely fibrillated YE8

was fitted with two Gaussian functions following linear baseline subtraction. The ratio of the resultant Gaussian peak areas S_{1377}/S_{1392} = 0.56 is consistent with the 3/5 ratio of the peptide bonds composing a turn (three peptide bonds per single turn) and β-sheet (five peptide bonds per single β-stand) motifs. The formation of β-strands and turns is highly correlated and occurs concurrently. This finding is consistent with the published kinetics of β-hairpin folding obtained by monitoring the relaxation times of tryptophan fluorescence upon nanosecond laser temperature jumps.[76] Munoz et al. demonstrated that folding of the hairpin involves the formation of a turn and β-strands on the microsecond time scale at room temperature.[76]

Kinetic Mechanism of FY8 Fibrillation. The existence of a certain structure in unordered polypeptide YE8 is consistent with recent studies of intrinsically disordered proteins and polypeptides, which show that many proteins or protein parts, lacking a unique, stable 3D structure in solution, represent a functional dynamic ensemble of conformations.[7,8,77] All of our results are consistent with the kinetic mechanism of fibrillation shown in Scheme 1. Taking into consideration the conformational ensemble property of initially unordered polypeptide, this mechanism could be further modified.

Here nucleus means formation of the species necessary for fibril formation to be thermodynamically favorable.[78] This scheme emphasizes the presence of global structural transformations without detectable secondary structure formation in the very early stages of fibrillation. Although it is yet to be established whether the discovered phenomenon is typical for all or the majority of IDPs, this development may be an important addition to the general mechanism of fibrillation proposed for IDP.[8]

Scheme 2 accentuates the difference between two distinct levels of polypeptide chain organization, ensembles of global and secondary structures. We demonstrated that structural changes at these levels are not necessarily completely correlated and can occur asynchronously. Nucleation in Scheme 3 can be associated with the partially folded conformation. An altered ensemble of conformations leads to an intermediate state between the initial and partially folded states being formed as a result of global structural rearrangements. The residual structure of YE8 in water and the initiation of YE8 fibrillation at global level of organization with a lack of any secondary structure are important and novel observations. During the course of fibrillation, exposure of tyrosine residues to water continuously increases, suggesting that the partial folding of YE8 is probably not driven by hydrophobic collapse. In contrast, specific intermolecular interactions are the likely driving forces of YE8 fibrillation.

CONCLUSIONS

The fibrillation mechanism of initially unordered polypeptide has been investigated using 2D correlation deep UV resonance Raman spectroscopy. The very first step of the fibrillation process was demonstrated to involve global structural rearrangements of YE8 polypeptide manifested by the change in tyrosine local environment with no detectable appearance of secondary structural elements. The formation of β-sheet followed the global structural changes and proceeded via a simultaneous formation of turns and β-strands. The described mechanism affords new insights into the general mechanism of IDP fibrillation.[8]

Genetic engineering offers an opportunity to model the behavior of specific classes of globular proteins. YE8 polypeptide is an excellent example simulating fibrillation of intrinsically disordered proteins. One important aspect of YE8 fibrillation is not yet fully understood. The strong dependence of YE8 fibrillation on the polypeptide concentration[29,30] indicates that intermolecular interaction is a driving force for fibrillation.[30] The question that remains is whether this intermolecular interaction initiates the discovered global structural rearrangements of YE8 in early stage of fibrillation. This investigation is in progress in our laboratory.

AUTHOR INFORMATION

Corresponding Author

*Phone: 518-591-8863. Fax: 518-442-3462. E-mail: ilednev@albany.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Victor Shashilov, Ludmila Popova, and Aliaksandra Sikirzhytskaya for helpful discussions. We also thank Aliaksandra Sikirzhytskaya and Joseph DeRubertis for assistance in manuscript preparation. This material is based on work supported by the National Science Foundation under grant no. CHE-0809525 and CHE-1152752 (I.K.L.). This work was also partially supported by the Program of the Russian Academy of Sciences for the "Molecular and Cellular Biology" (to V.N.U.).

REFERENCES

(1) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527−533.
(2) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35−41.
(3) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* **2000**, *41*, 415−427.
(4) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.
(5) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(6) Dobson, C. M. *Protein Pept. Lett.* **2006**, *13*, 219−227.
(7) Uversky, V. N. *Front Biosci.* **2009**, *14*, 5188−5238.
(8) Uversky, V. N. *Curr. Alzheimer Res.* **2008**, *5*, 260−287.
(9) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.
(10) Pedersen, J. S. *J. Diabetes Sci. Technol.* **2010**, *4*, 1357−1367.
(11) Stefani, M. *FEBS J.* **2010**, *277*, 4602−4613.
(12) Straub, J. E.; Thirumalai, D. *Curr Opin Struct Biol* **2010**, *20*, 187−195.
(13) Surguchev, A.; Surguchov, A. *Brain Res. Bull.* **2010**, *81*, 12−24.
(14) Topilina, N. I.; Sikirzhytski, V.; Higashiya, S.; Ermolenkov, V. V.; Welch, J. T.; Lednev, I. K. In *Instrumental Analysis of Intrinsically Disordered Proteins*; Uversky, V. N., Longhi, S., Eds.; John Wiley & Sons: Hoboken, NJ, 2010; pp 253−302.
(15) Receveur-Brechot, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. *Proteins* **2006**, *62*, 24−45.
(16) Abedini, A.; Raleigh, D. P. *Protein Eng., Des. Sel.* **2009**, *22*, 453−459.
(17) Dobson, C. M. *Nature* **2003**, *426*, 884−890.
(18) Ohnishi, S.; Takano, K. *Cell. Mol. Life Sci.* **2004**, *61*, 511−524.
(19) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16764−16769.
(20) Ferrone, F. *Methods Enzymol.* **1999**, *309*, 256−274.
(21) Librizzi, F.; Rischel, C. *Protein Sci.* **2005**, *14*, 3129−3134.
(22) Bhak, G.; Choe, Y. J.; Paik, S. R. *BMB Rep* **2009**, *42*, 541−551.
(23) Stefani, M.; Dobson, C. M. *J. Mol. Med. (Heidelberg, Ger.)* **2003**, *81*, 678−699.
(24) Pedersen, J. S.; Andersen, C. B.; Otzen, D. E. *FEBS J.* **2010**, *277*, 4591−4601.
(25) Pedersen, J. S.; Dikov, D.; Flink, J. L.; Hjuler, H. A.; Christiansen, G.; Otzen, D. E. *J. Mol. Biol.* **2006**, *355*, 501−523.
(26) Saunders, H. M.; Bottomley, S. P. *Protein Eng., Des. Sel.* **2009**, *22*, 447−451.
(27) Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2001**, *40*, 6036−6046.
(28) Uversky, V. N.; Fink, A. L. *Biochim. Biophys. Acta* **2004**, *1698*, 131−153.

(29) Topilina, N. I.; Sikirzhytsky, V.; Higashiya, S.; Ermolenkov, V. V.; Lednev, I. K.; Welch, J. T. *Biomacromolecules* **2010**, *11*, 1721−1726.
(30) Topilina, N. I.; Ermolenkov, V. V.; Higashiya, S.; Welch, J. T.; Lednev, I. K. *Biopolymers* **2007**, *86*, 261−264.
(31) Lednev, I. K.; Ermolenkov, V. V.; Higashiya, S.; Popova, L. A.; Topilina, N. I.; Welch, J. T. *Biophys. J.* **2006**, *91*, 3805−3818.
(32) Shashilov, V. A.; Lednev, I. K. *Chem Rev* **2010**, *110*, 5692−5713.
(33) Shashilov, V. A.; Sikirzhytski, V.; Popova, L. A.; Lednev, I. K. *Methods* **2010**, *52*, 23−37.
(34) Shashilov, V. A.; Lednev, I. K. *J. Am. Chem. Soc.* **2008**, *130*, 309−317.
(35) Xu, M.; Shashilov, V.; Lednev, I. K. *J. Am. Chem. Soc.* **2007**, *129*, 11002−11003.
(36) Xu, M.; Shashilov, V. A.; Ermolenkov, V. V.; Fredriksen, L.; Zagorevski, D.; Lednev, I. K. *Protein Sci.* **2007**, *16*, 815−832.
(37) Shashilov, V.; Xu, M.; Ermolenkov, V. V.; Fredriksen, L.; Lednev, I. K. *J. Am. Chem. Soc.* **2007**, *129*, 6972−6973.
(38) Shashilov, V. A.; Ermolenkov, V. V.; Lednev, I. K. *Inorg. Chem.* **2006**, *45*, 3606−3612.
(39) Sikirzhytski, V.; Topilina, N. I.; Higashiya, S.; Welch, J. T.; Lednev, I. K. *J. Am. Chem. Soc.* **2008**, *130*, 5852−5853.
(40) Noda, I. *J. Mol. Struct.* **2010**, *974*, 3−24.
(41) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42*, 38−48.
(42) Li, X.; Romero, P.; Rani, M.; Dunker, A. K.; Obradovic, Z. *Genome Inf. Ser.* **1999**, *10*, 30−40.
(43) Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. *J. Bioinf. Comput. Biol.* **2005**, *3*, 35−60.
(44) Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A. K.; Obradovic, Z. *BMC Bioinf.* **2006**, *7*, 208.
(45) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. *Biochim. Biophys. Acta* **2010**, *1804*, 996−1010.
(46) Prilusky, J.; Felder, C. E.; Zeev-Ben-Mordehai, T.; Rydberg, E. H.; Man, O.; Beckmann, J. S.; Silman, I.; Sussman, J. L. *Bioinformatics* **2005**, *21*, 3435−3438.
(47) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *Bioinformatics* **2005**, *21*, 3433−3434.
(48) Campen, A.; Williams, R. M.; Brown, C. J.; Meng, J.; Uversky, V. N.; Dunker, A. K. *Protein Pept. Lett.* **2008**, *15*, 956−963.
(49) Zhang, T.; Faraggi, E.; Xue, B.; Dunker, A. K.; Uversky, V. N.; Zhou, Y. *J. Biomol. Struct. Dynam.* **2012**, *29*, 799−813.
(50) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44*, 1989−2000.
(51) Mohan, A.; Sullivan, W. J., Jr; Radivojac, P.; Dunker, A. K.; Uversky, V. N. *Mol. Biosyst.* **2008**, *4*, 328−340.
(52) Xue, B.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. *FEBS Lett.* **2009**, *583*, 1469−1474.
(53) Lednev, I. K.; Ermolenkov, V. V.; He, W.; Xu, M. *Anal. Bioanal. Chem.* **2005**, *381*, 431−437.
(54) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627−1639.
(55) Noda, I. In *Handbook of Vibrational Spectroscopy*; Chalmers, J., Griffiths, P., Eds.; John Wiley & Sons, Ltd: New York, 2002; Vol. *3*, pp 2123−2134.
(56) Noda, I. *Anal. Sci.* **2007**, *23*, 139−146.
(57) Noda, I.; Ozaki, Y. *Two-Dimensional Correlation Spectroscopy - Applications in Vibrational and Optical Spectroscopy*; John Wiley and Sons: Chichester, U.K., 2004.
(58) *Two-Dimensional Correlation Spectroscopy: Kobe-Sanda, Japan 29 August−1 September 1999*; Ozaki, Y., Noda, I., Eds.; AIP Conference Proceedings *503*; American Institute of Physics: Melville, NY, 2000.
(59) Parkhe, A. D.; Cooper, S. J.; Atkins, E. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Int. J. Biol. Macromol.* **1998**, *23*, 251−258.
(60) Langer, R.; Tirrell, D. A. *Nature* **2004**, *428*, 487−492.
(61) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44*, 12454−12470.
(62) Cheng, Y.; Oldfield, C. J.; Meng, J.; Romero, P.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2007**, *46*, 13468−13477.
(63) Chi, Z.; Asher, S. A. *J. Phys. Chem. B* **1998**, *102*, 9595−9602.
(64) Hildebrandt, P. G.; Copeland, R. A.; Spiro, T. G.; Otlewski, J.; Laskowski, M., Jr; Prendergast, F. G. *Biochemistry* **1988**, *27*, 5426−5433.
(65) Xu, M.; Ermolenkov, V. V.; He, W.; Uversky, V. N.; Fredriksen, L.; Lednev, I. K. *Biopolymers* **2005**, *79*, 58−61.
(66) Asher, S. A. In *Handbook of Vibrational Spectroscopy*; Chalmers, J., Griffiths, P., Eds.; John Wiley & Sons, Ltd: New York, 2001; Vol. *1*, p 557−571.
(67) Asher, S. A.; Ianoul, A.; Mix, G.; Boyden, M. N.; Karnoup, A.; Diem, M.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2001**, *123*, 11775−11781.
(68) Garrido, M.; Rius, F. X.; Larrechi, M. S. *Anal. Bioanal. Chem.* **2008**, *390*, 2059−2066.
(69) Ianoul, A.; Boyden, M.; Asher, S. A. *J. Am. Chem. Soc.* **2001**, *123*, 7433−7434.
(70) Mirkin, N. G.; Krimm, S. *J. Phys. Chem. A* **2002**, *106*, 3391−3394.
(71) *Biochemical Applications of Raman and Resonance Raman Spectroscopies*; Carey, P. R., Ed.; Academic Press: New York, 1982.
(72) Chi, Z.; Chen, X. G.; Holtz, J. S.; Asher, S. A. *Biochemistry* **1998**, *37*, 2854−2864.
(73) Cardoso, J. F. *Neural Comput.* **1999**, *11*, 157−192.
(74) Nuzillard, D.; Nuzillard, J.-M. *Signal Process.* **2003**, *83*, 627−631.
(75) Choi, S.; Cichocki, A.; Beloucharni, A. *J. VLSI Signal Process.* **2002**, *32*, 93−104.
(76) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(77) *Instrumental Analysis of Intrinsically Disordered Proteins: Assesing Structure and Conformation*; Uversky, V. N., Longhi, S., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2010.
(78) Kodaka, M. *Biophys. Chem.* **2004**, *107*, 243−253.