Aza-Glycine Induces Collagen Hyperstability

Yitao Zhang, Roy M. Malamakal, and David M. Chenoweth*

Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104-6323, United States

S *Supporting Information*

ABSTRACT: Hydrogen bonding is fundamental to life on our planet, and nature utilizes H-bonding in nearly all biomolecular interactions. Often, H-bonding is already maximized in natural biopolymer systems such as nucleic acids, where Watson–Crick H-bonds are fully paired in double-helical structures. Synthetic chemistry allows molecular editing of biopolymers beyond nature's capability. Here we demonstrate that substitution of glycine (Gly) with aza-glycine in collagen may increase the number of interfacial cross-strand H-bonds, leading to hyperstability in the triple-helical form. Gly is the only amino acid that has remained intolerant to substitution in collagen. Our results highlight the vital importance of maximizing H-bonding in higher order biopolymer systems using minimally perturbing alternatives to nature's building blocks.

Biomolecular structure is governed by a delicate balance of non-covalent intra- and intermolecular interactions.[1] These interactions drive macromolecular assembly and intermolecular recognition events critical to all life processes.[2,3] The hydrogen bond is a ubiquitous non-covalent interactions in nature.[2,4,5] Hydrogen bonds are often found at subunit interfaces in helical biopolymer assemblies. Historically, speculative model building by pioneers such as Pauling, Watson, and Crick has led to some of the greatest achievements in structural biology, many of which hinged on the correct pairing or maximizing of H-bonds in biopolymers.[6] Understanding how molecular systems respond to perturbations in their building blocks can provide insight into biomolecule self-assembly, protein folding, drug design, materials and catalysis, in addition to modulating or enhancing the properties of biopolymers to create new biomimetic materials with superior properties.[7] Collagen adopts a densely packed H-bonded structure requiring a precise backbone conformation for proper self-assembly into its right handed triple-helical form.[8] The tertiary structural motif is composed of three staggered polypeptide chains, each chain adopting a left-handed poly-proline type-II helical conformation. A single cross-strand H-bond is present at each amino acid triplet repeat. In recent decades, collagen has been the focus of many research efforts aimed at developing a molecular-level understanding of its self-assembly properties[9] for designing new materials.[10] Collagen triple-helix assembly depends on a balance of non-covalent interactions and amino acid side-chain modifications can modulate the stability.[11] In contrast, collagen backbone modifications, with the exception of a peptoid residue,[12] have typically resulted in significantly destabilized structures.[13] To date, the glycine (Gly) residue in collagen has remained largely
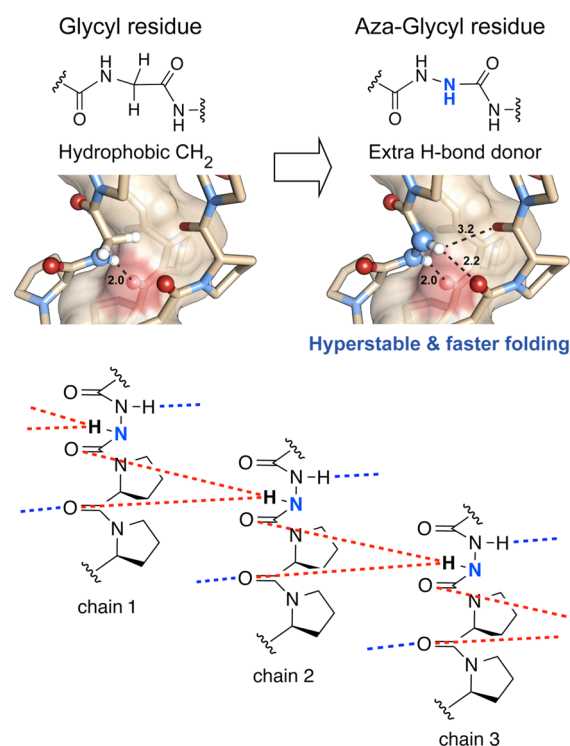

Figure 1. (Top) Glycyl versus aza-glycyl residues: azGly-containing collagen triple helix with an expanded view of the -Pro-Pro-azGly- repeat showing the proposed additional cross-strand H-bonding interactions. (Bottom) H-bond map of collagen containing the unnatural azGly modification. Natural collagen H-bonds are shown as blue dashed lines between strands. Proposed additional H-bonds from azGly substitution are shown as red dashed lines. AzGly N-atom is shown in blue.

intolerant to substitution, barring a recent thioamide substitution by Raines et al.[13d]

Recently, we provided new insight into the fundamental importance of stereochemistry as a pre-organizing element in biomolecular folding using stereodynamic probes.[14] We demonstrated that the rate of triple-helix self-assembly in a stereodynamic collagen model peptide is dramatically altered, with little to no effect on the thermal unfolding. Here we replace the α-C of Gly with a N-atom and show that addition of an extra H-bond donor in the backbone can lead to a hyperstable collagen triple helix (Figure 1). This Communication details the first atomic modification to the main-chain backbone of collagen that has resulted in significant triple-helix hyperstability and

significantly faster folding kinetics. These results have important implications for the design of new biomimetic materials.

We synthesized peptides **1−4** to assess the effects of substituting Gly with aza-glycine (azGly)[15,16] in the context of a collagen model peptide system. Our design constitutes replacement of a single CH with a single N-atom at the central location in a 260-atom (21 amino acid) peptide system (Figure 2). We used circular dichroism (CD) spectroscopy to evaluate
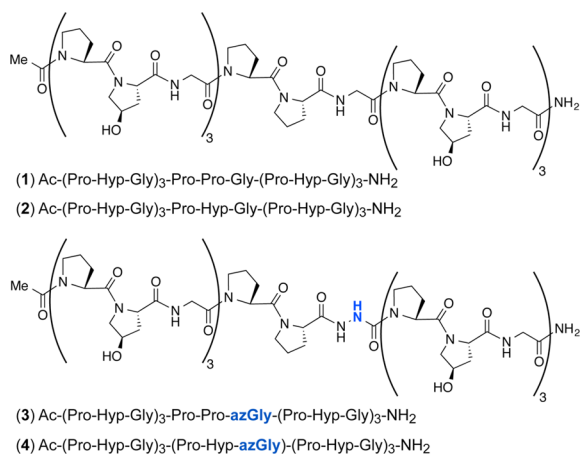

(1) Ac-(Pro-Hyp-Gly)₃-Pro-Pro-Gly-(Pro-Hyp-Gly)₃-NH₂

(2) Ac-(Pro-Hyp-Gly)₃-Pro-Hyp-Gly-(Pro-Hyp-Gly)₃-NH₂

(3) Ac-(Pro-Hyp-Gly)₃-Pro-Pro-**azGly**-(Pro-Hyp-Gly)₃-NH₂

(4) Ac-(Pro-Hyp-Gly)₃-(Pro-Hyp-**azGly**)-(Pro-Hyp-Gly)₃-NH₂

Figure 2. Chemical structure of collagen control peptides **1** and **2** and azGly-containing peptides **3** and **4**.

self-assembly of the triple helix for each collagen peptide. CD spectra of solutions containing **3** and **4** both exhibit characteristic maxima at ~224 nm, indicating the presence of triple-helix structure. Thermal denaturation experiments showed cooperative unfolding transitions for all peptides (see Supporting Information (SI)). The data from CD thermal denaturation experiments were fitted to a two-state model, as previously described.[17] The melting temperatures (T_m), at which 50% of the triple helix unfolds, are shown in Figure 3a for **1−4**. A striking increase in T_m of ~10 °C was observed for **3** and **4** compared to the corresponding control compounds **1** and **2**. The results show that replacement of a single Gly α-CH with a N-atom results in a significant increase in triple-helix thermal stability. In addition to being the first favorable replacement for a Gly residue in a collagen model peptide,[8c,18] this substitution results in the highest stabilizing effect of any single-residue mutation in a collagen PPG or POG peptide system.[19] Raines et al. reported a seminal study where replacement of a single proline (Pro) residue with a fluorinated Pro residue resulted in a 5 °C increase in thermal stability.[20] Pairing the azGly residue with the fluoroproline discovery of Raines could result in extremely stable collagen peptide systems with unique, new material properties.

A model based on a known collagen crystal structure in which a Gly was substituted by azGly shows the azGly α-NH is 2.2 Å from the carbonyl of Gly and 3.2 Å from the carbonyl of Pro in a neighboring peptide chain. This distance is similar to the key canonical interstrand H-bond from the Gly amide NH to the carbonyl preceding Pro in the Yaa position. Additional H-bonding from azGly could increase the number of interchain H-bonds within a triplet of Xaa-Yaa-Gly, providing a connection between all three peptide chains through multiple dynamic H-bonds (Figure 1, bottom).

Next, we assessed the kinetics of triple-helix formation for peptides **1−4**. The peptides, at 0.2 mM in PBS buffer, were

peptide	central triplet	Tm (°C) 12 °C/h	Tm (°C) 36 °C/h	ΔG (kcal mol⁻¹)	folding t₁/₂ (min)
1	-Pro-Pro-Gly-	37	41	-11	33 ± 2
2	-Pro-Hyp-Gly-	40	44	-12	24 ± 6
3	-Pro-Pro-**azGly**-	47	51	-13	27 ± 6
4	-Pro-Hyp-**azGly**-	51	55	-15	14 ± 3

SEC-MALS

peptide	avg. MW & range (kDa)	oligomeric state
1	5.4 5.2-5.4	trimer
2	5.6 5.7-5.8	trimer
3	5.8 5.6-5.8	trimer⁺
4	5.6 5.5-5.7	trimer⁺
***5**	1.9 1.7-2.0	monomer

*Peptide **5** contains a D-proline residue at the central triplet (-DPro-Pro-Gly-) and serves as a monomeric standard.

⁺Presence of trimeric species also verified by sedimentation equilibrium analysis.
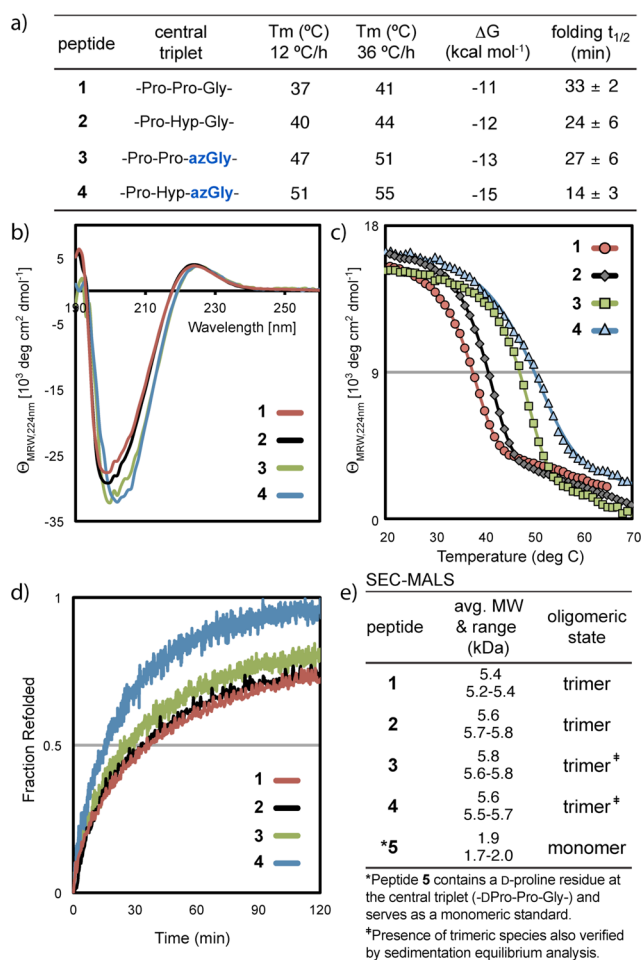
Figure 3. (a) Table of unfolding and refolding data for collagen model peptides. Values of T_m (standard error <1 °C) were determined in triplicate by CD spectroscopy at scan rates of 12 and 36 °C/h. (b) CD wavelength scans of homotrimers formed by **1−4**. (c) Unfolding curves for the thermal transition of **1−4** at 12 °C/h. (d) Refolding of peptides was observed by monitoring the recovery of ellipticity after thermal denaturation. (e) SEC-MALS analysis confirming the presence of the trimeric peptide assemblies for **1−4**. The presence of the trimeric state for **3** and **4** was also verified by sedimentation equilibrium analysis using AUC. Control peptide **5** contains a central D-Pro and serves as a monomeric standard for SEC-MALS analysis.

denatured at 80 °C for 15 min, and their CD profiles were monitored at 4 °C until they recovered (>50%) ellipticity at 224 nm. The refolding rate of peptides containing the azGly moiety was enhanced in both comparisons. The presence of trimers was confirmed for **1−4** by size exclusion chromatography with multi-angle light scattering (SEC-MALS) analysis using a D-Pro-containing collagen peptide (**5**) as a monomeric control (Figure 3e).[14] The trimeric state of azGly-containing **3** and **4** was also verified by sedimentation equilibrium analysis using analytical ultracentrifugtion (AUC).

A hysteresis study was performed to gain further insight into the stability of **3** and **4**. The free energy differences between the peptides were in accord with the differences in T_m and previous reports using this method: $\Delta G = -11, -12, -13$, and -15 kcal/mol for peptides **1−4**, respectively. The origin of ΔG is primarily consistent with an increase in the enthalpic gain from the ability of azGly to form extra H-bonds, although this will require verification by calorimetric methods. The large differences in half-time values for triple-helix self-assembly for azGly-

containing peptides are striking in comparison to our previous results with aza-proline, where we elucidated the role of stereochemistry with respect to biopolymer pre-organization and self-assembly.[14]

Our model suggests the incorporation of azGly into the triplet adds the possibility for extra H-bonding between the new α-NH and two different amide carbonyls on an adjacent peptide strand in addition to the already present amide H-bond. To gain further insight into the azGly substitution, we performed molecular dynamics (MD) calculations using GROMACS (see SI for details). Simulations on full triple-helical models of 1−4 revealed heavy-atom root-mean-square deviation (RMSD) values of <0.07 for the central 5-amino-acid triplets of 1 and 2 and slightly higher RMSD values for the N- (0.2) and C-terminal (0.14) triplets compared to a common starting model (Figure 4).
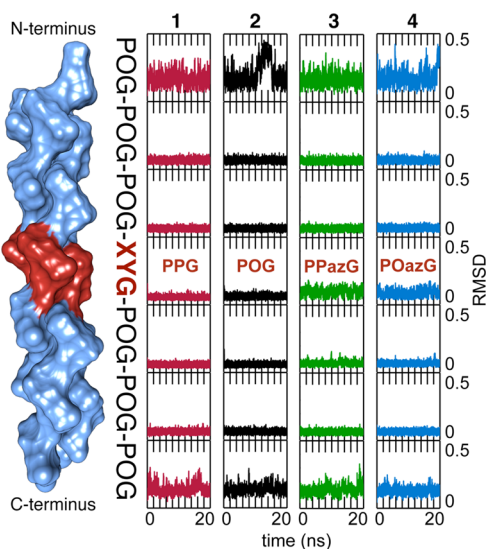

Figure 4. Model of the collagen triple helix, with the central variable triplet highlighted in red. Plot of RMSD for the heavy atoms of each trimeric triplet of each simulated collagen peptide system compared to a common reference crystal structure used to build all starting MD models. Reference crystal structure used was PDB 3B0S.

The azGly-containing triple-helical structures 3 and 4 showed RMSD values similar to those of the parent systems, except at the azGly position near the central triplet, where the RMSD increased to 0.10 for 3 and 0.08 for 4. Analysis of the azGly H-bonding parameters revealed the possibility of three different H-bonds at each azGly residue, with the new α-NH participating in up to two H-bonds (Figure 5). H-bond distances are shown in Figure 5c with angles shown in Figure 5d, revealing the possibility for three dynamic H-bonds with slightly less optimal parameters than the standard amide H-bond present in collagen. The MD simulation data imply that multiple dynamic but weak H-bonds with non-optimal distance and angle parameters may be more favorable in certain cases than strong H-bonds.

Peptide backbone substitutions have provided a wealth of insightful information regarding protein structure, and have led to the discovery of fundamentally important interactions such as the gauche effect in collagen and n−π* interactions. Previous reports of heteroatom replacement in the collagen peptide backbone have resulted in either severe destabilization or a complete lack of triple-helix formation. Amide-to-ester substitutions have provided a wealth of information regarding H-bond strength and have a detrimental impact on collagen triple-
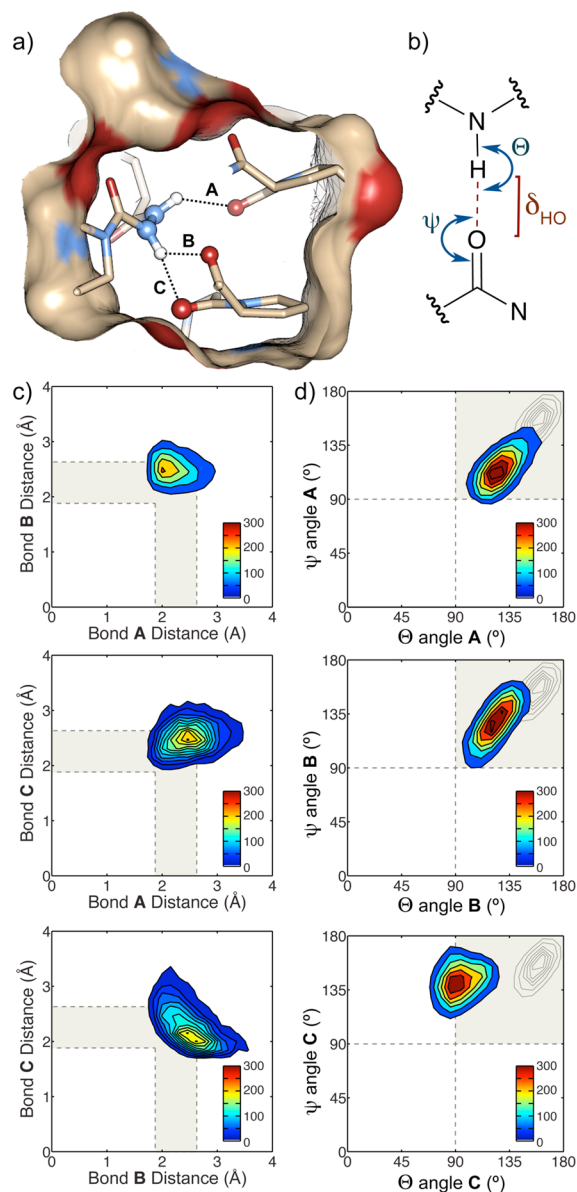

Figure 5. Hydrogen bond analysis for MD simulations of collagen model peptide 4. (a) Cross section through collagen triple helix showing possible azGly H-bonds. (b) Schematic of H-bond parameters. (c) H-bond distance measurements and correlations for A, B, and C from (a). (d) H-bond angle measurements for A, B, and C. Gray watermark on plots designates H-bond angles for non-azGly peptide system. Shaded regions on plots designate optimal H-bond parameters.

helix stability. In addition, trans alkene amide bond isosteres greatly destabilize the triple-helical structure of collagen irrespective of positioning and involvement in H-bonding. To date, these efforts have demonstrated an intolerance of the collagen peptide backbone to molecular editing.[13]

Our study suggests that nature's limited set of building blocks are not sufficient for optimizing the stability of self-assembled biopolymer systems such as collagen, and there is much to be gained from judicious synthetic modifications such as azGly incorporation. In addition to insight into the fundamental importance of hydrogen bonding as a stabilizing element in natural systems, these studies may provide insight into optimization of self-assembling biomimetic materials. Beyond collagen, these studies suggest the opportunity for protein

stabilization in a broader context via azGly scanning, which could identify unique positions for increasing thermal stability in addition to decreasing proteolytic degradation, as already reported for aza-amino acids. Current studies incorporating multiple azGly residues and assessing synergistic effects with other unnatural residues in addition to structural studies are under way and will be reported in due course.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b04590.

Methods, CD protocols, syntheses, and spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*dcheno@sas.upenn.edu
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14258.

(2) Anfinsen, C. B. *Science* **1973**, *181*, 223.

(3) (a) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291. (b) Dunitz, J. D. *Chem. Biol.* **1995**, *2*, 709. (c) Schneider, H.-J. *Angew. Chem., Int. Ed.* **2009**, *48*, 3924.

(4) (a) Mirsky, A. E.; Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1936**, *22*, 439. (b) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205. (c) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 729.

(5) (a) Hubbard, R. E.; Kamran Haider, M. In *eLS*; John Wiley & Sons, Ltd: New York, 2001. (b) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101. (c) Pace, C. N. *Nat. Struct. Mol. Biol.* **2009**, *16*, 681. (d) Bowie, J. U. *Curr. Opin. Struct. Biol.* **2011**, *21*, 42.

(6) (a) Pauling, L.; Corey, R. B. *Nature* **1951**, *168*, 550. (b) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 241. (c) Crick, F. H. C. *Nature* **1952**, *170*, 882. (d) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(7) (a) He, M.; Bode, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14752. (b) Barrett, K. T.; Metrano, A. J.; Rablen, P. R.; Miller, S. J. *Nature* **2014**, *509*, 71. (c) Wang, J.; Feringa, B. L. *Science* **2011**, *331*, 1429.

(8) (a) Ramachandran, G. N.; Kartha, G. *Nature* **1954**, *174*, 269. (b) Rich, A.; Crick, F. H. *J. Mol. Biol.* **1961**, *3*, 483. (c) Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009**, *78*, 929.

(9) (a) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75. (b) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666. (c) Fields, G. B.; Prockop, D. J. *Biopolymers* **1996**, *40*, 345. (d) Schumacher, M.; Mizuno, K.; Bächinger, H. P. *J. Biol. Chem.* **2005**, *280*, 20397. (e) Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. *Biochemistry* **2005**, *44*, 15812. (f) Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2002**, *11*, 262. (g) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683. (h) Lee, S.-G.; Lee, J. Y.; Chmielewski, J. *Angew. Chem., Int. Ed.* **2008**, *47*, 8429.

(10) (a) Kusebauch, U.; Cadamuro, S. A.; Musiol, H.-J.; Lenz, M. O.; Wachtveitl, J.; Moroder, L.; Renner, C. *Angew. Chem., Int. Ed.* **2006**, *45*, 7015. (b) Cejas, M. A.; Kinney, W. A.; Chen, C.; Leo, G. C.; Tounge, B. A.; Vinter, J. G.; Joshi, P. P.; Maryanoff, B. E. *J. Am. Chem. Soc.* **2007**, *129*, 2202. (c) Rele, S.; Song, Y.; Apkarian, R. P.; Qu, Z.; Conticello, V. P.; Chaikof, E. L. *J. Am. Chem. Soc.* **2007**, *129*, 14780. (d) Wang, A. Y.; Mo, X.; Chen, C. S.; Yu, S. M. *J. Am. Chem. Soc.* **2005**, *127*, 4130. (e) Przybyla, D. E.; Chmielewski, J. *J. Am. Chem. Soc.* **2008**, *130*, 12610. (f) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851.

(11) (a) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666. (b) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 777. (c) Shoulders, M. D.; Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2006**, *128*, 8112. (d) Cadamuro, S. A.; Reichold, R.; Kusebauch, U.; Musiol, H.-J.; Renner, C.; Tavan, P.; Moroder, L. *Angew. Chem.* **2008**, *120*, 2174. (e) Umashankara, M.; Babu, I. R.; Ganesh, K. N. *Chem. Commun.* **2003**, 2606. (f) Erdmann, R. S.; Wennemers, H. *J. Am. Chem. Soc.* **2010**, *132*, 13957. (g) Shoulders, M. D.; Satyshur, K. A.; Forest, K. T.; Raines, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 559. (h) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. *Nat. Chem. Biol.* **2010**, *6*, 615. (i) Choudhary, A.; Gandla, D.; Krow, G. R.; Raines, R. T. *J. Am. Chem. Soc.* **2009**, *131*, 7244. (j) Erdmann, R. S.; Wennemers, H. *Angew. Chem., Int. Ed.* **2011**, *50*, 6835. (k) Erdmann, R. S.; Wennemers, H. *J. Am. Chem. Soc.* **2012**, *134*, 17117. (l) Siebler, C.; Erdmann, R. S.; Wennemers, H. *Angew. Chem., Int. Ed.* **2014**, *53*, 10340. (m) Newberry, R. W.; Bartlett, G. J.; VanVeller, B.; Woolfson, D. N.; Raines, R. T. *Protein Sci.* **2014**, *23*, 284.

(12) Goodman, M.; Bhumralkar, M.; Jefferson, E. A.; Kwak, J.; Locardi, E. *Biopolymers* **1998**, *47*, 127.

(13) (a) Shah, N. K.; Brodsky, B.; Kirkpatrick, A.; Ramshaw, J. A. M. *Biopolymers* **1999**, *49*, 297−302. (b) Jenkins, C. L.; Vasbinder, M. M.; Miller, S. J.; Raines, R. T. *Org. Lett.* **2005**, *7*, 2619. (c) Dai, N.; Etzkorn, F. A. *J. Am. Chem. Soc.* **2009**, *131*, 13728. (d) Newberry, R. W.; VanVeller, B.; Raines, R. T. *Chem. Commun.* **2015**, *51*, 9624.

(14) Zhang, Y.; Malamakal, R. M.; Chenoweth, D. M. *Angew. Chem., Int. Ed.* **2015**, *54*, 10826.

(15) Sabatino, D.; Proulx, C.; Klocek, S.; Bourguet, C. B.; Boeglin, D.; Ong, H.; Lubell, W. D. *Org. Lett.* **2009**, *11*, 3650.

(16) (a) Zega, A. *Curr. Med. Chem.* **2005**, *12*, 589. (b) Proulx, C.; Sabatino, D.; Hopewell, R.; Spiegel, J.; Garcia Ramos, Y.; Lubell, W. D. *Future Med. Chem.* **2011**, *3*, 1139.

(17) Engel, J.; Chen, H. T.; Prockop, D. J.; Klump, H. *Biopolymers* **1977**, *16*, 601.

(18) (a) Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4273. (b) Horng, J. C.; Kotch, F. W.; Raines, R. T. *Protein Sci.* **2006**, *16*, 208. (c) Chen, Y. S.; Chen, C. C.; Horng, J. C. *Biopolymers* **2011**, *96*, 60.

(19) Natural amino acids do not outcompete Pro in X and Hyp in Y: (a) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960. (b) Jenkins, C. L.; Bretscher, L. E.; Guzei, I. A.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 6422.

(20) Shoulders, M. D.; Kamer, K. J.; Raines, R. T. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3859.