The Role of Proline-Containing Peptide Triads in beta-Sheet Formation: A Kinetic Study

ARTICLE *in* BIOPOLYMERS · FEBRUARY 2015

Impact Factor: 2.39 · DOI: 10.1002/bip.22622

The Role of Proline-Containing Peptide Triads in β-Sheet Formation: A Kinetic Study

Gaius A. Takor, Seiichiro Higashiya, Vitali K. Sikirzhytski, Jason P. Seeley, Igor K. Lednev, John T. Welch

Department of Chemistry, University at Albany, State University of New York, Albany, NY 12222

Received 22 July 2014; revised 31 December 2014; accepted 31 December 2014
Published online 6 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22622

ABSTRACT:

*The design of biomimetic materials through molecular self-assembly is a growing area of modern nanotechnology. With problems of protein folding, self-assembly, and sequence–structure relationships as essential in nanotechnology as in biology, the effect of the nucleation of β-hairpin formation by proline on the folding process has been investigated in model studies. Previously such studies were limited to investigations of the influence of proline on the formation of turns in short peptide sequences. The effect of proline-based triads on the folding of an 11-kDa amyloidogenic peptide $GH_6[(GA)_3GY(GA)_3GE]_8$-GAH_6 (**YE8**) was investigated by selective substitution of the proline-substituted triads at the γ-turn sites. The folding and fibrillation of the singly proline-substituted polypeptides, e.g., GH6—[(GA)3GY(GA)3GE]7(GA)3-GY(GA)3PD—GAH6 (**8PD**), and doubly proline-substituted polypeptides, e.g., GH6—[(GA)3GY(GA)3-GE]3(GA)3GY(GA)3PD[(GA)3GY(GA)3GE]3(GA)3-GY(GA)3PD—GAH6 (**4,8PD**), were directly monitored by circular dichroism and deep UV resonance Raman and fluorescence spectroscopies. These findings were used to identify the essential folding domains, i.e., the minimum number of β-strands necessary for stable folding. These experimental findings may be especially useful in the design and construction of peptidic materials for a wide range of applications as well as in understanding the mechanisms of folding critical to fibril formation.*

Keywords: *synthetic amyloidogenic protein; β-sheet; proline-based triads; β-sheet nucleation*

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of any preprints from the past two calendar years by emailing the Biopolymers editorial office at biopolymers@wiley.com.

Correspondence to: John T. Welch; e-mail jwelch@albany.edu
Contract grant sponsor: National Science Foundation
Contract grant number: CHE-0809525

INTRODUCTION

Protein folding and misfolding have been intensely studied over the past decades, not only because many debilitating diseases are associated with misfolding but also because the exploitation of peptide self-assembly has enabled the preparation of novel biomaterials. Robust self-assembling protein fibers defined by a common cross-β core structure,[1,2] amyloid fibrils are implicated in an array of human ailments ranging from neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and transmissible spongiform encephalopathy,[3–5] to systemic amyloidosis. A hallmark of these diseases, also known as protein misfolding disorders, is the change in the secondary and/or tertiary structure of a native protein without alteration of the primary structure. Common structural changes include increases in β-sheet content, in oligomerization and in formation of fibrillar amyloid-like polymers.[5] Amyloid-like structures also impart a number of unique properties potentially useful in materials science, including compressive

strength and resistance to high pressure or proteolytic digestion, conditions that typically destabilize globular proteins.[6–8] Using genetic engineering, numerous repetitive polypeptides[9–16] have been synthesized to investigate the utility of the repetitive polypeptides in tissue engineering, drug delivery, and fabrication of nanodevices.[12]

Because an increase in β-sheet content is a hallmark of amyloid diseases, modern molecular biology and medicine have focused on understanding the general mechanism of folding and aggregation. Effective folding modulation may be useful in the development of therapeutic agents for the treatment and prevention of amyloidogenic and protein misfolding diseases.

The use of simple models to study the folding of globular proteins has led to the identification of a number of factors that influence and enhance β-sheet folding. β-Hairpins, such as type I + G1 β-bulge, type I' β-turn, or type II' β-turn, can act as folding nuclei. These hairpin structures are dependent upon the primary sequence as well as the number and location of the stabilizing hydrogen bonds.[17] Although there is usually cooperativity in β-sheet formation in at least one dimension, the essential interactions are often perpendicular to the strand direction. Implicitly, the stability of the β-sheet increases along with an increase in the number of strands and, therefore, the number of hydrogen bonds. On the other hand, cooperativity can induce an asymmetrical twist that unfavorably affects neighboring strand side-chain interactions. Environmental control of folding and fibrillation is modulated by pH and electrostatic interactions. Innate properties, such as net charge, hydrophobicity, and the tendency of certain side chains to promote β-sheet formation, determine the propensity of proteins and peptides to form β-sheet containing assemblies.[18]

Short model polypeptides with repetitive sequences have been used in computational studies of the conformational dynamics of both the component of β-sheets and hairpin motifs.[19–22] In an investigation of the structural influences of β-turns and hairpin motifs, Perczel et al.[21] used short model peptides to determine that the relative stability of the antiparallel model sheets increases in a nonlinear fashion with an increase in the number of turn or hairpin motifs. Proline, frequently found in tight turns,[23,24] differs from all other amino acids in that the five-member pyrrolidine ring profoundly limits side chain conformations. Proline often acts as a helix-breaker because of the absence of the *N*-hydrogen required for formation of a helix interior hydrogen bond.[25] The thermostability of some enzymes can be cumulatively enhanced by increasing the frequency of proline occurrences at tight turns without significantly altering the secondary structure or the function.[26,27]

The alanyl–glycyl-rich polypeptides composed of repeating —$(AG)_3$PEG— units examined by McGrath et al.[28] were inspired by a silk-like protein composed of repetitive units with alanyl–glycyl-rich sequences, e.g., $(—GGX—)_n$, $(—GGGX—)_n$, and $(—GAGAGAS—)_n$ (where $1 \leq n \geq 15$ and "*X*" = S, Y, L, Q, or other hydrophobic residues) that form β-sheets.[29] The design was based on the following observations : (i) poly (glycylalanine) adopts an antiparallel β-sheet arrangement in the solid state, (ii) as mentioned earlier, proline frequently initiates β- or γ-turns in globular proteins, and (iii) proline and glutamic acid are poor β-strand residues, and as such could be predicted to induce strand reversal and hence hairpin formation.[28] However, characterization of these synthetic polypeptides by X-ray and Fourier transform infrared spectroscopy offered no evidence for a β-sheet structure, instead amorphous glassy aggregates formed at room temperature. It was deduced that β-strands composed of an odd number of amino acids would not readily form β-sheets with an alternating pattern of hydrogen bonds as interstrand hydrogen bonding might not be complementary. Because "normal" two-residue long β-turns could form only at one end of the above strands, the overall fold would be destabilized, leading to formation of the amorphous glassy structures.[22] Subsequent model peptides, therefore, consisted of potential strands containing an even number of amino acids but lacking proline.[30] Nonetheless, the influence of proline on the folding of even-numbered strands remained elusive in studies of this new generation of polypeptides. Strategic proline substitutions in short 26-residue, amyloid-forming model peptides have been investigated and characterized by circular dichroism (CD) spectroscopy and thioflavin T (ThT) fluorescence staining.[24,31]

Selective single and double proline substitutions at turns proximal to the N- and C-termini of an 11-kDa (143-residue) *de novo* designed polypeptide, **YE8**, GH_6-$[(GA)_3GY(GA)_3GE]_8$-GAH_6 influence the rate of folding. Glycylalanyl diad $(GA)_3$ repeats were selected for the formation of identical, weakly interacting β-strands (an even number of residues per β-strand), whereas aromatic amino acids and/or charged amino acid residues, e.g., Y and E, respectively, which are sensitive to environmental changes, were chosen for the γ-turn construct. **YE8** is known to be an amyloidogenic polypeptide forming well-defined fibrillar structures.[18,32] In this work, turn-inducing proline-containing triads such as proline–aspartic acid–glycine (PDG)[17,33] or proline–glutamic acid–glycine (PEG) were incorporated at turn positions to enable investigation of the effect of single and double "proline substitution(s)" on folding kinetics. Figure 1 illustrates the representative structures of the doubly proline-substituted polypeptides: **1,5PE**, GH_6—$(GA)_3GY(GA)_3$-$PE[(GA)_3GY(GA)_3GE]_3(GA)_3GY(GA)_3PE[(GA)_3GY(GA)_3GE]_3$-$GAH_6$, **1,5 PD**, GH_6—$(GA)_3GY(GA)_3PD[(GA)_3GY(GA)_3GE]_3$ $(GA)_3GY(GA)_3PD[(GA)_3GY(GA)_3GE]_3$—$GAH_6$, **4,8PE**, GH_6 —$[(GA)_3GY(GA)_3GE]_3(GA)_3GY(GA)_3PE[(GA)_3GY(GA)_3GE]3$ $(GA)_3GY(GA)_3PE$—GAH_6, and **4,8PD** GH_6—$[(GA)_3GY(GA)_3$-$GE]_3(GA)_3GY(GA)_3PD[(GA)_3GY(GA)_3GE]3(GA)_3GY(GA)_3PD$ —GAH_6 and singly proline-substituted polypeptides: **1PD**,

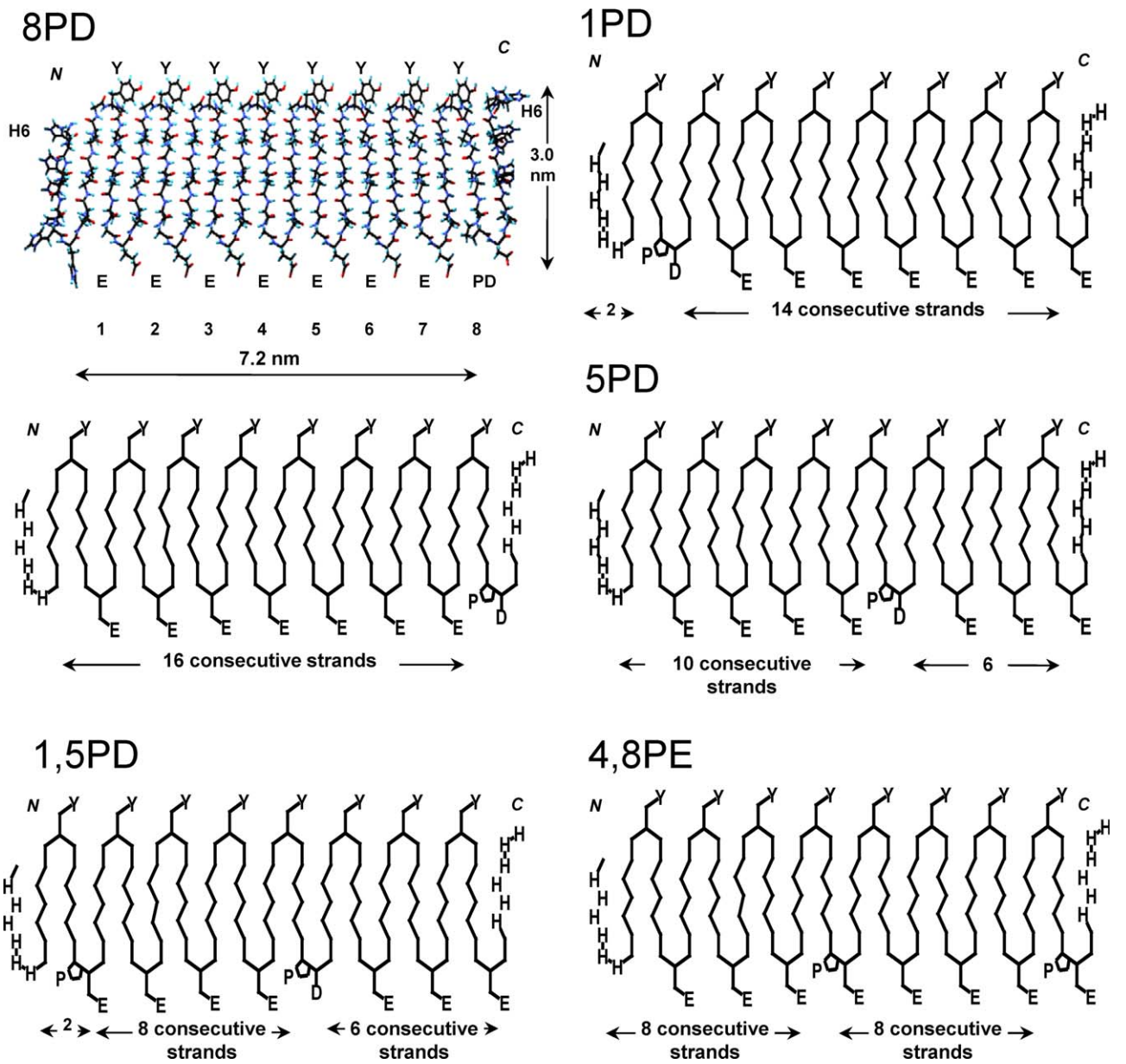

FIGURE 1 Schematic representations of **H6-YE8-H6** derivatives. The values below each polypeptide refer to the number of consecutive $(GA)_3$ β-sheet-forming strands uninterrupted by proline-substituted turns.

GH_6—$(GA)_3GY(GA)_3PD[(GA)_3GY(GA)_3GE]_7$—$GAH_6$, **5PD**, GH_6—$[(GA)_3GY(GA)_3GE]_4(GA)_3GY(GA)_3PD[(GA)_3GY(GA)_3$-GE]3—$GAH_6$, and **8PD**, GH_6—$[(GA)_3GY(GA)_3GE]_7(GA)_3GY(GA)_3PD$—$GAH_6$. In contrast to published studies of short peptide sequences, the model peptides of this study have molecular weights around 11 kDa. Peptide folding was directly monitored by CD and deep UV resonance Raman (DUVRR) spectroscopies. Fibrillation was followed by ThT fluorescence staining. On the basis of the results, it was possible to determine the minimum number of adjacent strands necessary for folding (the folding domain) and hence the importance of energy reduced by strand–strand hydrogen bond formation.

MATERIALS AND METHODS

Materials

*Bsa*I isozyme, *Eco*31I, and RAPID DNA Ligation Kit were purchased from Fermentas (Hanover, MD). Benzonase was purchased from Novagen (Madison, WI). Inoue ultracompetent and electrocompetent cells of XL1-Blue (Stratagene, La Jolla, CA) were prepared according

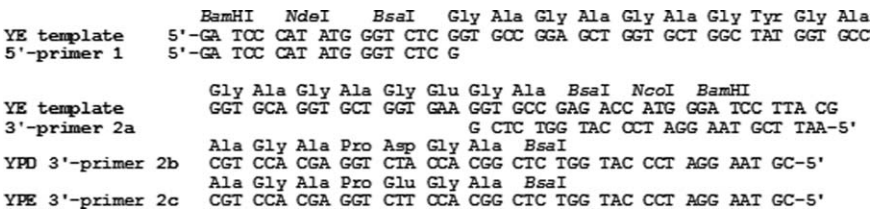

FIGURE 2 Primer sequences for YE, YPD, and YPE used in this study.

to standard methods.[34] BLR(DE3)pLysSRARE was prepared from BLR(DE3) (Novagen) transformed by pLysSRARE isolated from Rosetta(DE3)pLysS (Novagen) as previously reported.[35,36] A frozen stock of chemically competent cells of BLR(DE3)pLysSRARE was prepared by a modified standard calcium chloride method (0.1*M* $CaCl_2$ containing 15% glycerol) and was stored at −80°C.[34] Plasmids pUC18 and pET-28a-c were obtained from Bayou Biolabs (Metairie, LA) and Novagen, respectively. Plasmids and DNA fragments separated by agarose gel electrophoresis were purified using QIAprep Spin Miniprep Kit and QIAquick Gel Extraction Kit (Qiagen, Valencia, CA), respectively. Western blot signals were detected using SuperSignal West HisProbe kit (Pierce/Thermo Fisher Scientific, Rockford, IL). Large-scale purification of polyhistidine-tagged, repetitive polypeptides was effected using Ni-NTA Superflow affinity columns (Qiagen).

Methods

Construction of Coding Sequences. Modified **YE8**-coding sequences were prepared in expression vectors by concatenation of proline-coding DNA sequences (YPE and YPD) and the previously prepared YE coding sequence following earlier reports[35,36] (Figure 2). The 5′-primer **1** and either the YPD or YPE 3′-primer (**2b** or **2c**) (Figure 2) were used to amplify YE ($(GA)_3GY(GA)_3GE$)-, YPD ($(GA)_3GY(GA)_3PD$)-, and YPE ($(GA)_3GY(GA)_3PE$)-coding DNA fragments, respectively. Each PCR product was purified with a QIAquick Gel Extraction Kit and was digested by *Eco*31I (a 1:1 mixture of YE and either YPD or YPE fragments can also be used for *Eco*31I digestion for the following concatenation). The resulting unidirectionally concatenatable DNA fragments (~50 bp) were purified by agarose gel (low melting, 3.5%) electrophoresis. YE and either YPD or YPE concatenatable fragments were mixed at 1:1 ratio and concatenated with A2 adapters as previously published.[35] The purified concatenated and adapted fragments were cloned into pUC18, which was previously digested by *Bam*HI and *Eco*RI (pUC18/*Bam*HI-*Eco*RI). The plasmids coding for the desired fragments, **YPDYE**, **YEYPD**, **YPEYE**, and **YEYPE**, were identified by size on an agarose gel and purified. Further concatenation was repeated with **YE2** and either of the above identified fragments to obtain plasmids with the **YPDYE3**, **YE3YPD**, **YPEYE3**, and **YE3YPE** coding sequences. Concatenation was repeated with **YE4** and either **YPDYE3**, **YE3YPD**, **YPEYE3**, or **YE3YPE** to give plasmids encoding **1PD** (YPDYE7), **5PD** (YE4YPDYE3), **1,5PD** ((YPDYE3)2), **8PD** (YE7YPD), **4,8PD** ((YE3YPD)2), **1PE** (YPEYE7), **5PE** (YE4YPEYE3), **1,5PE** ((YPEYE3)2), and **8PE** (YE7YPE), **4,8PE** ((YE3YPE)2), respectively. The DNA fragments of **YE8** analogs were further incorporated into pET-28 digested by *Nco*I and *Bam*HI (pET-28/*Nco*I-*Bam*HI) with A3 (H6-H6) adapters permitting protein expression.[35] Subsequently, the BLR(DE3)pLysSRARE expression hosts were transformed with the expression vectors harboring the desired constructs.

Expression and Purification of Polypeptides. Ni-NTA Superflow affinity columns were used for the large-scale purification of the expressed polyhistidine-tagged repetitive polypeptides as described previously.[35,36] Expression was induced by 1 m*M* isopropyl β-D−1-thiogalactopyranoside for 4 h at 37°C. The cells were pelleted by centrifugation at 3500*g* for 30 min at 4°C, were resuspended in H_2O (15 mL per 800 mL culture), and were then frozen. The frozen cells were lysed by freeze–thaw sonication for 30 min in the presence of benzonase and phenylmethanesulfonyl fluoride (2 m*M*). Urea (11 g per 800 mL culture, ~8*M* at final concentration) was added to the lysate, and the solution was heated in a boiling water bath for 1 h with occasional mixing. The resultant lysate was centrifuged at 20,000*g* for 2 h at 4°C. The supernatant was applied to Ni-NTA column, and the column then was washed with 40 m*M* imidazole in 8*M* urea and phosphate-buffered saline. The desired polypeptide was eluted with 300 m*M* imidazole in 8*M* urea in phosphate-buffered saline. The eluent containing a purified polypeptide was dialyzed against doubly distilled H_2O at 4°C using a 3500-Da cutoff dialysis membrane (Fisher Scientific, Waltham, MA).

Sample Preparation. After dialysis, the polypeptides were centrifuged for 20 min at 15,000*g* and 4°C. After centrifugation, the supernatant of each sample was separated from the pelleted aggregate. The polypeptide concentration was adjusted to 100 μ*M* using DI water and then to 44 μ*M* using formate buffer and a final pH of 3.5.

Fluorescence, UV Absorption, and CD Spectroscopy. Far-UV CD spectra were measured in a 0.05-cm quartz cell using J-720 CD spectropolarimeter (JASCO, Tokyo, Japan). A bandwidth of 1 nm, a scan speed of 100 nm/min with a resolution of 0.5 nm, and a response time of 0.3 s was used in the acquisition of spectra. Normally, six CD spectra from 190 to 300 nm were averaged for each time point and sample. The UV absorption spectra were measured in a 0.05-cm quartz cell using Lambda 950 spectrometer (PerkinElmer, Waltham, MA) from 400 to 200 nm.

Fluorescence spectra were measured in a 1-cm rectangular quartz cell using a Jobin Yvon Fluoromax-3 spectrofluorometer (Jobin Yvon, Edison, NJ). Following excitation at 450 nm, the emission spectrum from 465 to 560 nm was recorded. Initially, the fluorescence of 2 mL of ThT solution (20 m*M* in 100 m*M* potassium phosphate buffer, pH 7.4) was taken as background, and then 100 μL of the sample solution was added and quickly mixed using a pipette. The fluorescent intensity at 480 nm was obtained by subtraction of the background from the sample–ThT mixture spectrum.

Deep UV Resonance Raman Spectroscopy Measurement. The DUVRR instrumentation has been described in detail elsewhere.[37] Briefly, a carbonyl group-absorbing, 197-nm laser beam (1

Table I Predicted Charges of YE8, 1PD, and 4,8PD Polypeptides, the β-Sheet Core, and Different Amino Acids at pH 3.5[a]

H6-YE8-H6		H6-1PD-H6		H6-1,5PD-H6	
Strand	Charge	Strand	Charge	Strand	Charge
8E	−0.6	1D	0	2D	0
8Y	0	1P	0	2P	0
		7E	−0.5	6E	−0.4
		8Y	0	8Y	0
YE8	−0.6	1PD	−0.5	1,5PD	−0.4
2·H6	12	2·H6	12	2 H6	12
Total	11.4		11.4[b]		11.4[b]

[a] The prediction was based on the pKa values for isolated amino acid residues using Protein Calculator v3.3.

[b] The calculated total charge of the full-length polypeptide with both hexahistidine tracks is still 11.4.

mW, Indigo-S laser system from Coherent) was focused on the samples. Scattered radiation enhanced by resonance Raman effects was collected in backscattering geometry, dispersed using a home-built double monochromator, and detected with a liquid nitrogen-cooled CCD camera (Roper Industries, Sarasota, FL). GRAMS/AI (7.01) software (Thermo Scientific, Salem, NH) was used for Raman spectroscopic data processing. The contribution of Suprasil and buffer to the spectra was quantitatively subtracted.

Principal Component Analysis of CD Spectrum. The software used in this work for the principal component analysis (PCA) and partial least squares-data analysis was the PLS toolbox 3.5 for MATLAB® from Eigenvector Research.

RESULTS AND DISCUSSION

Experimental Conditions for YE8 and Derivatives

Many amyloidogenic polypeptides fold under acidic conditions where significant charge could be present.[38,39] **YE8** has been found to fold and form fibrils at pH 3.5.[18,32] After dialysis against doubly distilled water at 4°C, the pH of the **YE8**-containing solution was lowered from neutral to 3.5 using formate buffer. As a consequence, the net charge decrease on **YE8** was accompanied by a concomitant decrease in electrostatic repulsion within the β-sheet-forming domain of **YE8**. The resultant folded peptide aggregated into well-defined fibrils.[18] At neutral pH, the significant charge on dissociated glutamic acid residues (E) of the polypeptide was most easily accommodated in the unfolded monomer. The singly proline-substituted derivatives (**1PD**, **5PD**, and **8PD**) and doubly substituted derivatives (**1,5PE**, **1,5PD**, **4,8PE**, and **4,8PD**) were expected to fold similarly under acidic conditions. In Table I, the predicted net charges at pH 3.5 of the representative polypeptides (**1PD** for single substitution and **1,5PD** for double substitution) are equal to that of **YE8** because all the polypeptides have only one

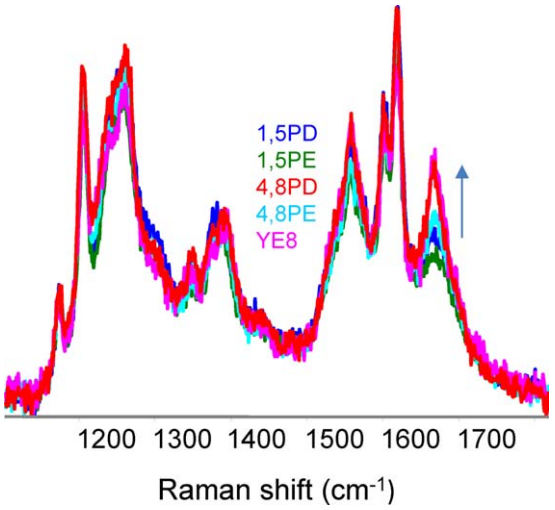

FIGURE 3 DUVRR spectra recorded on Day 45. (a–e) DUVRR spectra excited at 197 nm. Increase in the amide I peaks at 1680 cm^{-1} of **4,8PD** and **YE8** spectra indicate folded polypeptides.

charged species: negatively charged, dissociated glutamic acid. DUVRR spectra collected (data not shown) for all samples showed that the polypeptides remained unordered after more than 2 months of incubation at neutral pH and room temperature.

Doubly Proline-Substituted YE8

The initial solutions of **YE8** and the doubly proline-substituted derivatives **1,5PD**, **1,5PE**, **4,8PD**, and **4,8PE** were incubated at pH 3.5 and at room temperature and were probed frequently by DUVRR and CD spectroscopies. On the first day, DUVRR and CD spectra showed that the polypeptide solutions consisted of unordered polypeptides (Figure 3).

Analysis of Fibrillation of YE8 and Derivatives by DUVRR Spectroscopy. The peptide solutions were probed frequently by DUVRR over a period of 45 days. The DUVRR spectra were dominated by contributions from the amide chromophores and tyrosine. The Raman signature of the amide chromophore is extraordinarily sensitive to the polypeptide backbone conformation and provides direct quantitative information about the secondary structure of proteins. Amide I mode (Am I) consists of carbonyl C═O stretching, with a small contribution from C—N stretching and N—H bending. Amide II and amide III bands involve significant C—N stretching, N—H bending, and C—C stretching. The C_α—H bending vibrational mode involves C_α—H symmetric bending and C—C_α stretching. The appearance of a narrow and intense amide I band can clearly indicate β-sheet formation.[40] Figure 3 shows the DUVRR spectra of the polypeptides after 45 days incubation. There was no marked increase in the amide I band of **1,5PD** and **1,5PE**

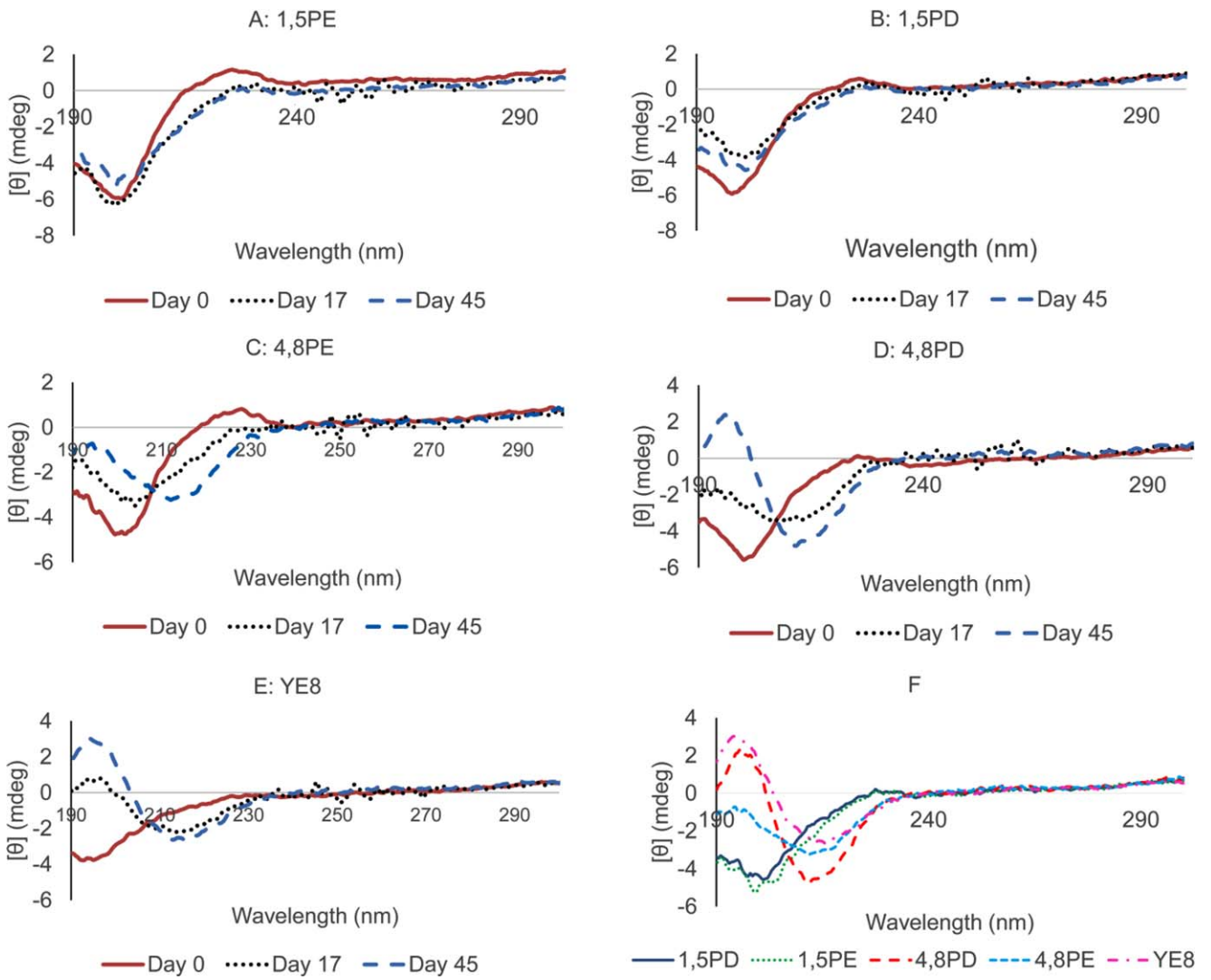

FIGURE 4 CD spectra of polypeptides on selected days of incubation (A–E). Comparisons of β-sheet structure of polypeptides after 45 days of incubation monitored by CD (F) spectroscopy. **4,8PD** and **YE8** exhibited folded β-sheet structures indicated by the positive maximum at ~196 nm and negative maximum at ~215 nm. **4,8PE** is only partially folded, whereas **1,5PD** and **1,5PE** remained unfolded after 45 days.

polypeptides after this incubation period. As expected, **YE8** showed an increase in the amide I peak but there was also an increase in the amide I band of **4,8PD**. Previously, we have shown that an increase in amide I intensity can be correlated with the extent of folding to assume a β-sheet conformation.[40]

Proline cannot be a hydrogen bond donor but can participate as a hydrogen bond acceptor. Known to disrupt the secondary structures of α-helices and β-sheets, in nature, proline is usually found at turn positions. Double substitution of **YE8** was found to, depending upon the site of substitution, either promote folding/fibrillation ultimately increasing the β-sheet content, or disrupt folding. Substitution of proline at the first and fifth turn positions inhibited folding in agreement with earlier work,[28] whereas substitution at the fourth and eighth turn positions tended to enhance folding as in the case of **4,8PD** having **PDG** triads. The "critical" turn position for folding/fibrillation enhancement was found to be the eighth turn position, i.e., the turn nearest the C-terminus. The additional mutation at the fourth position (within the polypeptide sequence) did not independently accelerate folding as will be explained later.

Analysis by CD. The CD spectra corroborated the Raman spectra. After 45 days, **1,5PD** and **1,5PE** remained unordered

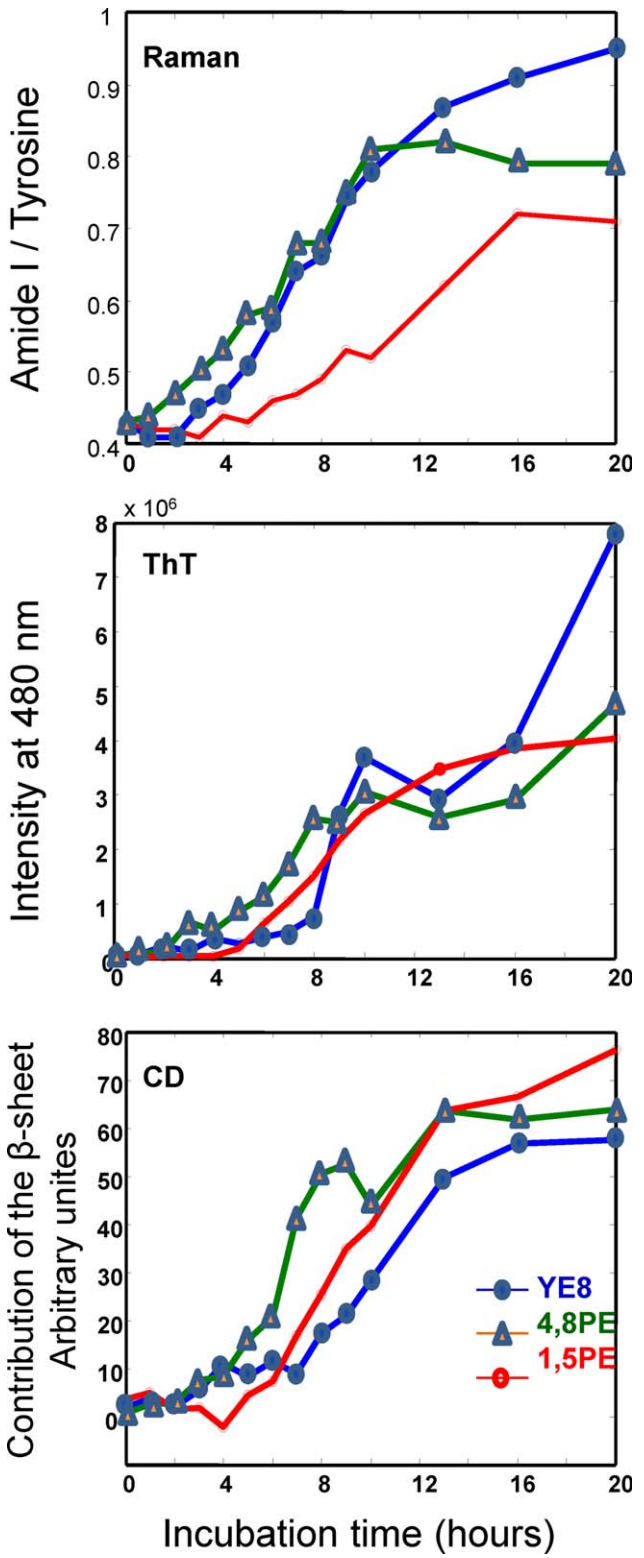

FIGURE 5 Kinetics of **YE8**, **4,8PE-YE8**, and **1,5PE** monitored by DUVRR, CD, and ThT fluorescence spectroscopies. Generally, all the methods allow a sigmoidal kinetics. The fibrillation "inhibitor" was effectively removed by shearing brought about by stirring the polypeptides for more than 16 h.

as can be seen in the CD spectra where a broad negative shoulder around 197 nm is formed on Day 1 and remained unchanged on Day 45. The appearance of a pronounced positive maximum at ∼197 nm and a modest maximum around 215–218 nm is consistent with a polyproline II type helix associated with a disordered state.[41–44] In contrast, after 45 days of incubation, the spectrum of **4,8PE** demonstrated a small positive maximum at 196 nm and a small negative minimum around 215 nm. The modest amplitude of the maxima implied that **4,8PE** was only partially folded after 45 days. The CD spectrum of **4,8PD** on the other hand had both a pronounced positive maximum and minimum at 197 nm and 215 nm, respectively, on Day 45. A comparison of the β-sheet content based on CD of all polypeptides after 45 days is summarized in Figure 4F. The highest β-sheet content was found with **4,8PD**, whereas **1,5PE** and **1,5PD** remained largely unfolded.

The mechanism by which proteins assume the associated three-dimensional structure has proven elusive. Computation of energy surfaces or landscapes allows the folding dynamics to be described and visualized in a meaningful manner. Dinner and Karplus [45] used computer simulations in comparative studies with experimental observations.[46]

As mentioned earlier, folded polypeptides, α-helices or β-sheets, are stabilized by the formation of hydrogen bonds between carbonyl and amide groups of the polypeptide backbone. Glycyl-alanyl diads are known to form β-strands,[14,47] where the sheets are stabilized by the intrastrand H-bonding. In this study, **1,5PD**, **1,5PE**, and **4,8PE** exhibited slower folding/fibrillation compared with **YE8** (the lead polypeptide), with **1,5PE** and **1,5PD** remaining unfolded in the absence of agitation for 2 months yet the total number of β-strand tracts remained unchanged. To determine whether the incorporation of PEG/PDG triads at turn positions permanently inhibited folding or simply raised the overall energy landscape and therefore retarded folding, **1,5PE**, **4,8PE**, and **YE8** were agitated at room temperature. The propensity of folding and fibrillation was monitored by CD and DUVRR spectroscopies and ThT fluorescence (Figure 5). Under these conditions, the polypeptides were completely folded after 20 h. It is possible that the difference in the propensity to fold is related to the propensity of D to be protonated under these conditions; in contrast, E cannot be protonated and hence charged.[48–50] In Figure 5, **YE8** had more β-sheet character and had undergone more extensive fibrillation after 64 h as determined by Raman and ThT spectroscopies. Clearly, the introduction of PEG triad or the PDG triad at the N- and C- terminal γ-turns did not disrupt the H-bonds between the strands but did raise the energy landscape. Once the hydrogen bond network is formed, it enables folding and ultimately fibrillation. The overall inhibition of fibrillation, if not folding as well, was

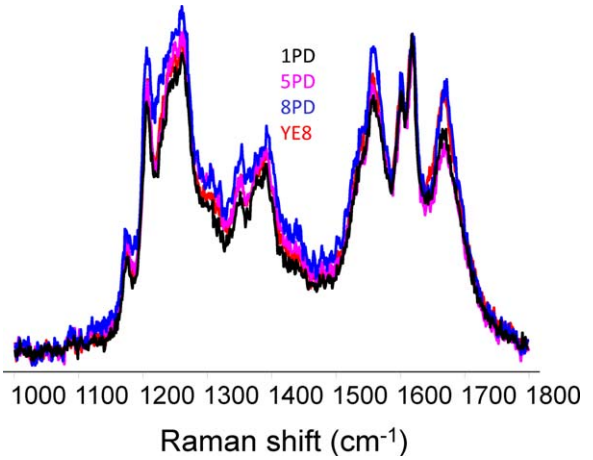

FIGURE 6 Raman spectra excited at 197 nm of **YE8**, **1PD**, **5PD**, and 8PD polypeptides at 40 days of incubation.

effectively reduced by sheer induced by agitation of the peptide solutions.

The kinetics of amyloid formation are also dependent on the relative stabilities of the intermediate states and on the barriers between conformational transitions that are strongly influenced by the environment.[51] In a related work (unpublished), barriers to folding and fibrillation can be altered by precipitously decreasing the pH of an alkaline solution of the polypeptides to pH 3.5, ultimately resulting in fibrillar polymorphism by the various polypeptides.

Single Proline Substitutions in YE8

It was anticipated that a single prolyl substitution could influence the folding/fibrillation process. Because the PDG triad replacement led to a significant increase in folding and fibrillation as seen earlier, the properties of **1PD**, **5PD**, and **8PD** were compared with those of **YE8**.

Using two-dimensional Correlation DUVRR, we previously applied a two-component model approach to describe the high level of complexity of fibrillation events in the transition of unordered **YE8** to an amyloid-like β-sheet structure.[52] In this fibrillation mechanism model, it was revealed that initially unordered **YE8** underwent a global structural rearrangement of tyrosine residues. The establishment of the β-sheet follows this global rearrangement occurring concurrently with β-strand formation. A two-component model was also applied to the polypeptide systems under investigation. PCA,[53–55] a good chemometric tool for the extraction of valuable information from data, was used to deduce recognizable patterns from the compressed data.

The polypeptides were incubated for 45 days, and after this period, as seen from the DUVRR spectra in Figure 6, **8PD** and **YE8** folded whereas the less-pronounced amide I bands of

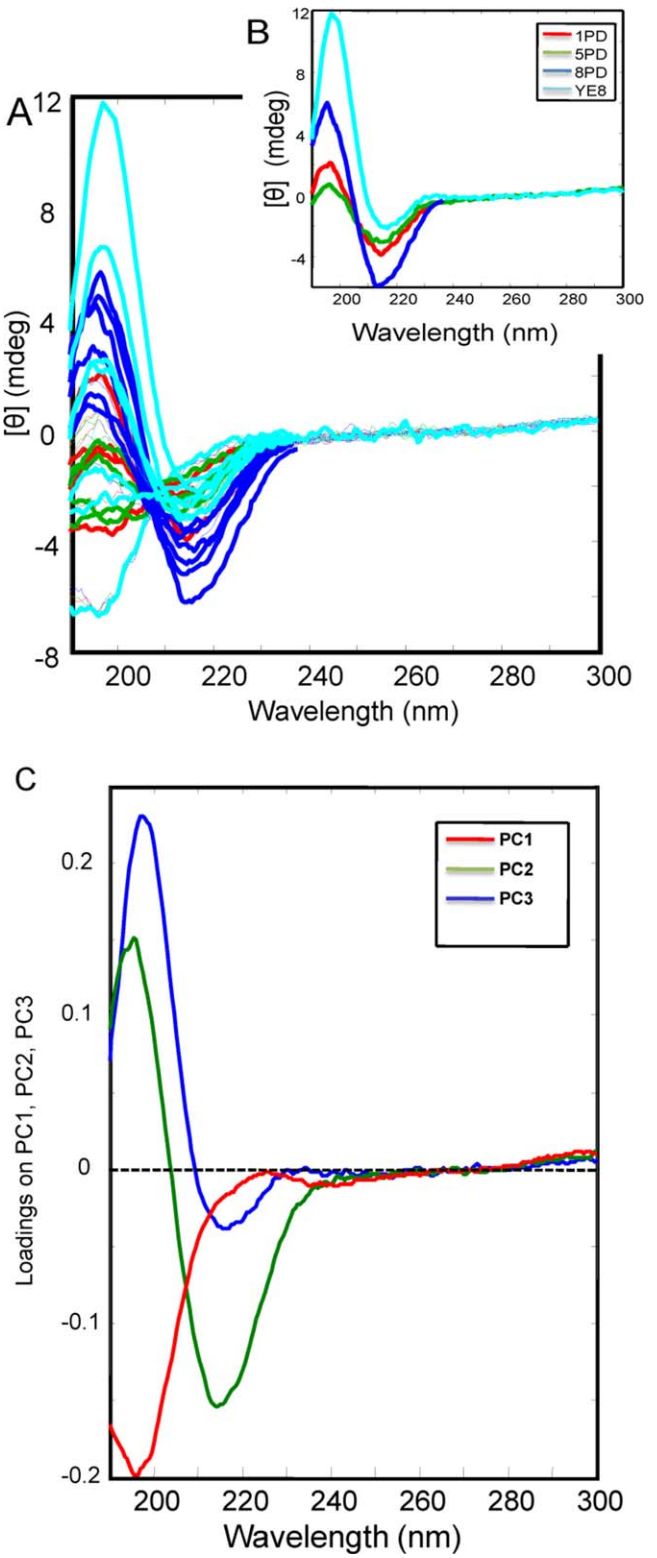

FIGURE 7 (a) Combined CD spectra of all polypeptides. (b) CD of all polypeptides on Day 40. (c) Principal component analysis yielded three principal components.

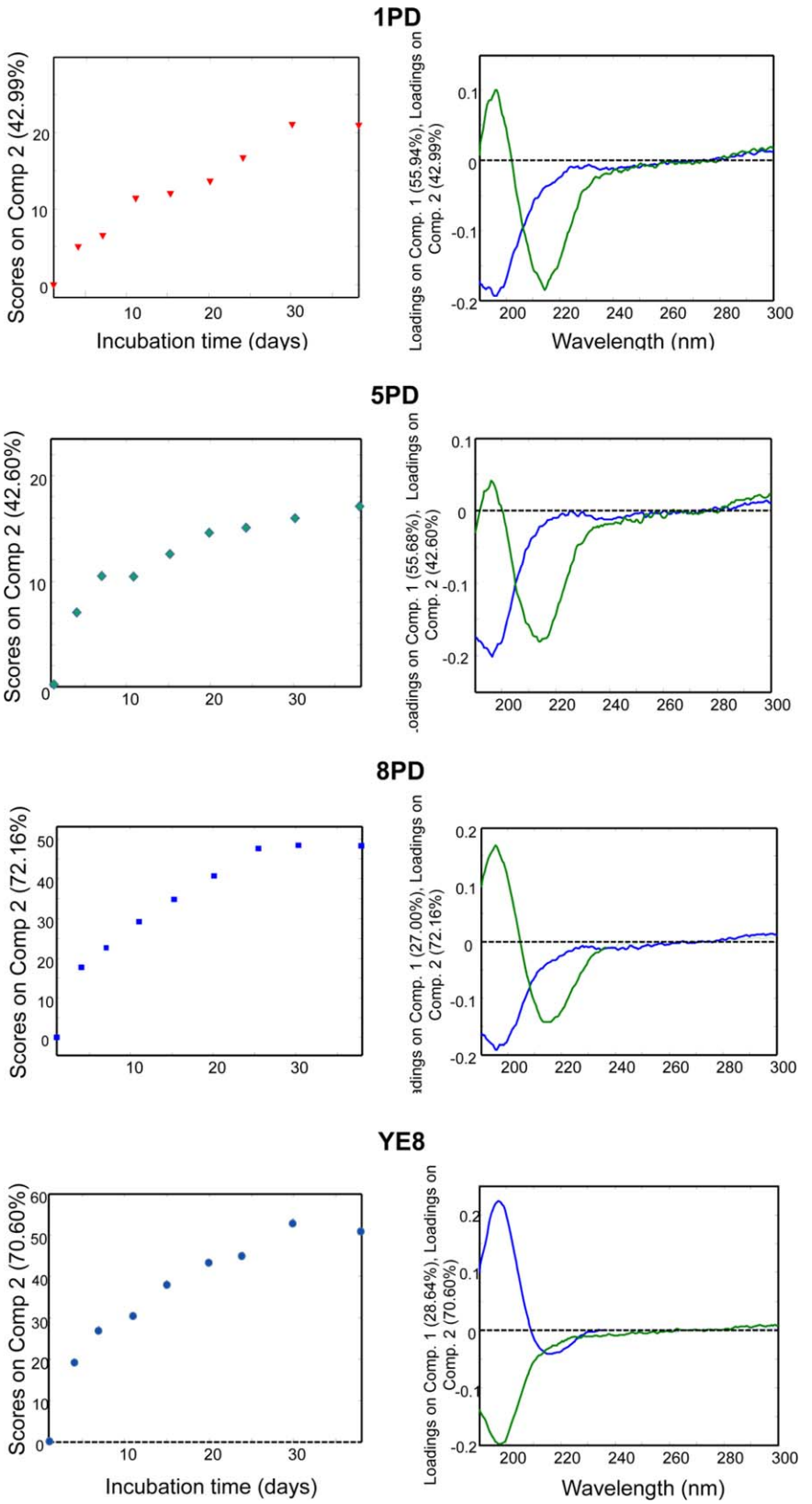

FIGURE 8 Principal component analysis: weighting of the principal components of each polypeptide. Basically, the folding kinetics of the each polypeptide is resolved by two principal components. Principal Component 2 increases over time, denoting transition from unordered to folded polypeptide.

1PD and **5PD** suggested those peptides to be only partially folded. The combined CD spectra of all the polypeptides were subsequently treated[52] and yielded three principal components, with Component 1 (PC1) contributing 16.01% and Components 2 (PC2) and 3 (PC3) contributing 32% and 51.91%, respectively (Figure 7). The three principal component models can rationalize the major variance of the data, and overfitting can be avoided. The first principal component separates a spectrum with a negative broad shoulder around 190 nm from the other spectra. The component spectral modeling can be associated with the CD spectrum of an unordered polypeptide. The second component models a spectrum that has a minimum with a deeper trough at ~215 nm and a maximum at ~197 nm. This spectrum of PC2 was assigned to the spectrum of the folded proline-substituted polypeptides especially that of **8PD**. Finally, the third component, a spectrum with a minimum with a more shallow trough (compared with PC2) at ~215 nm and a maximum at ~197 nm was assigned to folded **YE8**. Treated separately, the behavior of each polypeptide fits a two-component system as can be clearly seen in Figure 8, fitting the folding pattern of each polypeptide characterizing the transition from an unordered species to a folded polypeptide. For the folded proline-substituted polypeptides, after 45 days of incubation, **8PD** had 72% of β-sheet character, whereas **1PD** and **5PD** both had 43% β-sheet character as revealed by the PC2 contributions seen in Figure 8.

N-Terminus Vs. C-terminus. It was anticipated that these polypeptides would form antiparallel β-sheets with the interdigitated steric zipper structure commonly found in amyloid fibrils.[56] PDG or PEG triads introduced at specific turns revealed propensities to enhance or inhibit folding. As shown in Figure 1, the introduction of PEG and PDG divides the polypeptide into domains where there are contiguous $(GA)_3$ strands separated by PEG or PDG turns, e.g., **1PD** has two domains comprised of two strands and 14 strands, whereas **8PD** has 16 successive strands. For double proline substitution, **1,5PD** has three domains made up of two, eight, and six consecutive strands, whereas **4,8PD** still has two domains of eight consecutive strands.

For folding to occur, there should be at least eight uninterrupted strands. Additionally, the location of these eight successive strands, or the "folding domain," is important. The "folding domain" should be proximal to either the N- or C- terminus of the peptide rather than be centrally located. This rationale can explain why **1PD**, **5PD**, and **8PD**, after 40 days, had at least 43% β-sheet character. **1,5PD** with a folding domain located midway along the primary structure remained largely unfolded in unstirred conditions. The same phenomenon occurs with the PEG triad substituted polypeptides. Multiple proline substitutions in **YE8** should result in the inhibition of folding because the possibility of having a folding domain comprised of eight consecutive strands is diminished.

CONCLUSIONS

The minimum number of β-strands necessary for the folding of the model peptides was determined by the substitution of proline-based triads at various turn positions. The progress of the folding of the substituted molecules was monitored by CD, DUVRR, and ThT fluorescence spectroscopies. The results were in agreement with the published finding that a PDG triad rather than PEG triad is better β-sheet nucleus.[17,33] Incorporation of PDG triads at turns proximal to the N- or C-terminus as in **8PD** enhanced folding and fibrillation.

In the case of doubly proline-substituted peptides, **4,8PD** with two sets of eight contiguous β-strands folded more quickly than **4,8PE**, **1,5PE**, and **1,5PD**. The relative efficiency of folding of **4,8PD** relative to **1,5PD** is consistent with the hypothesis that a minimum of successive β-strands (eight contiguous strands) must be present for folding to occur expeditiously. With **1,5PD**, the disruption of one eight-strand tract is sufficient to measurably retard folding. The absence of eight contiguous β-strands results in a higher barrier to folding that can be overcome by agitation of the polypeptide solution. This finding confirmed that the primary structures of all the investigated peptides have an intrinsic ability to fold into β-sheet-containing assemblies.

Even without stirring, the singly substituted peptides (**1PD**, **5PD**, and **8PD**) had at least 43% β-sheet character after 45 days incubation at room temperature. With **8PD**, where the proline substitution is proximal to the C-terminus, the peptide had up to 72% β-sheet character after the same incubation period. Again, folding seems to be strongly promoted by the number of contiguous β-strands uninterrupted by a PDG triad. By analogy to published findings,[52] the folding of the polypeptides likely requires an initial global structural rearrangement resulting in change in tyrosine local environment, that is subsequently followed by simultaneous formation of turns and β-strands. Further aggregation leads to the growth of protofilaments, protofibrils, and finally mature fibrils.[52]

The sensitivity of folding to the primary structure is well-precedented; however, the sensitivity of the rate of folding to the position of proline-containing β-turns was not anticipated, especially differences in folding that occur when the prolyl turn was close to the C-terminus. PCA was shown to be especially helpful to identifying the steps of the folding process. Not only is an understanding of the factors

influencing folding valuable to revealing the importance of turn groups on conformational changes in amyloid diseases but also to the utilization of designed peptides in biotechnology and materials science. The selection of peptide sequences for use as building blocks requires molecular engineering with an appreciation for the control elements involved in programmed assembly.

REFERENCES

1. Bongiovanni, M. N.; Scanlon, D. B.; Gras, S. L. Biomaterials 2011, 32, 6099–6110.
2. Dobson, C. M. Nature 2003, 426, 884–890.
3. Dobson, C. M. Protein Pept Lett 2006, 13, 219–227.
4. Uversky, V. N.; Fink, A. L. Biochim Biophys Acta Proteins Proteomics 2004, 1698, 131–153.
5. Soto, C. FEBS Lett 2001, 498, 204–207.
6. Dirix, C.; Meersman, F.; MacPhee, C. E.; Dobson, C. M.; Heremans, K. J Mol Biol 2005, 347, 903–909.
7. Meersman, F.; Cabrera, R. Q.; McMillan, P. F.; Dmitriev, V. Biophys J 2011, 100, 193–197.
8. Zurdo, J.; Guijarro, J. I.; Dobson, C. M. J Am Chem Soc 2001, 123, 8141–8142.
9. Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. Biotechnol Prog 1990, 6, 198–202.
10. Deming, T. J. Adv Mater 1997, 9, 299–311.
11. Higashiya, S.; Topilina, N. I.; Ngo, S. C.; Zagorevskii, D.; Welch, J. T. Biomacromolecules 2007, 8, 1487–1497.
12. Zhao, X.; Zhang, S.; Werner, C. In Advances in Polymer Science; Springer: Berlin, 2006; pp 145–170.
13. Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Macromolecules 1991, 24, 1213–1214.
14. Parkhe, A. D.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Intl J Biol Macromol 1998, 23, 251–258.
15. Tirrell, D. A.; Fournier, M. J.; Mason, T. L. MRS Bull 1991, 16, 23–28.
16. McGrath, K. P.; Tirrell, D. A.; Kawai, M.; Mason, T. L.; Fournier, M. J. Biotechnol Prog 1990, 6, 188–192.
17. Blandl, T.; Cochran, A. G.; Skelton, N. J. Protein Sci 2003, 12, 237–247.
18. Topilina, N. I.; Sikirzhytsky, V.; Higashiya, S.; Ermolenkov, V. V.; Lednev, I. K.; Welch, J. T. Biomacromolecules 2010, 11, 1721–1726.
19. Vasquez, M.; Nemethy, G.; Scheraga, H. A. Chem Rev 1994, 94, 2183–2239.
20. Czinki, E.; Császár, A. G.; Perczel, A. Chem Eur J 2003, 9, 1182–1191.
21. Perczel, A.; Gáspári, Z.; Csizmadia, I. G. J Comput Chem 2005, 26, 1155–1168.
22. Pohl, G.; Beke, T.; Borbély, J.; Perczel, A. J Am Chem Soc 2006, 128, 14548–14559.
23. Raghavender, U. S.; Aravinda, S.; Rai, R.; Shamala, N.; Balaram, P. Org Biomol Chem 2010, 8, 3133–3135.
24. Raghothama, S. R.; Awasthi, S. K.; Balaram, P. J Chem Soc Perkin Trans 2 1998, 137–144.
25. O'Neil, K.; DeGrado, W. Science 1990, 250, 646–651.
26. Watanabe, K.; Suzuki, Y. J Mol Catal B: Enzym 1998, 4, 167–180.
27. Suzuki, Y. Proc Jpn Acad Ser B 1999, 75, 133–137.
28. McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. J Am Chem Soc 1992, 114, 727–733.
29. Winkler, S.; Kaplan, D. L. Rev Mol Biotechnol 2000, 74, 85–93.
30. Cantor, E. J.; Atkins, E. D.; Cooper, S. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. J Biochem 1997, 122, 217–225.
31. Gerling, U. I. M.; Brandenburg, E.; Berlepsch, H. V.; Pagel, K.; Koksch, B. Biomacromolecules 2011, 12, 2988–2996.
32. Topilina, N. I.; Ermolenkov, V. V.; Sikirzhytski, V.; Higashiya, S.; Lednev, I. K.; Welch, J. T. Biopolymers 2010, 93, 607–618.
33. Jourdan, M.; Griffiths-Jones, S. R.; Searle, M. S. Eur J Biochem 2000, 267, 3539–3548.
34. Sambrook, J.; Russell, D. Molecular Cloning: A Laboratory Manual, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.
35. Topilina, N. I.; Ermolenkov, V. V.; Higashiya, S.; Welch, J. T.; Lednev, I. K. Biopolymers 2007, 86, 261–264.
36. Topilina, N. I.; Higashiya, S.; Rana, N.; Ermolenkov, V. V.; Kossow, C.; Carlsen, A.; Ngo, S. C.; Wells, C. C.; Eisenbraun, E. T.; Dunn, K. A.; Lednev, I. K.; Geer, R. E.; Kaloyeros, A. E.; Welch, J. T. Biomacromolecules 2006, 7, 1104–1111.
37. Lednev, I. K.; Ermolenkov, V. V.; He, W.; Xu, M. Anal Bioanal Chem 2005, 381, 431–437.
38. Frare, E.; Polverino, d. L. P.; Zurdo, J.; Dobson, C. M.; Fontana, A. J Mol Biol 2004, 340, 1153–1165.
39. Kad, N. M.; Myers, S. L.; Smith, D. P.; Smith, D. A.; Radford, S. E.; Thomson, N. H. J Mol Biol 2003, 330, 785–797.
40. Lednev, I. K.; Ermolenkov, V. V.; Higashiya, S.; Popova, L. A.; Topilina, N. I.; Welch, J. T. Biophys J 2006, 91, 3805–3818.
41. Gokce, I.; Woody, R. W.; Anderluh, G.; Lakey, J. H. J Am Chem Soc 2005, 127, 9700–9701.
42. Shi, Z.; Woody, R. W.; Kallenbach, N. R. Adv Protein Chem 2002, 62, 163–240, 161 plate.
43. Sreerama, N.; Woody, R. W. Biochemistry 1994, 33, 10022–10025.
44. Woody, R. W. J Am Chem Soc 2009, 131, 8234–8245.
45. Dinner, A. R.; Karplus, M. J Mol Biol 1999, 292, 403–419.
46. Dinner, A. R.; Åali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. Trends Biochem Sci 2000, 25, 331–339.
47. Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. Science 1998, 281, 389–392.
48. Rybka, K.; Toal, S. E.; Verbaro, D. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. Proteins: Struct, Funct, Bioinf 2013, 81, 968–983.
49. Duitch, L.; Toal, S.; Measey, T. J.; Schweitzer-Stenner, R. J Phys Chem B 2012, 116, 5160–5171.
50. Hagarman, A.; Mathieu, D.; Toal, S.; Measey, T. J.; Schwalbe, H.; Schweitzer-Stenner, R. Chem Eur J 2011, 17, 6789–6797, S6789/6781–S6789/6789.
51. Miller, Y.; Ma, B.; Nussinov, R. Chem Rev 2010, 110, 4820–4838.
52. Sikirzhytski, V.; Topilina, N. I.; Takor, G. A.; Higashiya, S.; Welch, J. T.; Uversky, V. N.; Lednev, I. K. Biomacromolecules 2012, 13, 1503–1509.

53. Jaumot, J.; Eritja, R.; Navea, S.; Gargallo, R. Anal Chim Acta 2009, 642, 117–126.
54. Rudd, T. R.; Skidmore, M. A.; Guimond, S. E.; Cosentino, C.; Torri, G.; Fernig, D. G.; Lauder, R. M.; Guerrini, M.; Yates, E. A. Glycobiology 2009, 19, 52–67.
55. Navea, S.; Tauler, R.; Goormaghtigh, E.; de, J. A. Proteins: Struct, Funct, Bioinf 2006, 63, 527–541.
56. 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.; Madsen, A. O.; Riekel, C.; Eisenberg, D. Nature 2007, 447, 453–457.

Reviewing Editor: Stephen Blacklow