

Insights on the conformational stability of collagen

Cara L. Jenkins and Ronald T. Raines*

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

Received (in Cambridge, UK) 28th August 2001

First published as an Advance Article on the web 10th December 2001

Covering: 1994 to mid-2001

This review describes work on the conformational stability of the collagen triple helix. In 1994, the structure of collagen was determined at high resolution. Since then, much work has been done on synthetic mimics of collagen that contain host–guest peptides, tethers, peptoid residues, or analogs of the prevalent 4(*R*)-hydroxy-L-proline residues. This work has revealed much about the chemical basis for collagen stability, and could spawn useful new biomaterials. The literature from 1994 to mid 2001 is reviewed, and 116 references are cited.

- 1 Introduction
- 2 Collagen mimics
 - 2.1 Structural studies
 - 2.2 Host–guest studies
 - 2.3 Tethered triple-helical peptides
 - 2.4 Peptoid residues
- 3 4-Substituted proline residues
 - 3.1 Hydroxyproline residues
 - 3.2 Aminoproline residues
 - 3.3 Fluoroproline residues
- 4 Collagen as a biomaterial
- 5 Envoi
- 6 Acknowledgment
- 7 References

1 Introduction

Collagen is the most abundant protein in animals, being a

predominant component of connective tissues such as basement membranes, tendons, ligaments, cartilage, bone and skin. Over 19 different types of collagen have been identified, with several new ones being characterized currently. In addition, at least 15 other proteins have been shown to contain collagenous domains.¹

One of the defining features of collagen is its unique tertiary structure, consisting of three parallel left-handed polyproline II-type strands wound around a common axis to form a triple helix with a shallow right-handed superhelical pitch (Fig. 1). The packing of this coiled-coil structure requires that every third residue be glycine (Gly), resulting in a repeating Gly-Xaa-Yaa sequence. The residue in the Xaa position of these triplets is often L-proline (Pro), and the residue in the Yaa position is often 4(*R*)-hydroxy-L-proline (Hyp). 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



Cara L. Jenkins was born in 1972 in Provo, UT. She received BS and MS degrees in chemistry from Brigham Young University. At BYU, she worked with Steven A. Fleming to develop covalently and non-covalently tethered [2 + 2] photocycloaddition reactions. She is now a graduate student under the guidance of Ronald T. Raines in the chemistry department at the University of Wisconsin–Madison. As a trainee of the National Institutes of Health, Jenkins is using the methods of synthetic chemistry and biophysics to reveal the basis for the stability of collagen triple helices. She has earned both a McElvain Fellowship and the 2000 Hirschfelder Award from the chemistry department at Wisconsin.



Ronald T. Raines was born in 1958 in Montclair, NJ. He received ScB degrees in chemistry and biology from the Massachusetts Institute of Technology. At MIT, he worked with Christopher T. Walsh to reveal the reaction mechanisms of pyridoxal 5'-phosphate-dependent enzymes. Raines was a National Institutes of Health predoctoral fellow in the chemistry department at Harvard University. There, he worked with Jeremy R. Knowles to elucidate the reaction energetics of triosephosphate isomerase. Raines was a Helen Hay Whitney postdoctoral fellow in the biochemistry and biophysics department at the University of California, San Francisco. At UCSF, he worked with William J. Rutter to clone, express, and mutate the cDNA that codes for ribonuclease A. Raines then joined the faculty at the University of Wisconsin–Madison, where he is now professor of biochemistry and chemistry. His honors include the 1998 Pfizer Award in Enzyme Chemistry from the American Chemical Society and a 2001 Guggenheim Fellowship. His research group uses techniques that span the chemistry–biology interface to reveal protein structure–function relationships in vitro and in vivo. He has published over 100 papers on this topic.

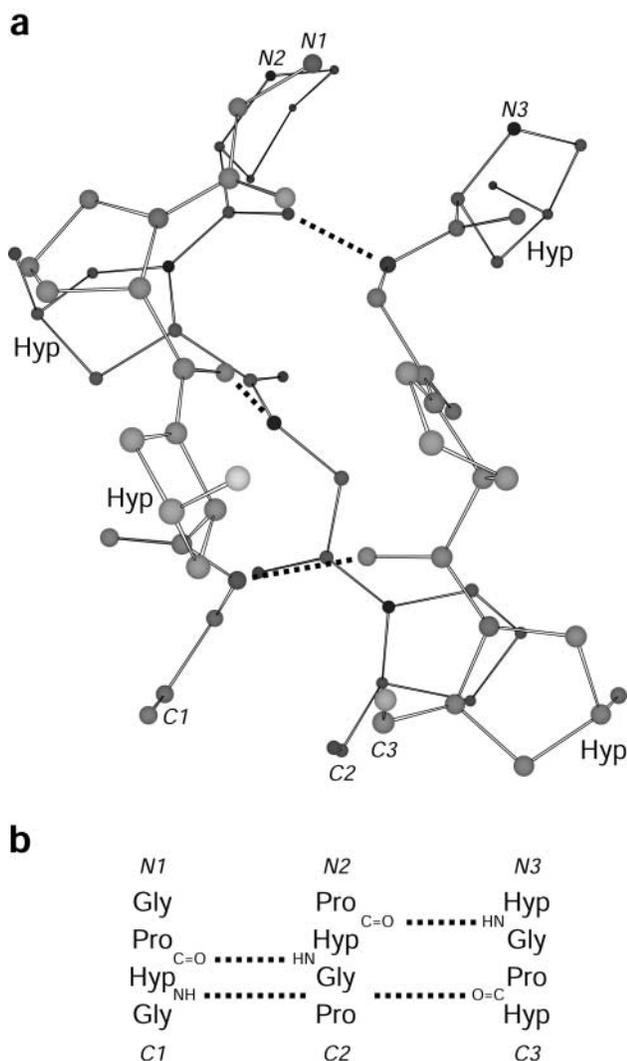


Fig. 1 Segment of a $(\text{Pro-Hyp-Gly})_n$ triple helix. (a) Ball-and-stick representation indicating 4-hydroxy-L-proline residues and XaaC=O \cdots H-NGly hydrogen bonds. (b) Register of the residues in the three strands of panel a. Atomic coordinates are from ref. 18 (PDB entry 1CAG).

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, scurvy, and some types of osteoporosis and arthritis.¹⁻⁵ A complete understanding of the basis for collagen stability (and instability) could lead to new biomaterials and effective therapies for these and other disorders.

2 Collagen mimics

The most abundant type of collagen is Type I, in which each strand consists of approximately 300 Gly-Xaa-Yaa triplets. Discerning the chemical basis for the conformational stability of such a large molecule is difficult. In 1973, Prockop and coworkers first synthesized small mimics of the collagen triple helix. This seminal work revealed that Hyp greatly stabilizes a triple helix when its 4(*R*) diastereomer, but not its 4(*S*) diastereomer, is present in the Yaa position.^{6,7} Subsequently, synthetic collagen mimics have been used to reveal the basis for this and other determinants of triple-helix structure and stability.⁸

2.1 Structural studies

To understand the stability of collagen, it is important to understand its structure. X-Ray diffraction analyses of biological samples provided the first glimpse into the unique structure of collagen.⁹⁻¹¹ In the 1950's, a model for the collagen triple

helix was proposed by Kartha and Ramachandran¹²⁻¹⁴ and refined by Rich and Crick¹⁵⁻¹⁷ to one that is essentially correct.

In 1994, Berman, Brodsky and coworkers used X-ray diffraction analysis to determine the first high-resolution structure of a triple-helical collagen mimic.¹⁸ The strands in this mimic had the sequence $(\text{Pro-Hyp-Gly})_4\text{-Pro-Hyp-Ala-(Pro-Hyp-Gly)}_5$, and the resulting triple helix is designated here as "Gly→Ala". This landmark structure revealed not only the positions of the atoms of the triple helix, but also a regular network of water molecules surrounding the triple helices in the crystal lattice. The structure appeared to lend support to a hypothesis^{19,20} that Hyp stabilizes collagen by forming hydrogen bonds with water molecules surrounding the triple helix.^{18,21,22}

Subsequent high-resolution structures of other collagen mimics led to a variety of conclusions about the factors that are most important in collagen stability. For example, Berman, Brodsky, Zagari, Mazzarella, and coworkers obtained crystals of $(\text{Pro-Pro-Gly})_{10}$ under two different conditions, from which they were able to refine short sections of triple helix (21 residues each) to a resolution of 1.7 and 2.0 Å.²³ The two structures have similar molecular structure and hydration patterns, but different crystal packing. In the two structures, as in the Gly→Ala structure, the pyrrolidine ring in each Xaa position has a *C'*-endo (or "down"²⁴) pucker, and Pro in each Yaa position has a *C'*-exo (or "up") pucker, with only one exception. The main-chain torsion angles and the first hydration shell around the peptide are similar to those in the Gly→Ala structure, despite the differences in crystal packing. Because the structures of the two peptides (Gly→Ala and $(\text{Pro-Pro-Gly})_{10}$) are so similar, these workers concluded that Hyp does not affect the triple-helix structure directly and that the contribution of Hyp to triple-helix stability arises only from Hyp-water interactions.

Okuyama and coworkers came to a different conclusion. Independently, they obtained a structure of crystalline $(\text{Pro-Pro-Gly})_{10}$, from which they were able to refine a triple helix of 21 residues to a resolution of 1.9 Å.²⁵ They found that the main-chain dihedral angles of their structure had no significant differences from those determined by the Berman group. Using different refinement procedures, the Okuyama group was able to locate only 15 water molecules, rather than the 40 solvent molecules in the $(\text{Pro-Pro-Gly})_{10}$ structure of the Berman group. In addition, the Okuyama group found that only 5 of 7 Pro residues in the Xaa position have a *C'*-endo pucker, and only 4 of 7 Pro residues in the Yaa position have a *C'*-exo pucker. This randomness is in conflict with the uniform pattern of Pro pucker observed by the Berman group.

Okuyama and coworkers also determined a structure for $(\text{Pro-Hyp-Gly})_{10}$ to a resolution of 1.9 Å.²⁶ In this structure, the pattern of pyrrolidine ring pucker is similar to that in Gly→Ala. Yet, only 17 water molecules are apparent, and only 3 of 7 Hyp hydroxy groups participate in hydrogen bonds with water. Okuyama and coworkers concluded that the close similarity of their $(\text{Pro-Pro-Gly})_{10}$ and $(\text{Pro-Hyp-Gly})_{10}$ structures indicates that Hyp does not affect directly the molecular structure, and that the nearly equal number of well-defined water molecules in the two structures indicates that the contribution of Hyp to stability is probably not due to an extensive network of water bridges.^{25,26} This conclusion is consistent with an earlier study in which Engel, Prockop, and coworkers showed that Hyp confers extra stability upon a triple helix even in anhydrous solution.²⁷

A 1.3 Å-resolution structure of crystalline $(\text{Pro-Pro-Gly})_{10}$ by Zagari, Mazzarella, and coworkers engendered an alternative hypothesis for the contribution of Hyp to collagen thermostability.^{28,29} These workers showed that Pro has a distinct φ torsion angle ($C'_{i-1}\text{-N}_i\text{-C}'_i\text{-C}'_i$) in the Xaa and Yaa positions: $\varphi = (-75 \pm 3)^\circ$ and $\varphi = (-60 \pm 2)^\circ$, respectively. They noted that the different angles correlated with the *C'*-endo pucker in the Xaa position and the *C'*-exo pucker in the Yaa position. Because Hyp is more rigid than Pro and favors the *C'*-exo

pucker, Hyp in the Yaa position reduces the number of conformations available to the unfolded state and gives it a higher propensity to fold into a triple helix. Hyp in the Xaa position, however, cannot adopt the *C^γ-endo* pucker, and is thus a residue with an unfavorable ϕ angle for triple-helix formation. Furthermore, 4(*S*)-Hyp adopts the *C^γ-endo* pucker, making its ϕ angle inappropriate for the Yaa position. In the Xaa position, the hydroxy group of 4(*S*)-Hyp is likely to destabilize the triple helix by a steric clash with a Pro residue in the Yaa position of another strand.

The structure of a crystalline triple helix containing a sequence from Type III collagen revealed that Pro and Hyp can alter the helical pitch. Each strand of this triple helix has the sequence (Pro-Hyp-Gly)₃-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-(Pro-Hyp-Gly)₄. In the refinement process, no one model was found to fit the electron density map of the entire triple helix.³⁰ Rather, the terminal (Pro/Hyp-rich) regions fit to a model with 7-fold symmetry, as in (Pro-Hyp-Gly)₁₀, but the central (Pro/Hyp-poor) regions fit to a model with 10-fold symmetry, as in natural collagen.^{16,25} This result indicates that the helical pitch is sequence-dependent, and that the main chain can accommodate small variations in torsion angles and still maintain triple-helicity. This finding forewarns of a potential complication in the interpretation of data from host-guest studies (*vide infra*), as the guest could prefer a different helical pitch than the host.

In addition to X-ray diffraction analysis, nuclear magnetic resonance (NMR) spectroscopy has been used to probe the structure of the collagen triple helix in solution,³¹⁻³³ as well as measure its dynamics^{34,35} and folding kinetics.³⁶⁻³⁸ In general, the structural data obtained by NMR spectroscopy and X-ray diffraction analysis are in gratifying agreement.

2.2 Host-guest studies

“Host-guest” peptides and proteins have been used to measure the propensity of individual amino acid residues to form an α -helix³⁹⁻⁴³ or β -sheet.⁴⁴⁻⁴⁷ In these studies, a parent sequence is chosen that is known to have the desired secondary structure. A central amino acid residue is then replaced systematically with other residues, and the conformational stability of the resulting structure is measured. This approach has been used to determine the contribution of individual as well as pairs of residues to triple-helix stability.⁴⁸

The frequency with which a given triplet appears in natural

collagen has been discerned by examining 4040 triplets from human fibril- and non-fibril-forming collagens.⁴⁹ Approximately 49% of all possible triplets are never or rarely found in fibrillar collagens, 41% are never found in non-fibrillar collagens, and 32% are never found in either type of collagen. The residues Trp and Cys are never found in triple-helical regions of collagen, Tyr is rarely found in the Xaa position, and Phe is never found in the Yaa position. A schematic of the combined triplet distribution of several human collagens is shown in Fig. 2. It is interesting to note that no triplet appears with high frequency except for Gly-Pro-Hyp, which is the most stabilizing triplet found in natural collagen.^{27,50,51} This survey provides a context for the host-guest studies of collagen mimics, and reduces the number of triplets that need to be studied to obtain a complete picture of the contribution of natural triplets to conformational stability.

A comprehensive host-guest study of individual residues in the Xaa and Yaa positions was reported by Brodsky and coworkers.⁵² The guest peptides had the sequence Ac-(Gly-Pro-Hyp)₃-Gly-Xaa-Hyp-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂ and Ac-(Gly-Pro-Hyp)₃-Gly-Pro-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂, where Xaa and Yaa were the 19 proteinogenic amino acids other than Pro and Hyp, respectively. They found that the triple helix with Xaa = Pro and Yaa = Hyp was the most stable, with a T_m (which is the temperature at the midpoint of the thermal transition) of 47.3 °C. Surprisingly, replacement of Hyp with Arg in the Yaa position resulted in a triple helix of nearly equal stability ($T_m = 47.2$ °C), whereas all the other amino acid substitutions resulted in triple helices with T_m values that are at least 5 °C lower (Table 1). This study revealed that there is a moderate correlation between the contribution of a given residue to triple-helix stability and its propensity to adopt a polyproline II-like conformation. This correlation was shown to be better for the Xaa position than the Yaa position, perhaps because of the greater solvent exposure of the residue in the Xaa position.⁵³ Trp, which is not found in natural collagen, was the most destabilizing residue in both the Xaa and Yaa positions; other aromatic amino acid residues (Phe and Tyr) were also destabilizing in both positions.

The unexpected stability conferred by the guest triplet Gly-Xaa-Arg, which accounts for nearly 10% of the triplets in common human collagens (Fig. 2),⁴⁹ was examined further with the peptides listed in Table 2.⁵⁴ The stabilization is not due to cationic charge alone, as triple helices containing Lys are considerably less stable than are those containing Arg (Table 1).

Xaa/Yaa	A	C	D	E	F	G	H	I	K	L	M	N	O	Q	R	S	T	V	W	Y
A	0.9		0.5	hatched				hatched	1.2	hatched		0.3	3.4	0.4	1.1	0.4	hatched	0.4		
C				hatched				hatched	hatched	hatched			1.5	0.5	1.2	0.3	hatched	0.3		
D	0.4		0.3	hatched				hatched	1.0	hatched	hatched		2.8	0.8	2.7	0.3	hatched	0.4		
E	1.2		0.4	hatched		hatched		0.4	2.5	hatched	0.3	0.3	2.5	0.3	hatched		0.6	0.4		
F	hatched		hatched					hatched	hatched	hatched			0.3	hatched			hatched			
G			hatched					hatched	0.3				0.5	hatched			hatched			hatched
H	hatched		hatched					hatched				hatched	0.5	hatched			hatched			hatched
I	0.4		hatched		hatched			0.3				hatched	1.5	hatched			hatched	hatched	hatched	hatched
K	hatched		0.5	hatched				hatched		hatched		hatched	1.4	hatched			hatched			hatched
L	0.5		0.4			hatched		hatched	1.0	hatched	hatched	hatched	5.5	0.7	0.6	hatched	0.3	hatched		hatched
M			hatched					hatched	hatched				0.6	hatched			hatched			hatched
N			hatched					hatched	hatched				0.7	hatched			hatched			hatched
P	3.4		0.4	0.4				0.9	2.7	0.5	0.4	0.3	10.5	2.5	2.6	1.4	0.8	1.3		
Q	0.3		hatched	hatched				hatched	0.6		hatched		1.1	0.3	0.3	hatched	hatched	hatched	hatched	hatched
R	0.3		0.5	hatched				hatched	hatched				1.1	hatched						
S	0.4		hatched	hatched				hatched	0.4			hatched	2.3	hatched	0.5	hatched	hatched	hatched	hatched	0.3
T	hatched		hatched					hatched	hatched				0.8	hatched			hatched	hatched	hatched	hatched
V	hatched		hatched		hatched			hatched	0.5	hatched	hatched		1.1	0.3	hatched	hatched	hatched	hatched	hatched	hatched
W																				
Y													0.5	hatched						

Fig. 2 Distribution of triplets in human collagen types I, II, III, V, XI, IV, VI, VII, VIII, IX, X, and XIII.⁴⁹ Rows and columns refer to the Xaa and Yaa positions of 4040 Gly-Xaa-Yaa triplets. Black boxes indicate triplets found fewer than 3 times (<0.074%), hatched boxes indicate triplets found 3–10 times (0.074–0.25%), and numbers in white boxes represent the occurrence of the most common triplets, rounded to the nearest 0.1%. All Pro residues in the Yaa position are assumed to be hydroxylated. All residues are indicated by their single letter codes, with O indicating Hyp.

A triple helix containing the guest triplet Gly-Pro-homoArg, which was synthesized from the guest triplet Gly-Pro-Lys by reaction with a guanidino group transfer reagent, had stability between that of triple helices containing Gly-Pro-Arg and Gly-Pro-Lys. Thus, some property of the guanidino group is important in stabilization, but the longer homoArg side chain decreases stability somewhat. Data on triple helices containing the guest triplets Gly-Ala-Hyp, Gly-Ala-Arg, and Gly-Ala-Lys show that Arg confers stability even when the residue in the Xaa position is not Pro, possibly by forming either an intra- or inter-chain hydrogen bond with a main-chain carbonyl group. If two Gly-Pro-Arg triplets are adjacent, unfavorable Coulombic interactions destabilize the triple helix somewhat, but separation by one or more intervening triplets abolishes this effect. Still, triple helices with two Gly-Pro-Arg triplets separated by one or two Gly-Pro-Hyp triplets have T_m values slightly lower than does a triple helix containing one Gly-Pro-Arg triplet. The peptide Ac-(Gly-Pro-Arg)₈Gly-Gly-NH₂ could not be folded into a triple helix until its concentration was 3 mg mL⁻¹ and the NaCl concentration of the solution was raised to 2 M. Apparently, unfavorable Coulombic interactions between the Arg residues are considerable.

The contribution of pairs of common non-polar residues to

Table 1 Conformational stability of host-guest triple helices and ranking of frequency of appearance in polyproline II helices⁵²

Xaa ^a	T_m /°C	Yaa ^a	T_m /°C	Polyproline II frequency ^b
Pro	47.3	Hyp	47.3	Pro
Glu	42.9	Arg	47.2	Gln
Ala	41.7	Met	42.6	Arg
Lys	41.5	Ile	41.5	Lys
Arg	40.6	Gln	41.3	Thr
Gln	40.4	Ala	40.9	Leu
Asp	40.1	Val	40.0	Asp
Leu	39.0	Glu	39.7	Met
Val	38.9	Thr	39.7	Ala
Met	38.6	Cys	37.7	Cys
Ile	38.4	Lys	36.8	Val
Asn	38.3	His	35.7	Glu
Ser	38.0	Ser	35.0	Asn
His	36.5	Asp	34.0	Phe
Thr	36.2	Gly	32.7	Ser
Cys	36.1	Leu	31.7	Ile
Tyr	34.3	Asn	30.3	Trp
Phe	33.5	Tyr	30.2	Tyr
Gly	33.2	Phe	28.3	His
Trp	31.9	Trp	26.1	Gly

^a Peptide strands of host-guest triple helices have the sequence Ac-(Gly-Pro-Hyp)₃-Gly-Xaa-Hyp-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂ and Ac-(Gly-Pro-Hyp)₃-Gly-Pro-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂. ^b Frequency of occurrence of amino acids in polyproline II regions in globular proteins.¹¹⁵

triple-helical stability has been examined by comparing a set of host-guest peptides.⁵⁵ The residues Pro and Hyp, along with Ala, Phe, and Leu, which are the most common non-polar residues found in collagen, were included in various combinations in peptides of the form Ac-(Gly-Pro-Hyp)₃-Gly-Xaa-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂. The Gly-Xaa-Yaa guest triplets and the T_m values for their triple helices are listed in Table 3. Leu and Phe are more destabilizing in the Yaa than in the Xaa position, and Phe is usually more destabilizing than Leu. It is also apparent from the data that hydrophobic residues do not lend any special stability to collagen triple helices, in contrast to globular proteins, which have hydrophobic cores.⁵⁶ In the Xaa position of collagen, non-polar residues could participate in intermolecular interactions with other collagen strands or with other proteins that bind to collagen,⁵⁵ as the Xaa position is somewhat more solvent-exposed than is the Yaa position.⁵³

Several pairs of charged residues have been examined in a similar manner.⁵⁷ Glu, Asp, Arg, and Lys were placed individually and in oppositely-charged pairs into peptides with the general sequence Ac-(Gly-Pro-Hyp)₃-Gly-Xaa-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂. The T_m values of the resulting triple helices are listed in Table 4. Comparing the stabilities of triple helices containing pairs of charged residues to those of triple helices containing individual charged residues at neutral pH indicates that, except for the triplets Gly-Lys-Asp and Gly-Arg-Asp, there is no stabilization from favorable Coulombic interactions. Rather, differences in stability can be accounted for by the effects of individual charged residues. Moreover, charged residues in the Xaa position have little effect on the T_m values of triple helices, and these charged residues decrease conformational stability only marginally compared to Pro and not at all compared to Ala. The difference is much larger when the charged residues are in the Yaa position. An explanation for this phenomenon is that side chains in the Yaa position are less solvent-accessible and closer to side chains in other chains,⁵³ leading to unfavorable steric and Coulombic interactions.

Table 3 Effect of non-polar residues on the conformational stability of host-guest triple helices⁵⁵

Gly-Xaa-Yaa	T_m /°C ^a	Gly-Xaa-Yaa	T_m /°C ^a
Gly-Pro-Hyp	44.5	Gly-Leu-Ala	29.5
Gly-Ala-Hyp	39.9	Gly-Ala-Ala	29.3
Gly-Leu-Hyp	39.0	Gly-Pro-Phe	28.3
Gly-Pro-Ala	38.3	Gly-Ala-Leu	27.8
Gly-Phe-Hyp	33.5	Gly-Phe-Ala	23.4
Gly-Pro-Leu	32.7	Gly-Ala-Phe	20.7

^a Values of T_m are for triple helices of Ac-(Gly-Pro-Hyp)₃-Gly-Xaa-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂ and were measured at pH 7.4.

Table 2 Effect of Arg on the conformational stability of host-guest triple helices⁵⁴

Peptide sequence	T_m /°C ^a
Ac-(Gly-Pro-Hyp) ₈ -Gly-Gly-NH ₂	45.5
Ac-(Gly-Pro-Hyp) ₃ -Gly-Pro-Arg-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	45.5
Ac-(Gly-Pro-Hyp) ₃ -Gly-Pro-homoArg-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	42.8
Ac-(Gly-Pro-Hyp) ₂ -Gly-Pro-Arg-(Gly-Pro-Hyp) ₂ -Gly-Pro-Arg-(Gly-Pro-Hyp) ₂ -Gly-Gly-NH ₂	42.8
Ac-(Gly-Pro-Hyp) ₃ -Gly-Pro-Arg-Gly-Pro-Hyp-Gly-Pro-Arg-(Gly-Pro-Hyp) ₂ -Gly-Gly-NH ₂	42.2
Ac-(Gly-Pro-Hyp) ₃ -Gly-Arg-Hyp-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	40.6
Ac-(Gly-Pro-Hyp) ₄ -Gly-Pro-Arg-Gly-Pro-Arg-(Gly-Pro-Hyp) ₂ -Gly-Gly-NH ₂	40.4
Ac-(Gly-Pro-Hyp) ₃ -Gly-Ala-Hyp-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	39.9
Ac-(Gly-Pro-Hyp) ₃ -Gly-Ala-Arg-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	38.2
Ac-(Gly-Pro-Hyp) ₃ -Gly-Pro-Lys-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	36.8
Ac-(Gly-Pro-Arg) ₈ -Gly-Gly-NH ₂	32.6 ^b
Ac-(Gly-Pro-Hyp) ₃ -Gly-Ala-Lys-(Gly-Pro-Hyp) ₄ -Gly-Gly-NH ₂	30.8

^a Values of T_m were measured in phosphate-buffered saline (pH 7.0, 150 mM NaCl). ^b Value of T_m was measured in 10 mM sodium phosphate buffer (pH 7.0) containing NaCl (2 M).

Table 4 Effect of charged residues on the conformational stability of host-guest triple helices⁵⁷

Gly-Xaa-Yaa	pH	T_m /°C ^a	Gly-Xaa-Yaa	pH	T_m /°C ^a
Gly-Ala-Hyp ^a	7.4	39.9	Gly-Pro-Ala ^b	7.4	38.3
Gly-Pro-Hyp ^a	7.4	44.5	Gly-Ala-Ala ^b	7.4	29.3
Gly-Asp-Hyp	2.7	37.6	Gly-Pro-Asp	2.7	33.1
	7.0	40.1		7.0	30.1
	12.2	38.0		12.2	30.1
Gly-Glu-Hyp	2.7	39.7	Gly-Pro-Glu	2.7	41.9
	7.0	42.9		7.0	39.7
	12.2	40.9		12.2	38.5
Gly-Lys-Hyp	2.7	40.4	Gly-Pro-Lys	2.7	37.1
	7.0	41.5		7.0	36.8
	12.2	38.3		12.2	38.8
Gly-Arg-Hyp	2.7	39.4	Gly-Pro-Arg	2.7	45.5
	7.0	40.6		7.0	44.5
	12.2	38.0		12.2	43.1
Gly-Asp-Lys	2.7	26.5	Gly-Lys-Asp	2.7	30.5
	7.0	30.9		7.0	35.8
	12.2	29.9		12.2	30.2
Gly-Asp-Arg	2.7	33.4	Gly-Arg-Asp	2.7	28.8
	7.0	37.1		7.0	35.0
	12.2	34.4		12.2	31.9
Gly-Glu-Lys	2.7	29.5	Gly-Lys-Glu	2.7	36.5
	7.0	35.0		7.0	35.3
	12.2	33.1		12.2	31.6
Gly-Glu-Arg	2.7	37.3	Gly-Arg-Glu	2.7	35.0
	7.0	40.4		7.0	33.8
	12.2	39.1		12.2	32.2

^a Values of T_m are for triple helices of Ac-(Gly-Pro-Hyp)₃-Gly-Xaa-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-NH₂. ^b The triplets Gly-Pro-Hyp, Gly-Ala-Hyp, Gly-Pro-Ala, and Gly-Ala-Ala are included for comparison.⁵⁵

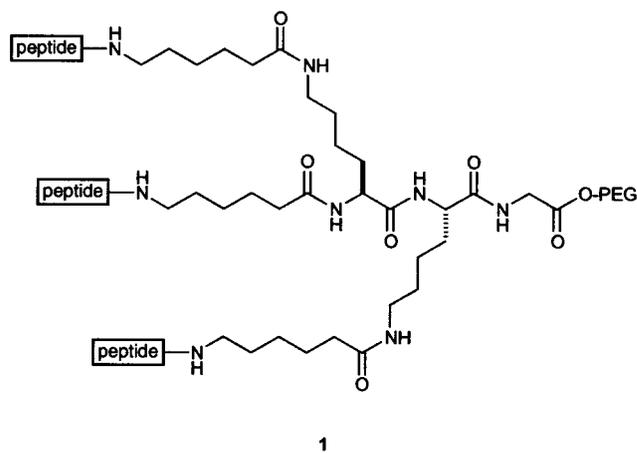
The host-guest studies outlined above help to clarify how a particular residue affects triple-helix stability in a relatively stable triple-helical environment. It is not clear, however, whether this understanding can be applied to regions of collagen lacking Pro and Hyp residues, which can have a different helical pitch.³⁰ The frequency with which a given residue appears in natural collagen (Fig. 2) does not necessarily correlate with the stability that that residue imparts to the triple helix in host-guest studies (Table 1). Of course, some residues could be present in collagen for other purposes, such as participating in interactions with other biomolecules.

Replacement of Gly with other amino acids has also been studied in host-guest fashion. A review⁴⁸ of these studies leads to the conclusion that the instability caused by these Gly substitutions varies, and that the degree of flexibility in the surrounding peptide can influence the severity of the destabilization. The identity and location of the Gly substitution also seems to correlate somewhat with the severity of osteogenesis imperfecta, which is caused by Gly substitutions in collagen.

2.3 Tethered triple-helical peptides

Several groups have used synthetic methods to tether three collagen-related peptides so as to enhance their triple helicity. For example, Heidemann and coworkers used a di-Lys-based construction with aminohexanoic acid linkers between the three Lys amino groups (one α and two ϵ) and their peptides (1).^{58,51} Fields and coworkers used a similar tether with an orthogonal protecting group strategy to produce triple helices of biological interest.^{59,60}

Tanaka and coworkers used a lysine dimer, similar to those of Heidemann and Fields, but having a β -alanine linker. In addition, the Tanaka group took a different approach to assembling a triple helix. Whereas Fields and Heidemann built their peptides onto the tether one strand at a time, Tanaka



and coworkers synthesized a peptide with a Cys residue at its N-terminus. The sulfhydryl group was alkylated with the tether **2**.⁶¹

The Tanaka group also devised a method to cross-link a collagen peptide at both the N- and C-termini.⁶² The N-terminal tether is **2**, and the C-terminal linker is **3**. The collagen mimic was synthesized to include a Lys residue at the C-terminus of the peptide with a Ser residue attached to N⁶. NaIO₄ oxidation of the peptide generated an aldehyde from this Ser, which then formed an oxime linkage with the aminoxy group on the linker. The N-linked peptides form more stable triple helices than do the unlinked peptides, and the dually N,C-linked peptides **4** form even more stable triple helices than do the N-linked ones.

Several groups have designed tethers based on motifs found in nature. Fields has attached lipids to the N-termini of collagen mimics and relied on self-assembly processes to drive triple-helix folding and stabilization.⁶³ Engel, Bächinger, and coworkers used a homotrimeric globular protein in a similar manner.⁶⁴ They created a plasmid that directs *Escherichia coli* to produce a chimeric protein in which (Gly-Pro-Pro)₁₀ is fused to the 27-residue C-terminal domain of bacteriophage T4 fibrin protein (termed “foldon”). The chimerae formed trimers with a high degree of conformational stability. Moroder and coworkers used a simplified version of the disulfide bridges found in the C-terminal domain of procollagen to stabilize collagen mimics.^{65–69} They designed a “cystine knot” derived from two pairs of differentially protected cysteine residues such that three strands are tethered in a selective manner, as shown in Fig. 3. Many human collagens contain two or three different strands,¹ and the cystine knot provides a facile means to stabilize heterotrimeric triple helices.

Goodman and coworkers used a cyclic tether based on *cis,cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid, also known as the Kemp triacid,⁷⁰ to create triple-helical collagen mimics.⁷¹ They chose this template because it is rigid and because its three carboxylic acid groups are fixed on the same face of the cyclohexane ring. Condensation of each carbonyl group of the triacid with a Gly residue provides enough flexibility for the collagen-like strands to adopt the requisite register (Fig. 1) as well as the larger diameter of the triple helix. In their first use of the Kemp triacid template,⁷² Goodman and coworkers showed that templated (Gly-Pro-Hyp)_n strands form much more stable triple helices than do acylated strands of equivalent length, as listed in Table 5. Indeed, the Kemp triacid template enabled incipient triple-helix formation from strands with only three Gly-Pro-Hyp triplets, which is the shortest triple helix reported to date.

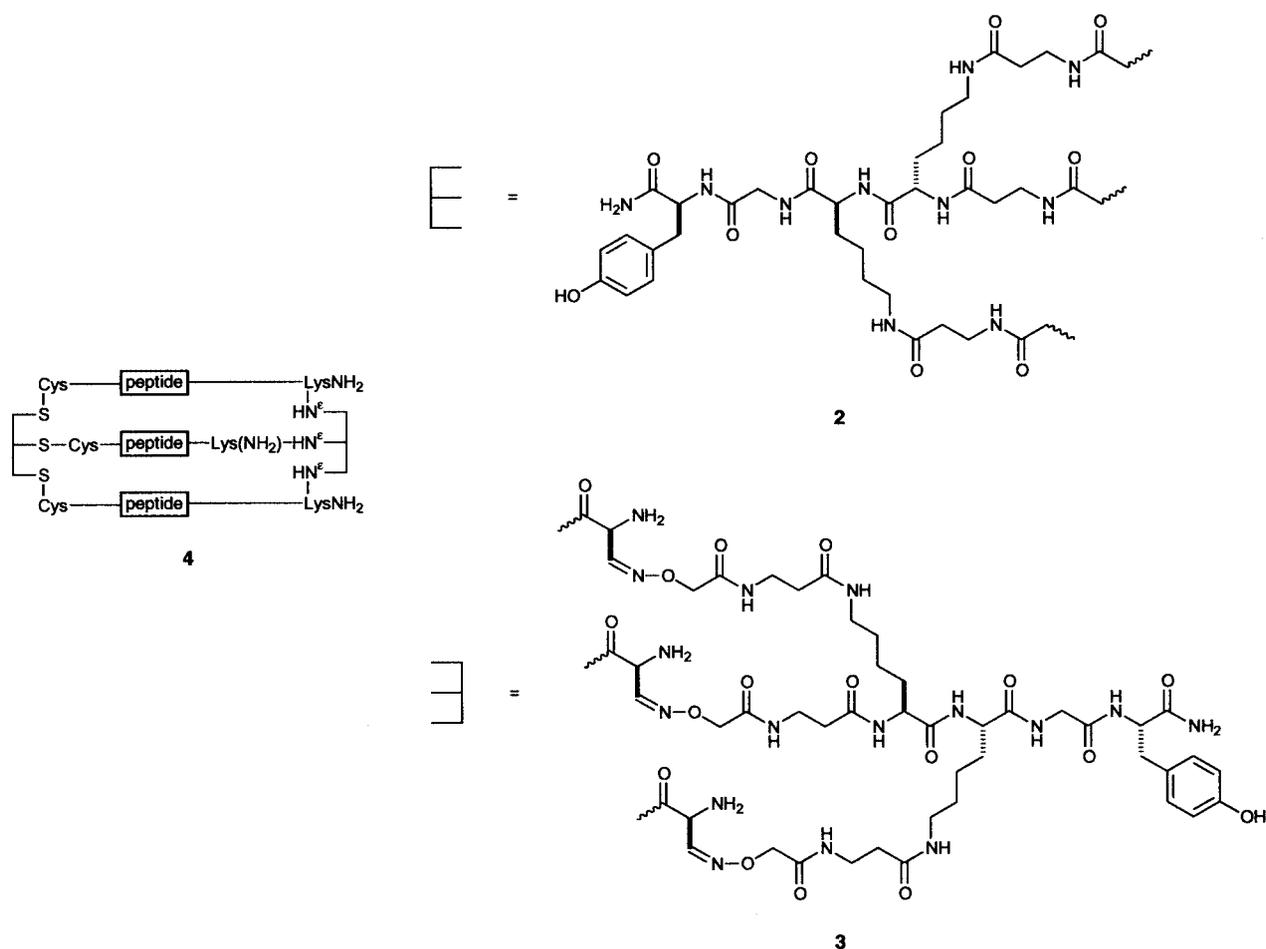
2.4 Peptoid residues

Having shown the stabilizing ability of the Kemp triacid as a template, Goodman and coworkers incorporated N-substituted

Table 5 Conformational stability of Kemp triacid (KTA)- and related acetyl-terminated triple helices⁷¹

Peptide	$T_m/^\circ\text{C}$	
	H ₂ O	Ethylene glycol–H ₂ O (2 : 1)
KTA-[Gly-(Gly-Pro-Hyp)-NH ₂] ₃	No transition	No transition
KTA-[Gly-(Gly-Pro-Hyp) ₃ -NH ₂] ₃	30	50
KTA-[Gly-(Gly-Pro-Hyp) ₅ -NH ₂] ₃	70	(Not determined)
KTA-[Gly-(Gly-Pro-Hyp) ₆ -NH ₂] ₃	81	(Not determined)
Ac-Gly-Pro-Hyp-NH ₂	No transition	No transition
Ac-(Gly-Pro-Hyp) ₃ -NH ₂	No transition	No transition
Ac-(Gly-Pro-Hyp) ₅ -NH ₂	18	32
Ac-(Gly-Pro-Hyp) ₆ -NH ₂	36–37 ^a	59

^a M. Goodman, personal communication.



glycine (peptoid) residues into their collagen mimics. Their initial work focused on a single peptoid residue, *N*-isobutylglycine (Nleu), which they chose because of its bulky hydrophobic side chain.⁷¹ They found that the sequences (Gly-Pro-Nleu)_{*n*} and (Gly-Nleu-Pro)_{*n*} (*n* ≥ 9 and *n* ≥ 6, respectively) form stable triple helices, whereas Gly-Nleu-Nleu has to be included in a host-guest fashion within sequences such as (Gly-Pro-Hyp)_{*n*} to adopt a triple-helical conformation. In addition, (Gly-Nleu-Pro)_{*n*} forms more stable triple helices than does (Gly-Pro-Nleu)_{*n*}, as listed in Table 6. The workers reasoned, with the aid of molecular modeling studies, that the isobutyl group of Nleu can form more hydrophobic contacts with Pro in other chains in triple helices composed of (Gly-Nleu-Pro)_{*n*} than in those composed of (Gly-Pro-Nleu)_{*n*}.

The Goodman group synthesized a series of host-guest peptides with the sequence Ac-(Gly-Nleu-Pro)₃-(Gly-N_{*x*}-Pro)₂-(Gly-Nleu-Pro)₃-NH₂, where N_{*x*} refers to a peptoid residue.⁷³ The various N_{*x*} residues and the T_m values of their triple helices

are listed in Table 7. It is interesting to note that when the side chain of the N_{*x*} residue is 2-hydroxyethyl or 2-aminoethyl, triple helices are not formed; whereas when the side chain is 2(*R*)-hydroxypropyl, the resulting triple helix is quite stable. One explanation is that the functionalized linear alkyl chains lack the steric bulk necessary to stabilize the triple helix. Another explanation is that the solvation of the hydroxy and amino groups is especially disruptive to triple-helix formation. The workers conclude that Gly-Nleu-Pro is an effective triple-helix promoter, that a variety of peptoid residues can be incorporated into collagen mimics, and that hydrophobic effects are important in the inter- and intrachain interactions that lead to stable peptoid-containing triple helices in aqueous solution.

Other work⁷⁴ examined the contribution of a tertiary amide other than Pro, Hyp, or a peptoid residue to the conformational stability of a triple helix. In their study, Kersten and Raines determined the conformational stability of triple helices of

Table 6 Effect of peptoid residues on the conformational stability of Kemp triacid (KTA)- and acetyl-terminated triple helices⁷¹

Peptide	$T_m/^\circ\text{C}$	
	H ₂ O	Ethylene glycol–H ₂ O (2 : 1)
Ac-(Gly-Pro-Nleu) ₆ -NH ₂	No transition	35
Ac-(Gly-Pro-Nleu) ₉ -NH ₂	39	58
(Gly-Pro-Nleu) ₅ -NH ₂	No transition	No transition
(Gly-Pro-Nleu) ₆ -NH ₂	No transition	28
(Gly-Pro-Nleu) ₇ -NH ₂	No transition	39
(Gly-Pro-Nleu) ₉ -NH ₂	28	50
KTA-[Gly-(Gly-Pro-Nleu)-NH ₂] ₃	No transition	No transition
KTA-[Gly-(Gly-Pro-Nleu) ₃ -NH ₂] ₃	No transition	12
KTA-[Gly-(Gly-Pro-Nleu) ₆ -NH ₂] ₃	33	52
KTA-[Gly-(Gly-Pro-Nleu) ₉ -NH ₂] ₃	47	69
Ac-(Gly-Nleu-Pro) ₃ -NH ₂	No transition	No transition
Ac-(Gly-Nleu-Pro) ₆ -NH ₂	26	43
Ac-(Gly-Nleu-Pro) ₉ -NH ₂	^a	^a
KTA-[Gly-(Gly-Nleu-Pro) ₃ -NH ₂] ₃	No transition	22
KTA-[Gly-(Gly-Nleu-Pro) ₆ -NH ₂] ₃	36 ^b	57 ^b
Ac-(Gly-Pro-Hyp) ₂ -(Gly-Nleu-Nleu) ₂ -(Gly-Pro-Hyp) ₂ -NH ₂	No transition	25
KTA-[Gly-(Gly-Pro-Hyp) ₂ -(Gly-Nleu-Nleu) ₂ -(Gly-Pro-Hyp) ₂ -NH ₂] ₃	20	43 ^b

^a Solution became cloudy at 35 °C. ^b M. Goodman, personal communication.

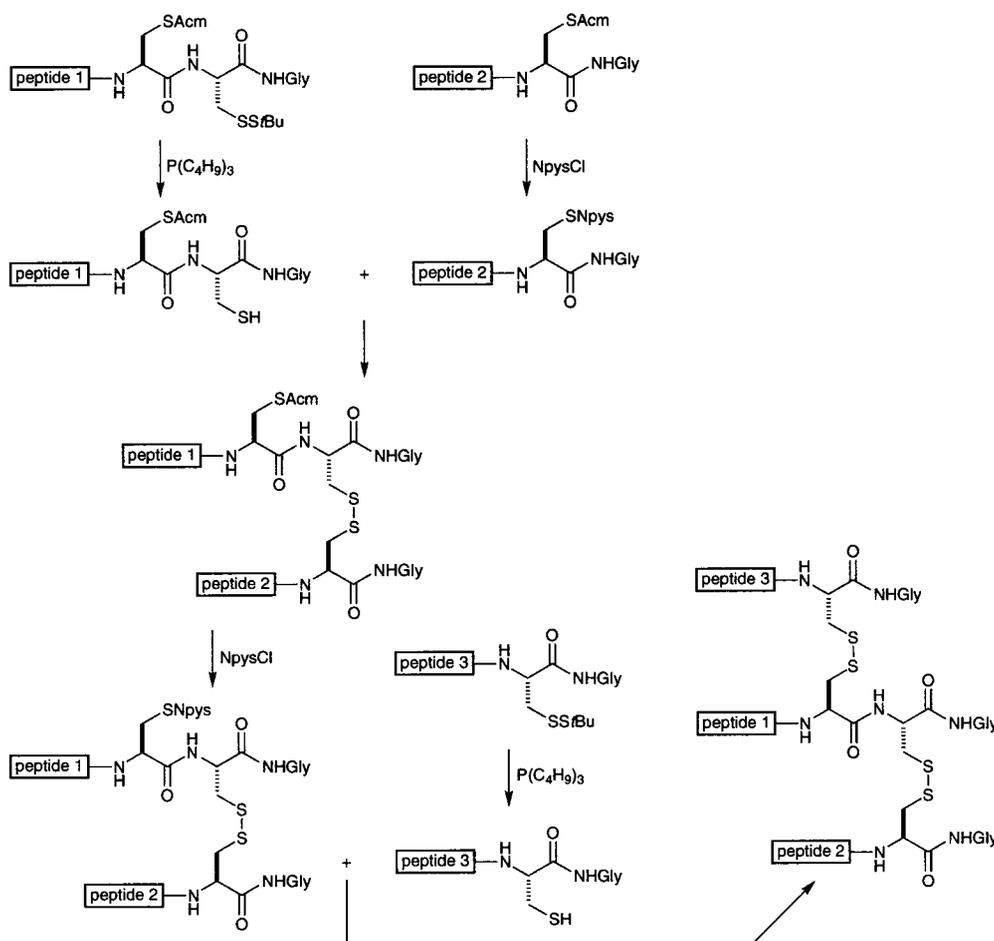


Fig. 3 Scheme for the synthesis of a collagen mimic with a cystine knot tether.^{65–69} NpysCl is 3-nitropyridyl-2-sulfonyl chloride.

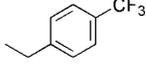
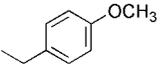
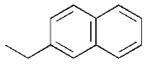
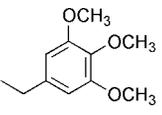
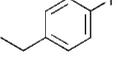
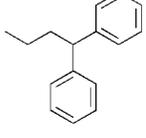
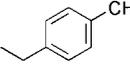
host–guest peptides with the sequence (Pro-Hyp-Gly)₃-Xaa-Yaa-Gly-(Pro-Hyp-Gly)₃. In order of decreasing stability, the central triplets are Pro-Hyp-Gly, Pro-Pro-Gly, Ala-Hyp-Gly, Pro-Ala-Gly, Pro-meAla-Gly, and meAla-Pro-Gly, where meAla refers to *N*-methyl-L-alanine. This residue is identical to Pro and Hyp except for the absence of the –C^γH₂– and –C^γHOH– groups of the pyrrolidine ring, respectively. These workers concluded that the mere presence of tertiary amides in collagen does not make a major contribution to triple-helix

stability. Rather, the conformational restrictions imposed by the pyrrolidine rings of Pro and Hyp are critical.

3 4-Substituted proline residues

In common forms of human collagen, Gly-Xaa-Hyp triplets account for nearly 40% of the amino acid sequence (Fig. 2).⁴⁹ Hyp residues are not incorporated into collagen by ribosomes. Rather, this post-translation modification of Pro residues is

Table 7 Effect of peptoid residues on the conformational stability of host-guest triple helices⁷³

Side chain of N _x	T _m /°C ^a	Side chain of N _x	T _m /°C ^a
	49		42
	47		41
	47		>40 ^b
	46		38
	45	—CH ₃	37
	44		
	44		24
	43		Not triple helical

^a Values of T_m are for triple helices of Ac-(Gly-Nleu-Pro)₃-(Gly-N_x-Pro)₂-(Gly-Nleu-Pro)₃-NH₂ in H₂O. ^b Peptide became insoluble at 40 °C.

mediated by prolyl 4-hydroxylase⁷⁵ after collagen biosynthesis but before the chains form a triple helix. 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.^{76–79} The absence of prolyl 4-hydroxylase is lethal to the nematode *Caenorhabditis elegans*.^{80,81}

In the 1970's, workers began to notice that the Hyp content of a collagen triple helix correlates with its conformational stability.^{82,83} In 1973, Prockop and coworkers demonstrated that triple-helix stability decreases in the order: (Pro-4(*R*)-Hyp-Gly)₁₀ >> (Pro-Pro-Gly)₁₀ >> (4(*R*)-Hyp-Pro-Gly)₁₀, (Pro-4(*S*)-Hyp-Gly)₁₀, or (4(*S*)-Hyp-Pro-Gly)₁₀.^{6,7,84} Understanding the chemical basis for this finding has motivated much work.

Raines and coworkers showed that attaching an electro-negative atom to C^γ has substantial effects on the chemical properties of a proline residue. For example, the nitrogen pK_a of the conjugate acid of 4(*R*)-fluoro-L-proline (FlpOH; 9.23) is lower than that of HypOH (9.68) and ProOH (10.8).⁸⁵ The nitrogen of AcFlpOME is more pyramidal than that of AcHypOME or AcProOME.⁸⁶ This result indicates that the nitrogen of AcFlpOME has greater sp³ character and hence higher electron density. The amide I vibrational mode, which results primarily from the C=O stretching vibration, decreases in the order: AcFlpOME > AcHypOME > AcProOME.⁸⁵ The value of ΔH[‡] for amide bond isomerization is smaller for AcFlpOME than for AcProOME.⁸⁵ Each of these results is consistent with the traditional picture of amide resonance⁸⁷ coupled with an inductive effect that increases the bond order in the amide C=O bond and decreases the bond order in the amide C–N bond. Raines and coworkers suggested that this inductive effect is the basis for the contribution of Hyp residues to the conformational stability of collagen.

3.1 Hydroxyproline residues

New insight on the contribution of Hyp residues to triple-helix stability was inspired by an unusual collagen. Based on the amino acid sequence of cuticle collagen in the hydrothermal

Table 8 Effect of 4(*R*)-hydroxy-L-proline (Hyp) in the Xaa position on the conformational stability of triple helices⁸⁸

Peptide	T _m /°C ^a
Ac-(Gly-Pro-Thr) ₁₀ -NH ₂	No transition
Ac-(Gly-Hyp-Thr) ₁₀ -NH ₂	19.2
Ac-(Gly-Pro-Thr(β-Gal)) ₁₀ -NH ₂	39.2
Ac-(Gly-Hyp-Thr(β-Gal)) ₁₀ -NH ₂	54.8

^a Values of T_m were measured in H₂O.

vent worm *Riftia pachyptila*, Bann and Bächinger made a pair of peptides with Hyp in the Xaa position: Ac-(Gly-Hyp-Thr)₁₀-NH₂ and Ac-(Gly-Hyp-Thr(β-Gal))₁₀-NH₂, where Thr(β-Gal) refers to a threonine residue with a galactose moiety on its side chain.⁸⁸ They compared the ability of these peptides to form triple helices with that of analogous peptides in which Pro replaces Hyp in the Xaa position. The T_m values for these triple helices are listed in Table 8. Hyp in the Xaa position is more stabilizing than is Pro. Most dramatically, Ac-(Gly-Pro-Thr)₁₀-NH₂ does not form triple helices to any appreciable extent, whereas Ac-(Gly-Hyp-Thr)₁₀-NH₂ has a T_m value of 19 °C. Glycosylation of the Thr residues adds additional stability. A possible explanation for the stabilization by Hyp in these peptides is that additional hydrogen bonds are forming with water, that Hyp increases the content of *trans* peptide bonds via an inductive effect,⁸⁵ or both. Why then does (Hyp-Pro-Gly)₁₀ not form stable triple helices?⁸⁴ Bann and Bächinger suggested that having Hyp and Pro adjacent to each other forces the pyrrolidine rings to adopt unfavorable puckers, which could change the ψ torsion angle (N_r-C^α_r-C^γ_r-N_{i+1}) to one that is unfavorable for triple-helix formation. They explained the additional triple-helix stabilization achieved upon glycosylation of the Thr residues by invoking a decrease in the activity of water surrounding the peptides, which in turn decreases the ability of water to form hydrogen bonds with main-chain amides and increases the favorability of amide–amide hydrogen bonds. The authors observed that a triple helix of Ac-(Gly-Hyp-Thr)₁₀-NH₂ has a much larger ΔH_m^o than does a triple helix of (Pro-Hyp-Gly)₁₀ or (Pro-Pro-Gly)₁₀. The strong enthalpic interaction between adjacent Hyp and Thr is presumably due to a hydrogen bond between side-chain hydroxy groups within a chain, between chains, or with water.

3.2 Aminoproline residues

Babu and Ganesh reported on the synthesis and stability of collagen mimics containing 4(*R*)-amino-L-proline (Amp) residues in place of Hyp (5–8).⁸⁹ They found that triple helices containing Amp residues are more stable than are those containing Hyp, and that the differential stability depends upon pH. In some instances, the difference in T_m values is remarkable—over 30 °C—as is listed in Table 9. The effect of Amp residues is, however, a complex function of solution conditions. For example, protonating the Amp amino groups produces both more favorable inductive effects and unfavorable Coulombic interactions. Those unfavorable Coulombic interactions will be more pronounced in a solution of low salt concentration.

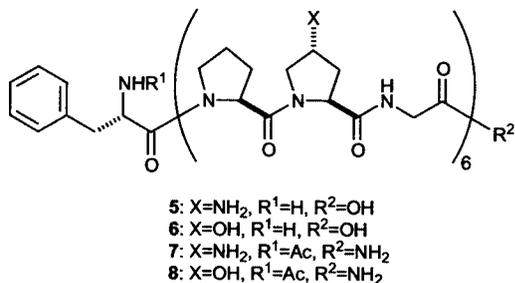
3.3 Fluoroproline residues

The 4(*R*) and 4(*S*) diastereomers of Flp were first synthesized by Witkop and coworkers in 1965 for the purpose of studying whether Flp can be incorporated into proteins by biosynthesis, and if so, whether the Flp incorporated into procollagen is defluorinated to yield Hyp.⁹⁰ *In vivo* studies were carried out by others in 1966.⁹¹ Apparently, the 4(*S*)-Flp diastereomer inhibits protein synthesis to some extent, but is incorporated into proteins in place of Pro. The 4(*R*)-Flp diastereomer is incorporated into proteins to a larger extent and is converted subsequently to

Table 9 Effect of 4(*R*)-amino-L-proline (Amp) on the conformational stability of triple helices⁸⁹

Peptide	$T_m/^\circ\text{C}$			
	pH 3.0 ^a	pH 7.0 ^a	pH 9.0 ^a	pH 12.0 ^a
5	42.5	31.3	18.5	56.6
6	23.0	21.6	15.5	39.6
7	60	56.5	26	49
8	27	28	27	27

^a Buffers: pH 3.0, 20 mM acetate; pH 7.0, 20 mM phosphate; pH 9.0 and 12.0, 20 mM borate. All buffers contained 0.1 M NaCl.



Hyp in collagen strands. Prockop and coworkers later showed that collagen containing 4(*S*)-Flp cannot be exported from cells.^{92,93}

The first incorporation of 4(*R*)-Flp into a collagen mimic was reported by Raines and coworkers.^{94,95} The T_m values of triple helices of (Pro-Pro-Gly)₁₀, (Pro-Hyp-Gly)₁₀, and (Pro-Flp-Gly)₁₀ are listed in Table 10. Flp imparts remarkable stability to the collagen triple helix. Indeed, three strands of (Pro-Flp-Gly)₁₀ form the most stable collagen mimic of similar size known to date. Moreover, because organic fluorine does not form hydrogen bonds,^{96,97} these data confirm that an electron-withdrawing substituent in the 4(*R*) position of Pro can stabilize collagen by a means other than a network of water bridges.⁹⁸

The remarkable stability imparted by Flp to (Pro-Flp-Gly)₁₀ derives from the interplay of several factors,⁹⁹ all of which arise from the inductive effect of the fluorine atom.⁸⁵ First, the *gauche* effect¹⁰⁰ dictates the pyrrolidine ring pucker.⁹⁹ The *gauche* effect arises when two vicinal carbons bear electronegative substituents. These electronegative substituents prefer to reside *gauche* (60°) to each other so that there is maximum overlap between the σ orbitals of more electropositive substituents, such as hydrogen, and the σ^* orbitals of the electronegative substituents, as shown in Fig. 4. As expected from the

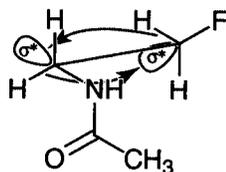


Fig. 4 The *gauche* effect in AcNHCH₂CH₂F.¹⁰⁰ The lobes opposite the C–N and C–F bonds represent σ^* antibonding orbitals, which overlap with the σ bonding orbitals of the indicated C–H bonds.

manifestation of the *gauche* effect, the *C'*-*exo* ring pucker is predominant in Hyp residues in the Yaa position of collagen-like peptides,²⁹ as well as in small-molecule structures of AcHyp-OME and AcFlpOME.⁸⁶ O'Hagan and colleagues showed that the fluorine–amide *gauche* effect is especially strong.¹⁰⁰

The *C'*-*exo* ring pucker preorganizes the main-chain torsion angles of Flp residues. The ϕ angle is a function of ring pucker, as described above.²⁹ Likewise, the ψ angle in crystalline AcFlp-OME is 141°,⁸⁶ which is close to the $\psi = (150 \pm 9)^\circ$ found for residues in the Yaa position of collagen.¹⁸ This value of ψ provides a favorable geometry for an interaction between a non-bonding electron pair of an amide oxygen (O'_{*i*-1}) and the π

Table 10 Effect of 4(*R*)-, 4(*S*)-fluoro-L-proline (Flp), 4(*R*)- and 4(*S*)-hydroxy-L-proline (Hyp) on the conformational stability of triple helices

Peptide	$T_m/^\circ\text{C}^a$
(Pro-Pro-Gly) ₇	6–7 ¹¹⁶
(Pro-4(<i>R</i>)-Hyp-Gly) ₇	36 ⁹⁹
(Pro-4(<i>R</i>)-Flp-Gly) ₇	45 ⁹⁹
(Pro-4(<i>S</i>)-Flp-Gly) ₇	<2 ⁹⁹
(Pro-Pro-Gly) ₁₀	41 ⁹⁴
(Pro-4(<i>R</i>)-Hyp-Gly) ₁₀	69 ⁹⁴
(Pro-4(<i>R</i>)-Flp-Gly) ₁₀	91 ⁹⁴
(Pro-4(<i>S</i>)-Hyp-Gly) ₁₀	<5 ⁷

^a Values of T_m were measured in dilute solutions of aqueous acetic acid.

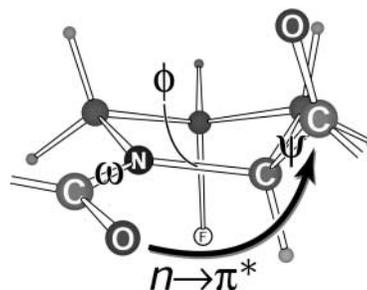


Fig. 5 Main-chain ω , ϕ , and ψ torsion angles of a Flp residue.⁸⁶ The *gauche* effect (Fig. 4) fixes the pyrrolidine ring pucker (*C'*-*exo*). In that ring pucker, the ω , ϕ , and ψ angles are preorganized at values close to those in the Yaa position of a collagen triple helix.⁹⁹ The indicated $n \rightarrow \pi^*$ interaction contributes to that preorganization.

antibonding orbital of an amide carbon (*C'*_{*i*}), as shown in Fig. 5. The O \cdots C=O angle in AcFlpOME is 98°, which is close to the ideal angle for such an $n \rightarrow \pi^*$ interaction.^{101–103} Moreover, the O \cdots C=O distance in AcFlpOME is only 2.76 Å, which predicates a meaningful interaction. Indeed, the ester carbonyl stretching vibration is lower by 6 cm⁻¹ in Ac-4(*R*)-Flp-OME than in Ac-4(*S*)-Flp-OME, presumably because the $n \rightarrow \pi^*$ interaction decreases the C=O bond order.

The $n \rightarrow \pi^*$ interaction stabilizes not only the ideal ψ angle for triple-helix formation, but also the requisite *trans* conformation ($\omega = 180^\circ$) of the Flp peptide bond. In the *cis* conformation ($\omega = 0^\circ$), C'_{*i*-1} rather than O'_{*i*-1} would be proximal to C'_{*i*}, and no $n \rightarrow \pi^*$ interaction can occur. Accordingly, as the electronegativity of the substituent in the 4(*R*)-position increases, the *trans* : *cis* ratio of the amide bond also increases.⁸⁵ The reverse trend is true for electronegative 4(*S*) substituents, which impose a *C'*-*endo* pucker and do not stabilize triple helices.^{7,99} The association of ω angle with pyrrolidine ring pucker provides an explanation for the observation that *cis* prolyl peptide bonds tend to have *endo* ring pucker in crystalline proteins.¹⁰⁴ In summary, Flp in the Yaa position stabilizes collagen by a stereoelectronic effect—the *gauche* effect—which fixes the pyrrolidine ring pucker and thus preorganizes all three main-chain torsion angles: ω , ϕ , and ψ (Fig. 5). This same reasoning applies to Hyp and Amp residues.

Moroder and coworkers have since used Flp in another context.¹⁰⁵ They incorporated 4(*R*)-Flp and 4(*S*)-Flp residues in place of a Pro residue with a *cis* peptide bond in barstar. They observed that 4(*S*)-Flp, which favors the *cis* conformation more than does Pro, stabilizes the protein and 4(*R*)-Flp, which favors the *trans* conformation more than does Pro, destabilizes the protein. Thus, 4(*R*)- and 4(*S*)-Flp residues can be useful tools for protein engineers.

4 Collagen as a biomaterial

Collagen is an important biomaterial.^{106,107} For example, collagen is the principal component of biodegradable sutures and

Table 11 Cytotoxicity and cell-binding of triple helices¹¹⁴

Peptide	n	Cytotoxicity	Inhibition of cell attachment ^c	
			Epithelial cells	Fibroblasts
(Gly-Pro-Nleu) _n -NH ₂	1	—	ND	—
	2	—	ND	ND
	3	—	—	—
	5	—	—	—
	7	—	ND	ND
(Gly-Pro-Nleu) _n -Gly-Pro-NH ₂ ^a	10	—	+	+
(Gly-Pro-Nleu) _n -Gly-Pro-NH ₂ ^b	10	—	+	ND
Ac-(Gly-Pro-Nleu) _n -NH ₂	1	—	—	—
	9	—	+	+
	10	—	—	ND
(Gly-Nleu-Pro) _n -NH ₂	10	—	—	—
Ac-(Gly-Nleu-Pro) _n -NH ₂	3	—	ND	—
	6	—	ND	—
	10	—	Insoluble	Insoluble
(Gly-Pro-Hyp) _n -NH ₂	9	—	—	—
KTA-[Gly-(Gly-Pro-Nleu) _n -NH ₂] ₃	9	—	+	+
KTA-[Gly-(Gly-Pro-Hyp) _n -NH ₂] ₃	5	—	ND	—
RGES		ND	—	—
GRGDSPK		ND	+	+

^a Peptide was left in solution at 4 °C for a minimum of 7 days before use. ^b Peptide solution was used immediately after preparation. ^c +, inhibition of cell attachment; —, no inhibition; ND, not determined.

artificial heart valves. Some of the advantages of using collagen as a biomaterial are its low immunogenicity and high durability. One disadvantage is that it is difficult to obtain collagen in high purity without degrading its structural integrity. In addition, chemical methods used to cross-link purified collagen can cause cytotoxicity. Furthermore, the collagen that is used most often as a biomaterial is bovine collagen, which can engender allergic and immunological side effects in humans, as well as other health risks. Several different constructs for producing recombinant human fibrillar collagens have been reported. For example, human procollagen II has been isolated from a stably transfected human tumor cell line¹⁰⁸ and from yeast,¹⁰⁹ human type III collagen has been expressed in baculovirus¹¹⁰ and yeast¹⁰⁹ systems; and human type I collagen has been produced in transgenic tobacco plants,¹¹¹ mouse mammary glands,¹¹² and two different yeast strains.^{109,113}

Despite the numerous studies on collagen mimics, few have been tested as biomaterials. One report by Goodman and coworkers has, however, provided encouraging results.¹¹⁴ In this report, collagen mimics containing Nleu were tested for the ability to inhibit fibroblast or epithelial cell attachment to polystyrene. The results are listed in Table 11. Triple-helical (Gly-Pro-Nleu)_n, but neither (Gly-Nleu-Pro)_n nor (Gly-Pro-Hyp)_n, inhibited cell attachment. It is important to note that none of the peptides tested showed any cytotoxicity. Hence, (Gly-Pro-Nleu)-containing sequences show promise for development into biomaterials.

5 Envoi

The appearance of high-resolution structures of crystalline collagen triple helices has led to a resurgence of interest in chemical aspects of this ubiquitous protein. With these structures as a guide, chemists, biochemists, and biophysicists have made much progress in delineating the forces responsible for the conformational stability of the triple helix. Still, important questions remain without answers. For example, how much does the ladder of XaaC=O...HNGly hydrogen bonds (Fig. 1) contribute to stability? In what contexts does Hyp, Amp, or Flp in the Xaa position stabilize a triple helix? Which main-chain torsion angle (ω , ϕ , or ψ ; Fig. 5) is best to pre-organize so as to enhance stability? Do changes in helical pitch affect the results of host-guest studies? Does the rare 3-hydroxy-L-proline residue, which is subject to a *gauche* effect in its pyrrolidine ring, contribute to conformational stability?

Which template provides the most accurate collagen mimic? Does greater triple-helix stability translate to greater fibril stability? And, most importantly, how can we use our knowledge of the collagen triple helix to solve real problems in biomedicine?

6 Acknowledgment

Work on collagen in the Raines laboratory is supported by grant AR44276 (NIH).

7 References

- 1 J. Myllyharju and K. I. Kivirikko, *Ann. Med.*, 2001, **33**, 7.
- 2 D. J. Prockop and K. I. Kivirikko, *Annu. Rev. Biochem.*, 1995, **64**, 403.
- 3 D. J. Prockop, *Matrix Biol.*, 1998, **16**, 519.
- 4 D. J. Prockop, *Biochem. Soc. Trans.*, 1999, **27**, 15.
- 5 P. H. Byers, *Clin. Genet.*, 2000, **58**, 270.
- 6 R. A. Berg and D. J. Prockop, *Biochem. Biophys. Res. Commun.*, 1973, **52**, 115.
- 7 K. Inouye, S. Sakakibara and D. J. Prockop, *Biochim. Biophys. Acta*, 1976, **420**, 133.
- 8 G. B. Fields and D. J. Prockop, *Biopolymers*, 1996, **40**, 345.
- 9 M. A. Rougvie and R. S. Bear, *J. Am. Leather Chem. Ass.*, 1953, **48**, 735.
- 10 G. N. Ramachandran, *Chemistry of Collagen*, Academic Press, London, 1967.
- 11 R. D. B. Fraser, T. P. MacRae and E. Suzuki, *J. Mol. Biol.*, 1979, **129**, 463.
- 12 G. N. Ramachandran and G. Kartha, *Nature*, 1954, **174**, 269.
- 13 G. N. Ramachandran and G. Kartha, *Nature*, 1955, **176**, 593.
- 14 C. Ramakrishnan, *Protein Sci.*, 2001, **10**, 1689.
- 15 A. Rich and F. H. C. Crick, *Nature*, 1955, **176**, 915.
- 16 A. Rich and F. H. C. Crick, *J. Mol. Biol.*, 1961, **3**, 483.
- 17 G. Riddihough, *Nat. Struct. Biol.*, 1998, **5**, 858.
- 18 J. Bella, M. Eaton, B. Brodsky and H. M. Berman, *Science*, 1994, **266**, 75.
- 19 K. H. Gustavson, in *Connective Tissue*, ed. R. E. Tunbridge, Blackwell, Oxford, 1957, p. 185.
- 20 E. Suzuki, R. D. B. Fraser and T. P. MacRae, *Int. J. Biol. Macromol.*, 1980, **2**, 54.
- 21 J. Bella, B. Brodsky and H. M. Berman, *Structure*, 1995, **3**, 893.
- 22 R. Z. Kramer and H. M. Berman, *J. Biomol. Struct. Dyn.*, 1998, **16**, 367.
- 23 R. Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari and H. M. Berman, *J. Mol. Biol.*, 1998, **280**, 623.
- 24 F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga, *J. Phys. Chem.*, 1975, **79**, 2361.
- 25 K. Okuyama, V. Nagarajan and S. Kamitori, *Proc. Indian Acad. Sci. (Chem. Sci.)*, 1999, **111**, 19.

- 26 V. Nagarajan, S. Kamitori and K. Okuyama, *J. Biochem.*, 1999, **125**, 310.
- 27 J. Engel, H.-T. Chen, D. J. Prockop and H. Klump, *Biopolymers*, 1977, **16**, 601.
- 28 R. Berisio, L. Vitagliano, G. Sorrentino, L. Carotenuto, C. Piccolo, L. Mazarella and A. Zagari, *Acta Crystallogr., Sect. D*, 2000, **56**, 55.
- 29 L. Vitagliano, R. Berisio, L. Mazarella and A. Zagari, *Biopolymers*, 2001, **58**, 459.
- 30 R. Z. Kramer, J. Bella, P. Mayville, B. Brodsky and H. M. Berman, *Nat. Struct. Biol.*, 1999, **6**, 454.
- 31 C. G. Long, M. H. Li, J. Baum and B. Brodsky, *J. Mol. Biol.*, 1992, **225**, 1.
- 32 M.-H. Li, P. Fan, B. Brodsky and J. Baum, *Biochemistry*, 1993, **32**, 7377.
- 33 G. Melacini and M. Goodman, *Chirality*, 1998, **10**, 28.
- 34 P. Fan, M. H. Li, B. Brodsky and J. Baum, *Biochemistry*, 1993, **32**, 13299.
- 35 G. Melacini, A. M. J. J. Bonvin, M. Goodman, R. Boelens and R. Kaptein, *J. Mol. Biol.*, 2000, **300**, 1041.
- 36 X. Liu, D. L. Siegel, P. Fan, B. Brodsky and J. Baum, *Biochemistry*, 1996, **35**, 4306.
- 37 X. Liu, S. Kim, Q. H. Dai, B. Brodsky and J. Baum, *Biochemistry*, 1998, **35**, 15528.
- 38 A. B. Buevich, Q. H. Dai, X. Liu, B. Brodsky and J. Baum, *Biochemistry*, 2000, **39**, 4299.
- 39 K. T. O'Neil and W. F. DeGrado, *Science*, 1990, **250**, 646.
- 40 M. Blaber, X. Zhang and B. W. Matthews, *Science*, 1993, **260**, 1637.
- 41 K. Groebke, P. Renold, K. Y. Tsang, T. J. Allen, K. F. McClure and D. S. Kemp, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 4025.
- 42 J. K. Myers, C. N. Pace and J. M. Scholtz, *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 2833.
- 43 J. X. Yang, E. J. Spek, Y. X. Gong, H. X. Zhou and N. R. Kallenbach, *Protein Sci.*, 1997, **6**, 1264.
- 44 C. W. A. Kim and J. M. Berg, *Nature*, 1993, **362**, 267.
- 45 D. L. Minor and P. S. Kim, *Nature*, 1994, **371**, 264.
- 46 D. E. Otzen and A. R. Fersht, *Biochemistry*, 1995, **34**, 5718.
- 47 A. G. Street and S. L. Mayo, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 9074.
- 48 A. V. Persikov, J. A. M. Ramshaw and B. Brodsky, *Biopolymers*, 2000, **55**, 436.
- 49 J. A. M. Ramshaw, N. K. Shah and B. Brodsky, *J. Struct. Biol.*, 1998, **122**, 86.
- 50 S. Sakakibara, K. Inouye, K. Shudo, Y. Kishida, Y. Kobayashi and D. J. Prockop, *Biochim. Biophys. Acta.*, 1973, **303**, 198.
- 51 H.-P. Germann and E. Heidemann, *Biopolymers*, 1988, **27**, 157.
- 52 A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick and B. Brodsky, *Biochemistry*, 2000, **39**, 14960.
- 53 E. Y. Jones and A. Miller, *J. Mol. Biol.*, 1991, **218**, 209.
- 54 W. Yang, V. C. Chan, A. Kirkpatrick, J. A. M. Ramshaw and B. Brodsky, *J. Biol. Chem.*, 1997, **272**, 28837.
- 55 N. K. Shah, J. A. M. Ramshaw, A. Kirkpatrick, C. Shah and B. Brodsky, *Biochemistry*, 1996, **35**, 10262.
- 56 W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1.
- 57 V. C. Chan, J. A. M. Ramshaw, A. Kirkpatrick, K. Beck and B. Brodsky, *J. Biol. Chem.*, 1997, **272**, 31441.
- 58 S. Thakur, D. Vadolas, H. P. Germann and E. Heidemann, *Biopolymers*, 1986, **25**, 1081.
- 59 C. G. Fields, C. M. Lovdahl, A. J. Miles, V. L. Matthias Hagen and G. B. Fields, *Biopolymers*, 1993, **33**, 1695.
- 60 C. G. Fields, D. J. Mickelson, S. L. Drake, J. B. McCarthy and G. B. Fields, *J. Biol. Chem.*, 1993, **268**.
- 61 T. Tanaka, Y. Wada, H. Nakamura, T. Doi, T. Imanishi and T. Kodama, *FEBS Lett.*, 1993, **334**, 272.
- 62 Y. Tanaka, K. Suzuki and T. Tanaka, *J. Peptide Res.*, 1998, **51**, 413.
- 63 G. B. Fields, *Bioorg. Med. Chem.*, 1999, **7**, 75.
- 64 S. Frank, R. A. Kammerer, D. Mechling, T. Schulthess, R. Landwehr, J. Bann, Y. Guo, A. Lustig, H. P. Bächinger and J. Engel, *J. Mol. Biol.*, 2001, **308**, 1081.
- 65 J. Ottl, R. Battistuta, M. Pieper, H. Tschesche, W. Bode, K. Kuhn and L. Moroder, *FEBS Lett.*, 1996, **398**, 31.
- 66 J. Ottl and L. Moroder, *J. Am. Chem. Soc.*, 1999, **121**, 653.
- 67 J. Ottl, H. J. Musiol and L. Moroder, *J. Pept. Sci.*, 1999, **5**, 103.
- 68 J. C. Muller, J. Ottl and L. Moroder, *Biochemistry*, 2000, **39**, 5111.
- 69 J. Ottl, D. Gabriel, G. Murphy, V. Knauper, Y. Tominaga, H. Nagase, M. Kroger, H. Tschesche, W. Bode and L. Moroder, *Chem. Biol.*, 2000, **7**, 119.
- 70 D. S. Kemp and K. S. Petrakis, *J. Org. Chem.*, 1981, **46**, 5140.
- 71 M. Goodman, M. Bhumralkar, E. A. Jefferson, J. Kwak and E. Locardi, *Biopolymers*, 1998, **47**, 127.
- 72 M. Goodman, Y. Feng, G. Melacini and J. P. Taulane, *J. Am. Chem. Soc.*, 1996, **118**, 5156.
- 73 J. Kwak, E. A. Jefferson, M. Bhumralkar and M. Goodman, *Bioorg. Med. Chem.*, 1999, **7**, 153.
- 74 E. A. Kersteen and R. T. Raines, *Biopolymers*, 2001, **59**, 24.
- 75 N. A. Guzman, *Prolyl hydroxylase, protein disulfide isomerase, and other structurally related proteins*, Marcel Dekker, New York, 1998.
- 76 N. J. Bulleid, R. Wilson and J. F. Lees, *Biochem. J.*, 1996, **317**, 195.
- 77 A. R. Walmsley, M. R. Batten, U. Lad and N. J. Bulleid, *J. Biol. Chem.*, 1999, **274**, 14884.
- 78 A. Snellman, M.-R. Keranen, P. O. Hagg, A. Lamberg, J. K. Hiltunen, K. I. Kivirikko and T. Pihlajaniemi, *J. Biol. Chem.*, 2000, **275**, 8936.
- 79 P. H. Byers, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 2001, **356**, 151.
- 80 L. Friedman, J. J. Higgin, G. Moulder, R. Barstead, R. T. Raines and J. Kimble, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 4736.
- 81 A. D. Winter and A. P. Page, *Mol. Cell. Biol.*, 2000, **20**, 4084.
- 82 T. V. Burjanadze, *Biopolymers*, 1979, **18**, 931.
- 83 T. V. Burjanadze, *Biopolymers*, 2000, **53**, 523.
- 84 K. Inouye, Y. Kobayashi, Y. Kyogoku, Y. Kishida, S. Sakakibara and D. J. Prockop, *Arch. Biochem. Biophys.*, 1982, **219**, 198.
- 85 E. S. Eberhardt, N. Panasik Jr. and R. T. Raines, *J. Am. Chem. Soc.*, 1996, **118**, 12261.
- 86 N. Panasik Jr., E. S. Eberhardt, A. S. Edison, D. R. Powell and R. T. Raines, *Int. J. Pept. Protein Res.*, 1994, **44**, 262.
- 87 L. Pauling, *The Nature of the Chemical Bond*, 3rd edn., Cornell University Press, Ithaca, NY, 1960.
- 88 J. G. Bann and H. P. Bächinger, *J. Biol. Chem.*, 2000, **275**, 24466.
- 89 I. R. Babu and K. N. Ganesh, *J. Am. Chem. Soc.*, 2001, **123**, 2079.
- 90 A. A. Gottlieb, Y. Fujita, S. Udenfriend and B. Witkop, *Biochemistry*, 1965, **4**, 2507.
- 91 S. Bakerman, R. L. Martin, A. W. Burgstahler and J. W. Hayden, *Nature*, 1966, **212**, 849.
- 92 T. Takeuchi and D. J. Prockop, *Biochim. Biophys. Acta*, 1969, **175**, 142.
- 93 T. Takeuchi, J. Rosenbloom and D. J. Prockop, *Biochim. Biophys. Acta*, 1969, **175**, 156.
- 94 S. K. Holmgren, K. M. Taylor, L. E. Bretscher and R. T. Raines, *Nature*, 1998, **392**, 666.
- 95 S. K. Holmgren, L. E. Bretscher, K. M. Taylor and R. T. Raines, *Chem. Biol.*, 1999, **6**, 63.
- 96 J. A. K. Howard, V. J. Hoy, D. O'Hagan and G. T. Smith, *Tetrahedron*, 1996, **52**, 12613.
- 97 J. D. Dunitz and R. Taylor, *Chem. Eur. J.*, 1997, **3**, 89.
- 98 J. Engel and D. J. Prockop, *Matrix Biol.*, 1998, **17**, 679.
- 99 L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider and R. T. Raines, *J. Am. Chem. Soc.*, 2001, **123**, 777.
- 100 D. O'Hagan, C. Bilton, J. A. K. Howard, L. Knight and D. J. Tozer, *J. Chem. Soc., Perkin Trans. 2*, 2000, 605.
- 101 H. B. Bürgi, J. D. Dunitz and E. Shefter, *J. Am. Chem. Soc.*, 1973, **95**, 5065.
- 102 H. B. Bürgi, J. D. Dunitz, J. M. Lehn and G. Wipff, *Tetrahedron*, 1974, **30**, 1563.
- 103 H. B. Bürgi, J. M. Lehn and G. Wipff, *J. Am. Chem. Soc.*, 1974, **96**, 1965.
- 104 J. E. Milner-White, L. H. Bell and P. H. Maccallum, *J. Mol. Biol.*, 1992, **228**, 725.
- 105 C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber and L. Moroder, *Angew. Chem., Int. Ed.*, 2001, **40**, 923.
- 106 J. A. Werkmeister and J. A. M. Ramshaw, *Collagen Biomaterials*, Elsevier Science, Barking, UK, 1992.
- 107 J. A. M. Ramshaw, J. A. Werkmeister and V. Glattauer, *Biotechnol. Genet. Eng. Rev.*, 1995, **13**, 335.
- 108 A. Fertala, A. L. Sieron, A. Ganguly, S.-H. Li, L. Ala-Kokko, K. R. Anumula and D. J. Prockop, *Biochem. J.*, 1994, **298**, 31.
- 109 J. Myllyharju, M. Nokelainen, A. Vuorela and K. I. Kivirikko, *Biochem. Soc. Trans.*, 2000, **28**, 353.
- 110 A. Lamberg, T. Helaakoski, J. Myllyharju, S. Peltonen, H. Notbohm, T. Pihlajaniemi and K. I. Kivirikko, *J. Biol. Chem.*, 1996, **271**, 11988.
- 111 F. Ruggiero, J.-Y. Exposito, P. Bournat, V. Gruber, S. Perret, J. Comte, B. Olagnier, R. Garrone and M. Thiesen, *FEBS Lett.*, 2000, **469**, 132.
- 112 N. J. Bulleid, D. C. A. John and K. E. Kadler, *Biochem. Soc. Trans.*, 2000, **28**, 350.
- 113 P. D. Toman, G. Chisholm, H. McMullin, L. M. Giere, D. R. Olsen, R. J. Kovach, S. D. Leigh, B. E. Fong, R. Chang, G. A. Daniels, R. A. Berg and R. A. Hitzeman, *J. Biol. Chem.*, 2000, **275**, 23303.
- 114 G. Johnson, M. Jenkins, K. M. McClean, H. J. Griesser, J. Kwak, M. Goodman and J. G. Steele, *J. Biomed. Mater. Res.*, 2000, **51**, 612.
- 115 B. J. Stapley and T. P. Creamer, *Protein Sci.*, 1999, **8**, 587.
- 116 B. R. Shaw and J. M. Schurr, *Biopolymers*, 1975, **14**, 1951.