NIH Public Access
Author Manuscript
Annu Rev Biochem. Author manuscript; available in PMC 2010 March 29.

Published in final edited form as:
Annu Rev Biochem. 2009 ; 78: 929–958. doi:10.1146/annurev.biochem.77.032207.120833.

COLLAGEN STRUCTURE AND STABILITY

Matthew D. Shoulders[1] and **Ronald T. Raines**[1,2]
Ronald T. Raines: raines@biochem.wisc.edu
[1]Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706
[2]Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

Abstract

Collagen is the most abundant protein in animals. This fibrous, structural protein comprises a right-handed bundle of three parallel, left-handed polyproline II-type helices. Much progress has been made in elucidating the structure of collagen triple helices and the physicochemical basis for their stability. New evidence demonstrates that stereoelectronic effects and preorganization play a key role in that stability. The fibrillar structure of type I collagen–the prototypical collagen fibril–has been revealed in detail. Artificial collagen fibrils that display some properties of natural collagen fibrils are now accessible using chemical synthesis and self-assembly. A rapidly emerging understanding of the mechanical and structural properties of native collagen fibrils will guide further development of artificial collagenous materials for biomedicine and nanotechnology.

Keywords

biomaterials; extracellullar matrix; fibrillogenesis; proline; stereoelectronic effects; triple helix

INTRODUCTION

Collagen is an abundant structural protein in all animals. In humans, collagen comprises one-third of the total protein, accounts for three-quarters of the dry weight of skin, and is the most prevalent component of the extracellular matrix (ECM). Twenty-eight different types of collagen composed of at least 46 distinct polypeptide chains have been identified in vertebrates, and many other proteins contain collagenous domains (1,2). Remarkably, intact collagen has been discovered in soft tissue of the fossilized bones of a 68 million-year-old *Tyrannosaurus rex* fossil (3,4), by far the oldest protein detected to date. That discovery is, however, under challenge (5,6).

The defining feature of collagen is an elegant structural motif in which three parallel polypeptide strands in a left-handed, polyproline II-type (PPII) helical conformation coil about each other with a one-residue stagger to form a right-handed triple helix (Figure 1). The tight packing of PPII helices within the triple helix mandates that every third residue be Gly, resulting in a repeating XaaYaaGly sequence, where Xaa and Yaa can be any amino acid. This repeat occurs in all types of collagen, although it is disrupted at certain locations within the triple-

DISCLOSURE STATEMENT
The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

NOTE ADDED IN PROOF
A twenty-ninth form of vertebrate collagen has been found in skin, lung, and intestine. Söderhäll C, Marenholz I, Kerscher T, Rüschendorf F, Esparza-Gordillo J, et al. 2007. Variants in a novel epidermal collagen gene (*COL29A1*) are associated with atopic dermatitis. *PLoS Biol*. 5:e242

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

helical domain of nonfibrillar collagens (8). The amino acids in the Xaa and Yaa positions of collagen are often (2*S*)-proline (Pro, 28%) and (2*S*,4*R*)-4-hydroxyproline (Hyp, 38%), respectively. ProHypGly is the most common triplet (10.5%) in collagen (9). In animals, individual collagen triple helices, known as tropocollagen (TC), assemble in a complex, hierarchical manner that ultimately leads to the macroscopic fibers and networks observed in tissue, bone, and basement membranes (Figure 2).

The categories of collagen include the classical fibrillar and network-forming collagens, the FACITs (fibril-associated collagens with interrupted triple helices), MACITs (membrane-associated collagens with interrupted triple helices), and MULTIPLEXINs (multiple triple-helix domains and interruptions). Collagen types, their distribution, composition, and pathology are listed in Table 1. It is noteworthy that, although the three polypeptide chains in the triple helix of each collagen type can be identical, heterotrimeric triple helices are more prevalent than are homotrimeric triple helices.

STRUCTURE OF THE COLLAGEN TRIPLE HELIX

In 1940, Astbury & Bell (11) proposed that the collagen molecule consists of a single extended polypeptide chain with all amide bonds in the *cis* conformation. A significant advance was achieved when, in the same 1951 issue of the *Proceedings of the National Academy of Sciences* in which he and coworkers put forth the correct structures for the α-helix and β-sheet, Pauling & Corey (12) proposed a structure for collagen. In that structure, three polypeptide strands were held together in a helical conformation by hydrogen bonds. Within each amino acid triplet, those hydrogen bonds engaged four of the six main chain heteroatoms, and their formation required two of the three peptide bonds to be in the *cis* conformation. In 1954, Ramachandran & Kartha (13,14) advanced a structure for the collagen triple helix on the basis of fiber diffraction data. Their structure was a right-handed triple helix of three staggered, left-handed PPII helices with all peptide bonds in the *trans* conformation and two hydrogen bonds within each triplet. In 1955, this structure was refined by Rich & Crick (15–16) and by North and coworkers (17) to the triple-helical structure accepted today–, which has a single interstrand $N–H_{(Gly)}\cdots O=C_{(Xaa)}$ hydrogen bond per triplet and a tenfold helical symmetry with a 28.6-Å axial repeat (10/3 helical pitch) (Figure 1).

Fiber diffraction studies cannot reveal the structure of collagen at atomic resolution. Exacerbating this difficulty, the large size, insolubility, repetitive sequence, and complex hierarchical structure of native collagen thwart most biochemical and biophysical analyses. Hence, a reductionist approach using triple-helical, collagen-related peptides (CRPs) has been employed extensively since the late 1960s (18).

In 1994, Berman and coworkers (19) reported the first high-resolution crystal structure of a triple-helical CRP (Figure 1*a*). This structure confirmed the existence of interstrand $N–H_{(Gly)}\cdots O=C_{(Xaa)}$ hydrogen bonds (Figure 1*c*,*d*) and provided additional insights, including that $C^{\alpha}–H_{(Gly/Yaa)}\cdots O=C_{(Xaa/Gly)}$ hydrogen bonds could likewise stabilize the triple helix (20). Using CRPs and X-ray crystallography, the structural impact of a single Gly → Ala substitution was observed (19), the effects of neighboring charged residues in a triple helix were analyzed (21), and a snapshot of the interaction of a triple-helical CRP with the I– domain of integrin $\alpha_2\beta_1$ was obtained (Figure 3) (22).

Most X-ray crystallographic studies on CRPs have been performed on proline-rich collagenous sequences. All of the resulting structures have a 7/2 helical pitch (20.0-Å axial repeat), in contrast to the 10/3 helical pitch (28.6-Å axial repeat) predicted for natural collagen by fiber diffraction (17). On the basis of X-ray crystal structures of proline-rich CRPs, and in accordance with an early proposal regarding the helical pitch of natural triple helices (23), Okuyama and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

coworkers (24) postulated that the correct average helical pitch for natural collagen is 7/2. The generality of this hypothesis is unclear, as few regions of natural collagen are as proline rich as the CRPs analyzed by X-ray crystallography. The actual helical pitch of collagen likely varies across the domains and types of natural collagen. Specifically, the helical pitch could be 10/3 in proline-poor regions and 7/2 in proline-rich regions. This proposal is supported by the observation that proline-poor regions within crystalline CRPs occasionally display a 10/3 helical pitch (25,26). Variability in the triple-helical pitch of native collagen could play a role in the interaction of collagenous domains with other biomolecules (22,27–29).

UNDERSTANDING TRIPLE-HELIX STRUCTURE AND STABILITY

The vital importance of collagen as a scaffold for animals demands a manifold of essential characteristics. These characteristics include thermal stability, mechanical strength, and the ability to engage in specific interactions with other biomolecules. Understanding how such properties are derived from the fundamental structural unit of collagen, the triple helix, necessitates a comprehensive knowledge of the mechanisms underlying triple-helix structure and stability.

Interstrand Hydrogen Bonds

The ubiquity of collagen makes the ladder of recurrent $N–H_{(Gly)}\cdots O=C_{(Xaa)}$ hydrogen bonds that form within the triple helix (Figure 1*c*,*d*) the most abundant amide–amide hydrogen bond in kingdom *Animalia*. Replacing the Yaa–Gly amide bond with an ester in a host-guest CRP (Figure 4*a*,*b*) enabled estimation of the strength of each amide–amide hydrogen bond as $\Delta G° = -2.0$ kcal/mol (30). Boryskina and coworkers (31) used a variety of other experimental techniques to assess this same parameter, estimating the strength of each amide–amide hydrogen bond within a poly(GlyProPro) CRP as $\Delta G° = -1.8$ kcal/mol and within native collagen as $\Delta G° = -1.4$ kcal/mol.

Glycine Substitutions

Numerous collagen-related diseases are associated with mutations in both triple-helical and nontriple-helical domains of various collagens (Table 1). These diseases have been reviewed in detail elsewhere (32) and are not discussed extensively herein.

The Gly residue in the XaaYaaGly repeat is invariant in natural collagen, and favorable substitutions are unknown in CRPs (33). A computational study suggested that replacing the obligate Gly residues of collagen with D-alanine or D-serine would stabilize the triple helix (34) and thus that the Gly residues in collagen are surrogates for nonnatural D-amino acids. Subsequent experimental data demonstrated, however, that this notion was erroneous (35).

Many of the most damaging mutations to collagen genes result in the substitution of a Gly residue involved in the ladder of hydrogen bonds within the triple helix (Figure 1*c*,*d*). Both the identity of the amino acid replacing Gly and the location of that substitution can impact the pathology of, for example, osteogenesis imperfecta (OI) (33, 36). Substitutions for Gly in proline-rich portions of the collagen sequence (Figure 3*a*) are far less disruptive than those in proline-poor regions, a testament to the importance of Pro derivatives for triple-helix nucleation (37). In vivo, triple helices fold in a C-terminal→N-terminal manner (38). The time delay between disruption of triple-helix folding by a Gly substitution and renucleation of the folding process N-terminal to the substitution site is much shorter when triple-helix nucleating, proline-rich sequences are immediately N-terminal to the substitution site (37). Any delay in triple-helix folding results in overmodification of the protocollagen chains [in particular, inordinate hydroxylation of Lys residues N-terminal to the Gly substitution and excessive glycosylation of the resultant hydroxylysine residues (Figure 2)], thereby perturbing triple-helical structure

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and contributing to the severity of OI (39). Thus, the severity of OI correlates with the abundance of triple-helix nucleating, proline-rich sequences immediately N-terminal to the substitution site (36).

Prolines in the Xaa and Yaa Positions

In the strands of human collagen, ~22% of all residues are either Pro or Hyp (9). The abundance of these residues preorganizes the individual strands in a PPII conformation, thereby decreasing the entropic cost for collagen folding (40). Despite their stabilizing properties, Pro derivatives also have certain deleterious consequences for triple-helix folding and stability that partially offset their favorable effects. For example, Pro has a secondary amino group and forms tertiary amides within a peptide or protein. Tertiary amides have a significant population of both the *trans* and the *cis* isomers (Figure 5), whereas all peptide bonds in collagen are *trans*. Thus, before a $(\text{ProHypGly})_n$ strand can fold into a triple helix, all the *cis* peptide bonds must isomerize to *trans*. *N*-Methylalanine (an acyclic, tertiary amide missing only C^{γ} of Pro) decreases triple-helix stability when used to replace Pro or Hyp in CRPs, presumably because it lacks the preorganization imposed by the pyrrolidine ring of Pro derivatives (41). In contrast, avoiding the issue of *cis-trans* isomerization altogether by replacing a Gly–Pro amide bond with a *trans*-locked alkene isostere also results in a destabilized triple helix, despite leaving all interchain hydrogen bonds intact (42). Clearly, the factors dictating triple-helix structure and stability are intertwined in a complex manner (*vide infra*).

Pro residues in the Yaa position of protocollagen triplets are modified by prolyl 4-hydroxylase (P4H), a nonheme iron enzyme that catalyzes the posttranslational and stereoselective hydroxylation of the unactivated γ-carbon of Pro residues in the Yaa position of collagen sequences to form Hyp (Figure 6). P4H activity is required for the viability of both the nematode *Caenorhabditis elegans* and the mouse *Mus musculus* (43,44). Thus, Hyp is essential for the formation of sound collagen in vivo.

Role of Hyp

The hydroxylation of Pro residues in the Yaa position of collagen increases dramatically the thermal stability of triple helices (Table 2). This stabilization occurs when the resultant Hyp is in the Yaa position (45,46) but not in the Xaa position, nor when the hydroxyl group is installed in the 4*S* configuration as in (2*S*,4*S*)-4-hydroxyproline (hyp) (Table 2) (47,48). These findings led to the proposal that the 4*R* configuration of a prolyl hydroxyl group is privileged in alone enabling the formation of water-mediated hydrogen bonds that stitch together the folded triple helix (49). Indeed, such water bridges between Hyp and main chain heteroatoms were observed by Berman and coworkers (19,50) in their seminal X-ray crystallographic studies of CRPs. The frequency of Hyp in most natural collagen is, however, too low to support an extensive network of water bridges. For example, four or more repeating triads of Xaa–Hyp–Gly occur only twice in the amino acid sequence of human type I collagen.

The hypothesis that the water bridges observed in crystalline $(\text{ProHypGly})_n$ triple helices are meaningful was tested by replacing Hyp residues in CRPs with (2*S*,4*R*)-4-fluoroproline (Flp). As fluoro groups do not form strong hydrogen bonds (51), water bridges cannot play a major role in stabilizing a $(\text{ProFlpGly})_{10}$ triple helix. Nonetheless, $(\text{ProFlpGly})_{10}$ triple helices are hyperstable (Table 2) (52,53). Accordingly, water bridges cannot be of fundamental importance for triple-helix stability. How, then, does 4*R*-hydroxylation of Yaa-position Pro residues stabilize the triple helix?

A gauche effect—Replacing Hyp in the Yaa position with (2*S*,4*S*)-4-fluoroproline (flp), a diastereomer of Flp, prevents triple-helix formation (Table 2) (54). This discovery that the stereochemistry of electronegative substituents at the 4-position of the Pro ring is important

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

for the formation of stable triple helices suggests that Flp and Hyp in the Yaa position stabilize collagen via a stereoelectronic effect, rather than a simple inductive effect (54). Pro and its derivatives prefer one of two major pyrrolidine ring puckers, which are termed C^{γ}-*exo* and C^{γ}-*endo* (Figure 7). [The ring actually prefers two distinct twist, rather than envelope, conformations (55). As C^{γ} experiences a large out-of-plane displacement in the twisted rings, we refer to pyrrolidine ring puckers simply as C^{γ}-*exo* and C^{γ}-*endo*.] Pro itself has a slight preference for the C^{γ}-*endo* ring pucker (Table 3) (56). A key attribute of a 4*R* fluoro group on Pro (as well as the natural 4*R* hydroxyl group) is its imposition of a C^{γ}-*exo* pucker on the pyrrolidine ring via the gauche effect (Figure 8*a*,*b*) (56–58). The C^{γ}-*exo* ring pucker preorganizes the main chain torsion angles (φ, C'_{i-1}–N_i–C^{α}_i–C'_i; ψ, N_i–C^{α}_i–C'_i–N_{i+1}; and ω, C^{α}_i–C'_i–N_{i+1}–C^{α}_{i+1}) to those in the Yaa position of a triple helix (Table 4). Thus, 4*R*-hydroxylation of Pro residues in the Yaa position of collagen stabilizes the triple helix via a stereoelectronic effect. Flp is more stabilizing than Hyp because fluorine ($\chi_F = 4.0$) is more electronegative than oxygen ($\chi_O = 3.5$), and a fluoro group ($F_F = 0.45$) manifests a greater inductive effect than does a hydroxyl group ($F_{OH} = 0.33$). Thus, a 4*R* fluoro group enforces the C^{γ}-*exo* ring pucker of a Pro derivative more strongly than does a 4*R* hydroxyl group.

To probe further the role of Hyp in collagen stability, a (2*S*,4*R*)-4-methoxyproline residue (Mop) was incorporated into the Yaa position of a $(ProYaaGly)_{10}$ CRP (59). *O*-Methylation is perhaps the simplest possible covalent modification of a Hyp residue and reduces the extent of hydration without altering significantly the electron-withdrawing ability of the 4*R* substituent. Accordingly, Mop and Hyp residues have similar conformations (Table 4). Interestingly, reducing the hydration of $(ProHypGly)_{10}$ by methylation of Hyp residues enhances triple-helix stability significantly (Table 2). Moreover, alkylation with functional groups larger than a methyl group does not necessarily perturb triple-helix stability (60). Notably, (2*S*,4*R*)-4-chloroproline (Clp) residues also stabilize triple helices in the Yaa position (Table 2) (61). Like Flp, Clp has a strong preference for the C^{γ}-*exo* ring pucker, and a $(ProClpGly)_{10}$ triple helix is therefore more stable than a $(ProProGly)_{10}$ triple helix. Thus, a plethora of data indicate that the hydroxyl group of Hyp stabilizes collagen through a stereoelectronic effect. Water bridges provide little (if any) net thermodynamic advantage to natural collagen (59).

Surprisingly, a host-guest CRP of the form AcGly–$(ProHypGly)_3$–ProFlpGly–$(ProHypGly)_4$–$GlyNH_2$ actually forms a less stable triple helix than does AcGly–$(ProHypGly)_8$–$GlyNH_2$ (62). In contrast, a host-guest CRP of the form $(GlyProHyp)_3$–GlyProFlp–GlyValCys–GlyAspLys– GlyAsnPro–GlyTrpPro–GlyAlaPro–$(GlyProHyp)_4$-NH_2 forms a more stable triple helix than one containing Hyp rather than Flp (63). These results suggest that a fluoro group might disrupt the hydration induced by a long string of Hyp residues. Kobayashi and coworkers (64) used differential scanning calorimetry to demonstrate that $(ProHypGly)_{10}$ triple helices are stabilized by enthalpy, whereas $(ProFlpGly)_{10}$ triple helices are stabilized by entropy. These findings are consistent with Hyp decreasing the entropic cost for folding via main chain preorganization but increasing that cost by specific hydration. This interpretation is in accord with the stability of $(ProMopGly)_{10}$ triple helices arising from a nearly equal contribution of enthalpy and entropy (59).

A steric effect—Electronegative substituents on Pro rings are not the only means of enforcing an advantageous ring pucker. Pro ring pucker can also be dictated by steric effects, as in (2*S*,4*S*)-4-methylproline (Mep) (65) and (2*S*,4*R*)-4-mercaptoproline (Mpc) (Figure 7) (66). The 4-methyl substituent of Mep prefers the pseudoequatorial orientation and thus enforces the C^{γ}-*exo* ring pucker of Pro (analogous results are observed for Mpc) (Table 4). Indeed, triple helices formed from $(ProMepGly)_7$ have stability similar to those formed from $(ProHypGly)_7$ (Table 2) (65).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Proline Derivatives in the Xaa Position

The C^{γ}-*exo* ring pucker of Pro residues in the Yaa position enhances triple-helix stability. Likewise, the ring pucker of Pro in the Xaa position is important for triple-helix stability. Typically, Pro residues in the Xaa position of biological collagen are not hydroxylated and usually display the C^{γ}-*endo* ring pucker (67). By employing C^{γ}-substituents, both the gauche effect and steric effects can be availed to preorganize the C^{γ}-*endo* ring pucker (Figure 7 and Figure 8). Installation of flp, (2*S*,4*S*)-4-chloroproline (clp), or (2*S*,4*R*)-4-methylproline (mep) residues (all of which prefer the C^{γ}-*endo* ring pucker) (Table 3) in the Xaa position of collagen is stabilizing relative to Pro, but installation of Flp, Clp, or Hyp (which prefer the C^{γ}-*exo* ring pucker) is destabilizing (Table 2) (61,65,68–70). These results suggest that preorganizing the C^{γ}-*endo* ring pucker in the Xaa position of CRPs stabilizes triple helices. This conclusion is reasonable because Pro derivatives with a C^{γ}-*endo* ring-pucker have φ and ψ main chain torsion angles similar to those observed in the Xaa position of triple helices (Table 3).

Notably, replacing Pro in the Xaa position of $(ProProGly)_{10}$ with hyp, a Pro derivative that, like flp and clp, should prefer the C^{γ}-*endo* ring pucker owing to the gauche effect, yields CRPs that do not form triple helices (Table 2) (47). This anomalous result for hyp in the Xaa position could be attributable to deleterious hydration, idiosyncratic conformational preferences of hyp residues, or both (71).

Type IV collagen, which is the primary component of basement membranes, has a high incidence of (2*S*,3*S*)-3-hydroxyproline (3*S*-Hyp) in the Xaa position (72). This modification is present in some other collagen types and in invertebrate collagens. 3*S*-Hyp, which prefers a C^{γ}-*endo* ring pucker (73), is introduced almost exclusively within ProHypGly triplets via posttranslational modification of individual collagen strands by prolyl 3-hydroxylase (P3H), which is distinct from P4H (74). A recessive form of OI is associated with a P3H deficiency (75,76). Certain mutations to the gene encoding cartilage-associated protein, a P3H-helper protein, prevent 3*S*-hydroxylation of α1(I)Pro986 as well as 3*S*-hydroxylation of some other Xaa-position Pro residues, resulting in a phenotype nearly identical to classical OI. The underlying basis for the importance of 3*S*-hydroxylation of α1(I)Pro986 is unclear but could involve lower rates of triple-helix secretion (76). Replacing Pro with 3*S*-Hyp in the Xaa position of CRPs can enhance triple-helix stability slightly (73,77). A crystal structure of a triple helix containing 3*S*-Hyp substitutions reveals the maintenance of the prototypical triple-helix structure and the absence of unfavorable steric interactions (Figure 4*c*) (78). In contrast, replacing 3*S*-Hyp with (2*S*,3*S*)-3-fluoroproline destabilizes a triple helix markedly, possibly owing to a through-bond inductive effect that diminishes the ability of its main chain oxygen to accept a hydrogen bond (Figure 4*d*) (79).

An *n*→π* Interaction

A general principle in the design of CRPs is that Pro residues with either a C^{γ}-*endo* or C^{γ}-*exo* ring pucker will stabilize triple helices in the Xaa and Yaa positions, respectively (Table 2–Table 4). Appropriate ring pucker, enforced by a stereoelectronic or steric effect, preorganizes the φ and ψ torsion angles to those required for triple-helix formation.

Intriguingly, the stability of a $(flpProGly)_7$ or $(clpProGly)_{10}$ triple helix is significantly less than that of a $(ProFlpGly)_7$ or $(ProClpGly)_{10}$ triple helix, respectively (Table 2) (61,68). Likewise, a $(mepProGly)_7$ triple helix is less stable than a $(ProMepGly)_7$ triple helix (Table 2) (65). Two factors contribute to the lower stability of triple helices formed from CRPs with stabilizing Pro derivatives substituted in the Xaa position rather than the Yaa position. First, a C^{γ}-*endo* ring pucker is already favored in Pro (56); flp, clp, and mep merely enhance that preference (Table 3). In contrast, Flp, Clp, Hyp, and Mep have the more dramatic effect of reversing the preferred ring pucker of Pro, thereby alleviating the entropic penalty for triple-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

helix formation to a greater extent (Table 4). Second, Flp, Clp, and Mep in the Yaa position cause favorable preorganization of all three main chain torsion angles (φ, ψ, and ω) (Table 4). In contrast, flp, clp, and mep have a low probability of adopting a *trans* peptide bond (ω = 180°) (54,61,65) relative to Pro (Table 3), thereby mitigating the benefit accrued from proper preorganization of φ and ψ. Notably, ^{13}C-NMR studies on collagen in vitro show that 16% of Gly–Pro bonds in unfolded collagen are in the *cis* conformation, whereas only 8% of Xaa–Hyp bonds in unfolded collagen are *cis*, an observation that confirms the effect of C$^{\gamma}$-substitution on the conformation of the preceding peptide bond (80).

How does the effect of a 4-X substituent on Pro ring pucker influence the peptide bond isomerization equilibrium constant ($K_{trans/cis}$) (Figure 5 and Table 3 and Table 4)? The explanation stems from another stereoelectronic effect: an $n\rightarrow\pi^*$ interaction (56,81). In an $n\rightarrow\pi^*$ interaction, the oxygen of a peptide bond (O_{i-1}) donates electron density from its lone pairs into the antibonding orbital of the carbonyl in the subsequent peptide bond ($C_i'=O_i$) (Figure 8*c*,*d*). The C$^{\gamma}$-*exo* ring pucker of a Pro residue provides a more favorable $O_{i-1}\cdots C_i'=O_i$ distance and angle for an $n\rightarrow\pi^*$ interaction than does the C$^{\gamma}$-*endo* pucker (56). Importantly, $K_{trans/cis}$ for peptidyl prolyl amide bonds is determined by the pyrrolidine ring pucker and is not generally affected by the identity of substituents in the 4-position of the pyrrolidine ring (82). Because an $n\rightarrow\pi^*$ interaction can occur only if the peptide bond containing O_{i-1} is *trans*, the $n\rightarrow\pi^*$ interaction has an impact on the value of $K_{trans/cis}$ for main chains with appropriate torsion angles (Table 4). Thus, imposing a C$^{\gamma}$-*exo* pucker on a pyrrolidine ring in the Yaa position of a CRP preorganizes not only the φ and ψ angles for triple-helix formation, but also the ω angle. Indeed, a single $n\rightarrow\pi^*$ interaction can stabilize the *trans* conformation by $\Delta G° = -0.7$ kcal/mol (81,83).

Hyp in the Xaa Position

In the Xaa position, a Pro residue with a C$^{\gamma}$-*endo* pucker generally stabilizes a triple helix, whereas one with a C$^{\gamma}$-*exo* pucker destabilizes a triple helix. For example, $(HypProGly)_n$ triple helices are far less stable than $(ProProGly)_n$ triple helices (Table 2) (84) because Hyp prefers the C$^{\gamma}$-*exo* ring pucker and thus preorganizes the φ and ψ torsion angles improperly for the Xaa position of a collagen triple helix (Table 4). Surprisingly, $(HypHypGly)_{10}$ triple helices are actually slightly more stable than $(ProHypGly)_{10}$ triple helices (Table 2) (85,86) despite the Hyp residues in the Xaa position of $(HypHypGly)_{10}$ displaying the C$^{\gamma}$-*exo* ring pucker in the triple helix (87,88). It is noteworthy that crystal structures of $(HypHypGly)_{10}$ show that the main chain torsion angles in the Xaa position of a $(HypHypGly)_n$ triple helix adjust to accommodate a C$^{\gamma}$-*exo* ring pucker in that position (87,88).

The finding that Hyp can stabilize triple helices in the Xaa position in a context-dependent manner was presaged in a study by Gruskin and coworkers (89) on the global substitution of Hyp for Pro in recombinant type I collagen polypeptides that formed stable triple helices. Notably, Hyp is found in the Xaa position of some invertebrate collagens (90) and can be acceptable in CRPs in which the Yaa position residue is not Pro (86,91,92). Berisio and coworkers (93) have suggested that $(HypHypGly)_{10}$ triple helices might be hyperstable owing to interstrand dipole-dipole interactions between proximal C$^{\gamma}$–OH bonds of adjacent Hyp residues. Kobayashi and coworkers (87) have proposed that the stability of $(HypHypGly)_{10}$ triple helices is attributable to the high hydration level of the peptide chains in the single-coil state prior to triple-helix formation, which could reduce the entropic cost of water bridge formation. A combination of these factors is likely to be responsible for this anomaly.

Heterotrimeric Synthetic Triple Helices

Both flp and Flp greatly enhance triple-helix stability when in the Xaa and Yaa position, respectively. Nonetheless, $(flpFlpGly)_n$ forms much less stable triple helices than does

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(ProProGly)$_n$ (Table 2) (79,94). In such a helix, the fluorine atoms of flp and Flp residues in alternating strands would be proximal, and the C–F dipoles would interact unfavorably (Figure 9*a*) (79). These negative steric and electronic interactions presumably compromise triple-helix stability despite appropriate preorganization of main chain torsion angles. This hypothesis was confirmed by two other findings. First, a (clpClpGly)$_{10}$ triple helix does not even form at 4° C, whereas a (flpFlpGly)$_{10}$ triple helix has T_m = 30°C (Table 2) (61,94). The steric clash between chlorine atoms of opposing clp and Clp residues is exacerbated by the large size of chlorine relative to fluorine (Figure 9*b*). Second, (mepMepGly)$_7$ forms more stable triple helices than do either of the corresponding mono-substituted CRPs, (mepProGly)$_7$ and (ProMepGly)$_7$ (Table 2). The 4-methyl groups protrude radially from the triple helix (Figure 9*c*) and thus cannot interact detrimentally with each other (65).

The steric and stereoelectronic effects on triple-helix stability manifested in the (flpFlpGly)$_7$ CRP provided, for the first time, a means to generate noncovalently linked, heterotrimeric triple helices with defined stoichiometry. Analysis of triple-helix cross sections suggested a triple helix composed of (flpFlpGly)$_7$:(ProProGly)$_7$ in either a 1:2 or 2:1 ratio could be stable, as the presence of some Pro residues in the Xaa and Yaa positions would eliminate deleterious steric interactions between fluorine residues in opposing strands. A (flpFlpGly)$_7$:(ProProGly)$_7$ ratio of 2:1 yielded the most stable triple helices, thereby demonstrating the first instance of heterotrimeric assembly of triple helices with controlled stoichiometry (79) and suggesting the possibility of developing a “code” for triple-helix assembly along the lines of the Watson-Crick code for DNA assembly.

Gauba & Hartgerink (95) developed an alternative strategy that employs Coulombic interactions to guide the assembly of heterotrimeric triple helices. They observed that a 1:1:1 mixture of (ProArgGly)$_{10}$:(GluHypGly)$_{10}$:(ProHypGly)$_{10}$ produces triple helices containing one negatively charged, one positively charged, and one neutral CRP. Intriguingly, a (ProLysGly)$_{10}$:(AspHypGly)$_{10}$:(ProHypGly)$_{10}$ triple helix has a T_m value similar to that of a (ProHypGly)$_{10}$ homotrimer, even though Asp and Lys are known to destabilize significantly the triple helix relative to Pro and Hyp (Figure 9*d*). This finding demonstrates the utility of Coulombic interactions for stabilizing triple helices (96).

Synthetic collagen heterotrimers are appealing mimics of natural collagen strands, as most collagen types are themselves heterotrimers (Table 1). Gauba & Hartgerink (97) employed their Coulombic approach to generate mimics of type I collagen variants that lead to OI. Specifically, they studied the stability of triple-helical heterotrimers containing one, two, or three Gly→Ser substitutions. They observed that a Gly→Ser substitution in only one or two chains is not as debilitating for triple-helix stability and folding as is a Gly→Ser substitution in all three chains.

Nonproline Substitutions in the Xaa and Yaa Positions

Brodsky and coworkers (9) determined the frequency of occurrence of all possible tripeptides in a set of fibrillar and nonfibrillar collagen sequences. Only a few of the 400 possible triplets formed from the 20 natural amino acids are observed with any frequency in collagen. Additionally, they have examined exhaustively the incorporation of all 20 common amino acids in the Xaa and Yaa positions of CRPs using a host-guest model system wherein a single XaaYaaGly triplet is placed within a (ProProGly)$_n$ or (ProHypGly)$_n$ CRP (98). These host-guest studies revealed a correlation between the propensity of a particular residue to adopt a PPII conformation and its contribution to triple-helix stability (98). Notably, Arg in the Yaa position confers triple-helix stability similar to Hyp (99). The aromatic amino acid residues Trp, Phe, and Tyr are all strongly destabilizing to the triple helix (98), although the structural basis for this destabilization is unclear. Brodsky and coworkers (100) used their data on host-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

guest CRPs to develop an algorithm that enables a priori calculation of the effect of Xaa and Yaa substitutions on triple-helix stability.

HIGHER-ORDER COLLAGEN STRUCTURE

In vivo collagen has a hierarchical structure (Figure 2). Individual TC monomers self-assemble into the macromolecular fibers that are essential components of tissues and bones. The self-assembly processes involved in collagen fibrillogenesis are of enormous importance to ECM pathology and proper animal development (see sidebar for a discussion of how collagen self-assembly might be directed away from deleterious protein aggregates).

Fibril Structure

There are many classes of collagenous structures in the ECM, including fibrils, networks, and transmembrane collagenous domains. For brevity, we focus here on fibrils composed primarily of type I collagen.

TC monomers of type I collagen have the unique property of actually being unstable at body temperature (101); that is, the random coil conformation is the preferred one. How can stable tissue structures form from an unstable protein? The answer must be that collagen fibrillogenesis has a stabilizing effect on triple helices. Moreover, the assembly of strong macromolecular structures is essential to enable collagen to support stress in one, two, and three dimensions (102). The importance of collagen fibrillogenesis is underscored by the conclusion of Kadler and coworkers (103) that the fundamental principles underlying the formation of some types of modern collagen fibrils were established at least 500 Mya.

Collagen fibrillogenesis in situ occurs via assembly of intermediate-sized fibril segments, called microfibrils (Figure 2) (104). Thus, there are two important issues for understanding the molecular structure of the collagen fibril. First, what is the arrangement of individual TC monomers within the microfibril? Second, what is the arrangement of the individual microfibrils within the collagen fibril? These questions have proven difficult to answer, as individual natural microfibrils are not isolable and the large size and insolubility of mature collagen fibrils prevent the use of standard structure-determination techniques.

Collagen fibrils formed mainly from type I collagen (all fibrous tissues except cartilage) and fibrils formed largely from type II collagen (cartilage) have slightly different structures. Although we focus solely on type I collagen fibrils, recent data have enabled the determination of thin cartilage fibril structure to intermediate resolution (~4 nm). This structure suggests that cartilage collagen fibrils have a 10 + 4 heterotypic microfibril structure—meaning that the fibril surface presents ten equally spaced microfibrils and that there are four equally spaced microfibrils in the core of the fibril (105).

Fibrils of type I collagen in tendon are up to 1 cm in length (106) and up to ~500 nm in diameter. An individual triple helix in type I collagen is <2 nm in diameter and ~300 nm long. Clearly, fibrillogenesis on an extraordinary scale is necessary to achieve the structural dimensions of natural collagen fibrils. The most characteristic feature of collagen fibrils is that they are *D*-periodic with $D = 67$ nm. The banded structure observed in transmission electron microscopy (TEM) images of collagen fibrils occurs because the actual length of a TC monomer is not an exact multiple of D but $L = 4.46D$, resulting in gaps of $0.54D$ and overlaps of $0.46D$ (Figure 2). This regular array of gap and overlap regions must be accounted for in structural models of the collagen fibril and microfibril.

The initial proposal for the three-dimensional structure of fibrillar collagen was a simplified structural model for collagen microfibrils advanced by Hodge & Petruska (107) in 1963. Their

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

model consists of a two-dimensional stack in which five TC monomers within a microfibril are offset by D = 67 nm between neighboring strands (Figure 2). This model accounts for the gap and overlap regions apparent in mature collagen fibrils by TEM and atomic force microscopy (AFM). Many research groups began efforts to determine the three-dimensional structure of type I collagen fibrils at higher resolution. Numerous models were proposed to account for the features of fiber diffraction and of TEM and AFM images of such fibrils (108–111). Researchers generally agreed on a quasi-hexagonal unit cell containing five TC monomers as the basis for an accurate model of the collagen fibril, but important details were in dispute. Recent findings indicate that the fibril structure controversy is approaching resolution.

In 2001, Orgel and coworkers (112,113) reported the first electron-density map of a type I collagen fiber at molecular anisotropic resolution (axial: 5.16 Å; lateral: 11.1 Å) using synchrotron radiation. Their data confirm that collagen microfibrils have a quasi-hexagonal unit cell. The molecular packing of the TC monomers in this model results in TC neighbors arranged to form supertwisted, right-handed microfibrils that interdigitate with neighboring microfibrils–-leading to a spiral-like structure for the mature collagen fibril (113). Their model advances the provocative idea that the collagen fibril is a networked, nanoscale rope–-an idea also suggested by the AFM studies of Bozec and coworkers (111).

Orgel and coworkers determined the axial location of the N- and C-terminal collagen telopeptides and found that neighboring telopeptides within a TC monomer interact with each other and are cross-linked covalently subsequent to the action of lysyl oxidase (114). The cross-links can be both within and between microfibers. Intriguingly, the supertwisted nature of the collagen microfibril is maintained through the nonhelical telopeptide regions (113).

This new model of the fibril of type I collagen explains the failure of previous researchers to isolate individual collagen microfibrils from tissue samples: The microfibrils interdigitate and cross-link, thus preventing separation from each other in an intact form. The new model also justifies the observation that TC in fibrils is far more resistant to collagen proteolysis by matrix metalloproteinase 1 (MMP1) than is monomeric TC; the collagen fibril protects regions vulnerable to proteolysis by MMP1. Proteolysis of the C-terminal telopeptide of TC in a fibril is required before MMP1 can gain access to the cleavage site of a TC monomer (115).

Nucleation and Modulation of Collagen Fibrillogenesis

Collagen fibrillogenesis requires completion of two stages of self-assembly: nucleation and fiber growth. Collagen fibrillogenesis begins only after procollagen N- and C-proteinases cleave the collagen propeptides at each triple-helix terminus to generate TC monomers. The C-terminal propeptides are essential for proper triple-helix formation but prevent fibrillogenesis (116). After cleavage of the propeptides, TC monomers are composed of a lengthy triple-helical domain consisting of a repeating XaaYaaGly sequence flanked by short, nontriple-helical telopeptides (Figure 2).

The C-terminal telopeptides of TC are important for initiating proper fibrillogenesis. Prockop and Fertala (117) suggested that collagen self-assembly into fibrils is driven by the interaction of C-terminal telopeptides with specific binding sites on triple-helical monomers. The addition of synthetic telopeptide mimics can inhibit collagen fibrillogenesis, presumably by preventing the interaction between collagen telopeptides and TC monomers. Triple helices lacking the telopeptides can, however, assemble into fibrils with proper morphology (118). Thus, collagen telopeptides could accelerate fibril assembly and establish the proper register within microfibrils and fibrils but might not be essential for fibrillogenesis.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Collagen telopeptides have a second role in stabilizing mature collagen fibrils. Lys side chains in the telopeptides are cross-linked subsequent to fibril assembly, forming hydroxylysyl pyridinoline and lysyl pyridoline cross-links between Lys and hydroxylysine residues with the aid of lysyl oxidase (Figure 2) (119). The cross-linking process endows mature collagen fibrils with strength and stability, but is not involved in fibrillogenesis. Thus, although collagen telopeptides might not be essential for nucleating collagen fibrillogenesis, their absence greatly weakens the mature fibril owing to the lack of cross-links within and between triple helices (119).

MECHANICAL PROPERTIES OF COLLAGEN FIBRILS

The hierarchical nature of collagen structure theoretically enables evaluation of the mechanical properties of collagen at varying levels of structural complexity, including the TC monomer, individual collagen fibrils, and collagen fibers. Perhaps the most direct measures of the mechanical properties of collagen have been obtained by studying TC monomers and fibrils formed from type I collagen. Researchers have employed various biophysical and theoretical techniques over the past 20 years, and recent advances in AFM methodology have enabled more refined evaluations.

In 2006, Buehler estimated the fracture strength of a TC monomer to be 11 GPa, which is significantly greater than that of a collagen fibril (0.5 GPa) (102). This difference is reasonable, given that fracture of a TC monomer requires unraveling of the triple helix and ultimately breaking of covalent bonds, whereas fracture of a fibril does not necessarily require the disruption of covalent bonds. For comparison, the tensile strength of collagen in tendon is estimated to be 100 MPa (120).

The Young's modulus of a TC monomer is E = 6–7 GPa (102,121), whereas AFM measurements show that dehydrated fibrils of type I collagen from bovine Achilles tendon (122) and rat tail tendon (123) have $E \approx 5$ GPa and $E \leq 11$ GPa, respectively. Because collagen fibrils are anisotropic, the shear modulus (which is a measure of rigidity) is also an important measure of the strength of a collagen fibril. In 2008, AFM revealed that dehydrated fibrils of type I collagen from bovine Achilles tendon have G = 33 MPa (124). Hydration of these fibrils reduced their shear modulus significantly, whereas carbodiimide-mediated cross-linking increased their shear modulus. It is noteworthy that a certain level of cross-linking is favorable for the mechanical properties of collagen fibrils, but excessive cross-linking results in extremely brittle collagen fibrils (102), a common symptom of aging.

An analysis by Buehler (102) of the mechanical properties of collagen fibrils suggests that nature has selected a length for the TC monomer that maximizes the robustness of the assembled collagen fibril via efficient energy dissipation. Simulations indicate that TC monomers either longer or shorter than ~300 nm (which is the length of a type I collagen triple helix) would form collagen fibrils with less favorable mechanical properties.

COLLAGENOUS BIOMATERIALS

Research on the structure and stability of collagen triple helices has focused on blunt-ended triple helices composed of $(\text{XaaYaaGly})_{n\leq 10}$ CRPs. These short triple helices, although valuable for studies directed at understanding the physicochemical basis of triple-helix structure and stability, are not useful for many potential biomaterial applications because of their small size, which does not approach the scale of natural collagen fibers (Figure 2).

Bovine collagen is readily available and useful for some biomedical purposes, but it suffers from heterogeneity, potential immunogenicity, and loss of structural integrity during the isolation process. An efficient recombinant or synthetic source of collagen could avoid these

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

complications. The heterologous production of collagen is made problematic by the difficulty of incorporating posttranslational modifications, such as that leading to the essential Hyp residues (Figure 6), and by the need to use complex expression systems (125). These challenges underscore the need for synthetic sources of collagen-like proteins and fibrils.

Collagen via Chemical Synthesis

Early approaches to long synthetic collagen triple helices relied on the condensation (126, 127) or native chemical ligation of short CRPs (127). Interestingly, concentrated aqueous solutions of $(ProHypGly)_{10}$ self-assemble into highly branched fibrils (128). Brodsky and coworkers (129) have shown that the rate of $(ProHypGly)_{10}$ self-assembly and the morphology of the resultant fibrils are sequence dependent. CRPs containing a single Pro→Ala or Pro→Leu substitution display slower self-assembly; fibril morphology can be modified by a Gly→Ser substitution, or prevented by a single Gly→Ala substitution or global Hyp→Pro substitutions. Regardless, the higher-order structures formed by the self-assembly of $(ProHypGly)_{10}$ and related CRPs do not resemble natural collagen fibrils.

Long collagen triple helices have been prepared by using a design that takes advantage of the intrinsic propensity of individual CRP strands to form triple helices. Specifically, a cystine knot within short collagen fragments was utilized to set the register of individual collagen strands such that short, “sticky” ends preorganized for further triple-helix formation were displayed at the end of each triple-helical, monomeric segment (Figure 10*a*) (130, 131). Self-assembly of these short, triple-helix fragments was then mediated by association of the sticky ends, resulting in collagen assemblies as long as 400 nm–-significantly longer than natural TC monomers (131). Koide and coworkers (132) used this system to prepare tunable collagen-like gels with potential biomaterial applications.

Maryanoff and coworkers (133) developed another approach to long triple helices, one that relied on the predilection of electron-rich phenyl rings of C-terminal phenylalanine residues installed in a short CRP to engage in π-stacking interactions preferentially with electron-poor pentafluorophenyl rings of N-terminal pentafluorophenylalanine residues (Figure 10*b*). Their strategy produced micrometer-scale triple-helical fibers. This π-stacking approach has been used to generate thrombogenic collagen-like fibrils for applications in biomedicine (134). In addition, attachment of gold nanoparticles to these fibrils and subsequent electroless silver plating yielded collagen-based nanowires that conduct electricity (135).

Przybyla & Chmielewski (136) used metal-triggered self-assembly to obtain collagen fibrils from a CRP. A single Hyp residue in Ac-$(ProHypGly)_9$-NH_2 was replaced with a bipyridyl-modified Lys residue. Addition of Fe(II) to a solution of this CRP triggered self-assembly into morphologically diverse fibrils of up to 5 μm in length with a mean radius of 0.5 μm.

A major advance in the development of synthetic CRP assemblies with improved similarity to collagen fibrils was reported by Chaikof and coworkers (137). They synthesized a CRP with the sequence $(ProArgGly)_4$–$(ProHypGly)_4$–$(GluHypGly)_4$ and observed self-assembly in solution into fibrils 3–4 μm in length and 12–15 nm in diameter. Upon heating the peptide solution to 75°C for 40 min and then cooling to room temperature, they observed thicker fibrils (~70 nm in diameter). Importantly, these fibrils exhibited two key characteristics of natural collagen fibrils. First, the fibrils displayed tapered tips at their termini–-a feature observed in type I collagen fibers and thought to be important for fiber growth (138). Second, Chaikof and coworkers observed *D*-periodic structure in synthetic collagen fibrils, with $D \approx 18$ nm. The self-assembly process presumably relies on Coulombic interactions and hydrogen bonds between charged Arg and Glu residues in individual, axially staggered triple helices (Figure 10*c*).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The methodologies described above enable the creation of long, triple-helical, collagen-like fibrils. Despite major advances, synthetic collagen-mimetic fibrils still lack many of the characteristics of higher-order collagen structures. In addition, the mechanical properties of synthetic collagenous materials have not been studied to date. Synthetic collagens that closely mimic the length, girth, patterns, mechanical properties, and complexity of natural collagen fibrils remain to be developed, but rapid progress in the past few years engenders great optimism.

Biological and Biomedical Applications of Synthetic Collagen

Relatively few CRPs have been tested as biomaterials. Goodman and coworkers (139) showed that peptoid-containing CRPs have a notable ability to bind to epithelial cells and fibroblasts, particularly when displayed on a surface. CRPs are also useful for inducing platelet aggregation, which can aid the wound-healing process (140,141).

A key step toward utilizing collagenous biomaterials for therapeutic purposes is the development of CRPs that can either adhere to or bury themselves within biological collagen. Most efforts toward these objectives have relied on immobilization of CRPs on an unrelated substance. Yu and coworkers (142) prepared CRP-functionalized gold nanoparticles and demonstrated binding of the gold nanoparticles to the gap region of natural collagen. Maryanoff and coworkers found that CRPs displayed on latex nanoparticles can stimulate human platelet aggregation with a potency similar to that of type I collagen (140). In an important extension of this work, they demonstrated that triple-helical fibrils obtained via aromatic interactions had a similar level of thrombogenic activity to the CRPs immobilized on latex nanoparticles (134). Finally, single strands of CRPs and polyethylene glycol-conjugated CRPs bind to collagen films even without immobilization on nanoparticles (143) and are of potential use in collagen imaging (144) and wound-healing applications. The future of these approaches appears to be especially bright.

SUMMARY POINTS

1. High-resolution crystal structures and modern biophysical approaches have enabled detailed study of the structure and stability of collagen triple helices. The ladder of hydrogen bonds observed in these crystal structures is essential for holding the triple helix together, and its absence in natural collagen leads to a variety of pathological conditions.
2. Stereoelectronic effects impart significant structural stability to collagen by preorganizing individual polypeptides for triple-helix formation. For example, Hyp in the Yaa position stabilizes the triple helix via a stereoelectronic effect. Stereoelectronic effects are also important for the structure and stability of numerous other peptides and proteins.
3. Posttranslational modifications to protocollagen are of fundamental importance to the synthesis of a stable ECM. These modifications include hydroxylation and cross-linking reactions.
4. Collagen fibrillogenesis is an essential process for the formation of macromolecular biological scaffolds. Relatively high-resolution models of type I and type II collagen fibrils are now available and, for type I collagen, show that collagen fibrils can be described as nanoscale ropes.
5. Simple means to synthesize long collagen triple helices and fibrils have become apparent. The resultant materials are poised for use in biomedicine and nanotechnology.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

FUTURE ISSUES

1. Many important conclusions regarding the mechanisms underlying collagen triple-helix structure and stability derive from studies of triple-helical CRPs. It is important to evaluate the extent to which CRP triple-helix stability correlates with collagen fibril stability and mechanical properties.
2. The factors that affect triple-helix stability for Pro derivatives in the Yaa-position are now clear. In contrast, the Xaa position is poorly understood. In particular, the underlying reasons for the anomalous effects of hyp and Hyp on triple-helix stability in the Xaa position remain to be resolved.
3. Biomaterial and biomedical applications of collagen require improved methods to synthesize long collagen triple helices and mimic complex, hierarchichal collagen structures. Continued efforts in this field will be guided by recent findings regarding the structure of collagen fibrils and fibers.
4. Elucidation of the physicochemical basis for collagen fibril strength and stability as well as the functionalities in natural collagen that are important for fibril formation is an important research area that will aid understanding of ECM-related diseases and biomaterials efforts.
5. The domains of major types of collagen that engage in interactions with other proteins and biomolecules have been described. In combination with recent developments in understanding the three-dimensional structure of collagen fibrils, these results should enable detailed study and rational manipulation of these interactions.
6. The molecular structures and mechanisms of formation of the various types of nonfibrillar collagen assemblies are poorly understood. Further research must be undertaken to develop a thorough understanding of their structures and functions.
7. Effective use of collagen biomaterials effectively in materials science and medicine requires development of artificial collagens displaying the properties of natural collagen and identification of the areas where collagen biomaterials can be applied for the greatest benefit.

SIDEBAR: AVOIDING AGGREGATION

Long, unfolded polypeptides have an innate tendency to form aggregates (145), such as the amyloid fibrils implicated in neurodegenerative diseases. Interestingly, despite their long length and slow folding, protocollagen strands are not known to aggregate. Amyloid fibrils and other aggregates are composed largely of β-sheets (146). Pro and Gly are the two amino acid residues with the lowest propensity to form a β-sheet (147,148), and Gly residues are known explicitly to reduce protein aggregation rates (149).

We propose that the prevalence of Pro and Gly residues in protocollagen is necessary to avert the formation of harmful aggregates. This proposal is supported by the remarkably high Pro/Gly content of other fibrous, structural proteins in plants and animals, such as elastin, extensin, glycine-rich proteins, and proline-rich proteins. Molecular dynamics simulations of elastin polypeptides likewise support this proposal, as a minimum threshold of Pro/Gly content must be attained to realize elastomers instead of amyloid fibrils (150). Apparently, the molecular evolution of collagen and other fibrous, structural proteins has

availed Pro and Gly residues to avoid β-sheet formation and the consequent formation of harmful aggregates.

Acknowledgments

The authors acknowledge Dr. Jeet Kalia for critical reading of the manuscript and Amit Choudhary for creating Figure 8*d*. M.D.S. was supported by graduate fellowships from the Department of Homeland Security and the Division of Medicinal Chemistry, American Chemical Society. Collagen research in our laboratory is supported by Grant AR044276 (N.I.H.).

LITERATURE CITED

1. Brinckmann J. Collagens at a glance. Top. Curr. Chem 2005;247:1–-6.
2. Veit G, Kobbe B, Keene DR, Paulsson M, Koch M, Wagener R. Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. J. Biol. Chem 2006;281:3494–3504. [PubMed: 16330543]
3. Schweitzer MH, Suo Z, Avci R, Asara JM, Allen MA, et al. Analyses of soft tissue from *Tyrannosaurus rex* suggest the presence of protein. Science 2007;316:277–280. [PubMed: 17431179]
4. Asara JM, Schweitzer MH, Freimark LM, Phillips M, Cantley LC. Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. Science 2007;316:280–285. [PubMed: 17431180]
5. Buckley M, Walker A, Ho SYW, Yang Y, Smith C, et al. Comment on "Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry.". Science 2008;319:333. [PubMed: 18202290]
6. Pevzner PA, Kim S, Ng J. Comment on "Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry.". Science 2008;321:1040. [PubMed: 18719266]
7. Berisio R, Vitagliano L, Mazzarella L, Zagari A. Crystal structure of the collagen triple helix model $[(\text{Pro-Pro-Gly})_{10}]_3$. Protein Sci 2002;11:262–270. [PubMed: 11790836]
8. Brazel D, Oberbäumer I, Dieringer H, Babel W, Glanville RW, et al. Completion of the amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 nontriplet interruptions located within the collagenous domain. Eur. J. Biochem 1987;168:529–536. [PubMed: 3311751]
9. Ramshaw JAM, Shah NK, Brodsky B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. J. Struct. Biol 1998;122:86–91. [PubMed: 9724608]
10. Fitzgerald J, Rich C, Zhou FH, Hansen U. Three novel collagen VI chains, α4(VI), α5(VI), and α6 (VI). J. Biol. Chem 2008;283:20170–20180. [PubMed: 18400749]
11. Astbury WT, Bell FO. The molecular structure of the fibers of the collagen group. Nature 1940;145:421–422.
12. Pauling L, Corey RB. The structure of fibrous proteins of the collagen-gelatin group. Proc. Natl. Acad. Sci. USA 1951;37:272–281. [PubMed: 14834150]
13. Ramachandran GN, Kartha G. Structure of collagen. Nature 1954;174:269–270. [PubMed: 13185286]
14. Ramachandran GN, Kartha G. Structure of collagen. Nature 1955;176:593–595. [PubMed: 13265783]
15. Rich A, Crick FHC. The structure of collagen. Nature 1955;176:915–916. [PubMed: 13272717]
17. Cowan PM, McGavin S, North ACT. The polypeptide chain configuration of collagen. Nature 1955;176:1062–1064. [PubMed: 13272747]
16. Rich A, Crick FHC. The molecular structure of collagen. J. Mol. Biol 1961;3:483–-506. [PubMed: 14491907]
18. Fields GB, Prockop DJ. Perspectives on the synthesis and application of triple-helical, collagen-model peptides. Biopolymers 1996;40:345–357. [PubMed: 8765606]
19. Bella J, Eaton M, Brodsky B, Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science 1994;266:75–81. [PubMed: 7695699] NOTE: First high-resolution (1.9-Å) crystal structure of a collagen triple helix, formed from CRPs.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

20. Bella J, Berman HM. Crystallographic evidence for C^{α}-H–O=C hydrogen bonds in a collagen triple helix. J. Mol. Biol 1996;264:734–742. [PubMed: 8980682]
21. Kramer RZ, Venugopal MG, Bella J, Mayville P, Brodsky B, Berman HM. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. J. Mol. Biol 2000;301:1191–1205. [PubMed: 10966815]
22. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. Structural basis of collagen recognition by integrin $\alpha_2\beta_1$. Cell 2000;101:47–-56. [PubMed: 10778855]
23. Cohen C, Bear RS. Helical polypeptide chain configuration in collagen. J. Am. Chem. Soc 1953;75:2783–2784.
24. Okuyama K, Xu X, Iguchi M, Noguchi K. Revision of collagen molecular structure. Biopolymers 2006;84:181–191. [PubMed: 16206128]
25. Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM. Sequence dependent conformational variations of collagen triple-helical structure. Nat. Struct. Biol 1999;6:454–457. [PubMed: 10331873]
26. Boudko S, Engel J, Okuyama K, Mizuno K, Bächinger HP, Schumacher MA. Crystal structure of human type III collagen G991–G1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. J. Biol. Chem 2008;283:32580–32589. [PubMed: 18805790]
27. Sweeney SM, Guy CA, Fields GB, San Antonio JD. Defining the domains of type I collagen involved in heparin-binding and endothelial tube formation. Proc. Natl. Acad. Sci. USA 1998;95:7275–7280. [PubMed: 9636139]
28. Di Lullo GA, Sweeney SM, Körkkö J, Ala-Kokko L, San Antonio JD. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. J. Biol. Chem 2002;277:4223–4231. [PubMed: 11704682]
29. Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, et al. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J. Biol. Chem 2008;283:21187–21197. [PubMed: 18487200] NOTE: Thorough analysis of the cell interaction and matrix interaction domains of the collagen fibrils.
30. Jenkins CL, Vasbinder MM, Miller SJ, Raines RT. Peptide bond isosteres: ester or (*E*)-alkene in the backbone of the collagen triple helix. Org. Lett 2005;7:2619–2622. [PubMed: 15957905]
31. Boryskina OP, Bolbukh TV, Semenov MA, Gasan AI, Maleev VY. Energies of peptide-peptide and peptide-water hydrogen bonds in collagen: evidences from infrared spectroscopy, quartz piezogravimetry, and differential scanning calorimetry. J. Mol. Struct 2007;827:1–10.
32. Myllyharju J, Kivirikko KI. Collagens and collagen-related diseases. Ann. Med 2001;33:7–21. [PubMed: 11310942]
33. Beck K, Chan VC, Shenoy N, Kirkpatrick A, Ramshaw JAM, Brodsky B. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. Proc. Natl. Acad. Sci. USA 2000;97:4273–4278. [PubMed: 10725403]
34. Tsai MI-H, Xu Y, Dannenberg JJ. Completely geometrically optimized DFT/ONIOM triple-helical collagen-like structures containing the ProProGly, ProProAla, $ProPro^{D}Ala$, and $ProPro^{D}Ser$ triads. J. Am. Chem. Soc 2005;127:14130–14131. [PubMed: 16218576]
35. Horng J-C, Kotch FW, Raines RT. Is glycine a surrogate for a d-amino acid in the collagen triple helix? Protein Sci 2007;16:208–215. [PubMed: 17189476]
36. Bodian DL, Madhan B, Brodsky B, Klein TE. Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. Biochemistry 2008;47:5424–5432. [PubMed: 18412368]
37. Hyde TJ, Bryan MA, Brodsky B, Baum J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. J. Biol. Chem 2006;281:36937–36943. [PubMed: 16998200]
38. Khoshnoodi J, Cartailler J-P, Alvares K, Veis A, Hudson BG. Molecular recognition in the assembly of collagens: Terminal noncollagenous domains are key recognition modules in the formation of triple-helical protomers. J. Biol. Chem 2006;281:38117–38121. [PubMed: 17082192]
39. Raghunath M, Bruckner P, Steinmann B. Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. J. Mol. Biol 1994;236:940–949. [PubMed: 8114103]
40. Cram DJ. The design of molecular hosts, guests, and their complexes. Science 1988;240:760–767. [PubMed: 3283937]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

41. Kersteen EA, Raines RT. Contribution of tertiary amides to the conformational stability of collagen triple helices. Biopolymers 2001;59:24–28. [PubMed: 11343277]
42. Nan D, Wang XJ, Etzkorn FA. The effect of a *trans*-locked Gly–Pro alkene isostere on collagen triple helix stability. J. Am. Chem. Soc 2008;130:5396–5397. [PubMed: 18366169]
43. Friedman L, Higgin JJ, Moulder G, Barstead R, Raines RT, Kimble J. Prolyl 4-hydroxylase is required for viability and morphogenesis in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 2000;97:4736–4741. [PubMed: 10781079] NOTE: Demonstration that 4*R*-hydroxylation of Pro residues in the Yaa position of collagen strands is required for animal life.
44. Holster T, Pakkanen O, Soininen R, Sormunen R, Nokelainen M, et al. Loss of assembly of the main basement membrane collagen, type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice. J. Biol. Chem 2007;282:2512–2519. [PubMed: 17135260]
45. Berg RA, Prockop DJ. The thermal transition of a nonhydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen. Biochem. Biophys. Res. Commun 1973;52:115–120. [PubMed: 4712181]
46. Sakakibara S, Inouye K, Shudo K, Kishida Y, Kobayashi Y, Prockop DJ. Synthesis of (Pro–Hyp–Gly)$_n$ of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. Biochim. Biophys. Acta 1973;303:198–-202. [PubMed: 4702003]
47. Inouye K, Sakakibara S, Prockop DJ. Effects of the stereo-configuration of the hydroxyl group in 4-hydroxyproline on the triple-helical structures formed by homogenous peptides resembling collagen. Biochim. Biophys. Acta 1976;420:133–141. [PubMed: 1247577]
48. Jiravanichanun N, Nishino N, Okuyama K. Conformation of alloHyp in the Y position in the host-guest peptide with the Pro-Pro-Gly sequence: implication of the destabilization of (Pro-alloHyp-Gly)$_{10}$. Biopolymers 2006;81:225–233. [PubMed: 16273514]
49. Suzuki E, Fraser RDB, MacRae TP. Role of hydroxyproline in the stabilization of the collagen molecule via water molecules. Int. J. Biol. Macromol 1980;2:54–56.
50. Bella J, Brodsky B, Berman HM. Hydration structure of a collagen peptide. Structure 1995;3:893–906. [PubMed: 8535783]
51. Dunitz JD, Taylor R. Organic fluorine hardly ever accepts hydrogen bonds. Chem. Eur. J 1997;3:89–98.
52. Holmgren SK, Taylor KM, Bretscher LE, Raines RT. Code for collagen's stability deciphered. Nature 1998;392:666–667. [PubMed: 9565027] NOTE: Overturned the long-standing hypothesis that water bridges are important for the structure and stability of the collagen triple helix.
53. Holmgren SK, Bretscher LE, Taylor KM, Raines RT. A hyperstable collagen mimic. Chem. Biol 1999;6:63–70. [PubMed: 10021421]
54. Bretscher LE, Jenkins CL, Taylor KM, DeRider ML, Raines RT. Conformational stability of collagen relies on a stereoelectronic effect. J. Am. Chem. Soc 2001;123:777–778. [PubMed: 11456609]
55. Gilli, G. Molecules and molecular crystals. In: Giacovazzo, C., editor. Fundamentals of Crystallography. Oxford, UK: Oxford Univ. Press; 2002. p. 618-625.
56. DeRider ML, Wilkens SJ, Waddell MJ, Bretscher LE, Weinhold F, et al. Collagen stability: insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. J. Am. Chem. Soc 2002;124:2497–2505. [PubMed: 11890798]
57. Panasik N Jr, Eberhardt ES, Edison AS, Powell DR, Raines RT. Inductive effects on the structure of proline residues. Int. J. Pept. Protein Res 1994;44:262–269. [PubMed: 7822103]
58. Improta R, Benzi C, Barone V. Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. J. Am. Chem. Soc 2001;123:12568–12577. [PubMed: 11741421]
59. Kotch FW, Guzei IA, Raines RT. Stabilization of the collagen triple helix by *O*-methylation of hydroxyproline residues. J. Am. Chem. Soc 2008;130:2952–2953. [PubMed: 18271593]
60. Lee S-G, Lee JY, Chmielewski J. Investigation of pH-dependent collagen triple-helix formation. Angew. Chem. Int. Ed. Engl 2008;47:8429–8432. [PubMed: 18816568]
61. Shoulders MD, Guzei IA, Raines RT. 4-Chloroprolines: synthesis, conformational analysis, and effect on the collagen triple helix. Biopolymers 2008;89:443–454. [PubMed: 17937398]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

62. Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. Triple-helix propensity of hydroxyproline and fluoroproline: comparison of host-guest and repeating tripeptide models. J. Am. Chem. Soc 2003;125:11500–11501. [PubMed: 13129344]

63. Malkar NB, Lauer-Fields JL, Borgia JA, Fields GB. Modulation of triple-helical stability and subsequent melanoma cellular responses by single-site substitution of fluoroproline derivatives. Biochemistry 2002;41:6054–6064. [PubMed: 11994000]

64. Nishi Y, Uchiyama S, Doi M, Nishiuchi Y, Nakazawa T, et al. Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of the collagen triple helix. Biochemistry 2005;44:6034–6042. [PubMed: 15835892]

65. Shoulders MD, Hodges JA, Raines RT. Reciprocity of steric and stereoelectronic effects in the collagen triple helix. J. Am. Chem. Soc 2006;128:8112–8113. [PubMed: 16787056]

66. Cadamuro SA, Reichold R, Kusebauch U, Musiol H-J, Renner C, et al. Conformational properties of 4-mercaptoproline and related derivatives. Angew. Chem. Int. Ed. Engl 2008;47:2143–2146. [PubMed: 18228549]

67. Vitagliano L, Berisio R, Mazzarella L, Zagari A. Structural bases of collagen stabilization induced by proline hydroxylation. Biopolymers 2001;58:459–464. [PubMed: 11241217]

68. Hodges JA, Raines RT. Stereoelectronic effects on collagen stability: the dichotomy of 4-fluoroproline diastereomers. J. Am. Chem. Soc 2003;125:9262–9263. [PubMed: 12889933]

69. Doi M, Nishi Y, Uchiyama S, Nishiuchi Y, Nakazawa T, et al. Characterization of collagen model peptides containing 4-fluoroproline; (4(*S*)-fluoroproline–Pro–Gly)$_{10}$ forms a triple helix, but (4(*R*)-fluoroproline–Pro–Gly)$_{10}$ does not. J. Am. Chem. Soc 2003;125:9922–9923. [PubMed: 12914445]

70. Barth D, Milbradt AG, Renner C, Moroder L. A (4*R*)- or a (4*S*)-fluoroproline residue in position Xaa of the (Xaa–Yaa–Gly) collagen repeat severely affects triple-helix formation. ChemBioChem 2004;5:79–86. [PubMed: 14695516]

71. Lesarri A, Cocinero EJ, López JC, Alonso JL. Shape of 4*S*- and 4*R*-hydroxyproline in gas phase. J. Am. Chem. Soc 2005;127:2572–2579. [PubMed: 15725012]

72. Kefalides NA. Structure and biosynthesis of basement membranes. Int. Rev. Connect. Tissue Res 1973;6:63–104. [PubMed: 4198817]

73. Jenkins CL, Bretscher LE, Guzei IA, Raines RT. Effect of 3-hydroxyproline residues on collagen stability. J. Am. Chem. Soc 2003;125:6422–6427. [PubMed: 12785781]

74. Tryggvason K, Risteli J, Kivirikko K. Separation of prolyl 3-hydroxylase and 4-hydroxylase activities and the 4-hydroxyproline requirement for synthesis of 3-hydroxyproline. Biochem. Biophys. Res. Commun 1976;76:275–281. [PubMed: 194596]

75. Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, et al. CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. Cell 2006;127:291–304. [PubMed: 17055431]

76. Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, et al. Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat. Genet 2007;39:359–365. [PubMed: 17277775]

77. Mizuno K, Peyton DH, Hayashi T, Engel J, Bächinger HP. Effect of the -Gly-3(*S*)-hydroxyprolyl-4 (*R*)-hydroxyprolyl-tripeptide unit on the stability of collagen model peptides. FEBS J 2008;275:5830–5840. [PubMed: 19021759]

78. Schumacher MA, Mizuno K, Bächinger HP. The crystal structure of a collagen-like polypeptide with 3(*S*)-hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. J. Biol. Chem 2006;281:27566–27574. [PubMed: 16798737]

79. Hodges JA, Raines RT. Stereoelectronic and steric effects in the collagen triple helix: toward a code for strand association. J. Am. Chem. Soc 2005;127:15923–15932. [PubMed: 16277536]

80. Sarkar SK, Young PE, Sullivan CE, Torchia DA. Detection of *cis* and *trans* X-Pro bonds in proteins by ^{13}C NMR: application to collagen. Proc. Natl. Acad. Sci. USA 1984;81:4800–4803. [PubMed: 6589627]

81. Hinderaker MP, Raines RT. An electronic effect on protein structure. Protein Sci 2003;12:1188–1194. [PubMed: 12761389]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

82. Jenkins CL, Lin G, Duo J, Rapolu D, Guzei IA, et al. Substituted 2-azabicyclo[2.1.1]hexanes as constrained proline analogues: implications for collagen stability. J. Org. Chem 2004;69:8565–8573. [PubMed: 15575731]

83. Hodges JA, Raines RT. Energetics of an $n \rightarrow \pi^*$ interaction that impacts protein structure. Org. Lett 2006;8:4695–4697. [PubMed: 17020280]

84. Inouye K, Kobayashi Y, Kyogoku Y, Kishida Y, Sakakibara S, Prockop DJ. Synthesis and physical properties of (hydroxyproline-proline-glycine)$_{10}$. Hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. Arch. Biochem. Biophys 1982;219:198–203. [PubMed: 7181510]

85. Berisio R, Granata V, Vitagliano L, Zagari A. Imino acids and collagen triple helix stability: Characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats. J. Am. Chem. Soc 2004;126:11402–11403. [PubMed: 15366862]

86. Mizuno K, Hayashi T, Peyton DH, Bächinger HP. Hydroxylation-induced stabilization of the collagen triple helix. J. Biol. Chem 2004;279:38072–38078. [PubMed: 15231845]

87. Kawahara K, Nishi Y, Nakamura S, Uchiyama S, Nishiuchi Y, et al. Effect of hydration on the stability of the collagen-like triple-helical structure of [4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolylglycine]$_{10}$. Biochemistry 2005;44:15812–15822. [PubMed: 16313184]

88. Schumacher M, Mizuno K, Bächinger HP. The crystal structure of the collagen-like polypeptide (glycyl-4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl)$_9$ at 1.55 angstrom resolution shows up-puckering of the proline ring in the Xaa position. J. Biol. Chem 2005;280:20397–20403. [PubMed: 15784619]

89. Buechert DD, Paolella DN, Leslie BS, Brown MS, Mehos KA, Gruskin EA. Co-translational incorporation of *trans*-4-hydroxyproline into recombinant proteins in bacteria. J. Biol. Chem 2003;278:645–650. [PubMed: 12399455]

90. Mann K, Mechling DE, Bächinger HP, Eckerskorn C, Gaill F, Timpl R. Glycosylated threonine but not 4-hydroxyproline dominates the triple helix stabilizing positions in the sequence of a hydrothermal vent worm cuticle collagen. J. Mol. Biol 1996;261:255–266. [PubMed: 8757292]

91. Bann JG, Bächinger HP. Glycosylation/hydroxylation-induced stabilization of the collagen triple helix: 4-*trans*-hydroxyproline in the Xaa position can stabilize the triple helix. J. Biol. Chem 2000;275:24466–24469. [PubMed: 10827193]

92. Mizuno K, Hayashi T, Bächinger HP. Hydroxylation-induced stabilization of the collagen triple helix. J. Biol. Chem 2003;278:32373–32379. [PubMed: 12807876]

93. Improta R, Berisio R, Vitagliano L. Contribution of dipole-dipole interactions to the stability of the collagen triple helix. Protein Sci 2008;2008:955–961. [PubMed: 18369197]

94. Doi M, Nishi Y, Uchiyama S, Nishiuchi Y, Nishio H, et al. Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)$_{10}$ analogues: (4(*S*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-Gly)$_{10}$ and (4(*S*)-fluoroprolyl-4(*R*)-fluoroprolyl-Gly)$_{10}$. J. Pept. Sci 2005;11:609–616. [PubMed: 15880478]

95. Gauba V, Hartgerink JD. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. J. Am. Chem. Soc 2007;129:2683–2690. [PubMed: 17295489] NOTE: Formation of a 1:1:1 heterotrimeric triple-helix from a positively charged, a negatively charged, and a neutral CRP.

96. Gauba V, Hartgerink JD. Surprisingly high stability of collagen ABC heterotrimer: evaluation of side chain charge pairs. J. Am. Chem. Soc 2007;129:15034–15041. [PubMed: 17988128]

97. Gauba V, Hartgerink JD. Synthetic collagen heterotrimers: structural mimics of wild-type and mutant collagen type I. J. Am. Chem. Soc 2008;130:7509–7515. [PubMed: 18481852]

98. Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple helix. Biochemistry 2000;39:14960–14967. [PubMed: 11101312]

99. Yang W, Chan VC, Kirkpatrick A, Ramshaw JAM, Brodsky B. Gly–Pro–Arg confers stability similar to Gly–Pro–Hyp in the collagen triple-helix of host-guest peptides. J. Biol. Chem 1997;272:28837–28840. [PubMed: 9360948]

100. Persikov AV, Ramshaw JAM, Brodsky B. Prediction of collagen stability from amino acid sequence. J. Biol. Chem 2005;280:19343–19349. [PubMed: 15753081]

101. Leikina E, Mertts MV, Kuznetsova N, Leikin S. Type I collagen is thermally unstable at body temperature. Proc. Natl. Acad. Sci. USA 2002;99:1314–1318. [PubMed: 11805290]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

102. Buehler MJ. Nature designs tough collagen: explaining the nanostructure of collagen fibrils. Proc. Natl. Acad. Sci. USA 2006;103:12285–12290. [PubMed: 16895989] NOTE: Analysis of the molecular evolution of collagen fibrils for the purpose of achieving maximal strength and flexibility.
103. Kadler KE, Holmes DF, Trotter JA, Chapman JA. Collagen fibril formation. Biochem. J 1996;316:1–11. [PubMed: 8645190]
104. Birk DE, Zycband EI, Winkelmann DA, Trelstad RL. Collagen fibrillogenesis in situ: Fibril segments are intermediates in matrix assembly. Proc. Natl. Acad. Sci. USA 1989;86:4549–4553. [PubMed: 2734306]
105. Holmes DF, Kadler KE. The 10+4 microfibril structure of thin cartilage fibrils. Proc. Natl. Acad. Sci. USA 2006;103:17249–17254. [PubMed: 17088555] NOTE: Highest-resolution structure (~4 nm) of thin cartilage fibrils determined to date.
106. Craig AS, Birtles MJ, Conway JF, Parry DA. An estimate of the mean length of collagen fibrils in rat tail tendon as a function of age. Connect. Tissue Res 1989;19:51–62. [PubMed: 2477190]
107. Hodge, AJ.; Petruska, JA. Recent studies with the electron microscope on ordered aggregates of the tropocollagen macromolecule. In: Ramachandran, GN., editor. Aspects of Protein Structure. London: Academic Press; 1963. p. 289-300.
108. Hulmes DJS, Miller A. Quasi-hexagonal molecular packing in collagen fibrils. Nature 1979;282:878–880. [PubMed: 514368]
109. Trus BL, Piez KA. Compressed microfibril models of the native collagen fibril. Nature 1980;286:300–301. [PubMed: 7402317]
110. Hulmes DJS, Jesior J-C, Miller A, Berthet-Colominas C, Wolff C. Electron microscopy shows periodic structure in collagen fibril cross sections. Proc. Natl. Acad. Sci. USA 1981;78:3567–3571. [PubMed: 6943556]
111. Bozec L, van der Heijden G, Horton M. Collagen fibrils: nanoscale ropes. Biophys. J 2007;92:70–75. [PubMed: 17028135]
112. Orgel JPRO, Miller A, Irving TC, Fischetti RF, Hammersley AP, Wess TJ. The in situ supermolecular structure of type I collagen. Structure 2001;9:1061–1069. [PubMed: 11709170]
113. Orgel JPRO, Irving TC, Miller A, Wess TJ. Microfibrillar structure of type I collagen in situ. Proc. Natl. Acad. Sci. USA 2006;103:9001–9005. [PubMed: 16751282] NOTE: Structure of a type I collagen microfibril at molecular anisotropic resolution (5.16-Å axial; 11.1-Å equatorial).
114. Orgel JP, Wess TJ, Miller A. The in situ conformation and axial location of the intermolecular cross-linked nonhelical telopeptides of type I collagen. Structure 2000;8:137–142. [PubMed: 10673433]
115. Perumal S, Olga A, Orgel JPRO. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. Proc. Natl. Acad. Sci. USA 2008;105:2824–2829. [PubMed: 18287018]
116. Kadler KE, Hojima Y, Prockop DJ. Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. J. Biol. Chem 1987;262:15696–15701. [PubMed: 3316206]
117. Prockop DJ, Fertala A. Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides. J. Biol. Chem 1998;273:15598–15604. [PubMed: 9624151]
118. Kuznetsova N, Leikin S. Does the triple helical domain of type I collagen encode molecular recognition and fiber assembly while telopeptides serve as catalytic domains? J. Biol. Chem 1999;274:36083–36088. [PubMed: 10593890]
119. Eyre DR, Paz MA, Gallop PM. Cross-linking in collagen and elastin. Annu. Rev. Biochem 1984;53:717–748. [PubMed: 6148038]
120. Howard, J. Mechanics of Motor Proteins and the Cytoskeleton. Sunderland, MA: Sinauer; 2001.
121. in 't Veld PJ, Stevens MJ. Simulation of the mechanical strength of a single collagen molecule. Biophys. J 2008;95:33–39. [PubMed: 18359794]
122. van der Rijt JAJ, van der Werf KO, Bennink ML, Dijkstra PJ, Feijen J. Micromechanical testing of individual collagen fibrils. Macromol. Biosci 2006;6:697–702. [PubMed: 16967482]
123. Wenger MPE, Bozec L, Horton M, Mesquida P. Mechanical properties of collagen fibrils. Biophys. J 2007;93:1255–1263. [PubMed: 17526569]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

124. Yang L, van der Werf KO, Fitie CFC, Bennink ML, Dijkstra PJ, Feijen J. Mechanical properties of native and cross-linked type I collagen fibrils. Biophys. J 2008;94:2204–2211. [PubMed: 18032556]

125. Olsen D, Yang C, Bodo M, Chang R, Leigh S, et al. Recombinant collagen and gelatin for drug delivery. Adv. Drug Deliv. Rev 2003;55:1547–1567. [PubMed: 14623401]

126. Kishimoto T, Morihara Y, Osanai M, Ogata S, Kamitakahara M, et al. Synthesis of poly(Pro-Hyp-Gly)$_n$ by direct polycondensation of (Pro-Hyp-Gly)$_n$, where n = 1, 5, and 10, and stability of the triple helical structure. Biopolymers 2005;79:163–172. [PubMed: 16094625]

127. Paramonov SE, Gauba V, Hartgerink JD. Synthesis of collagen-like peptide polymers by native chemical ligation. Macromolecules 2005;38:7555–7561.

128. Kar K, Amin P, Bryan MA, Persikov AV, Mohs A, et al. Self-association of collagen triple-helix peptides into higher order structures. J. Biol. Chem 2006;281:33283–33290. [PubMed: 16963782]

129. Kar K, Wang Y-H. Sequence dependence of kinetics and morphology of collagen model peptide self-assembly into higher order structures. Protein Sci 2008;17:1086–1095. [PubMed: 18441232]

130. Koide T, Homma DL, Asada S, Kitagawa K. Self-complementary peptides for the formation of collagen-like triple helical supramolecules. Bioorg. Med. Chem. Lett 2005;15:5230–5233. [PubMed: 16185864]

131. Kotch FW, Raines RT. Self-assembly of synthetic collagen triple helices. Proc. Natl. Acad. Sci. USA 2006;103:3028–3033. [PubMed: 16488977] NOTE: Synthesis of lengthy collagen triple helices (up to 400 nm) by molecular self-assembly.

132. Yamazaki CM, Asada S, Kitagawa K, Koide T. Artificial collagen gels via self-assembly of de novo designed peptides. Biopolymers 2008;90:816–823. [PubMed: 18846567]

133. Cejas M, Kinney WA, Chen C, Leo GC, Tounge BA, et al. Collagen-related peptides: self-assembly of short, single strands into a functional biomaterial of micrometer scale. J. Am. Chem. Soc 2007;129:2202–2203. [PubMed: 17269769]

134. Cejas MA, Kinney WA, Chen C, Vinter JG, Almond HRJ, et al. Thrombogenic collagen-mimetic peptides: self-assembly of triple helix-based fibrils driven by hydrophobic interactions. Proc. Natl. Acad. Sci. USA 2008;105:8513–8518. [PubMed: 18559857]

135. Gottlieb DG, Morin S, Jin S, Raines RT. Self-assembled collagen-like peptide fibers as templates for metallic nanowires. J. Mater. Chem 2008;18:3865–3870. [PubMed: 20130788]

136. Przybyla DE, Chmielewski J. Metal-triggered radial self-assembly of collagen peptide fibers. J. Am. Chem. Soc 2008;130:12610–12611. [PubMed: 18763780]

137. Rele S, Song Y, Apkarian RP, Qu Z, Conticello VP, Chaikof EL. *D*-periodic collagen-mimetic microfibers. J. Am. Chem. Soc 2007;129:14780–14787. [PubMed: 17985903] NOTE: First self-assembly of CRPs into micrometer-scale fibrils that have *D*-periodicity—-a hallmark of natural collagen fibrils.

138. Holmes DF, Chapman JA, Prockop DJ, Kadler KE. Growing tips of type I collagen fibrils formed in vitro are near-paraboloidal in shape, implying a reciprocal relationship between accretion and diameter. Proc. Natl. Acad. Sci. USA 1992;89:9855–9859. [PubMed: 1409712]

139. Johnson G, Jenkins M, McLean KM, Griesser HJ, Kwak J, et al. Peptoid-containing collagen mimetics with cell binding activity. J. Biomed. Mater. Res 2000;51:612–624. [PubMed: 10880109]

140. Cejas MA, Chen C, Kinney WA, Maryanoff BE. Nanoparticles that display short collagen-related peptides. Potent stimulation of human platelet aggregation by triple helical motifs. Bioconjug. Chem 2007;18:1025–1027. [PubMed: 17583929]

141. Smethurst PA, Onley DJ, Jarvis GE, O'Connor MN, Knight CG, et al. Structural basis for the platelet-collagen interaction. J. Biol. Chem 2007;282:1296–1304. [PubMed: 17085439]

142. Mo X, An Y, Yun C-S, Yu SM. Nanoparticle-assisted visualization of binding interactions between collagen mimetic peptides and collagen fibers. Angew. Chem. Int. Ed. Engl 2006;45:2267–2270. [PubMed: 16511895]

143. Wang AY, Mo X, Chen CS, Yu SM. Facile modification of collagen directed by collagen mimetic peptides. J. Am. Chem. Soc 2005;127:4130–4131. [PubMed: 15783169]

144. Wang AY, Foss CA, Leong S, Mo X, Pomper MG, Yu SM. Spatio-temporal modification of collagen scaffolds mediated by triple helical propensity. Biomacromolecules 2008;9:1755–1763. [PubMed: 18547103]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

145. Dobson CM. Protein folding and misfolding. Nature 2003;426:884–890. [PubMed: 14685248]

146. Nelson R, Sawaya MR, Balbirnie M, Madsen AØ, Riekel C, et al. Structure of the cross-β spine of amyloid-like fibrils. Nature 2005;435:773–778. [PubMed: 15944695]

147. Kim CA, Berg JM. Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. Nature 1993;362:267–270. [PubMed: 8459852]

148. Minor DL Jr, Kim PS. Measurement of the β-sheet-forming propensities of amino acids. Nature 1994;367:660–663. [PubMed: 8107853]

149. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 2003;424:805–808. [PubMed: 12917692]

150. Rauscher S, Baud S, Miao M, Keeley FW, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. Structure 2006;14:1667–1676. [PubMed: 17098192]

RELATED RESOURCES

Dalgleish, R. A database of osteogenesis imperfecta and type III collagen mutations. 2009. http://www.le.ac.uk/genetics/collagen/

Khoshnoodi, J.; Cartailler, J-P.; Alvares, K.; Veis, A.; Hudson, BG. Computer-generated animation of assembly of type I and type IV collagen for Reference [38]. 2006. http://www.mc.vanderbilt.edu/cmb/collagen/

Ricard-Blum S, Ruggiero F, van der Rest M. The collagen superfamily. Top. Curr. Chem 2005;247:35–84.

Koide T, Nagata K. Collagen biosynthesis. Top. Curr. Chem 2005;247:85–114.

Greenspan DS. Biosynthetic processing of collagen molecules. Top. Curr. Chem 2005;247:149–183.

Birk DE, Bruckner P. Collagen suprastructures. Top. Curr. Chem 2005;247:185–205.

Franzke C-W, Bruckner P, Bruckner-Tuderman L. Collagenous transmembrane proteins: Recent insights into biology and pathology. J. Biol. Chem 2005;280:4005–4008. [PubMed: 15561712]

Myllyharju J. Prolyl-4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biol 2003;22:15–24. [PubMed: 12714038]

Raines RT. 2005 Emil Thomas Kaiser award. Protein Sci 2006;15:1219–1225. [PubMed: 16641494]

ACRONYMS AND KEY TERMS

D-Periodicity	the axial stagger of adjacent tropocollagen molecules by a distance, *D*, which is the sum of the gap and overlap regions.
Collagen fibrillogenesis	the process of tropocollagen monomers assembling into mature fibrils
Preorganization	the extent to which hosts and guests are organized for binding prior to their complexation, thereby increasing complex stability.”
Protocollagen	the hydroxylated form of collagen prior to collagen propeptide cleavage
Collagen propeptides	N- and C-terminal non–triple-helical domains of collagen strands that direct triple-helix folding prior to fibrillogenesis
Protocollagen	the nonhydroxylated, non–triple-helical form of collagen prior to the action of P4H, P3H, lysyl hydroxylase, and protein disulfide isomerase
Self-assembly	a process in which specific, local interactions between disordered components leads to an organized structure, without external direction

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Stereoelectronic effects	relationships between structure, conformation, energy, and reactivity that result from the alignment of filled or unfilled electronic orbitals
Collagen telopeptides	N- and C-terminal 11- to 26-residue non–triple-helical domains of tropocollagen strands involved in fibrillogenesis and cross-linking
Tropocollagen	the monomeric collagen triple helix after proteolysis of collagen propeptides
CRP	collagen-related peptide
ECM	extracellular matrix
Flp	(2*S*,4*R*)-4-fluoroproline
flp	(2*S*,4*S*)-4-fluoroproline
Hyp	(2*S*,4*R*)-4-hydroxyproline
hyp	(2*S*,4*S*)-4-hydroxyproline
OI	osteogenesis imperfecta
P4H	prolyl 4-hydroxylase
PPII	polyproline II-type

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

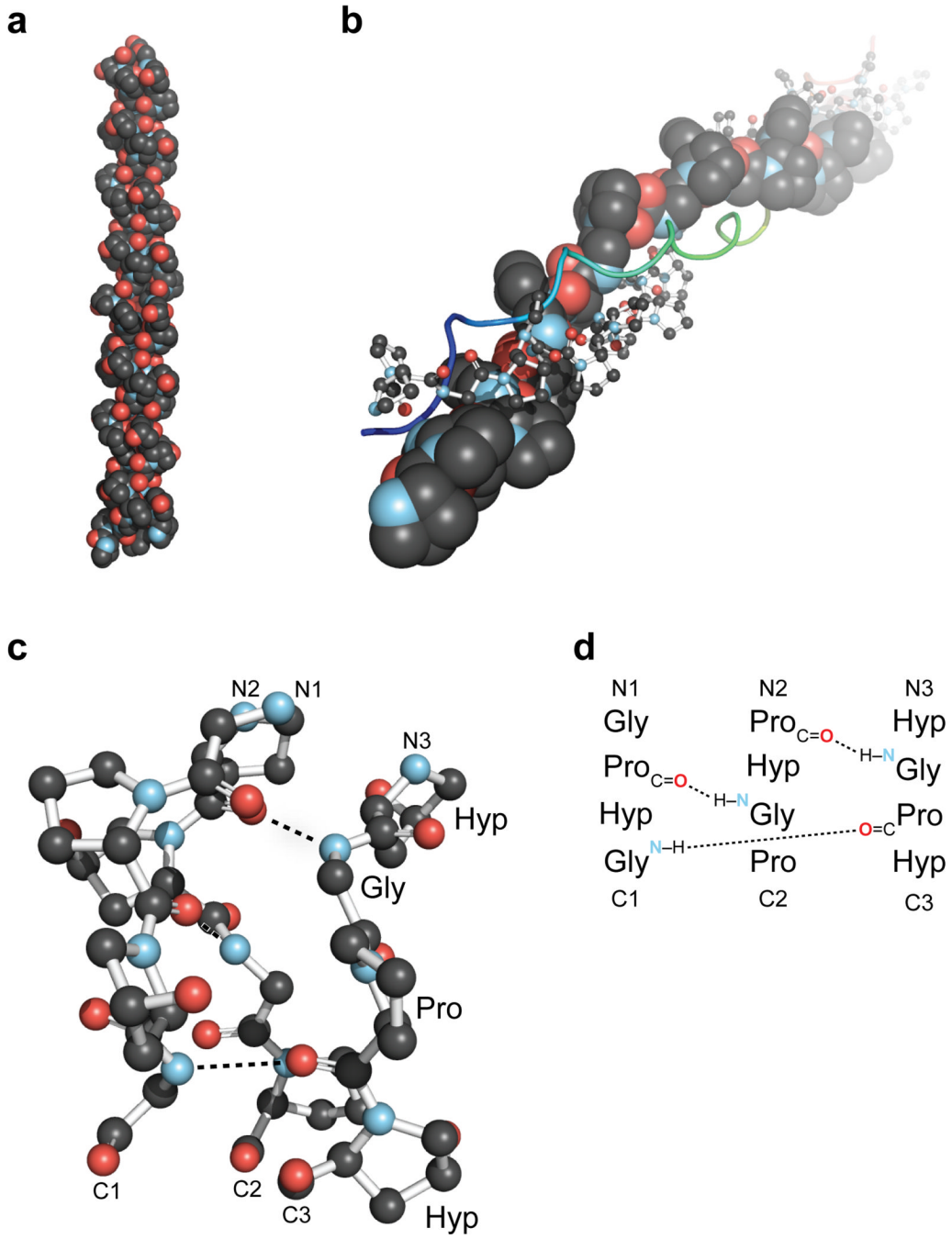


Figure 1.

Overview of the collagen triple helix. (*a*) First high-resolution crystal structure of a collagen triple helix, formed from $(ProHypGly)_4$–(ProHypAla)–$(ProHypGly)_5$ [Protein Data Bank (PDB) entry 1cag] (19). (*b*) View down the axis of a $(ProProGly)_{10}$ triple helix [PDB entry 1k6f (7)] with the three strands depicted in space-filling, ball-and-stick, and ribbon representation. (*c*) Ball-and-stick image of a segment of collagen triple helix [PDB entry 1cag (19)], highlighting the ladder of interstrand hydrogen bonds. (*d*) Stagger of the three strands in the segment in panel *c*.

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

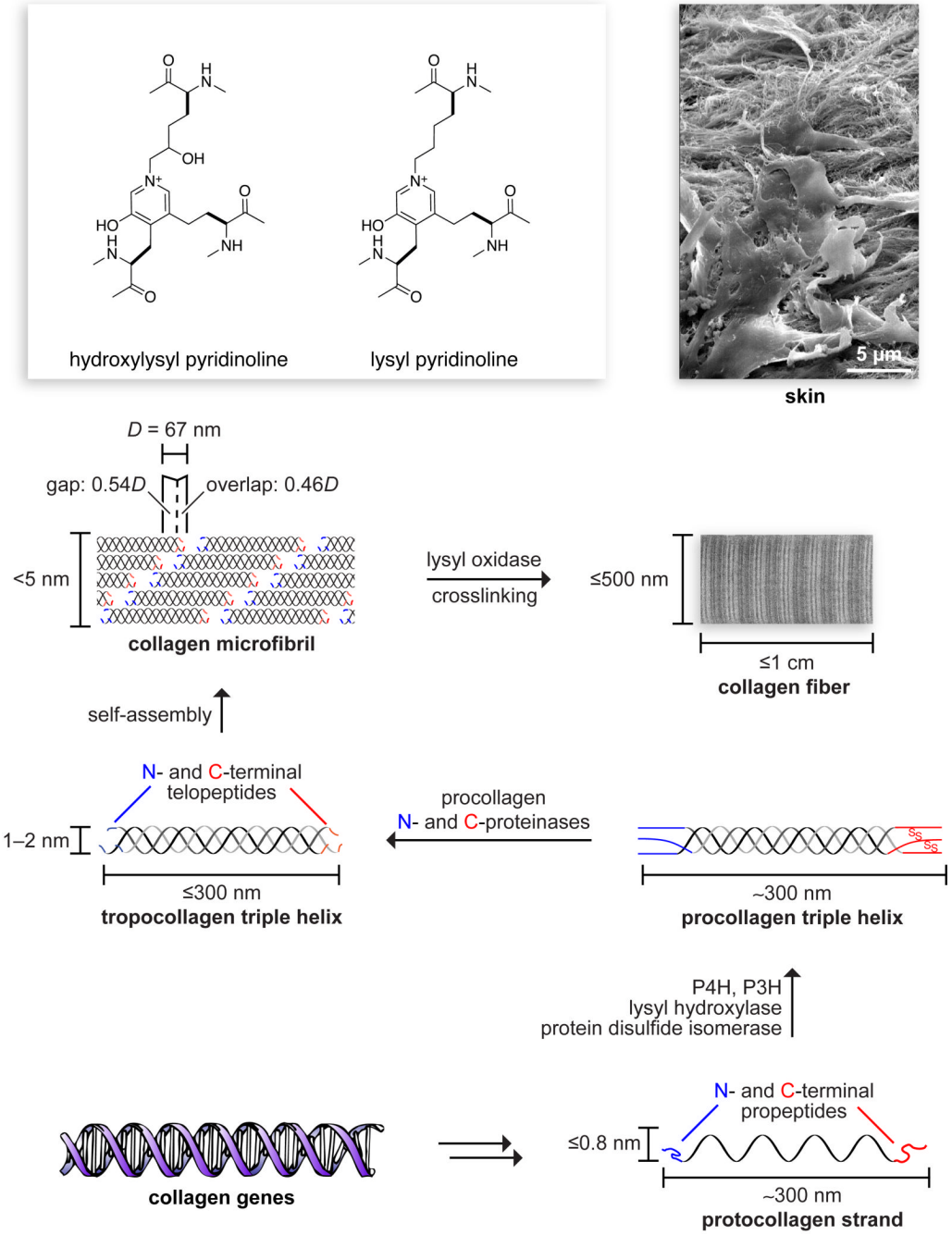


Figure 2.

Biosynthetic route to collagen fibers (110), which are the major component of skin. Size and complexity is increased by posttranslational modifications and self-assembly. Oxidation of lysine side chains leads to the spontaneous formation of hydroxylysyl pyridinoline and lysyl pyridinoline cross-links.

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

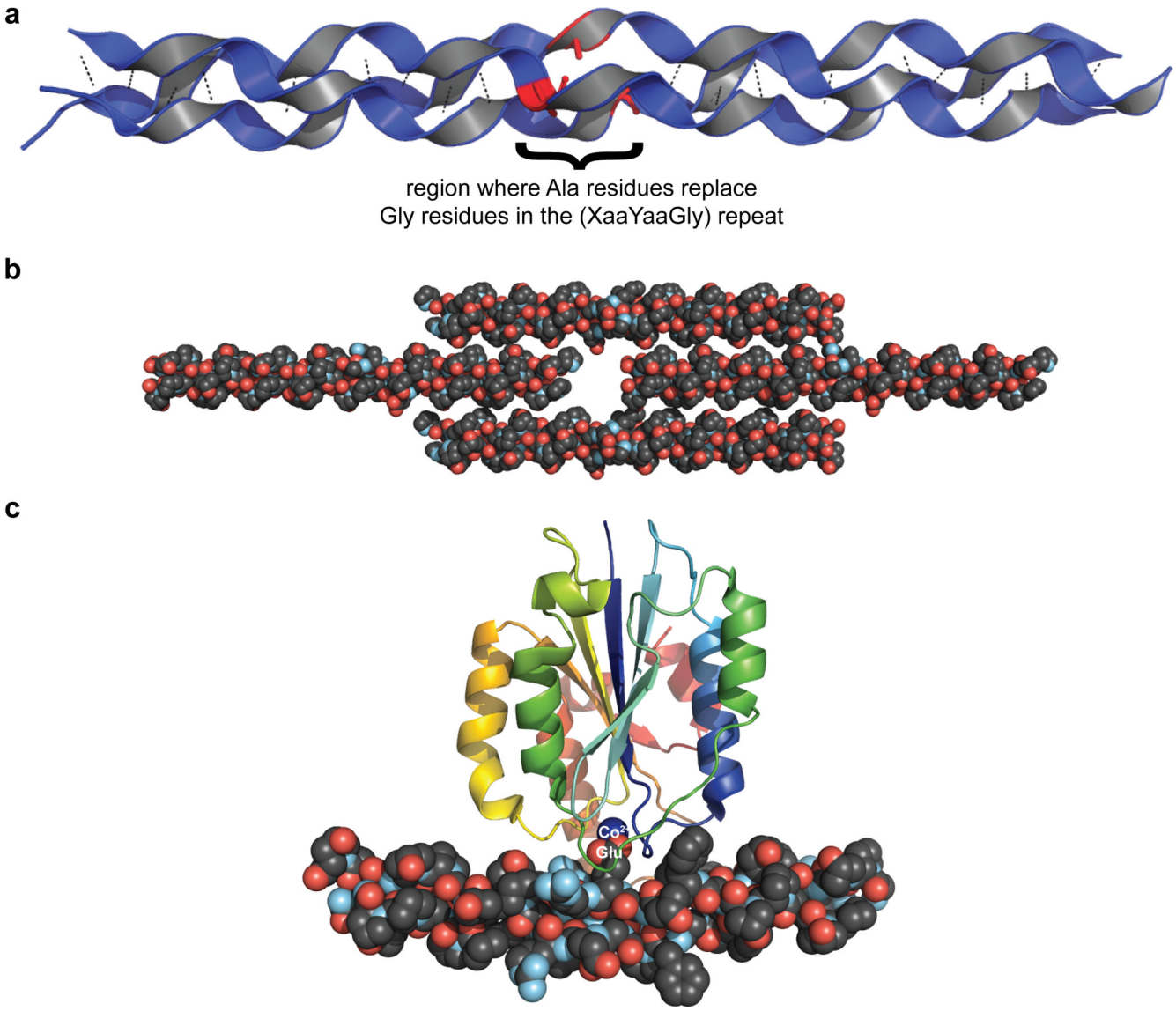


Figure 3.

Snapshots of interesting crystal structures of collagen triple helices. (*a*) Impact of a Gly→Ala substitution on the structure of a collagen triple helix formed from the collagen-related peptide (CRP) $(ProHypGly)_4$–(ProHypAla)–$(ProHypGly)_5$ [Protein Data Bank (PDB) entry 1cag (19)]. The Ala residues (*red*) disturb the structure. Mutations leading to such structural irregularities are common in osteogenesis imperfecta and can be lethal. (*b*) Depiction of the effect of a single GluLysGly triplet on the packing of neighboring triple-helical CRPs in crystalline $(ProHypGly)_4$–(GluLysGly)–$(ProHypGly)_5$ [PDB entry 1qsu (21)]. The axial stagger of the individual triple helices, which is presumably compelled by deleterious Coulombic interactions between charged residues, is reminiscent of the *D*-periodic structure in collagen fibrils. Similar interactions could contribute to the morphology of collagen fibrils. (*c*) Triple-helical CRP containing the integrin-binding domain GFOGER in complex with the I domain of integrin $\alpha_2\beta_1$ [PDB entry 1dzi (22)]. The bend in the triple helix is thought to arise from the protein-protein interaction. A Glu residue in the middle strand of the triple helix coordinates to Co(II) (*blue*) bound in the I domain of integrin $\alpha_2\beta_1$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

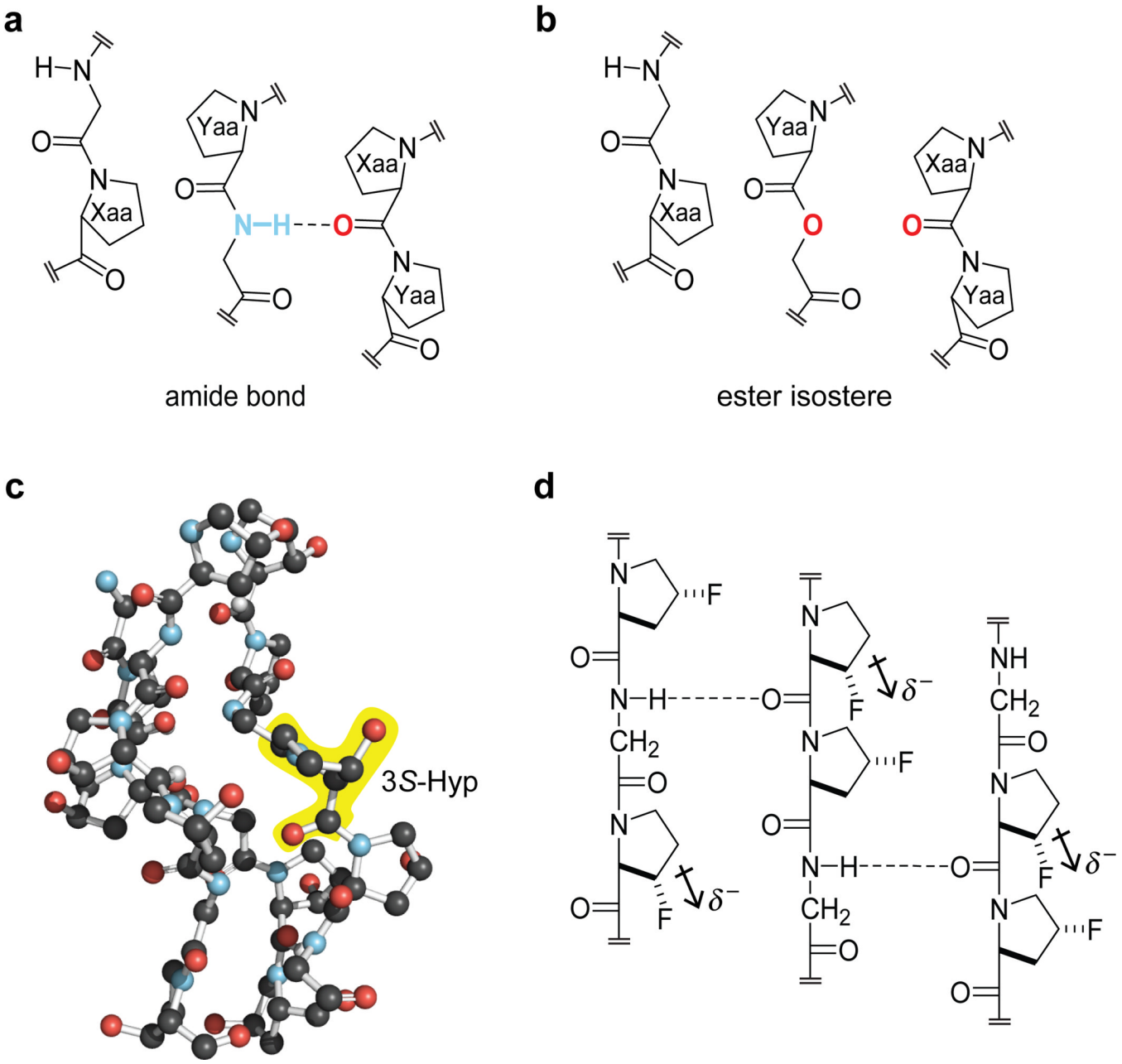


Figure 4.

Importance of interstrand hydrogen bonds for collagen triple-helix stability. (*a*) A segment of a $(ProProGly)_{10}$ triple helix. (*b*) Comparison of the stability of a triple helix formed from $(ProProGly)_4$–ProProOGly–$(ProProGly)_5$, wherein one Pro–Gly amide bond is replaced with an ester, with that in panel *a* revealed that each interstrand hydrogen bond contributes $\Delta G = -2.0$ kcal/mol to triple-helix stability (30). (*c*) Crystal structure of a triple helix formed from a collagen-related peptide that mimics a common sequence in type IV collagen, $(GlyProHyp)_3$–$(3S\text{-}HypHypGly)_2$–$(GlyProHyp)_4$, showing that 3*S*-Hyp in the Xaa position yields a prototypical collagen triple helix [PDB entry 2g66 (78)]. (*d*) (2*S*,3*S*)-3-Fluoroproline in the Xaa position destabilizes a collagen triple helix, perhaps by withdrawing electron density from the proximal Xaa carbonyl and thereby reducing the strength of the interstrand hydrogen bond (79).

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

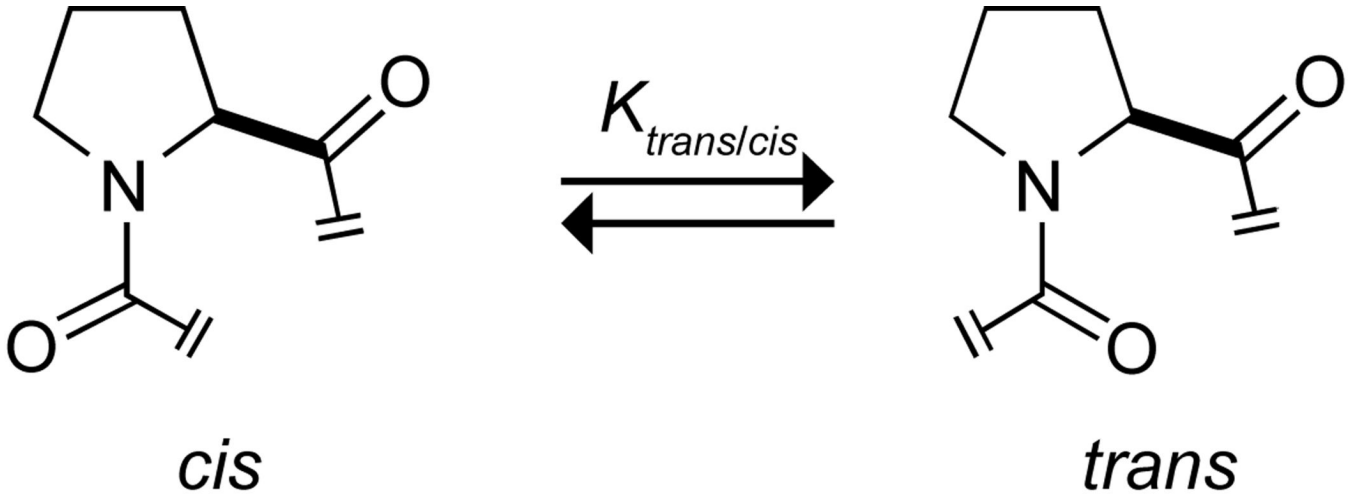


Figure 5.
Pro *cis–trans* isomerization. Unlike other proteinogenic amino acids, Pro forms tertiary amide bonds, resulting in a significant population of the *cis* conformation.

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

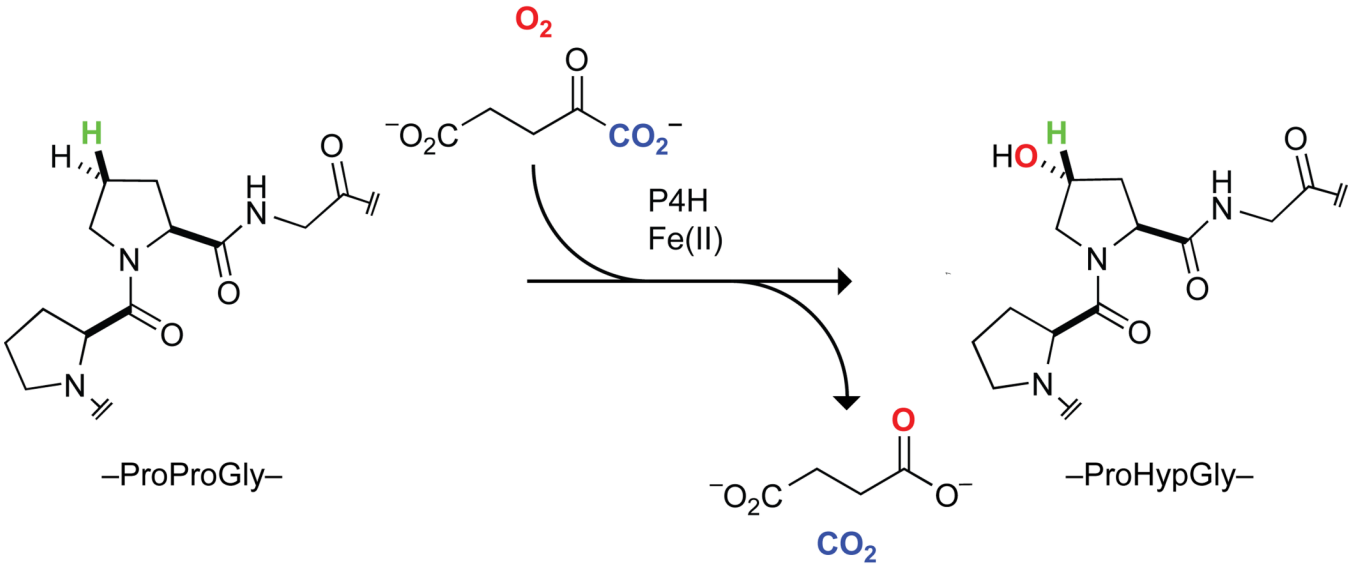


Figure 6.
Reaction catalyzed by prolyl 4-hydroxylase (P4H). Pro residues in the Yaa position of collagen strands are converted into Hyp prior to triple-helix formation.

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

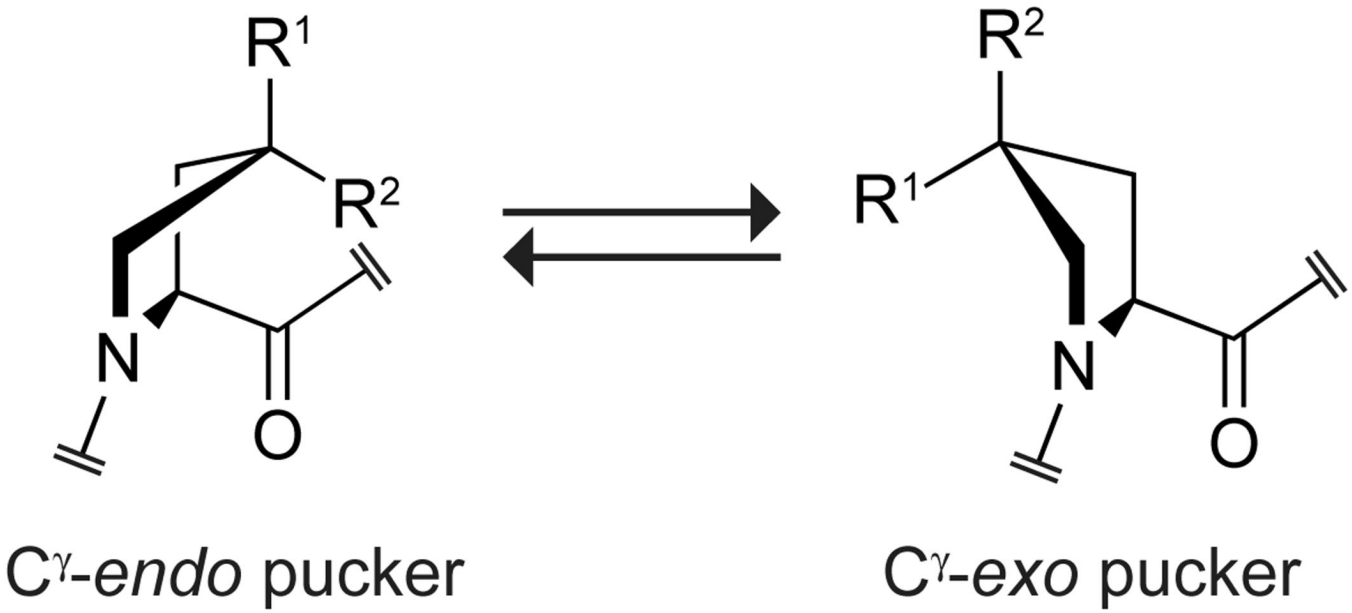


Figure 7.

Ring conformations of Pro and Pro derivatives. The C^{γ}-*endo* conformation is favored strongly by stereoelectronic effects when R^1 = H, R^2 = F (flp) or Cl (clp), and by steric effects when R^1 = Me (mep) or SH (mcp), R^2 = H. The C^{γ}-*exo* conformation is favored strongly by stereoelectronic effects when R^1 = OH (Hyp), F (Flp), OMe (Mop), or Cl (Clp), R^2 = H, and by steric effects when R^1 = H, R^2 = Me (Mep) or SH (Mcp). The C^{γ}-*endo*:C^{γ}-*exo* ratio is ~2 when R^1 = R^2 = H (56).

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

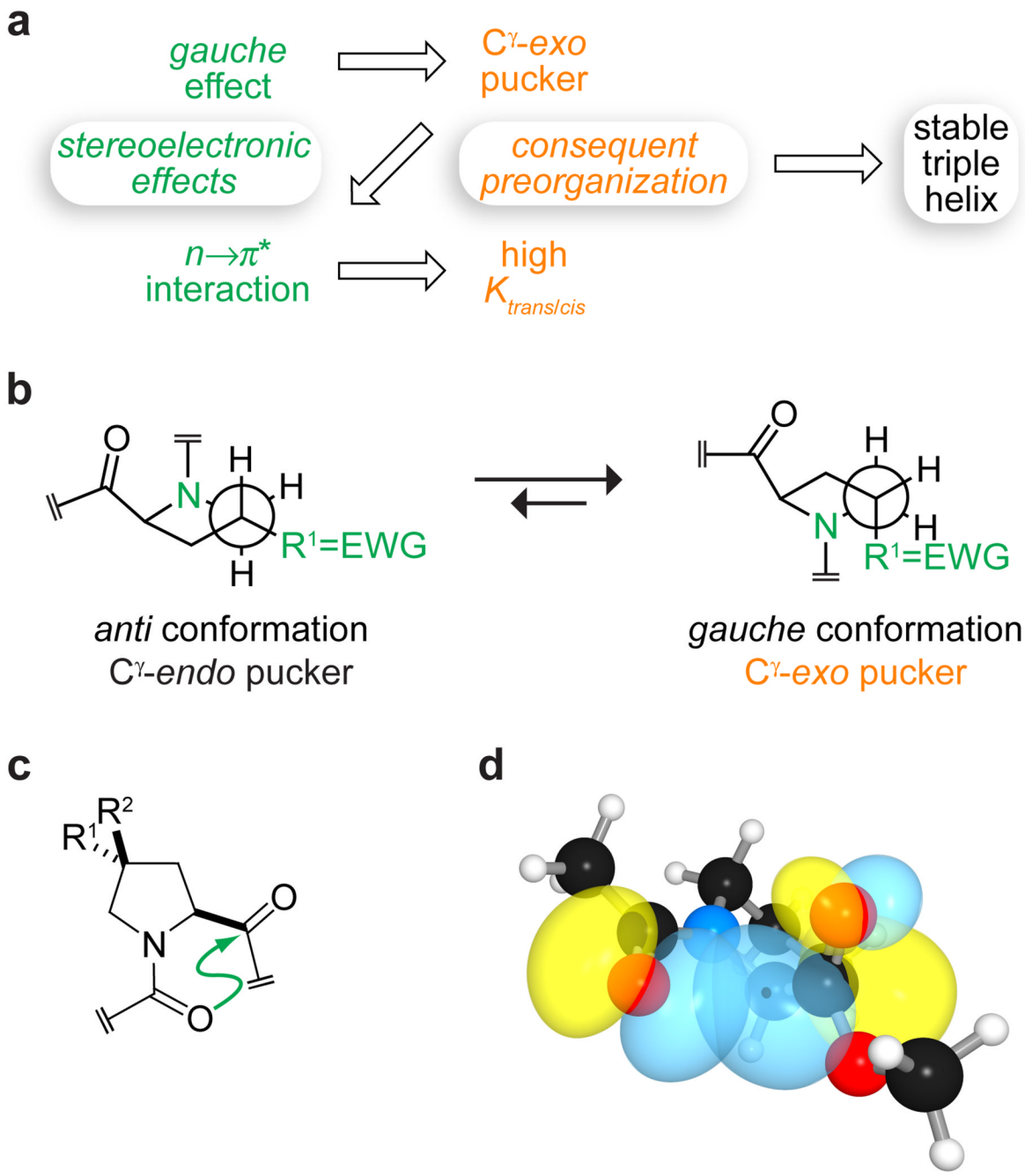


Figure 8.

Stereoelectronic effects that stabilize the collagen triple helix. (*a*) A gauche effect and an $n \rightarrow \pi^*$ interaction preorganize main chain torsion angles and enhance triple-helix stability. (*b*) A gauche effect, elicited by an electron-withdrawing group (EWG) in the 4*R* position, stabilizes the C^{γ}-*exo* ring pucker. (*c*) An $n \rightarrow \pi^*$ interaction stabilizes the *trans* isomer of the peptide bond but is substantial only when Pro derivatives are in the C^{γ}-*exo* ring pucker (e.g., R^1 = OH or F, R^2 = H). (*d*) Depiction of overlap between *n* and π^* natural bond orbitals (NBO*View*©) in a Pro residue with C^{γ}-*exo* pucker.

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

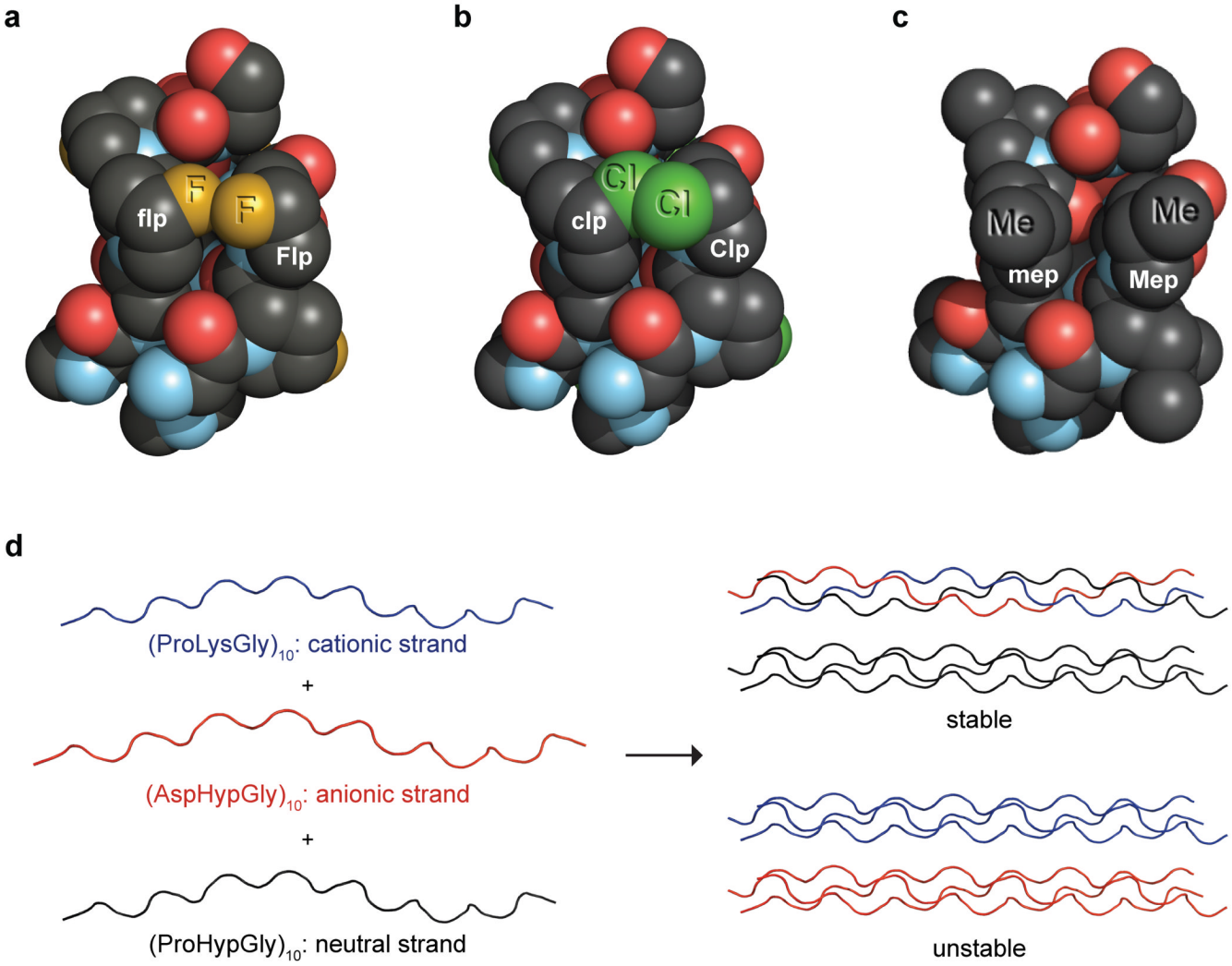


Figure 9.

Heterotrimeric synthetic collagen triple helices. (*a*–*c*) Steric approach. Space-filling models of triple-helix segments constructed from the structure of a $(ProHypGly)_n$ triple helix [PDB entry 1cag (19)] with the program SYBYL (Tripos, St. Louis, MO). In panel *a*, $r_{F\cdots F}$ = 2.4 Å in a $(flpFlpGly)_n$ triple helix (79). In panel *b*, $r_{Cl\cdots Cl}$ = 1.9 Å (61) in a $(clpClpGly)_n$ triple helix. In panel *c*, the methyl groups in a $(mepMepGly)_n$ triple helix are radial and distal. (*d*) Coulombic approach. Favorable Coulombic interactions drive the preferential assembly of triple helices having a 1:1:1 ratio of $(ProLysGly)_{10}$:$(AspHypGly)_{10}$:$(ProHypGly)_{10}$ (96).

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

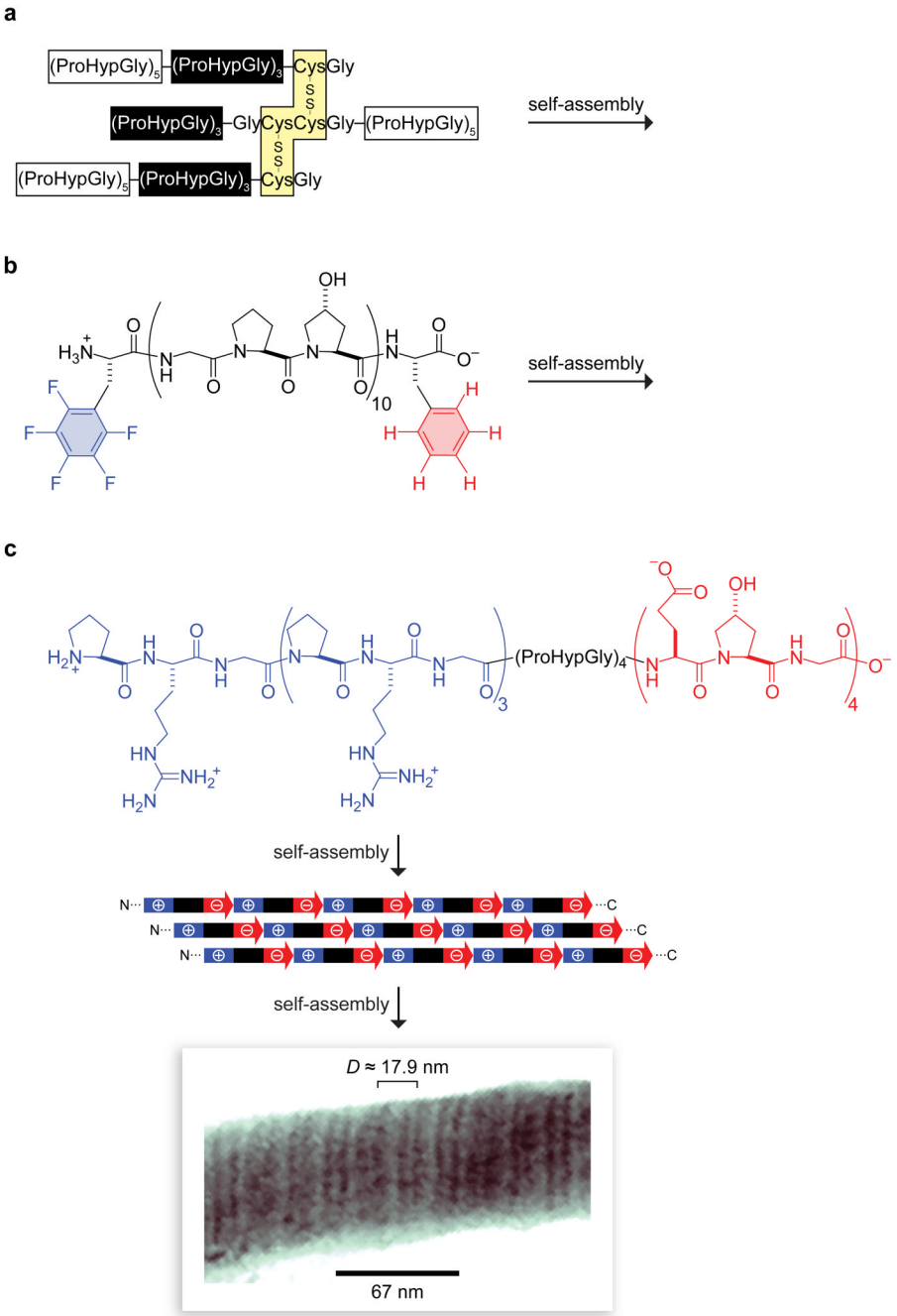


Figure 10.

Strategies for the self-assembly of long, synthetic collagen triple helices and fibrils. (*a*) Disulfide bonds enforce a strand register with sticky ends that self-assemble (131). (*b*) Stacking interactions between electron-poor pentafluorophenyl rings and electron-rich phenyl rings lead to self-assembly (133,134). (*c*) Coulombic forces between cationic and anionic blocks encourage self-assembly. TEM image of a resulting fiber shows *D*-periodicity with *D* = 17.9 nm (137). Natural type I collagen has *D* = 67 nm.

EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Vertebrate collagens[a]

Type	Class	Composition	Distribution[b]	Pathology[c]
I	Fibrillar	$\alpha1[I]_2\alpha2[I]$	Abundant and widespread: dermis, bone, tendon, ligament	OI, Ehlers–Danlos syndrome, osteoporosis
II	Fibrillar	$\alpha1[II]_3$	Cartilage, vitreous	Osteoarthrosis, chondrodysplasias
III	Fibrillar	$\alpha1[III]_3$	Skin, blood vessels, intestine	Ehlers–Danlos syndrome, arterial aneurysms
IV	Network	$\alpha1[IV]_2\alpha2[IV]$ $\alpha3[IV]\alpha4[IV]\alpha5[IV]$ $\alpha5[IV]_2\alpha6[IV]$	Basement membranes	Alport syndrome
V	Fibrillar	$\alpha1[V]_3$ $\alpha1[V]_2\alpha2[V]$ $\alpha1[V]\alpha2[V]\alpha3[V]$	Widespread: bone dermis, cornea, placenta	Ehlers–Danlos syndrome
VI	Network	$\alpha1[VI]\alpha2[VI]\alpha3[VI]$[d] $\alpha1[VI]\alpha2[VI]\alpha4[VI]$	Widespread: bone, cartilage, cornea, dermis	Bethlem myopathy
VII	Anchoring fibrils	$\alpha1[VII]_2\alpha2[VII]$	Dermis, bladder	Epidermolysis bullosa acquisita
VIII	Network	$\alpha1[VIII]_3$ $\alpha2[VIII]_3$ $\alpha1[VIII]_2\alpha2[VIII]$	Widespread: dermis, brain, heart, kidney	Fuchs endothelia corneal dystrophy
IX	FACIT[e]	$\alpha1[IX]\alpha2[IX]\alpha3[IX]$	Cartilage, cornea, vitreous	Osteoarthrosis, multiple epiphyseal dysplasia
X	Network	$\alpha1[X]_3$	Cartilage	Chondrodysplasia
XI	Fibrillar	$\alpha1[XI]\alpha2[XI]\alpha3[XI]$	Cartilage, intervertebral disc	Chondrodysplasia, osteoarthrosis
XII	FACIT	$\alpha1[XII]_3$	Dermis, tendon	—
XIII	MACIT	—	Endothelial cells, dermis, eye, heart	—
XIV	FACIT	$\alpha1[XIV]_3$	Widespread: bone dermis, cartilage	—
XV	MULTIPLEXIN	—	Capillaries, testis kidney, heart,	—
XVI	FACIT	—	Dermis, kidney	—
XVII	MACIT	$\alpha1[XVII]_3$	Hemidesmosomes in epithelia	Generalized atrophic epidermolysis bullosa
XVIII	MULTIPLEXIN	—	Basement membrane, liver	Knobloch syndrome
XIX	FACIT	—	Basement membrane	—
XX	FACIT	—	Cornea (chick)	—
XXI	FACIT	—	Stomach, kidney	—
XXII	FACIT	—	Tissue junctions	—
XXIII	MACIT	—	Heart, retina	—

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Type	Class	Composition	Distribution[b]	Pathology[c]
XXIV	Fibrillar	—	Bone, cornea	—
XXV	MACIT	—	Brain, heart, testis	Amyloid formation?
XXVI	FACIT	—	Testis, ovary	—
XXVII	Fibrillar	—	Cartilage	—
XXVIII[f]	—	—	Dermis, sciatic nerve	Neurodegenerative disease?

[a]Information is updated from References 1 and 32.

[b]Partial listing of tissues in which the relevant collagen type occurs.

[c]For a discussion of the role of specific collagen types in human disease, see Reference 32.

[d]α4[VI], α5[VI], and α6[VI] chains were reported in 2008 (10), but the composition of triple helices containing α5[VI] and α6[VI] is unknown.

[e]Abbreviations: FACIT, fibril-associated collagen with interrupted triple helices; MACIT, membrane-associated collagen with interrupted triple helices; MULTIPLEXIN, multiple triple helix domains and interruptions.

[f]Collagen XXVIII was reported in 2006 (2).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Values of T_m for triple-helical CRPs

(XaaYaaGly)$_n$	T_m (°C) [a]	References
(ProFlpGly)$_7$	45	(54)
(ProHypGly)$_7$	36	(54)
(mepMepGly)$_7$	36	(65)
(flpProGly)$_7$	33	(68)
(ProMepGly)$_7$	29	(65)
(ProClpGly)$_7$	23	(61)
(mepProGly)$_7$	13	(65)
(clpProGly)$_7$	No helix	(61)
(ProProGly)$_7$	No helix	(79)
(flpFlpGly)$_7$	No helix	(79)
(ProflpGly)$_7$	No helix	(54)
(FlpProGly)$_7$	No helix	(68)
(ProFlpGly)$_{10}$	91	(53)
(ProMopGly)$_{10}$	70	(59)
(HypHypGly)$_{10}$	65	(85)
(ProHypGly)$_{10}$	61–69	(53,85)
(flpProGly)$_{10}$	58	(69)
(ProClpGly)$_{10}$	52	(61)
(clpProGly)$_{10}$	33	(61)
(ProProGly)$_{10}$	31–41	(53,64)
(flpFlpGly)$_{10}$	30	(94)
(clpClpGly)$_{10}$	No helix	(61)
(HypProGly)$_{10}$	No helix	(84)
(ProhypGly)$_{10}$	No helix	(47)
(FlpProGly)$_{10}$	No helix	(69)
(ClpProGly)$_{10}$	No helix	(61)
(hypProGly)$_{10}$	No helix	(47)

[a] Values of T_m depend on both CRP concentration and heating rate. Hence, detailed comparisons require knowledge of experimental procedures.

 EMBZIM EX1017

NIH-PA Author Manuscript

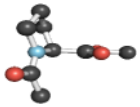
NIH-PA Author Manuscript

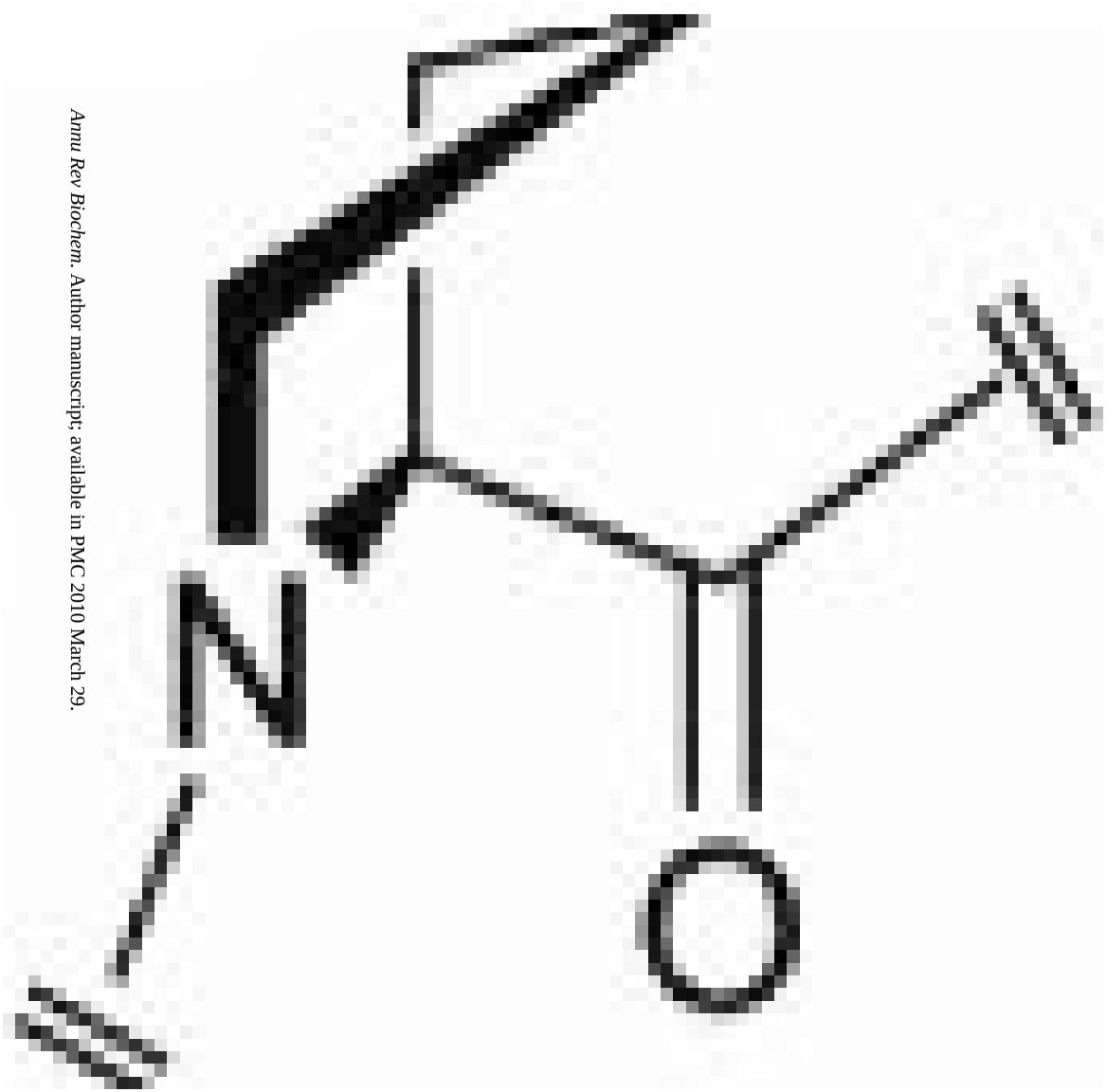
NIH-PA Author Manuscript

Table 3

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Conformation of Pro and its 4-substituted derivatives that prefer the Xaa position [φ = −75°, ψ = 164° (7)] in a collagen triple helix.

Residue (References)	Crystal structure	Ring pucker	$(E_{endo}-E_{exo})^a$ (kcal/mol)	φ (degrees)	ψ degrees)	$K_{trans/cis}{}^c$
Pro (5, 56,57)		C^{γ}-*endo*	−0.41	−79[b]	177[b]	4.6

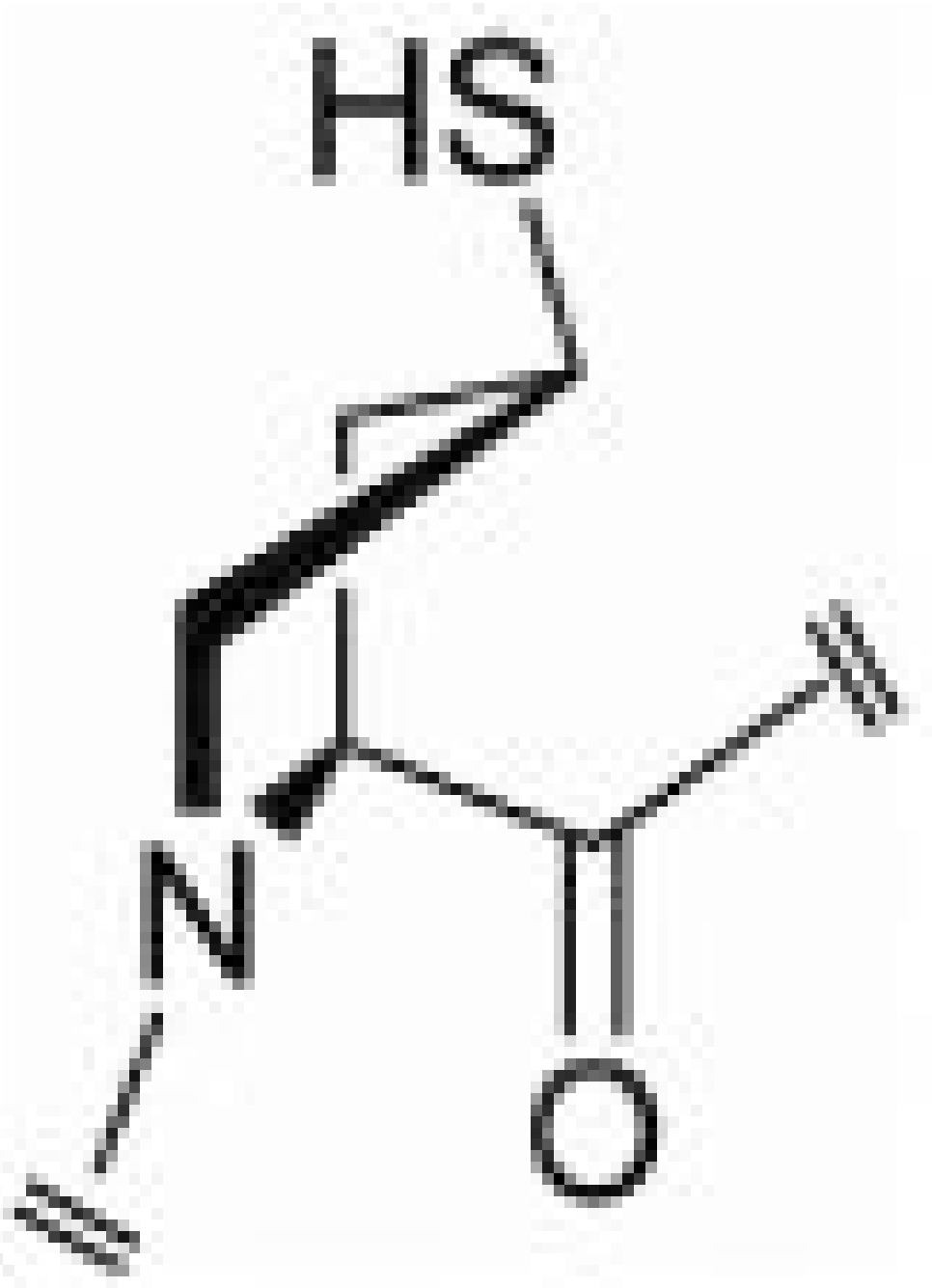


EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

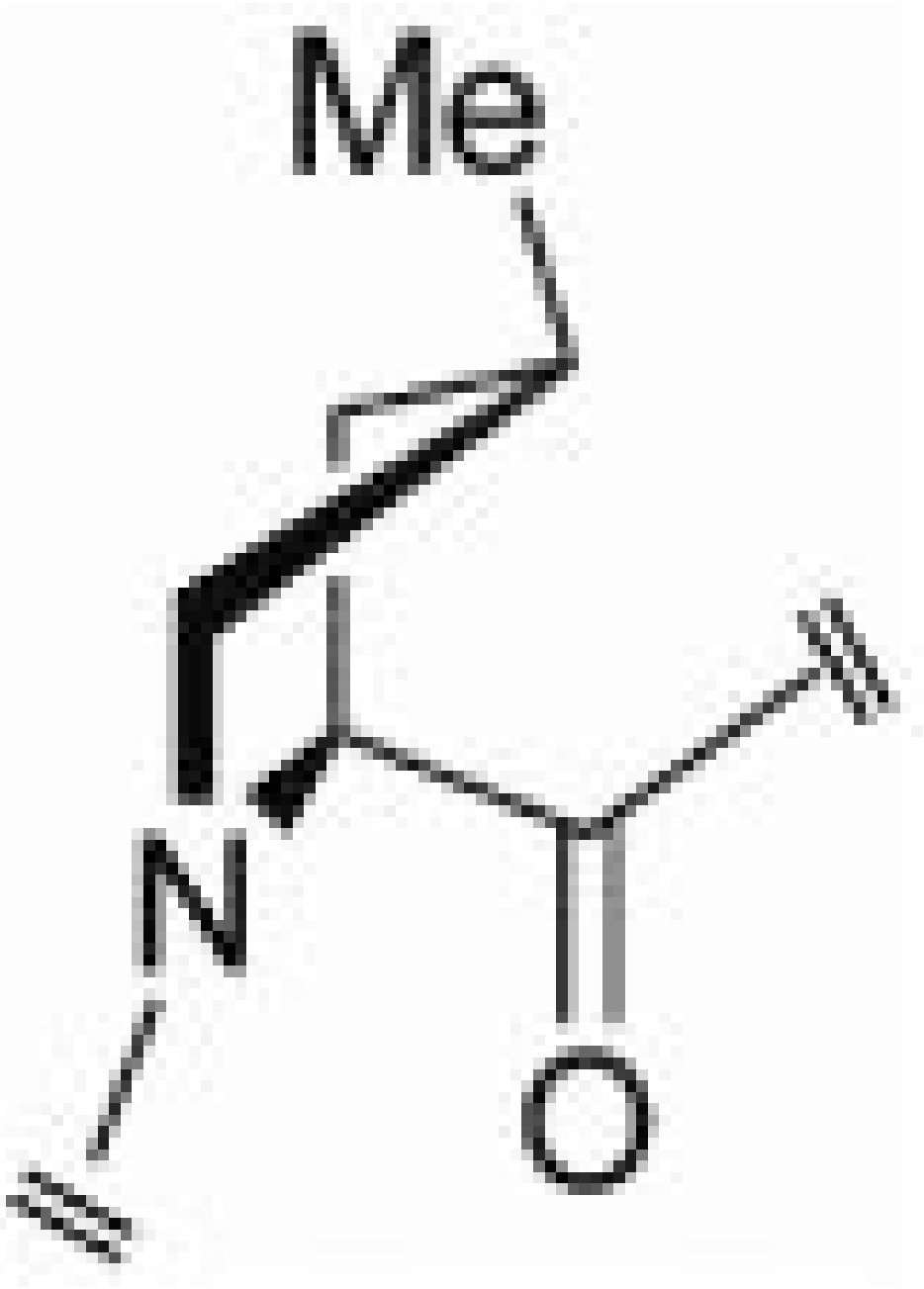
NIH-PA Author Manuscript

Residue (References)	Crystal structure	Ring pucker	$(E_{endo}-E_{exo})^a$ (kcal/mol)	φ (degrees)	ψ degrees)	$K_{trans/cis}{}^c$
mcp (66) 	—	C^{γ}-*endo*	—	—	—	4.7

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

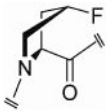
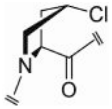
Residue (References)	Crystal structure	Ring pucker	$(E_{endo}-E_{exo})^{a}$ (kcal/mol)	φ (degrees)	ψ degrees)	$K_{trans/cis}{}^{c}$
mep (65) 	—	C^{γ}-*endo*	−1.4	—	—	3.7

EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Residue (References)		Crystal structure	Ring pucker	(E_{endo}−E_{exo})[a] (kcal/mol)	φ (degrees)	ψ degrees)	$K_{trans/cis}$[c]
flp (54–56,68)		—	C^{γ}-*endo*	−0.61	−76[a]	172[a]	2.5
clp (61)		—	C^{γ}-*endo*	—	—	—	2.2

[a]From DFT calculations.

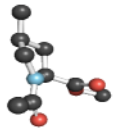
[b]Values of φ and ψ (here, N_i–$C^{\alpha}{}_i$–C'_i–O_{i+1}) are from the crystal structure of Ac-Pro-OMe, which has a *cis* peptide bond.

[c]Values of $K_{trans/cis}$ (Figure 5) were determined in solution by NMR spectroscopy.

EMBZIM EX1017

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

52° (7) in a collagen triple helix.

Ring pucker	$(E_{endo} - E_{exo})^a$ (kcal/mol)	φ [b] (degrees)	ψ[b] (degrees)	$K_{trans/cis}$[c]	Crystal structure
C^{γ}-*exo*	1.7	−62[d]	153[d]	7.4	

EMBZIM EX1017

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Crystal structure	Ring pucker	$(E_{endo} - E_{exo})^a$ (kcal/mol)	φ [b] (degrees)	ψ [b] (degrees)	$K_{trans/cis}$ [c]
	C^{γ}-*exo*	—	−58	148	6.7

 EMBZIM EX1017

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Crystal structure	Ring pucker	$(E_{endo} - E_{exo})^a$ (kcal/mol)	φ^b (degrees)	ψ^b (degrees)	$K_{trans/cis}{}^c$
	C^{γ}-*exo*	0.85	−55	141	6.7

EMBZIM EX1017

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Crystal structure	Ring pucker	$(E_{endo} - E_{exo})^a$ (kcal/mol)	φ^b (degrees)	ψ^b (degrees)	$K_{trans/cis}{}^c$
	C^{γ}-*exo*	—	−57	151	6.1

 EMBZIM EX1017

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Crystal structure	Ring pucker	$(E_{endo} - E_{exo})^a$ (kcal/mol)	φ^b (degrees)	ψ^b (degrees)	$K_{trans/cis}{}^c$
	C^{γ}-*exo*	—	−56	148	5.4

EMBZIM EX1017

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

$K_{trans/cis}$[c]	ψ[b] (degrees)	φ[b] (degrees)	(E_{endo} − E_{exo})[a] (kcal/mol)	Ring pucker	Crystal structure
5.4	—	—	—	C^{γ}-*exo*	—

EMBZIM EX1017

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[a] From DFT calculations

[b] Values of φ and ψ (here, N_i–$C^{\alpha}{}_i$–C'_i–O_{i+1}) are from crystal structures of Ac-Yaa-OMe.

[c] Values of $K_{trans/cis}$ (Figure 5) were determined in solution by NMR spectroscopy.

[d] M.D. Shoulders, I.A. Guzei, & R.T. Raines, unpublished data.