Stronger and (now) Longer Synthetic Collagen

Ronald T. Raines

Department of Biochemistry and Department of Chemistry
University of Wisconsin–Madison, Madison, WI 53706, USA

Introduction

Prevalence of Collagen. Collagen is the most abundant protein in modern animals, including humans.[1,2] The natural selection of collagen as an anatomical scaffold was made long ago. Collagen has been isolated from soft tissue in the fossilized bones of a 68-million-year-old *Tyrannosaurus rex* and a half-million-year-old mastodon.[3,4] No isolated DNA is that old.[5,6] The archaeological longevity of collagen is even more remarkable considering that a peptide bond has a half-life of 400 years, whereas a phosphodiester bond has a half-life of 30 million years.[7-10]

Collagen comprises $^1/_3$ of the human proteome and $^3/_4$ of the dry weight of human skin.[1,2] To date, 28 different types of collagen have been found in humans, and many other human proteins are known to contain collagenous domains.[2] A defining feature of all types of collagen is a unique tertiary structure in which three parallel strands, each in a polyproline II-type (PPII) helix (which is left-handed), are wound around a common axis to form a triple helix (which is right-handed) (Figure 1A). The packing of this coiled-coil structure requires that every third residue be glycine (Gly), resulting in a repeating Xaa–Yaa–Gly sequence. The residue in the Xaa position of these triplets is often 2*S*-proline (L-proline or Pro), and the residue in the Yaa position is often (2*S*,4*R*)-4-hydroxyproline (Hyp). In human type I collagen, which is the most abundant form, 28% of the Xaa residues are Pro, 38% of the Yaa residues are Hyp, and Pro–Hyp–Gly is the most common triplet.[11]

Individual triple helices of collagen are organized into fibrils of great tensile strength and flexibility. These fibrils can be arranged and cross-linked so as to support stress efficiently in one, two, or three dimensions in tissues such as tendon, skin, and cartilage, respectively. Abnormalities in collagen structure are associated with connective tissue diseases, such as osteogenesis imperfecta, Ehlers–Danlos syndrome, and some types of osteoporosis and arthritis.[1,12-15] A complete understanding of the basis for collagen stability (and instability) could lead to effective therapies for these and other disorders.

Collagen is also an important biomaterial.[16,17] For example, natural collagen is the principal component of biodegradable sutures and artificial heart valves. Obtaining natural collagen in high purity without degrading its structural integrity is difficult. Moreover, the natural collagen that is most readily available is bovine collagen, which can engender allergic and immunological side effects in humans.[18] Despite numerous studies on synthetic collagen, few have been tested as biomaterials.[19]

Role of Hyp Residues. The hydroxyl groups of the prevalent Hyp residues have an important role in collagen stability. Hyp residues are not incorporated into collagen by ribosomes.[20] Instead, this difficult post-translational modification is mediated by the enzyme prolyl 4-hydroxylase[21,22] after the strands are biosynthesized but before they form a triple helix. The prevalence of this post-translational modification is extraordinary: the abundance of Hyp in humans is 4.2%, a value calculated from the abundance of collagen in humans ($^1/_3$) and Hyp in collagen (38% $\times$ $^1/_3$). Thus, the abundance of Hyp exceeds that of seven "common" amino acids:

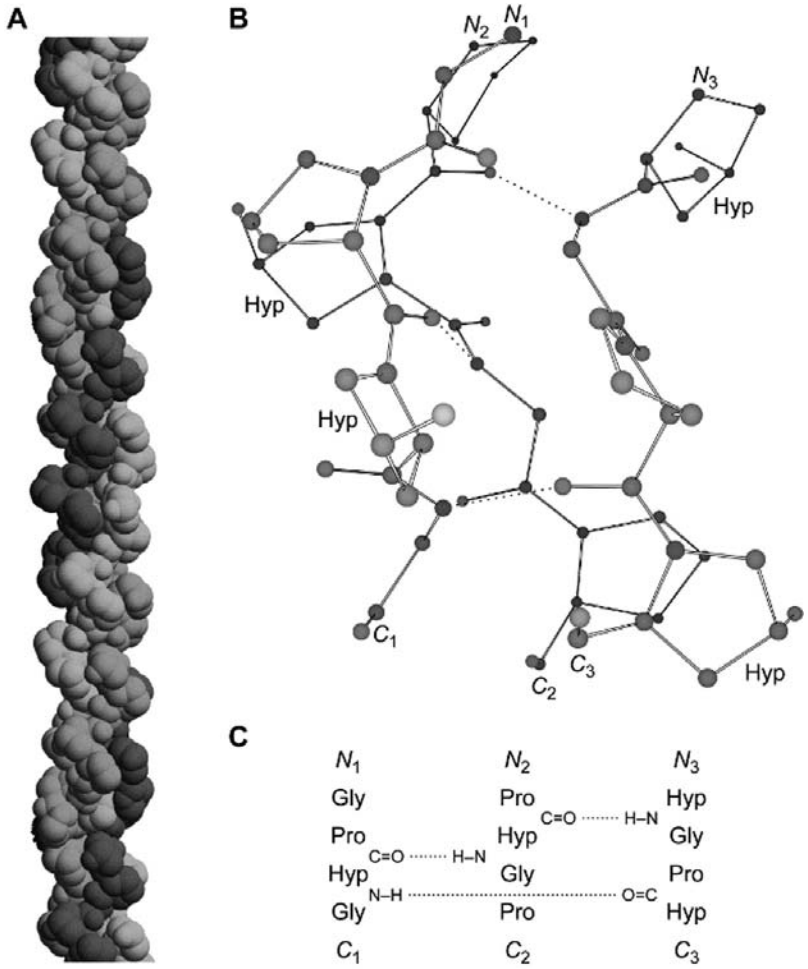


Fig. 1. Structure of a (Pro–Hyp–Gly)$_n$ triple helix. (A) Space-filling model of a long segment. (B) Ball-and-stick model of a short segment indicating Hyp residues and XaaC=O···H–NGly hydrogen bonds. (C) Register of the residues in the three strands of panel B. Atomic coordinates are from PDB entry 1CAG.[23]

Cys, Gln, His, Met, Phe, Trp, and Tyr.[24] Hydroxylation is critical for the folding of collagen, its secretion to the extracellular matrix, and its further processing and incorporation into fibrils or other structures.[25-28] In 2000, we discovered that the absence of prolyl 4-hydroxylase (and hence Hyp residues) is lethal to an animal—the nematode *Caenorhabditis elegans*.[29]

In type I collagen, each strand consists of ~300 Xaa–Yaa–Gly triplets. Discerning the chemical basis for the conformational stability of such a large molecule is difficult. Hence, a reductionist approach using peptide models has played a pivotal role and contributed much insight,[30-33] and led to the following landmark discoveries.

In 1973, Prockop and coworkers used synthetic peptides to demonstrate that the hydroxyl group of Hyp residues dramatically increases the thermal stability of triple-helical collagen.[34] The two decades after 1973 were dominated by the notion that the stability of collagen relies on water molecules that form bridges between the hydroxyl group of its Hyp residues and a main-chain oxygen. In 1994, X-ray diffraction analysis revealed the first truly high-resolution three-dimensional structure of a collagen triple helix.[23] In that triple helix (Figures 1A and 1B), the Hyp

residues do indeed have water molecules bound to their hydroxyl groups. These water molecules form bridges between the Hyp residues in one chain and a main-chain oxygen in another chain.[35]

Results and Discussion

Stronger Synthetic Collagen—Stereoelectronic Effects. We were skeptical of the "water-bridge" hypothesis, believing that the bridges were artifactual rather than meaningful. In 1998, we tested this hypothesis by replacing Hyp with (2*S*,4*R*)-4-fluoroproline (Flp). We found that Flp confers unprecedented stability upon a collagen triple helix.[36,37] Because a fluoro group exerts strong inductive effects but does not form strong hydrogen bonds, our result called into question the 25-year-old paradigm.

In 2001, we demonstrated that replacing Hyp with (2*S*,4*S*)-4-fluoroproline (flp), which is a diastereomer of Flp, does not allow for triple helix formation (Table 1).[38] This result indicated that the fluoro group was manifesting a stereoelectronic effect, rather than merely an inductive effect. In 2002, we discovered that the key attribute of the fluoro group (as well as the natural hydroxyl group) was its imposition of a C^{γ}-*exo* pucker on the pyrrolidine ring via the *gauche* effect (Figure 2).[39] This pucker preorganizes the main-chain dihedral angles (ϕ, ψ, and ω) to be those in the triple helix. This work not only demonstrated the role of the hydroxyl group of Hyp residues, but was the first example of a stereoelectronic effect conferring stability upon a protein.

In 2003, we attempted to enlist stereoelectronic effects in the Xaa position to enhance collagen stability. Again, we used a fluoro group, demonstrating that flp but not Flp enables the formation of a stable triple helix (Table 1).[40,41] Thus, stereoelectronic effects can operate adventitiously (or deleteriously) in the Xaa position of collagen. The stereochemical dichotomy (Xaa prefers flp, whereas Yaa prefers Flp) results from the preference for a C^{γ}-*endo* pucker in the Xaa position. In 2006, we showed how to fix the pyrrolidine ring pucker with reciprocal steric effects, thereby reiterating the conformational stability conferred by stereoelectronic effects.[42]

A conundrum led us to another insight. We could not explain why the pucker imposed by the 4-substituent had a marked effect on ω, which refers to the dihedral angle of the peptide bond. In AcFlpOMe, the amide bond has $K_{trans/cis} = 6.7$; in its diastereomer, AcflpOMe, $K_{trans/cis} = 2.5$.[38]

In 2002, we discovered that the explanation arose from another unappreciated stereoelectronic effect—an $n \rightarrow \pi^*$ interaction.[39,43] In an $n \rightarrow \pi^*$ interaction (which is not to be confused with an $n \rightarrow \pi^*$ electronic transition) the oxygen of a peptide bond

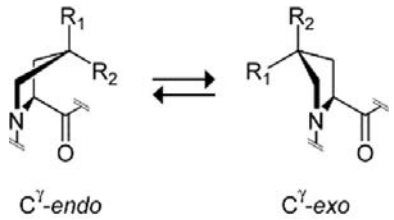
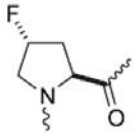
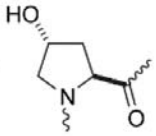
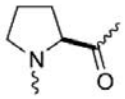
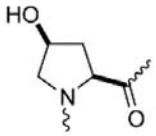
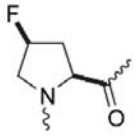


Fig. 2. Ring conformations of 4-substituted proline residues. The C^{γ}-endo conformation is favored strongly when R_1 = H and R_2 = F (as in flp). The C^{γ}-exo conformation is favored strongly when R_1 = OH (Hyp) or F (Flp) and R_2 = H. The C^{γ}-exo:C^{γ}-endo ratio is ~1:2 when $R_1 = R_2$ = H (Pro).[39]

(O*i*–1) donates electron density from one of its lone pairs into the antibonding orbital of the carbon in the subsequent peptide bond (C*i* ′=O*i*) (Figure 3A). The C^{γ}-exo pucker of a proline residue provides a more favorable O*i*–1···C*i* ′=O*i* distance and angle for an $n\rightarrow\pi^*$ interaction than does the C^{γ}-*endo* pucker.[39] Because the $n\rightarrow\pi^*$ interaction can occur only if the peptide bond containing O*i*–1 is trans (*i.e.*, *Z*), as opposed to cis (*E*), the $n\rightarrow\pi^*$ interaction has an impact on the trans/cis ratio. Thus, the two stereoelectronic effects influence peptide conformation as in the scheme:

The $n\rightarrow\pi^*$ interaction resembles the approach of a nucleophile to the electrophilic carbon of an acyl group. Accordingly, the $n\rightarrow\pi^*$ interaction is strongest when O*i*–1 is proximal to C*i* ′and along the Bürgi–Dunitz trajectory to C*i* ′=O*i*. In polypeptides, this geometry occurs in the right-and left-handed α-helix and the PPII helix (Figure 3B), which is the conformation assumed by the strands of the collagen triple helix (Figure 1A).

Table 1. Correlation of ring pucker with collagen triple-helix stability.[38-40,44] In a crystalline triple helix, proline residues have $\phi = -73°$, $\psi = 164°$ in the Xaa position and $\phi = -60°$, $\psi = 150°$ in the Yaa position.[23]

		Residue Ring Pucker		Triple Helix T_m	
		C^{γ}-*endo*	C^{γ}-*exo*	$(XaaProGly)_7$	$(ProYaaGly)_7$
Flp		14%	86% $\phi = -55°$ $\psi = 140°$	no helix	45 °C
Hyp (natural)				no helix	36°C
Pro (natural)		66%	34%	6 °C	6 °C
hyp				no helix	no helix
flp		95% $\phi = -76°$ $\psi = 172°$	5%	33 °C	no helix

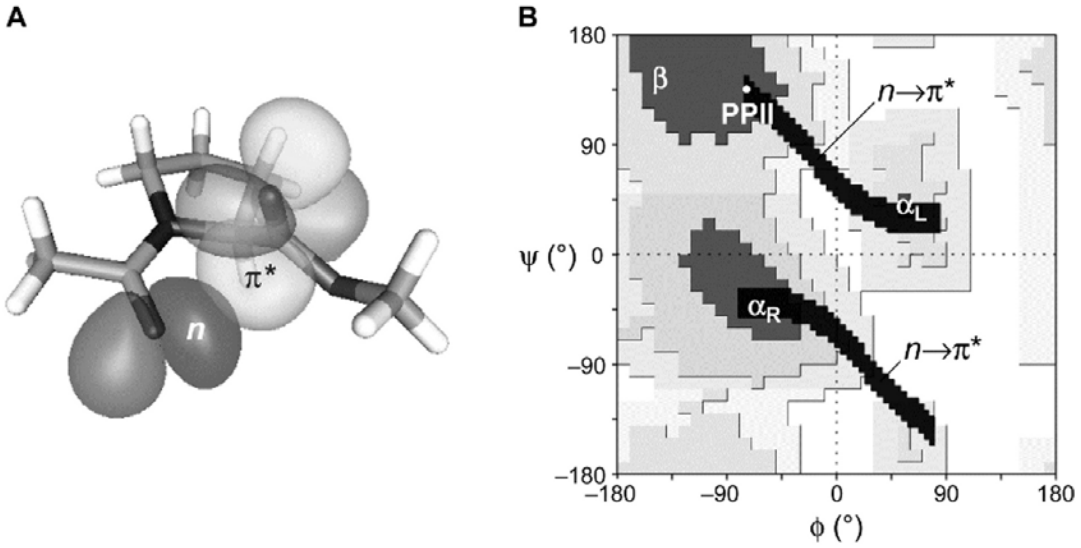


Fig. 3. The n→π interaction. (A) Natural bond orbitals depicting the n→π* interaction between O_{i-1} and C_i" in AcProOMe with a trans peptide bond and $C^{\#}$-exo ring pucker.[43] (B) Ramachandran plot showing the regions where the n→π* interaction will be strongest. In these regions, the $O_{i-1}\cdots C_i' \leq$ distance is 3.2 Å and the $O_{i-1}\cdots C_i' = O_i$ angle is ≥ 99° and ≤ 119° The white dot indicates the ϕ and ψ angles for the PPII helix assumed by the strands of a collagen triple helix.[43]*

The *n*→π* interaction can occur only if the peptide bond containing O_{i-1} is trans (*i.e.*, *Z*), as opposed to cis (*E*). Accordingly, the trans/cis ratio of an Xaa_{i-1}–Pro_i peptide bond can report on the strength of an *n*→π* interaction. We have used the trans/cis ratio of a peptide bond to quantify the energetics of an *n*→π* interaction.

We found that the trans/cis ratio of the amide bond in *N*-formylproline phenylesters correlates with electron-withdrawal by a *para* substituent (Figure 4).[45] The slope of the Hammett plot ($\rho = 0.26$) is indicative of a substantial effect. Density functional theory calculations and natural bond order analysis indicated that this

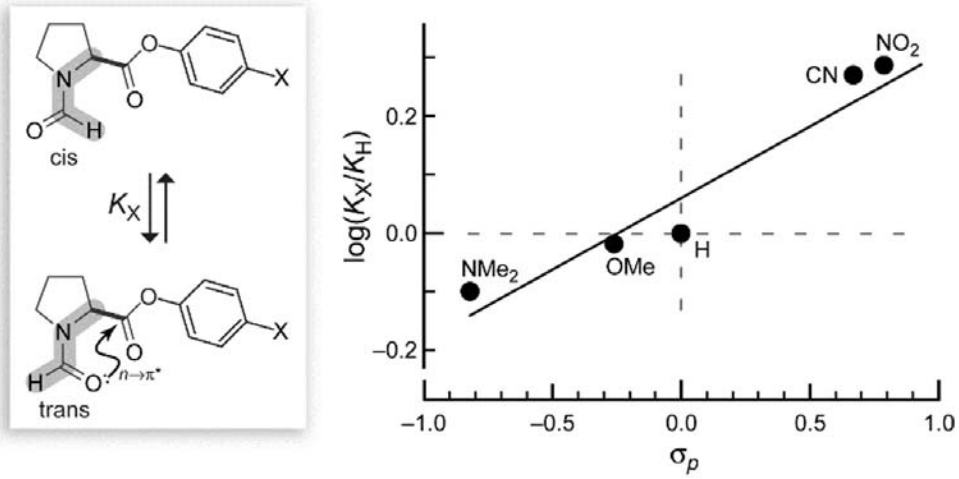


Fig. 4. Hammett plots depicting the relationship between the trans/cis ratio and electron-withdrawing ability of X in Fm-ProOC_6H_4-p-X.[45] (A) Values of K_X were determined by ^1H NMR spectroscopy, and yield $\rho = 0.26$.

effect arises from a favorable $n \rightarrow \pi^*$ interaction between the amide oxygen and ester carbon.

In textbooks, the well-known preference of the peptide bond for the *trans* conformation is attributed to steric effects. Our data indicate that this preference cannot be explained by steric effects alone. Rather, the $n \rightarrow \pi^*$ interaction, which is only extant in the trans isomer, contributes significantly to this preference. We found the $n \rightarrow \pi^*$ interaction to be worth $\Delta G^\circ \cong -0.7$ kcal/mol in a simple model system.[39,43] This free energy is now being exploited in the *de novo* design of peptides, peptoids, and proteins by other workers.[46-48] Hence, we propose that the $n \rightarrow \pi^*$ interaction should be added to the list of noncovalent interactions that direct a polypeptide chain to assume a particular folded structure.

Longer Synthetic Collagen—Self-Assembly. One long-term goal of our work is to develop collagen-based biomaterials with tunable attributes that can be used as both collagen surrogates and templates for nanotechnological applications. A major barrier in achieving this goal is that triple helices derived from synthetic peptides are much shorter (<10 nm) than natural collagen (~300 nm). Inspired by the self-assembly of double-helical fragments of DNA, we envisioned that sticky-ended fragments of synthetic collagen could self-assemble into long triple helices. Though unlike the situation with DNA, there is no "code" for the noncovalent association of collagen strands, other than the need for one Xaa, Yaa, and Gly residue to be in each cross section of a triple helix. We have demonstrated how to access long triple helices with molecular self-assembly.

We synthesized short collagen fragments in which the three strands were held in a staggered array by disulfide bonds (Figure 5).[49,50] These fragments were synthesized directly on a solid support by using a strategy based on the orthogonal deprotection of cysteine residues. Data from circular dichroism spectroscopy, dynamic light scattering, analytical ultracentrifugation, atomic force microscopy, and transmission electron microscopy indicated that these "sticky-ended" fragments self-assembled via intermolecular triple helix formation. The resulting fibrils resembled natural collagen, and some were longer (nearly 1 μm) than any known collagen. Assemblies with Yaa=Hyp tended to be longer and more stable than those with Yaa=Pro. Additional control over the length and stability was attained by modulating the temperature and solvent. Minimalist fragments like these can be elaborated by chemical synthesis to display motifs that promote cell adhesion for engineering tissues, lateral packing for accessing two- and three-dimensional architectures, and metal coordination for producing nanowires. Hence, we anticipate that our self-assembly strategy can provide synthetic collagen-mimetic materials for a variety of applications.

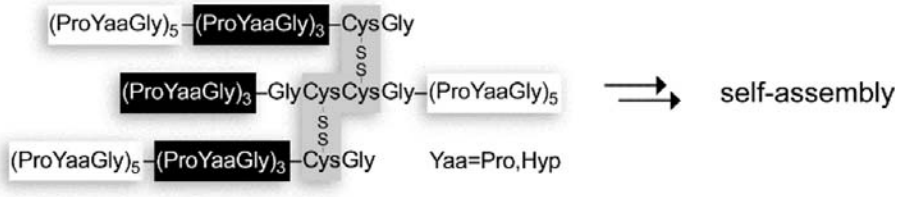


Fig. 5. Structure of a synthetic collagen fragment that self-assembles into fibrils of 1-nm width and nearly 1-μm length.[49,50]

Acknowledgments

The 2007 Makineni Lectureship is dedicated to the graduates students and postdoctorates who have worked with me on collagen, especially Dr. Frank W. Kotch, who performed the experiments in Ref. 49 and 50. This work was supported by grant AR44276 (NIH).

References

1. Myllyharju, J. and Kivirikko, K. I. *Ann. Med.* **33**, 7-21 (2001).
2. Ricard-Blum, S., Ruggiero, F. and van der Rest, M. *Top. Curr. Chem.* **247**, 35-84 (2005).
3. Schweitzer, M. H., Suo, Z., Avci, R., Asara, J. M., Allen, M. A., Arce, F. T. and Horner, J. R. *Science* **316**, 277-280 (2007).
4. Asara, J. M., Schweitzer, M. H., Freimark, L. M., Phillips, M. and Cantley, L. C. *Science* **316**, 280-285 (2007).
5. Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A. and Cooper, A. *Science* **300**, 791-795 (2003).
6. Binladen, J., Gilbert, M. T. and Willerslev, E. *Biol. Lett.* **3**, 55-56 (2007).
7. Radzicka, A. and Wolfenden, R. *J. Am. Chem. Soc.* **118**, 6105-6109 (1996).
8. Smith, R. M. and Hansen, D. E. *J. Am. Chem. Soc.* **120**, 8910-8913 (1998).
9. Wolfenden, R. and Snider, M. J. *Acc. Chem. Res.* **34**, 938-945 (2001).
10. Schroeder, G. K., Lad, C., Wyman, P., Williams, N. H. and Wolfenden, R. *Proc. Natl. Acad. Sci. USA* **103**, 4052-4055 (2006).
11. Ramshaw, J. A. M., Shah, N. K. and Brodsky, B. *J. Struct. Biol.* **122**, 86-91 (1998).
12. Prockop, D. J. and Kivirikko, K. I. *Annu. Rev. Biochem.* **64**, 403-434 (1995).
13. Prockop, D. J. *Matrix Biol.* **16**, 519-528 (1998).
14. Prockop, D. J. *Biochem. Soc. Trans.* **27**, 15-31 (1999).
15. Byers, P. H. *Clin. Genet.* **58**, 270-279 (2000).
16. Werkmeister, J. A. and Ramshaw, J. A. M. (Eds.) *Collagen Biomaterials*, Elsevier Science, Barking, Essex, England, 1992.
17. Ramshaw, J. A. M., Werkmeister, J. A. and Glattauer, V. *Biotechnol. Genet. Eng. Rev.* **13**, 335-382 (1995).
18. Lynn, A. K., Yannas, I. V. and Bonfield, W. *J. Biomed. Mater. Res. B Appl Biomater.* **71**, 343-354 (2004).
19. Johnson, G., Jenkins, M., McClean, K. M., Griesser, H. J., Kwak, J., Goodman, M. and Steele, J. G. *J. Biomed. Mater. Res.* **51**, 612-624 (2000).
20. Towe, K. M., In Billingham, J. (Ed.) *Life in the Universe*, MIT Press, Cambridge, MA, 1981.
21. Guzman, N. A. (Ed.) *Prolyl Hydroxylase, Protein Disulfide Isomerase, and Other Structurally Related Proteins*, Marcel Dekker, New York, 1998.
22. Myllyharju, J. *Matrix Biol.* **22**, 15-24 (2003).
23. Bella, J., Eaton, M., Brodsky, B. and Berman, H. M. *Science* **266**, 75-81 (1994).
24. McCaldon, P. and Argos, P. *Proteins: Struct. Funct. Genet.* **4**, 99-122 (1988).
25. Bulleid, N. J., Wilson, R. and Lees, J. F. *Biochem. J.* **317**, 195-202 (1996).
26. Walmsley, A. R., Batten, M. R., Lad, U. and Bulleid, N. J. *J. Biol. Chem.* **274**, 14884-14892 (1999).
27. Snellman, A., Keranen, M.-R., Hagg, P. O., Lamberg, A., Hiltunen, J. K., Kivirikko, K. I. and Pihlajaniemi, T. *J. Biol. Chem.* **275**, 8936-8944 (2000).
28. Byers, P. H. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 151-158 (2001).
29. Friedman, L., Higgin, J. J., Moulder, G., Barstead, R., Raines, R. T. and Kimble, J. *Proc. Natl. Acad. Sci. USA* **97**, 4736-4741 (2000).
30. Fields, G. B. and Prockop, D. J. *Biopolymers* **40**, 345-357 (1996).
31. Jenkins, C. L. and Raines, R. T. *Nat. Prod. Rep.* **19**, 49-59 (2002).
32. Engel, J. and Bächinger, H. P. *Top. Curr. Chem.* **247**, 7-33 (2005).
33. Raines, R. T. *Protein Sci.* **15**, 1219-1225 (2006).

34. Berg, R. A. and Prockop, D. J. *Biochem. Biophys. Res. Comm.* **52**, 115-120 (1973).
35. Bella, J., Brodsky, B. and Berman, H. M. *Structure* **3**, 893-906 (1995).
36. Holmgren, S. K., Taylor, K. M., Bretscher, L. E. and Raines, R. T. *Nature* **392**, 666-667 (1998).
37. Holmgren, S. K., Bretscher, L. E., Taylor, K. M. and Raines, R. T. *Chem. Biol.* **6**, 63-70 (1999).
38. Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L. and Raines, R. T. *J. Am. Chem. Soc.* **123**, 777-778 (2001).
39. DeRider, M. L., Wilkens, S. J., Waddell, M. J., Bretscher, L. E., Weinhold, F., Raines, R. T. and Markley, J. L. *J. Am. Chem. Soc.* **124**, 2497-2505 (2002).
40. Hodges, J. A. and Raines, R. T. *J. Am. Chem. Soc.* **125**, 9262-9263 (2003).
41. Hodges, J. A. and Raines, R. T., In Chorev, M. and Sawyer, T. K. (Eds.) *Peptide Revolution: Genomics, Proteomics & Therapeutics: Proceedings of the 18th American Peptide Symposium*, Kluwer Academic, Norwell, MA, 2004, p. 472-473.
42. Shoulders, M. D., Hodges, J. A. and Raines, R. T. *J. Am. Chem. Soc.* **128**, 8112-8113 (2006).
43. Hinderaker, M. P. and Raines, R. T. *Protein Sci.* **12**, 1188-1194 (2003).
44. Inouye, K., Sakakibara, S. and Prockop, D. J. *Biochim. Biophys. Acta* **420**, 133-141 (1976).
45. Hodges, J. A. and Raines, R. T. *Org. Lett.* **8**, 4695-4697 (2006).
46. Haduthambi, D. and Zondlo, N. J. *J. Am. Chem. Soc.* **128**, 12430-12431 (2006).
47. Kümin, M., Sonntag, L. S. and Wennemers, H. *J. Am. Chem. Soc.* **129**, 466-467 (2007).
48. Gorske, B. C., Bastian, B. L., Geske, G. D. and Blackwell, H. E. *J. Am. Chem. Soc.* **129**, 8928-8929 (2007).
49. Kotch, F. W. and Raines, R. T. *Proc. Natl. Acad. Sci. USA* **103**, 3028-3033 (2006).
50. Kotch, F. W. and Raines, R. T., In Blondelle, S. E. (Ed.) *Understanding Biology Using Peptides: Proceedings of the 19th American Peptide Symposium*, Springer, New York, NY, 2006, p. 688-689.