VIP Protein Substructures

DOI: 10.1002/ange.201102986

Impact of Strand Length on the Stability of Parallel-β-Sheet Secondary Structure**

*Felix Freire, Aaron M. Almeida, John D. Fisk, Jay D. Steinkruger, and Samuel H. Gellman**

The thousands of tertiary structural motifs known among natural proteins are built up from a handful of regular secondary structures, mostly α helices and β sheets, which are linked by β turns and larger loops. Sequence–stability relationships for α helices[1–3] and for antiparallel β sheets[4–7] have been widely explored using autonomously folding peptides, an experimental approach that eliminates the potentially confounding influence of a specific tertiary context. In contrast, a minimal parallel β sheet cannot readily be generated within a conventional peptide because the N terminus of one strand is necessarily distant from the C terminus of a neighboring strand. A continuous peptide design for a parallel β sheet would require a long interstrand loop, introducing significant tertiary packing effects or an entropic penalty for folding, either of which would compromise the analysis. Thus, novel strategies are required to elucidate the factors that govern parallel-β-sheet stability.[8,9]

Herein we use recently developed tools, including short nonpeptidic units for connecting peptide strands through their C termini,[10,11] to explore a fundamental question: does parallel-β-sheet secondary structure become more stable as the strands grow longer? The α helix is well known to be stabilized by increasing length.[12–15] α-Helix initiation (ordering of the first four residues, to enable formation of the first intrahelical H-bond) entails a large entropic cost, which can be offset by a small but favorable energetic gain from propagation (incorporation of individual residues into a preexisting helix).[16–18] The longer the α helix, the larger the net thermodynamic benefit from propagation, if the residues have a favorable helix propensity. In contrast, antiparallel-β-sheet secondary structure appears to be subject to an intrinsic limit on strand length.[19] The relationship between strand length and parallel-β-sheet stability has not previously been addressed. This gap in experimentation may underlie the relatively limited theoretical treatment of β-sheet formation,[20–22] especially for parallel β sheets,[20] which stands in contrast to the extensive theoretical analysis of α-helical folding.[16–18] Herein, we show that parallel β sheets becomes steadily more stable as the strands are lengthened, if the added residues have a high β-sheet propensity.

Our experimental design is outlined in Figure 1a. A diamine segment containing D-proline is used to link two strand-forming peptide segments through their C termini.[10,23] This linker promotes but does not strictly enforce β-sheet interactions between attached strands in aqueous solution. The extent of β-sheet formation is monitored by NMR spectroscopy[24,25] by measuring α-proton chemical shifts ($\delta C_\alpha H$) at "indicator" residues (black dots) within the core segment (black) as the strands are extended incrementally (blue). The indicator residues are insulated from the extension segments by an intervening residue; therefore, changes in $\delta C_\alpha H$ value induced by varying extension length should reflect a global change in β-sheet population. Our initial extension sequences contain valine (Val, V) and threonine (Thr, T), residues that have a high intrinsic β-sheet propensity because of branching in the side chain.[26–28] For the extensions shown in Figure 1a, each added Thr residue should form backbone H-bonds to the opposite strand, while each new Val should present its backbone H-bonding groups (C=O and N–H) to the solvent. Members of this series are designated $\mathbf{P(T\rightarrow V)}_n$, where n is the number of residues in each extension segment; the peptide is **P** when $n = 0$. Our design includes Coulombic attractions at the N termini of the strand segments[29] (red), which are intended to minimize fraying of the open end of the parallel hairpin. Sedimentation equilibrium analytical ultracentrifugation, variable concentration NMR spectroscopy, and diffusion oriented spectroscopy (DOSY) NMR studies suggest that **P** and the other peptides discussed do not aggregate under the conditions used to obtain the NMR data (see Figures S8–S12 in the Supporting Information).

Two-dimensional NMR analysis of parent peptide **P** indicates that the intended parallel-β-sheet conformation is formed in aqueous solution. Multiple NOE signals are observed between side chains that are distant from one another in terms of sequence, but that are expected to lie near one another in the parallel β sheet (see Figure S6 in the Supporting Information). No NOE signals inconsistent with the expected parallel-β-sheet conformation were observed.

A set of ten low-energy structures for **P** (backbone only) derived from NOE-restrained molecular dynamics simulations is shown in Figure 2a. The structures vary significantly from one another, but the presence of a consistent hairpin fold is evident from the overlay. This image reflects a feature that is probably common to many peptides and small

[*] Dr. F. Freire,[+] A. M. Almeida,[+] Dr. J. D. Fisk, J. D. Steinkruger, Prof. S. H. Gellman
Department of Chemistry, University of Wisconsin–Madison
1101 University Av., Madison, WI 53726 (USA)
E-mail: gellman@chem.wisc.edu

[+] These authors contributed equally to this work.

[**] This research was supported by the NIH (GM-061238). F.F. was supported in part by an MEC-Fulbright Postdoctoral Fellowship. J.D.F. was supported in part by a biophysics training grant from NIGMS. NMR spectrometers were purchased with partial support from NIH and NSF.

Supporting information for this article (experimental details, compound characterizations, SEAU data, NMR data, and thermodynamic analysis) is available on the WWW under http://dx.doi.org/10.1002/anie.201102986.

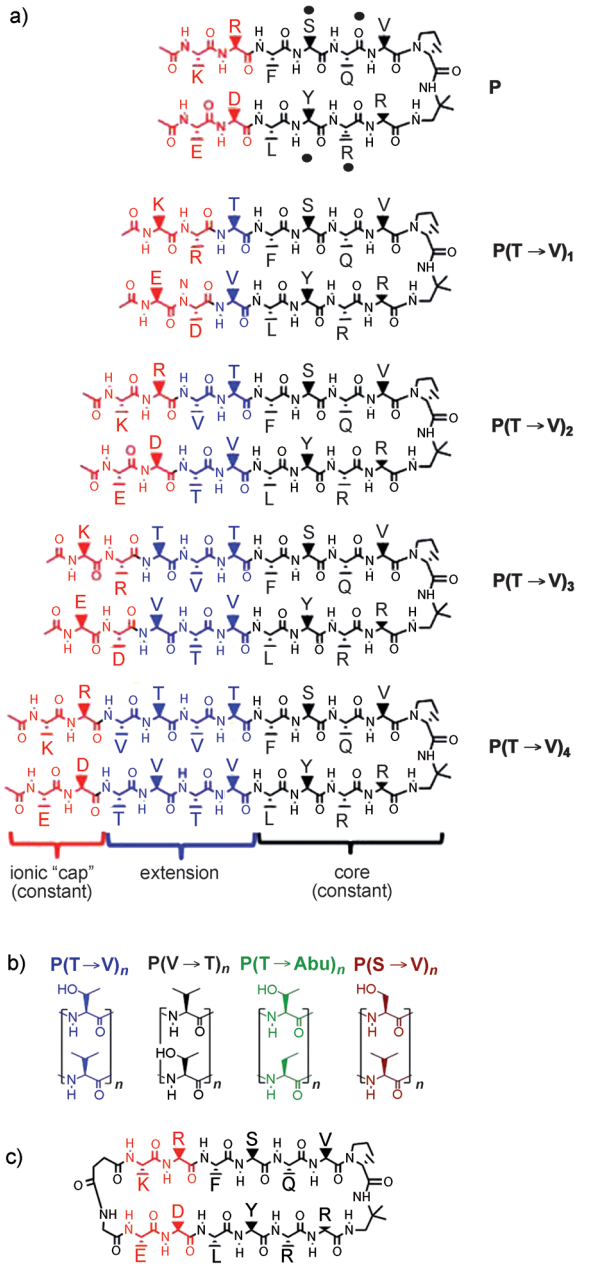

Figure 1. Experimental design for elucidating the impact of the strand length on the stability of the parallel-β-sheet secondary structure. a) Sequences of peptides **P** and the members of the series **P(T→V)$_n$**, with colors indicating key regions within each molecule: the core region (black), which is constant in all peptides; the extension segments (blue), which vary in length across the **P(T→V)$_n$** and vary in composition in other series; and the ionic capping segments (red), which are constant in all peptides and intended to ensure favorable Coulombic interactions between the N-termini of the strands. Black dots highlight the indicator residues, NMR data from which were used to calculate the extent of parallel-β-sheet formation in each molecule. b) Sequences used for the extension segments. c) Macrocyclic peptide **P-cyc**, which is used to obtain NMR data for the folded (parallel-β-sheet) state of linear peptides; color coding as in part (a). Capital letters represent residues of the following amino acids: D = aspartic acid, E = glutamic acid, F = phenylalanine, K = lysine, L = leucine, Q = glutamine, R = arginine, S = serine, T = threonine, V = valine, Y = tyrosine.

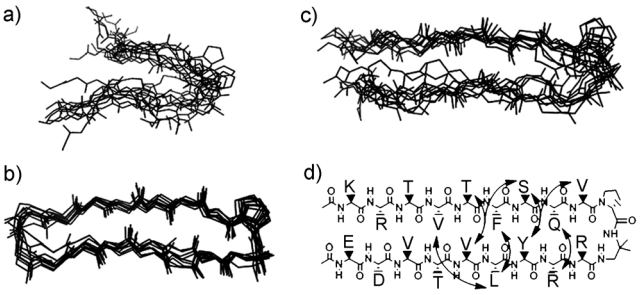

Figure 2. NMR-based structural analysis of peptides that adopt parallel-β-sheet conformations in aqueous solution. Overlays of the ten structures with lowest calculated energies from NOE-restrained molecular dynamics simulations for **P** (a), and **P-cyc** (b), and **P(T→V)$_3$** (c) based on data obtained for 2.5 mM peptide solutions in 100 mM [D$_3$]acetate, pH 3.8, at 4 °C. In each case, only backbone atoms are shown, except for the proline residue. For **P**, the root-mean-square deviation (rmsd) among backbone atoms is (2.5 ± 0.7) Å. For **P-cyc**, the rmsd among backbone atoms is (0.54 ± 0.25) Å. For **P(T→V)$_3$**, the rmsd among backbone atoms is (1.6 ± 0.5) Å. d) Graphical summary of interstrand NOE patterns observed for **P(T→V)$_3$**. Each curved arrow signifies that at least one NOE signal was observed between protons in the two residues indicated.

proteins: the folded state is not a single, precisely defined structure, but rather an ensemble of closely related structures.[30] Figure 2b shows the overlay of ten low-energy backbone structures derived from NOE data for **P-cyc** (Figure 1c), a macrocyclic analogue of **P**. The covalent constraint resulting from backbone closure is expected to stabilize the parallel-β-sheet state of **P-cyc**,[10,25] and the tighter definition of the low-energy structure set for **P-cyc** relative to **P** is consistent with this expectation.

A key assumption underlying our experimental design is that strand-extended derivatives of parent peptide **P** experience two-state folding behavior, involving an unfolded state and a parallel-β-sheet state with strand register matching that of **P**. We tested this assumption with 2D NMR studies. For **P(T→V)$_4$**, resonance overlap made it impossible to assign interstrand NOE signals conclusively; however, for **P(T→V)$_3$** we could identify a large number of interstrand NOE signals, all consistent with the expected parallel-β-sheet conformation (Figure 2c,d). These results suggest that strand extensions lead to propagation and enhancement of the β-sheet structure evident in **P**.

Peptide **P** is expected to equilibrate rapidly between unfolded and folded states on the NMR time scale, and $\delta C_\alpha H$ data for **P** should therefore reflect a population-weighted average of contributions from these two states (the $\delta C_\alpha H$ value for a residue in a β sheet occurs downfield from the $\delta C_\alpha H$ value for that residue in an unfolded segment[24]). The β-sheet population can be inferred from $\delta C_\alpha H$ data by using designed peptides that approximate the fully unfolded state and the fully folded state.[25,31] For **P**, the unfolded state is modeled by **P-L**, the diastereomer differing from **P** only in the configuration of the proline residue; previous work has demonstrated that L-Pro in the linker prevents sheet formation between L-residue strands.[10,25] The fully folded state is modeled by **P-cyc**. At 4 °C, in 100 mM acetate buffer, pH 3.8,

the β-sheet population of **P** can be estimated from $\delta C_\alpha H$ data for each of the four indicator residues, Ser (ca. 26%) and Gln (ca. 39%) in the upper strand, and Tyr (ca. 22%) and Arg (ca. 22%) in the lower strand. This variation in estimated parallel-β-sheet population reflects limitations inherent in conformational population analysis of dynamic peptides.[30] The "true" β-sheet population is presumably near the average of the four values (ca. 27%). Our experimental design requires that parallel-β-sheet population not be too high for **P**, so that the system can sensitively report on conformational stabilization resulting from strand extension.

Chemical shift data for the **P(T→V)$_n$** series suggest that the extent of parallel-β-sheet folding increases as the strands are lengthened. Figure 3a shows $\Delta\delta C_\alpha H$ data for the four indicator residues in each peptide. For each residue in a given peptide, $\Delta\delta C_\alpha H = [\delta C_\alpha H\,(\text{observed}) - \delta C_\alpha H\,(\textbf{P-L})]$, where $\delta C_\alpha H$ (**P-L**) is the chemical shift from the appropriate residue in unfolded reference **P-L**. Incremental extension is accompanied by a steady increase in the $\Delta\delta C_\alpha H$ value for each indicator residue, which qualitatively suggests that the extent of β-sheet formation increases steadily as the strands grow longer.[24]

Chemical shift data were used to assess β-sheet population for members of the **P(T→V)$_n$** series. The $\delta C_\alpha H$ value for each indicator residue in a member of the series was used along with the $\delta C_\alpha H$ value for the analogous residue in **P-L**, representing the fully unfolded state, and the $\delta C_\alpha H$ value for the analogous residue in **P-cyc**, representing the fully folded state (see Figure S7 in the Supporting Information), to estimate parallel-β-sheet population based on a two-state assumption (folded vs. unfolded).[25,31] The free energy for folding (ΔG_f) was calculated from the NMR-derived β-sheet population. Members of the **P(T→V)$_n$** series were compared in terms of the $\Delta\Delta G_f$ value, calculated relative to parent peptide **P** (Figure 3b). Although there is some variation in the $\Delta\Delta G_f$ values among indicator residues within each series member, overall agreement is good among these four sites in a given strand-extended peptide. The data show that strand extension increases population of the parallel-β-sheet state within this homologous peptide set.

Bioinformatics analysis of parallel β sheets in proteins suggests that the intrinsic stability of lateral pairs can vary depending upon residue orientation, that is, which residue orients its C=O and N–H groups toward the partner strand.[32] Therefore, we examined a series isomeric to that shown in Figure 1a, in which the positions of the Thr and Val residues in the extensions are exchanged (designated **P(V→T)$_n$**; black in Figure 1b). The $\Delta\Delta G_f$ data in Figure 3c show that strand extensions in this second series lead to parallel-β-sheet stabilization that is quantitatively comparable to the stabilization observed in the **P(T→V)$_n$** series. Relative to **P**, the four-residue extensions in these two isomeric series (**P(T→V)$_4$** and **P(V→T)$_4$**) enhance parallel-β-sheet stability by 1.0 to 1.5 kcal mol^{-1}.

We next examined extension segments containing residues with diminished β-sheet propensity. This end was achieved in two complementary ways, both of which involved removing a subset of side-chain branch points. In one extension series, we replaced Thr with Ser (**P(V→S)$_n$**; brown in Figure 1b), and in a second, isomeric series we replaced Val with the nonproteinogenic residue derived from α-aminobutyric acid (Abu; **P(T→Abu)$_n$**; green in Figure 1b). Within each series, the $\Delta\Delta G_f$ value relative to **P** was calculated from $\delta C_\alpha H$ data as described above. Figure 3d summarizes $\Delta\Delta G_f$ values, averaged over the four indicator residues, for members of these series along with comparable data from fully branched series **P(T→V)$_n$** and **P(V→T)$_n$**. The data show that simply removing one-half of the side-chain branch points within an extension series, as in **P(S→V)$_n$** or **P(T→Abu)$_n$**, leads to the loss of the length-dependent stabilization trend that is clearly manifested when all side chains are branched.

The data in Figure 3d suggest that the β-sheet propensities of component residues determine whether a given

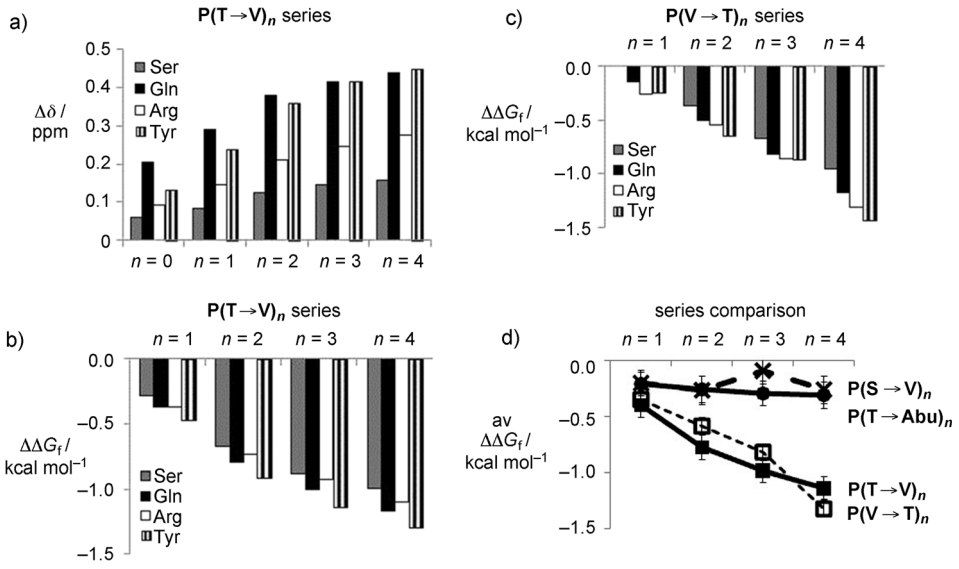

Figure 3. NMR-based analysis of the thermodynamics of parallel-β-sheet formation as a function of increasing strand length. a) ^{1}H NMR chemical shift data for protons attached to C_α ($\delta C_\alpha H$) of indicator residues in **P** ($n=0$) and members of the **P(T→V)$_n$** series. b) Difference in ΔG_f value for parallel-β-sheet formation in members of the **P(T→V)$_n$** series relative to ΔG_f value for parallel-β-sheet formation in **P** ($\Delta\Delta G_f$) for each of the four indicator residues. c) Difference in ΔG_f value for parallel-β-sheet formation in members of the **P(V→T)$_n$** series relative to ΔG_f value for parallel-β-sheet formation in **P** for each of the four indicator residues. d) Comparison of different extension strand compositions in terms of the effect of strand lengthening on the free energy of parallel-β-sheet formation, as manifested $\Delta\Delta G_f$ (relative to **P**), with $\Delta\Delta G_f$ averaged over the four indicator residues in each case. Lines connecting the symbols are arbitrary. The error bars indicate standard deviation derived from averaging the values for the four indicator residues. Arg = arginine, Gln = glutamine, Ser = serine, Tyr = tyrosine.

sequence can support parallel-β-sheet stabilization by strand extension. Many β-sheet-propensity scales have been constructed for the proteinogenic amino acid residues,[26,27] and most indicate that the propensity of Ser is substantially lower than that of Thr. Physicochemical principles suggest that the nonproteinogenic Abu residue should have a lower β-sheet propensity relative to Val because there is no branch point adjacent to the backbone.

Extrapolating from the standard explanation for length-dependent stabilization of the α helix,[14–18] we speculate that the turn-forming linker and neighboring residues initiate the two-stranded parallel-β-sheet formation, and incorporation of additional residues into the sheet constitutes propagation.[21,22] Adding four high-propensity residues (e.g., Ala) is expected to stabilize an α helix by 0.65 kcal mol^{-1},[33] and our data reveal a similar extent of stabilization from adding four favorable residues (two per strand) to a two-stranded parallel β sheet (**P** vs. **P(T→V)$_2$** or **P(V→T)$_2$**; Figure 3).

We previously examined the effect of strand extension on antiparallel-β-sheet stability through an approach related to the one employed here.[19] Extension of the two antiparallel strands from five to seven residues was moderately stabilizing (0.3 to 0.6 kcal mol^{-1}), whether high-propensity (pure Thr) or mixed-propensity (Thr–Ser) sequences were employed. However, extension of each strand by a further two residues did not lead to a clear additional stabilization of the antiparallel β sheet in any case. This pattern is similar to our observations for mixed-propensity strands in parallel β sheet (**P(S→V)$_n$** and **P(T→Abu)$_n$**; Figure 3d), but differs from our observations for high-propensity strands (**P(T→V)$_n$** and **P(V→T)$_n$**).

The results presented here provide fundamental insight on a common protein architectural motif by documenting the relationship between strand length and conformational stability for parallel-β-sheet secondary structure. Our experiments have been limited to two-stranded β sheets, but the trends we observe are presumably relevant to parallel β sheets containing three or more strands within proteins. Thus, the trends we have uncovered should contribute to the complex network of factors that determines protein folding preferences,[34] and may be of particular significance in families that are rich in parallel-β-sheet secondary structure, such as leucine-rich repeat proteins[35] and parallel-β-helix proteins.[36] Previous work suggests that antiparallel-β-sheets are less susceptible to length-dependent stabilization than we find to be true of parallel β sheets.[19] This apparent difference between strand topologies may underlie an apparent preference for parallel orientations among intermolecular β-sheet interactions within amyloid assemblies.[37–39]

Received: April 30, 2011
Published online: August 2, 2011

Keywords: β sheets · biophysics · peptides · protein folding · protein structures

[1] A. Chakrabartty, R. L. Baldwin, *Adv. Protein Chem.* **1995**, *46*, 141–176.
[2] H. A. Scheraga, *Pure Appl. Chem.* **1978**, *50*, 315–324.
[3] D. S. Kemp, J. G. Boyd, C. C. Muendel, *Nature* **1991**, *352*, 451–454.
[4] M. S. Searle, B. Ciani, *Curr. Opin. Struct. Biol.* **2004**, *14*, 458–464.
[5] R. M. Hughes, M. L. Waters, *Curr. Opin. Struct. Biol.* **2006**, *16*, 514–524.
[6] B. L. Kier, I. Shu, L. A. Eidenschink, N. H. Andersen, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10466–10471.
[7] J. M. Gao, D. A. Bosco, E. T. Powers, J. W. Kelly, *Nat. Struct. Mol. Biol.* **2009**, *16*, 684–690.
[8] J. S. Nowick, *Acc. Chem. Res.* **2008**, *41*, 1319–1330.
[9] P. Chitnumsub, W. R. Fiori, H. A. Lashuel, H. Dias, J. W. Kelly, *Bioorg. Med. Chem.* **1999**, *7*, 39–59.
[10] J. D. Fisk, S. H. Gellman, *J. Am. Chem. Soc.* **2001**, *123*, 343–344.
[11] F. Freire, S. H. Gellman, *J. Am. Chem. Soc.* **2009**, *131*, 7970–7972.
[12] J. M. Scholtz, Q. Hong, E. J. York, J. M. Stewart, R. L. Baldwin, *Biopolymers* **1991**, *31*, 1463–1470.
[13] C. A. Rohl, J. M. Scholtz, E. J. York, J. M. Stewart, R. L. Baldwin, *Biochemistry* **1992**, *31*, 1263–1269.
[14] B. H. Zimm, P. Doty, K. Iso, *Proc. Natl. Acad. Sci. USA* **1959**, *45*, 1601–1607.
[15] M. Goodman, A. S. Verdini, C. Toniolo, W. D. Phillips, F. A. Bovey, *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 444.
[16] B. H. Zimm, J. K. Bragg, *J. Chem. Phys.* **1959**, *31*, 526–535.
[17] S. Lifson, A. Roig, *J. Chem. Phys.* **1961**, *34*, 1963–1974.
[18] H. Qian, J. A. Schellman, *J. Phys. Chem.* **1992**, *96*, 3987–3994.
[19] H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Giriat, T. Muir, S. H. Gellman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12015–12020.
[20] S.-J. Chen, K. A. Dill, *J. Chem. Phys.* **1998**, *109*, 4602–4616.
[21] J. K. Sun, A. J. Doig, *J. Phys. Chem. B* **2000**, *104*, 1826–1836.
[22] L. Hong, *J. Chem. Phys.* **2008**, *129*, 225101.
[23] J. D. Fisk, D. R. Powell, S. H. Gellman, *J. Am. Chem. Soc.* **2000**, *122*, 5443–5447.
[24] D. S. Wishart, B. D. Sykes, F. M. Richards, *Biochemistry* **1992**, *31*, 1647–1651.
[25] J. D. Fisk, M. A. Schmitt, S. H. Gellman, *J. Am. Chem. Soc.* **2006**, *128*, 7148–7149.
[26] P. Y. Chou, G. D. Fasman, *Annu. Rev. Biochem.* **1978**, *47*, 251–276.
[27] L. Muñoz, L. Serrano, *Proteins Struct. Funct. Genet.* **1994**, *20*, 301–311.
[28] C. Toniolo, *Macromolecules* **1978**, *11*, 437–438.
[29] B. Ciani, M. Jourdan, M. S. Searle, *J. Am. Chem. Soc.* **2003**, *125*, 9038–9047.
[30] H. J. Dyson, P. E. Wright, *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519–538.
[31] F. A. Syud, J. F. Espinosa, S. H. Gellman, *J. Am. Chem. Soc.* **1999**, *121*, 11577–11578.
[32] H. M. Fooks, A. C. R. Martin, D. N. Woolfson, R. B. Sessions, E. G. Hutchinson, *J. Mol. Biol.* **2006**, *356*, 32–44.
[33] T. E. Creighton, *Proteins*, 2nd ed., W. H. Freeman & Co., New York, **1993**, pp. 185.
[34] K. A. Dill, *Biochemistry* **1990**, *29*, 7133–7155.
[35] J. Bella, K. L. Hindle, P. A. McEwan, S. C. Lovell, *Cell. Mol. Life Sci.* **2008**, *65*, 2307–2333.
[36] J. H. Choi, C. Govaerts, B. C. H. May, F. E. Cohen, *Proteins Struct. Funct. Bioinf.* **2008**, *73*, 150–160.
[37] R. Tycko, *Q. Rev. Biophys.* **2006**, *39*, 1–55.
[38] R. Tycko, R. Savtchenko, V. G. Ostapchenko, N. Makarava, I. V. Baskakov, *Biochemistry* **2010**, *49*, 9488–9497.
[39] C. Liu, M. R. Sawaya, P.-N. Cheng, J. Zheng, J. S. Nowick, D. Eisenberg, *J. Am. Chem. Soc.* **2011**, *133*, 6736–6744.

