

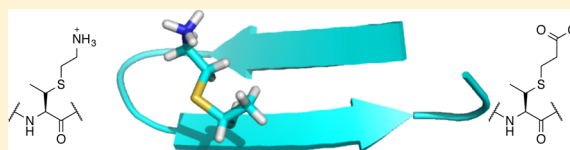
New Charge-Bearing Amino Acid Residues That Promote β -Sheet Secondary Structure

Stacy J. Maynard, Aaron M. Almeida, Yasuharu Yoshimi,[†] and Samuel H. Gellman*

Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, United States

S Supporting Information

ABSTRACT: Proteinogenic amino acid residues that promote β -sheet secondary structure are hydrophobic (e.g., Ile or Val) or only moderately polar (e.g., Thr). The design of peptides intended to display β -sheet secondary structure in water typically requires one set of residues to ensure conformational stability and an orthogonal set, with charged side chains, to ensure aqueous solubility and discourage self-association. Here we describe new amino acids that manifest substantial β -sheet propensity, by virtue of β -branching, and also bear an ionizable group in the side chain.



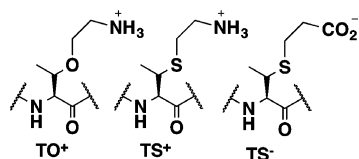
Here we describe new amino acids that manifest substantial β -sheet propensity, by virtue of β -branching, and also bear an ionizable group in the side chain.

INTRODUCTION

β -Sheet is a very common motif within folded proteins. Small, autonomously folding peptides that display β -sheet secondary structure in the absence of a specific tertiary context are useful tools for exploring intrinsic relationships between sequence and β -sheet stability¹ and for establishing characteristic spectroscopic signatures of this motif.² In addition, autonomously folding systems have been valuable for probing noncovalent interactions involving post-translationally introduced units.³ Engineered β -sheets have been used to explore preferred modes of intermolecular β -strand associations, phenomena that underlie amyloid formation;⁴ such β -sheet designs may ultimately lead to diagnostic or therapeutic tools for amyloid diseases.⁵

Design of β -sheet-forming peptides is challenging because the natural residues that promote this secondary structure are hydrophobic (Ile, Val, Phe, Trp, Tyr) or only moderately polar (Thr).⁶ Therefore, sequences intended to adopt β -sheet conformations but not self-associate in aqueous solution must be rich both in β -promoting residues and in solubility-promoting residues, the latter usually bearing charged side chains. The necessity of using distinct sets of residues for conformational propensity and solubility represents a significant design constraint. Here we explore the hypothesis that β -sheet propensity and peripheral charge can be combined in a single α -amino acid residue.

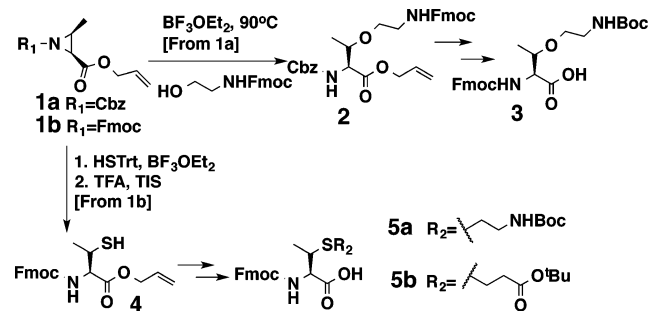
Side chain branching adjacent to the backbone (β -branching), as in Ile, Val and Thr, is correlated with high β -sheet propensity;⁷ therefore, we have synthesized and evaluated new amino acids that contain both a β -branch point and an ionizable group in the side chain (Figure 1).



RESULTS AND DISCUSSION

Initial Design: Ether Derivative of Thr. Our initial design hypothesis focused on derivatives of threonine in which the side chain hydroxyl is used to form an ether linkage to a unit bearing an ionizable group. Scheme 1 shows the synthesis of protected

Scheme 1. Synthesis of Amino Acids



amino acid 3 from aziridine 1, which was generated from L-threonine via well-precedented methods.^{8a} Vederas et al. have shown that nucleophiles can open closely related Thr-derived aziridines in the presence of BF_3 with high stereoselectivity,⁸ and this approach was useful in our case. The residue derived from 3, which we designate TO^+ , was readily incorporated into synthetic peptides via Fmoc-based solid-phase synthesis. TO^+ can be viewed as an analogue of Lys that contains a β -branch point in the side chain.

We used 12-mer peptide I as a benchmark for evaluating the β -sheet propensity of TO^+ and other new residues described below (Figure 1). The conformational behavior of I was previously characterized via NMR.⁹ In aqueous solution this peptide adopts an antiparallel two-stranded β -sheet conformation (β -hairpin) with a reverse turn at the DPro-Gly segment.

Received: October 6, 2014

Published: November 13, 2014

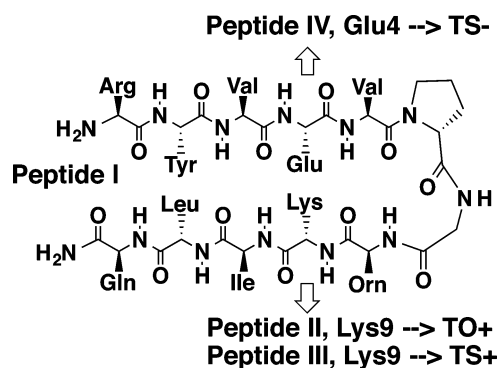


Figure 1. Sequences of hairpin peptides.

D-Proline was employed to promote a type II' reverse turn, which favors β -sheet interactions between flanking strands composed of L-residues.¹⁰ NMR data indicate ~68% population of the β -hairpin conformation at 4 °C in aqueous buffer (pH 3.8).^{9b}

Peptide II is the analogue of I in which Lys-9 has been replaced with TO⁺. 2D NMR analysis suggested significant population of the β -hairpin conformation at 4 °C in aqueous buffer (pH 3.8).¹¹ The NOEs observed for II are consistent with adoption of the expected β -hairpin conformation. For additional insight, we compared chemical shifts for α protons ($\delta C_{\alpha}H$) of residues within I and II. $\delta C_{\alpha}H$ values are sensitive to secondary structure, with residues that participate in α -helix displaying upfield shifts and residues that participate in β -sheet showing downfield shifts relative to random coil values.¹²

Figure 2a shows $\Delta\delta C_{\alpha}H$ data, the deviation from the random coil chemical shift at each residue, for peptides I and II. The LPro-6 diastereomers of I and II were used to provide the random coil $\delta C_{\alpha}H$ values because we have previously shown that replacing DPro with LPro abolishes β -hairpin folding in I and comparable designs.⁹ For both I and II, $\Delta\delta C_{\alpha}H \geq 0.2$ ppm for the segments Tyr-2 to Val-5 and Orn-8 to Ile-10, which is consistent with the expected locations of the two β -strand segments. The consistent negative $\Delta\delta C_{\alpha}H$ values for Leu-11 presumably arise from the proximity of the Tyr side chain to Leu-11 C α H in the β -hairpin conformation. The near-zero values at dPro-6 and Gly-7 are consistent with the expected turn, and the near-zero values at Arg-1 and Gln-12 are consistent with “fraying” at the termini.

The $\Delta\delta C_{\alpha}H$ comparison between I and II suggests that the new TO⁺ residue may have a slightly lower β -sheet propensity than does Lys. This surprising conclusion emerges because the absolute $\Delta\delta C_{\alpha}H$ values for II are smaller than those for I at most strand residues common to the two peptides.

Redesign: Thioether Derivatives of Thr. The unexpected behavior of TO⁺ led us to re-evaluate our design hypothesis, which focused exclusively on β -branching in the side chain. The side chain oxygen of Thr can form an H-bond with a nearby backbone N–H, which leads to Thr backbone torsion angles (ϕ and ψ) more compatible with the α -helical than β -sheet secondary structure.^{7a,b} An ether is a better H-bond acceptor than a comparable alcohol.^{13a} If the H-bond acceptor ability of the ether oxygen in TO⁺ works against β -sheet propensity, then we hypothesized that the desired conformational properties should be achieved by replacing this oxygen atom with sulfur, to generate TS⁺. Thioethers are much poorer H-bond acceptors than are ethers.¹³ A protected α -amino acid that could be used to incorporate TS⁺ residues was prepared as shown in Scheme

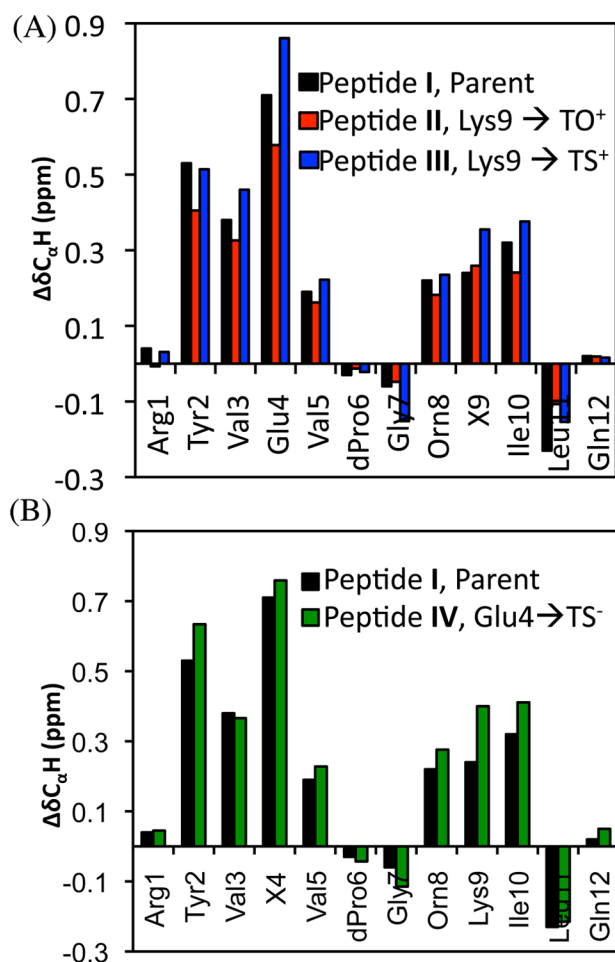


Figure 2. Comparison of $\Delta\delta C_{\alpha}H$ values for the (A) parent, TS⁺, and TO⁺ peptides and (B) parent and TS⁻ peptides. All peptides were referenced relative to unfolded controls of the same sequence. The $\Delta\delta C_{\alpha}H$ values at or adjacent to the substitution cannot be directly compared as the substitution changes the dynamic range. $\Delta\delta C_{\alpha}H$ values at hydrogen bonding positions in the core of the peptide have been shown to most accurately reflect the population of the β -hairpin.^{2b}

1; the key step was use of trityl thiol to open the Thr-derived aziridine.¹⁴ Protected amino acid **5a** allowed solid-phase synthesis of peptide **III**, which contains TS⁺ in place of Lys-9 of I.

2D NMR analysis of **III** revealed interstrand NOEs consistent with significant population of the expected β -hairpin conformation in aqueous buffer at 4 °C. Resonance overlap hindered identification of NOEs involving the Tyr-2 side chain at this temperature, but at 10 °C such NOEs could be detected (Figure 3). All NOEs are consistent with the expected β -sheet structure.

$\Delta\delta C_{\alpha}H$ data for **III** (Figure 2a) suggest that the extent of β -hairpin formation is higher for this peptide than for **II**, which contains the TO⁺ residue, because the absolute value of $\Delta\delta C_{\alpha}H$ is larger at each strand residue (positions 2–5 and 8–11) for **III** than for **II**. Moreover, the $\Delta\delta C_{\alpha}H$ values are larger at most strand positions for **III** relative to parent peptide **I**, which suggests that the TS⁺ residue stabilizes the β -hairpin conformation relative to Lys.

We explored the scope of this design strategy by evaluating the residue TS⁻, an analogue of TS⁺ that bears an acidic rather

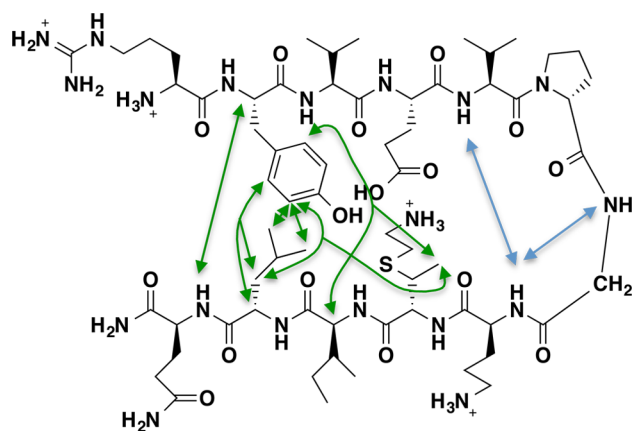


Figure 3. Cross-strand NOEs observed for peptide **III** at 10 °C in pH 3.8 aqueous buffer.

than a basic side chain along with a thioether-based β -branch point. Peptide **IV**, prepared using **5b**, is an analogue of **I** in which Glu-4 is replaced by TS^- . The NOEs observed at 4 °C are consistent with the formation of the expected β hairpin. $\Delta\delta C_{\alpha}H$ data for **IV** (Figure 2b) suggest that this peptide forms a more stable β -hairpin than that formed by **I**. Circular dichroism spectra of peptides **I**, **III**, and **IV** (phosphate buffer, pH = 7.0) are consistent with the formation of a β hairpin and the stability trends observed by NMR.¹¹

$\Delta\delta C_{\alpha}H$ -based population analysis was undertaken in order to estimate the thermodynamic impact of replacing Lys-9 with either TO^+ or TS^+ or replacing Glu-4 with TS^- . Residues 3, 5, 8, and 10 occupy hydrogen bonded positions within the β -hairpin conformation adopted by **I–IV**, and we previously found that these positions are optimal “indicators” for population analysis in **I** and related peptides.^{2b,9b} Such peptides are presumed to equilibrate rapidly between β -hairpin and unfolded states on the NMR time scale. The $\Delta\delta C_{\alpha}H$ value measured at each indicator residue represents a populated-weighted average of the contributions from the fully unfolded state and the fully folded (i.e., β -hairpin) state. For each sequence, we use the LPro-6 diastereomer to estimate $\Delta\delta C_{\alpha}H$ for each indicator residue in the fully unfolded state, and we use a macrocyclized analogue with DPro–Gly turns at both ends to estimate $\delta C_{\alpha}H$ in the fully folded state.¹¹ (The reference peptides for **III** are shown in Figure 4.) Table 1 shows the β -hairpin population deduced in this way at each of the four indicator positions in peptides **I–IV**. The variation among the four values for each molecule is a measure of the intrinsic uncertainty associated with this quantification strategy, which involves independent analysis at four distinct sites within each peptide. Despite this uncertainty, the data support the qualitative conclusion that replacing Lys-9 of **I** with TO^+ (**II**) has little effect on β -hairpin population, while replacing Lys-9 with TS^+ (**III**) or Glu-4 with TS^- (**IV**) leads to enhanced β -hairpin population.

Table 1 shows $\Delta\Delta G_{\text{fold}}$ calculated for peptides **II–IV** relative to parent **I**, and for **III** relative to **II**, based on the residue-specific population. This thermodynamic analysis shows that replacing Lys with TO^+ , the basic Thr derivative with an ether linkage at the branch point, fails to enhance β -hairpin stability, which invalidates our original design hypothesis. In contrast, the analogous basic residue containing a thioether linkage at the branch point, TS^+ , stabilizes the β -hairpin conformation. Although the $\Delta\Delta G_{\text{fold}}$ values for **II** vs **I** and for **III** vs **I** are

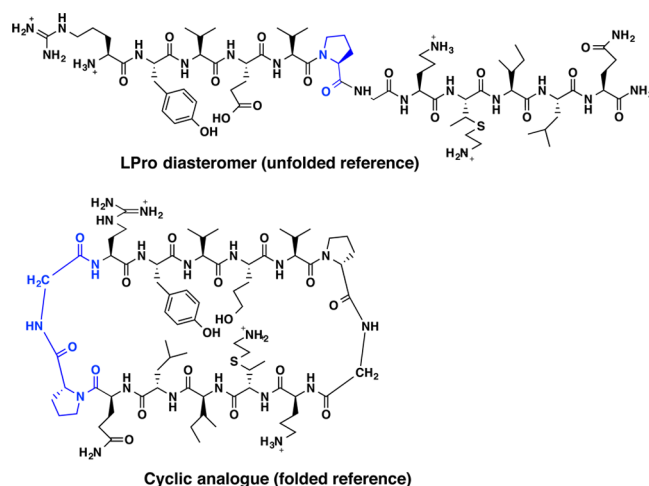


Figure 4. Reference peptides for thermodynamic analysis. Differences from the peptide of interest are highlighted in blue.

Table 1. Population and Thermodynamic Analysis for the Reporter Positions in Hairpins **I–IV**^a

peptide	Val 3 (%)	Val 5 (%)	Orn 8 (%)	Ile 10 (%)	$\Delta\Delta G_{\text{fold}}$ (kcal/mol)
I ^b	70	76	65	59	
II (TO^+) ^c	62	77	63	51	+0.1 \pm 0.2
III (TS^+) ^c	85	93	76	71	−0.5 \pm 0.4
IV (TS^-) ^d	93	96	84	77	−0.6 \pm 0.3

^aThe $\Delta\Delta G$ was calculated for each reporter position in the hairpin and then averaged. ^bData from Syud et al. (ref 9a). ^cReplacement for Lys-9. For **II** relative to **III** $\Delta\Delta G = -0.8 \pm 0.5$ kcal/mol. ^dReplacement for Glu-4.

not cleanly distinguished given the uncertainties in Table 1, direct comparison of **III** vs **II** shows unambiguously that TS^+ enhances β -hairpin stability relative to TO^+ . The magnitude of the β -sheet stabilization provided by TS^+ or TS^- relative to proteinogenic residues with unbranched side chains, Lys and Glu, respectively, is significant given that the natively folded state of a typical globular protein, containing a few hundred residues, is only ~ 5 kcal/mol more stable than the unfolded state.¹⁵

It should be noted that differences between $\Delta\delta C_{\alpha}H$ values (bar heights) from Figure 2a and % folded values from Table 1 are not directly comparable. Each $\Delta\delta C_{\alpha}H$ value in Figure 2a is determined by the difference between the chemical shift ($\delta C_{\alpha}H$) for a given residue in one of three peptides (**I**, **II** or **III**) and the analogous chemical shift from a reference peptide that has L-proline in place of D-proline. The L-proline peptide represents the fully unfolded state. The % folded values in Table 1 are generated from an algebraic expression based on three chemical shifts, the two mentioned above and the $\delta C_{\alpha}H$ value for the analogous cyclic peptide, which represents the fully folded state. Since this last value can vary among macrocyclic peptides with different sequences, quantitative correlations between $\Delta\delta C_{\alpha}H$ values (Figure 2a) and % folded values (Table 1) may not be evident at specific residues, particularly at or adjacent to substitution positions. In addition, another factor may prevent such quantitative correlations: the relationship between changes in $\Delta\delta C_{\alpha}H$ and changes in % folding is not linear. For these reasons, and to account for local factors that can influence $\delta C_{\alpha}H$ values at particular residues, we

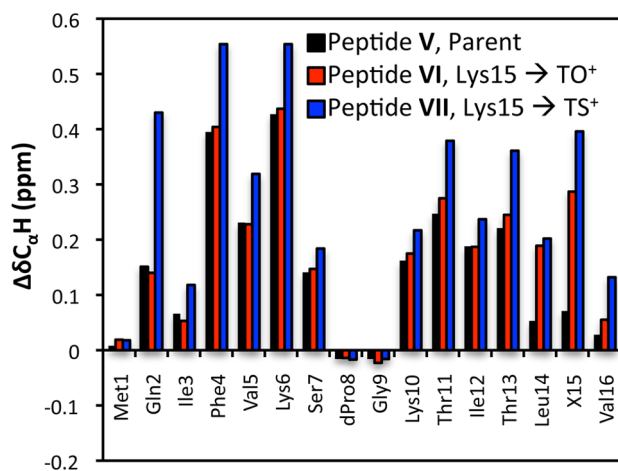


Figure 7. Comparison of $\Delta\delta C_{\alpha}H$ values for the parent, TS^+ , and TO^+ peptides. All peptides were referenced relative to unfolded controls of the parent. The $\Delta\delta C_{\alpha}H$ values at or adjacent to the substitution cannot be directly compared as the substitution changes the dynamic range. The $\Delta\delta C_{\alpha}H$ values at hydrogen bonding positions in the core of the peptide have been shown to most accurately reflect the population of the β -hairpin.^{2b}

useful design tool because one can match the charge provided by Glu while enhancing the folding tendency.

Hydrophilicity Assessments. Our design strategy is based on the assumption that new amino acid residues with a side chain that contains both a β -branch point and an ionizable group will be more hydrophilic than the proteinogenic β -branched residues (Thr, Ile and Val). To test this hypothesis, we evaluated the hydrophilicities of TS^+ and TS^- along with selected proteinogenic residues, Lys, Glu, Gly, Thr and Ile, using a previously described system.¹⁶ This method employs the *N*-4-nitrobenzoyl derivatives of the amino acids. Distribution coefficients are determined (at equilibrium) between equal volumes of octanol and aqueous buffer (100 mM phosphate, pH 7.0). The parameter of comparison, Π , is normalized: the logarithm of the distribution coefficient of glycine, $\log(D_{\text{Glycine}})$, is subtracted from the logarithm of distribution coefficient of the amino acid under consideration, $\log(D_{\text{amino acid}})$ to calculate Π for that amino acid. Table 2 shows Π values measured for TS^+ and TS^- and the selected proteinogenic residue.

Table 2. Normalized Octanol/Water Distribution Coefficients of Selected 4-Nitrobenzoyl Amino Acid Derivatives

amino acid	Π
Lys	-2.44 ± 0.02
TS^+	-1.47 ± 0.04
Glu	-1.85 ± 0.4
TS^-	-1.75 ± 0.2
Gly	0
Thr	0.07 ± 0.01
Ile	1.59 ± 0.08

As we predicted, TS^+ and TS^- cluster with Lys and Glu, all significantly preferring aqueous buffer relative to octanol. It is interesting to note that TS^- and Glu are very similar on this scale, while TS^+ is somewhat less hydrophilic than Lys. In contrast, Ile significantly prefers octanol to aqueous buffer,

while Thr is similar to Gly. The data clearly show that TS^+ and TS^- are significantly more hydrophilic than Thr or Ile.

CONCLUSIONS

We have identified a new family of unnatural α -amino acid residues featuring two properties, intrinsic conformational propensity and side chain charge, that have proven to be valuable for design of peptides that fold autonomously to β -sheet secondary structure in aqueous solution. These two properties are not paired in any proteinogenic amino acid residue, which has made it challenging to design small peptides that adopt β -sheet conformations but do not aggregate. High β -sheet propensity is generally associated with β -branching in a residue's side chain. Our initial attempt to build charge and side chain branching into a new residue, involving an ether linkage at the branch point, was not successful. Thioether-based branch points, on the other hand, lead to the desired properties. We have illustrated this approach with two new residues, TS^+ and TS^- , which feature basic and acidic side chains, respectively. The versatility of the synthetic route will enable preparation of many related thioether-containing α -amino acids.

ASSOCIATED CONTENT

Supporting Information

Experimental details including NMR spectra, NMR structure calculations, and chemical shift deviation analysis. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

gellman@chem.wisc.edu

Present Address

[†]Department of Applied Chemistry and Biotechnology, Graduate School of Engineering, University of Fukui, 3-9-1 Bunkyo, Fukui 910-8507, Japan.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the NIH (GM-061238). NMR spectrometers were purchased with partial support from NIH (Grant No. 1 S10 RR13866-01) and NSF. Y.Y. thanks the University of Fukui, Japan, for research sabbatical support.

REFERENCES

- (1) (a) Searle, M. S.; Ciani, B. *Curr. Opin. Struct. Biol.* **2004**, *14*, 458. (b) Hughes, R. M.; Waters, M. L. *Curr. Opin. Struct. Biol.* **2006**, *16*, 514. (c) Gao, J. M.; Bosco, D. A.; Powers, E. T.; Kelly, J. W. *Nat. Struct. Mol. Biol.* **2009**, *16*, 684. (d) Kier, B. L.; Shu, I.; Eidenschink, L. A.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10466. (e) Freire, F.; Almeida, A. M.; Fisk, J. D.; Steinkruger, J. D.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2011**, *50*, 8735. (f) Cheng, P. N.; Pham, J. D.; Nowick, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 5477.
- (2) (a) Woys, A. M.; Almeida, A. M.; Wang, L.; Chiu, C. C.; McGovern, M.; Pablo, J. J.; Skinner, J. L.; Gellman, S. H.; Zanni, M. T. *J. Am. Chem. Soc.* **2012**, *134*, 19118. (b) Syud, F. A.; Espinosa, J. F.; Gellman, S. H. *J. Am. Chem. Soc.* **1999**, *121*, 11577.
- (3) Riemen, A. J.; Waters, M. L. *J. Am. Chem. Soc.* **2010**, *132*, 9007.
- (4) Liu, C.; Sawaya, M. R.; Cheng, P. N.; Zheng, J.; Nowick, J. S.; Eisenberg, D. *J. Am. Chem. Soc.* **2011**, *133*, 6736.
- (5) (a) Cheng, P. N.; Liu, C.; Zhao, M.; Eisenberg, D.; Nowick, J. S. *Nat. Chem.* **2012**, *4*, 927. (b) Sievers, S. A.; Karanicolas, J.; Chang, H. W.; Zhao, A.; Jiang, L.; Zirafi, O.; Stevens, J. T.; Münch, J.; Baker, D.; Eisenberg, D. *Nature* **2011**, *475*, 96.

- (6) (a) Chou, P. Y.; Fasman, G. D. *Trends Biochem. Sci.* **1977**, *2*, 128. (b) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510. (c) Minor, D. L.; Kim, P. S. *Nature* **1994**, *371*, 264.
- (7) (a) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596. (b) Munoz, V.; Serrano, L. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 301. (c) Bai, Y.; Englander, S. W. *Proteins: Struct., Funct. Genet.* **1994**, *18*, 262.
- (8) (a) Liu, H.; Pattabiraman, V. R.; Vederas, J. C. *Org. Lett.* **2007**, *9*, 4211. (b) Liu, W.; Chan, A. S. H.; Liu, H.; Cochrane, S. A.; Vederas, J. C. *J. Am. Chem.* **2011**, *133*, 14216. (c) Tarrade, A.; Dauban, P.; Dodd, R. H. *J. Org. Chem.* **2003**, *68*, 9521.
- (9) (a) Syud, F. A.; Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **2001**, *123*, 8667. (b) Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236.
- (10) Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 2303.
- (11) See Supporting Information for details.
- (12) (a) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.
- (13) (a) Kebarle, P. *Annu. Rev. Phys. Chem.* **1977**, *28*, 445. (b) Arnett, E. M.; Mitchell, E. J.; Murty, T. S. S. R. *J. Am. Chem. Soc.* **1974**, *96*, 3875.
- (14) (a) Liu, W.; Chan, A. S. H.; Liu, H.; Cochrane, S. A.; Vederas, J. C. *J. Am. Chem. Soc.* **2011**, *133*, 14216. (b) Narayan, R. S.; VanNieuwenhze, M. S. *Org. Lett.* **2005**, *7*, 2655. (c) Carrillo, A. K.; VanNieuwenhze, M. S. *Org. Lett.* **2012**, *14*, 1034.
- (15) Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J.; Bona, D.; Miller, E. J.; Vallée-Bélisle, A.; Main, E. R. G.; Bemporad, F.; Qiu, L.; Teilum, K.; Vu, N.; Edwards, A. M.; Ruczinski, L.; Poulsen, F. M.; Kragelund, B. B.; Michnick, S. W.; Chiti, F.; Bai, Y.; Hagen, S. J.; Serrano, L.; Oliveberg, M.; Raleigh, D. P.; Wattung-Stafshede, P.; Radford, S. E.; Jackson, S. E.; Sosnick, T. R.; Marqusee, S.; Davidson, A. R.; Plaxco, K. W. *Protein Sci.* **2005**, *14*, 602.
- (16) (a) Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 15932. (b) Wilce, M. C. J.; Aguilar, M.; Hearn, T. W. *Anal. Chem.* **1995**, *67*, 1210.