

REVIEW ARTICLE

Collagen structure: new tricks from a very old dog

 Jordi Bella*¹

*Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

The main features of the triple helical structure of collagen were deduced in the mid-1950s from fibre X-ray diffraction of tendons. Yet, the resulting models only could offer an average description of the molecular conformation. A critical advance came about 20 years later with the chemical synthesis of sufficiently long and homogeneous peptides with collagen-like sequences. The availability of these collagen model peptides resulted in a large number of biochemical, crystallographic and NMR studies that have revolutionized our understanding of collagen structure. High-resolution crystal structures from collagen model peptides have provided a wealth of data on collagen conformational variability, interaction with water, collagen stability or the effects of interruptions. Furthermore, a large increase in the number of structures of collagen model peptides in complex with domains from receptors or collagen-binding proteins has shed light

on the mechanisms of collagen recognition. In recent years, collagen biochemistry has escaped the boundaries of natural collagen sequences. Detailed knowledge of collagen structure has opened the field for protein engineers who have used chemical biology approaches to produce hyperstable collagens with unnatural residues, rationally designed collagen heterotrimers, self-assembling collagen peptides, etc. This review summarizes our current understanding of the structure of the collagen triple helical domain (COL \times 3) and gives an overview of some of the new developments in collagen molecular engineering aiming to produce novel collagen-based materials with superior properties.

Key words: collagen recognition, collagen stability, collagen structure, hydrogen bonding, protein design and engineering.

COLLAGEN SUPERFAMILY

In its most widely used meaning, the term ‘collagen’ refers to the main structural protein in connective tissues such as skin, bone, tendons or cartilage. This cable-like protein forms elongated fibrils that provide mechanical stability to animals. Bone or skin collagen is mainly type I collagen, the most abundant protein in mammals, but in fact there are many types of collagens in any given animal [1]. Vertebrates have at least 45 distinct collagen genes that account for a family of 28 collagen proteins [2–5]. Invertebrates have their own *collagenomes* (used here to refer to the collection of collagens and collagen-like proteins from a given organism or taxonomic group), which contribute to the structural integrity of the animals. Thus, the cuticle of the *Caenorhabditis elegans* nematode is predominantly made of cross-linked collagens that are encoded by more than 170 collagen genes [6–8]. Collagens are already present in the most primitive animals, such as sponges, and therefore are considered intrinsic to the evolution of metazoans [9,10].

All collagens are trimeric proteins formed by association of three polypeptide chains. These can be identical (homotrimer collagens) or different (heterotrimer collagens). Different genes code for distinct collagen chains, and only chains from the same collagen type associate with each other (there are no trimers made from chains of different collagen types). Vertebrate collagen types are designated with roman numerals (I–XXVIII) in the chronological order of their discovery and the individual genetically different chains for each type are named with the letter α and an arabic numeral. Thus, the cartilage-predominant type II collagen is a homotrimer of three α_1 (II) chains, whereas the bone-

predominant type I collagen is a heterotrimer made of two α_1 (I) chains and one α_2 (I) chain [1–5].

Collagen types are broadly classified by their function, domain architecture and supramolecular organization. The main fibril-forming collagens (I, II, III, V and XI) account for 80–90% of the human collagens and are the principal source of tensile strength in animal skin, bones, cartilage, blood vessels, etc. Type IV collagen forms chicken-wire networks in basement membranes. Other collagen types form hexagonal lattices (VIII and X), beaded filaments (VI), anchoring fibrils (VII) or have transmembrane domains [1–5]. A number of additional proteins that share structural characteristics with the 28 collagen types are collectively classified as ‘collagen-like’ proteins rather than ‘collagens’, for not entirely clear reasons. The human collagenome includes collagen-like proteins such as acetylcholinesterase [11], adiponectin [12], collectins [13], C1q [14], class A scavenger receptors [15] and others [4].

All collagens and collagen-like proteins share a structural motif that defines the entire superfamily: the collagen triple helix. Each collagen has one or more occurrences of this particular triple helical domain (hereafter referred to as COL \times 3 domain), and its length varies across different collagen types. The mature form of the major fibrillar collagens is essentially a long COL \times 3 domain of just over 1000 amino acids and about 300 nm in length. In other collagen types, however, the COL \times 3 domain represents less than 10% of the number of amino acids of the mature protein [4]. Triple helical COL \times 3 domains are easily recognizable at the genomic level by a conspicuous –Gly-X-Y– repetitive sequence in which every third amino acid position is occupied by a glycine residue (Gly, G) and the X and Y positions are often occupied by

Abbreviations: Amp, aminoproline; CNA, collagen-binding protein from *Staphylococcus aureus*; COL \times 3, collagen triple helical domain; DDR2, discoidin domain receptor 2; GP0, GP1, GP2, triple helical steps with zero, one or two imino acids; Hsp47, heat-shock protein 47; Hyp/Hyp^{4R}/O, 4(R)-hydroxyproline; Hyp^{3S}/O^{3S}, 3(S)-hydroxyproline; Mse, σ M, selenomethionine; P3H, prolyl-3-hydroxylase; P4H, prolyl-4-hydroxylase; PGII, polyglycine II; PPII, polyproline II; Scl1/Scl2, *Streptococcus* collagen-like protein 1 and 2; T_m , melting (denaturation) temperature; κ , superhelical twist; τ , superhelical height.

¹ email jordi.bella@manchester.ac.uk

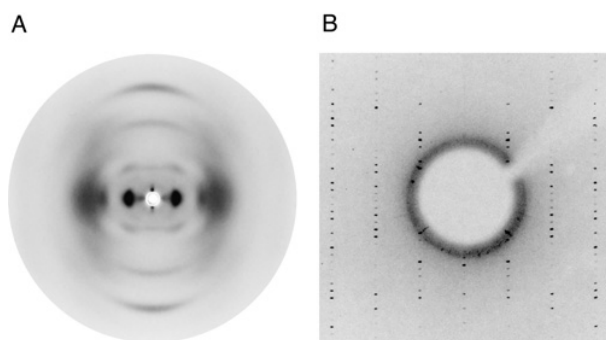


Figure 1 Examples of collagen X-ray diffraction

(A) Fibre diffraction photograph of dry adult rat tail tendon [26]. The vertical axis corresponds to the main direction of the aligned collagen molecules in the tendon. The strong outermost meridional reflection at 2.86 Å corresponds to the vertical separation of the repeating units in the collagen triple helix. The strongest equatorial reflection at 12 Å corresponds to the lateral separation between triple helical molecules, which increases with hydration. Obtained by A.C.T. North in 1952 (King's College London Archives, reproduced with permission). (B) Detail of an upper level precession photograph showing the $hk1$ reflections from a crystal of the Gly→Ala collagen peptide [55], obtained by the author in 1991 in the laboratory of H.M. Berman, Rutgers University. The horizontal spacing of 14 Å corresponds to the lateral separation between triple helical molecules along the b axis. The vertical spacing of 173 Å corresponds to the overall length of two peptide molecules aligned with the a axis in a C -centred cell. The asymmetric unit corresponding to one triple helix is 86.5 Å long.

proline residues (Pro, P). In animals, Pro residues in the Y position are often modified post-translationally to 4-hydroxyproline (Hyp, O). These sequence requirements respond to specific structural characteristics of the collagen triple helix, as will be reviewed later.

EARLY FIBRE DIFFRACTION STUDIES: DISCOVERY OF THE COLLAGEN TRIPLE HELIX

Tendons are primarily built from type I collagen, which represents 70–80% of their dry weight. Furthermore, the structure of tendon is such that the collagen fibres are essentially aligned to the long axis of the tendon [5]. This structural alignment is critical to provide the tensile strength required to connect bone to muscle while allowing for the mechanical mobility of the body [16]. Thus, the earliest structural information on collagen was obtained from fibre X-ray diffraction photographs of stretched tendons [17]. Given the sparse amount of information provided by these images (Figure 1A), it is remarkable that an essentially correct structure was eventually obtained after a few unsuccessful attempts at model building (see [18,19] for detailed historical accounts). Early models were based on single polypeptide chains containing *cis* or both *cis* and *trans* peptide bonds [17,20]. In 1951, Pauling and Corey [21] proposed a model for collagen as part of their classic series of papers on the stereochemistry of polypeptide structures. From density measurements and average residue weight, they concluded that the molecule of collagen consisted of three polypeptide chains. A triple helical structure with all peptide bonds *in trans* was first proposed in 1954 by Ramachandran and Kartha [22]. The same authors soon modified this model [23] by applying the concept of 'coiled-coil', where the three chains were wound around the common central axis, and introducing a vertical staggering following a superhelical path. The stereochemistry of this structural model (which became known as the 'Madras triple helix') was then refined by Rich and Crick [24,25] and Cowan et al. [26]. In parallel, physicochemical studies with soluble collagen confirmed that collagen molecules were rather rigid and

rod-like, with a length of 3000 Å ($1 \text{ Å} = 0.1 \text{ nm}$) and a diameter of 13.6 Å, and composed of three polypeptide chains which became separated upon temperature increase [18,27–29].

The models obtained from the combination of fibre diffraction, amino acid composition and physicochemical data already described the essential features of the collagen triple helix as we understand it today. The discovery had its controversy, in particular about the recognition of Ramachandran's contribution to the elucidation of the correct structure [30–32]. An interesting account of the historical events is given in Ramachandran's biography by Sarma [33]. Remarkably, the discussions about the stereochemistry of the collagen models led eventually to the publication of the Ramachandran map [34].

AVERAGE STRUCTURE OF THE COLLAGEN TRIPLE HELIX

Tendons used in fibre diffraction experiments have a degree of heterogeneity. Their oriented fibres can at best be classified as polycrystalline and more realistically as non-crystalline due to the large amount of disorder in their structure [35]. Thus, the models obtained from fibre diffraction were necessarily average structures of the repeating unit of the collagen triple helix [19]. Linked-atom least-squares refinement of early models against fibre diffraction data from highly stretched partially dehydrated kangaroo tail tendon produced probably the best average structure at the time [36].

The overall structure of the collagen molecule is a right-handed triple helix of three individual polypeptide strands, each in a left-handed helical conformation with three residues per turn and 3_2 helical symmetry known as polyproline II or polyglycine II (PPII/PGII) [37–40]. The three strands are supercoiled around each other in a right-handed manner, and a ladder of intermolecular backbone $N-H \cdots O=C$ hydrogen bonds links adjacent strands (Figure 2). These hydrogen bonds are transversal to the helical axis, as opposite to the ubiquitous α -helices where hydrogen bond directions are roughly parallel to the helical axis. To form this triple helical assembly two things need to happen: the three strands must be staggered by one residue with respect to each other (an approximate rise of 2.9 Å in the direction of the helical axis), and every third residue in each strand must be placed near the common helical axis, due to the resulting close packing of the three strands. This can only be achieved if the smallest amino acid Gly occurs at that position, which explains the repetitive $-Gly-X-Y-$ sequence seen in the primary structure of COL $\times 3$ domains. The one-residue-staggering in the direction of the axis has also important consequences as the three strands become topologically non-equivalent. This contrasts with the situation in, for instance, trimeric α -helical coiled coils, where the three strands are at the same axial level. A useful notation to distinguish the three strands as *trailing*, *intermediate* and *leading* (Figure 2) was introduced much later when it became important to describe the different interactions between each of the three collagen strands and an integrin receptor domain [41].

It is illustrative to stop for a moment on the PGII structure to visualize the intimate relation between collagen structure and sequence. The PGII/PPII conformation has 3-fold helical symmetry (3_1 or 3_2 for an achiral poly(G) sequence, oblique 3_2 for any other protein sequence). This symmetry places peptide bonds in three directions separated by 120° . Each polyglycine chain in the PGII structure is connected to its six closest neighbours via hydrogen bonds roughly perpendicular to the helix axis, resulting in a continuously connected hexagonal array [38]. This continuous structure, also observed in synthetic copolymers of Gly residues

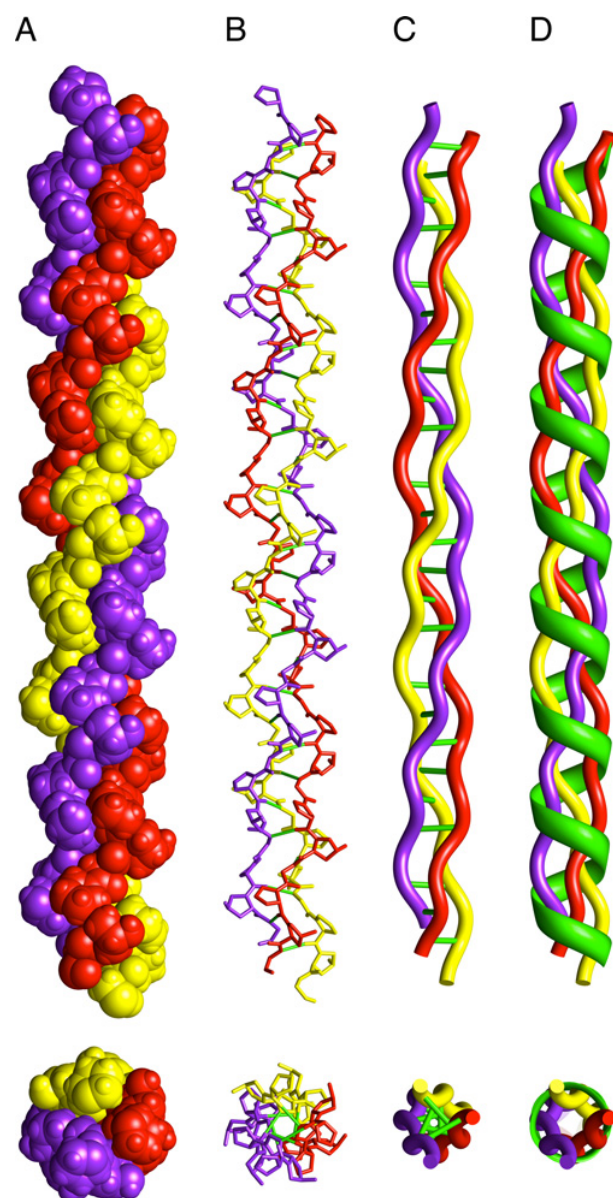


Figure 2 Different representations of a collagen triple helix with sequence $(\text{POG})_{10}$

The three polypeptide chains are shown in purple (trailing strand), red (middle strand) and yellow (leading strand), and each panel shows a side view and top view of the same representation. Strands are named here according to their vertical position in the direction from the N-terminus (top of the helix) to the C-terminus (bottom of the helix), although in some reports the opposite convention has been used. **(A)** Space-filling diagram, with the right-handed supercoil of each individual strand clearly visible. **(B)** Stick diagram where hydrogen atoms have been removed for clarity, showing hydrogen bonds as green sticks connecting the strands. The hydrogen bonding topology between Gly-N—H groups and Pro (X)-C=O groups follows a helical path: from trailing strand to middle strand, from middle strand to leading strand and from leading strand to one step (triplet) ahead on trailing strand (*TMLT* direction). **(C)** Ribbon diagram showing the ladder of hydrogen bonds perpendicular to the helix axis, following their own helical path and connecting alternate pairs of strands at each step. **(D)** Ribbon diagram including the left-handed superhelix (green), which describes mathematically the collagen triple helix as a continuous helix of repeating units across the three individual strands. The structural model shown here is an idealized 7_5 triple helix with the parameters shown in Table 1. Co-ordinates are available in the Supplementary Online Data (file *POG10_75.pdb*). Image prepared with UCSF Chimera [253].

with ω -amino acids [42,43], would be hardly compatible with a biological macromolecule that needs to avoid aggregation. In the consensus collagen sequence, only every third residue is Gly, which means that an individual chain can act as hydrogen bonding donor in only one direction, whereas it can still receive in any of the three directions (imino acids cannot donate hydrogen bonds and any side chain other than Gly would interfere with hydrogen bonding in that particular direction). This hydrogen bonding topology allows for the formation of a self-contained trimeric assembly, which ultimately is the basis for the triple helical structure of collagen. Interestingly, the constraints imposed by this self-contained structure, together with the high content of imino acids and Gly residues, prevent fibrillar collagen molecules from forming continuous harmful aggregate structures such as amyloid fibrils [44,45].

HIGH-RESOLUTION STRUCTURES FROM COLLAGEN MODEL PEPTIDES

The average structure of collagen was reviewed periodically without increasing significantly its level of detail. Synthetic polypeptides with a collagen-like structure, such as poly(PGP) or poly(GPO), did not improve the information obtained from native sources [19,46]. Collagen in general was considered 'non-crystallizable' and it was not until the introduction of solid-phase synthesis of collagen model peptides that this perception changed. Physicochemical analysis confirmed that these collagen peptides were trimeric in dilute aqueous solution, formed triple helical structures in which the three chains were parallel and in register, and showed sharp thermal transitions corresponding to the denaturation of the triple helices [47–53]. Chemically synthesized collagen peptides such as $(\text{PPG})_{10}$ or $(\text{POG})_{10}$ are homogeneous in molecular mass, have defined length and chemical composition, and thus are amenable to producing single crystals (Figure 1B). To date, more than 60 structures of collagen model peptides on their own or in complex with other proteins have been deposited in the PDB (Supplementary Table S1) with atomic or near-atomic resolution. These structures have confirmed the general features of the triple helical structure derived from fibre diffraction of tendons (with some important differences discussed later), and have provided a wealth of data on the conformational variability of the collagen triple helix, molecular details of the interaction between collagen and water or effects of collagen interruptions. They have also helped to clarify the role of Hyp and other residue types in collagen stability and resolved a number of long-on-going controversies about the topology of hydrogen bonding in the collagen triple helix.

The first successful crystals and high-resolution X-ray diffraction data were obtained by Okuyama et al. [50,54] with the $(\text{PPG})_{10}$ peptide. The highly repetitive structure of this peptide and a columnar arrangement of the molecules in the unit cell prevented a complete determination of the entire triple helix. Thus, an average atomic structure for the repeating unit of an infinite helix model was determined. The first full-length crystal structure was obtained with the Gly \rightarrow Ala peptide $(\text{POG})_4$ -POA- $(\text{POG})_5$ (PDB code 1CAG) [55]. In this peptide, the change of one central Gly residue to alanine (Ala, A) was sufficient to break the columnar arrangement seen in the $(\text{PPG})_{10}$ structure and facilitated structural determination. Better quality average structures for $(\text{PPG})_{10}$ and $(\text{POG})_{10/11}$ were later obtained (Supplementary Table S1), and the problem of highly repetitive sequences was eventually solved with very-high-resolution crystals for the peptides $(\text{PPG})_9$ (PDB code 3AH9), $(\text{PPG})_{10}$ (PDB code 1K6F) and $(\text{POG})_9$ (PDB code 3B0S) [56–58]. In the meantime, collagen model peptides

Table 1 Average helical parameters of different models for the collagen triple helix

The individual chains are right-handed, the superhelix is left-handed. *Average unit twist approximated to a rational helix with integer coefficients. †Vertical rise per turn. ‡Following the crystallographic notation (right-hand rule) for left-handed screw axes. §The relationship between the superhelical symmetry P_0 and that of the individual chains p_0 follows the relationship $3\frac{p_0}{P_0} - \frac{p_0}{P_0} = 2$ [252].

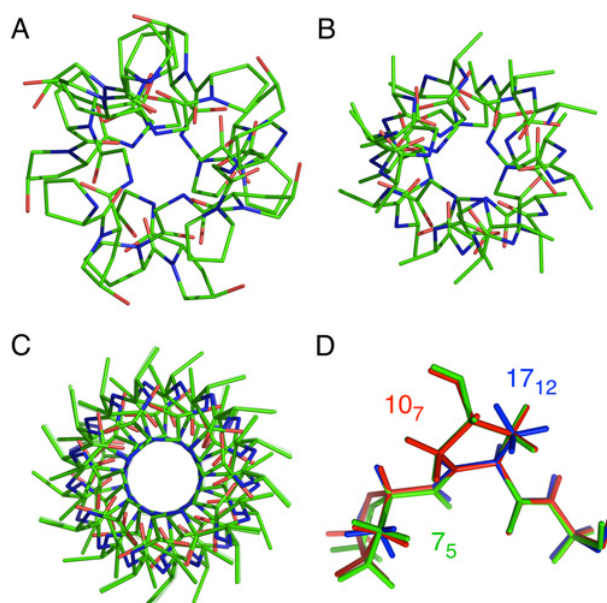
	10/3 helix	7/2 helix	Variable helix*
Individual chains			
Symmetry (p_0)	10_1	7_1	$\sim 17_2$
Unit twist (t)	36°	51.43°	42.4°
Unit height (h)	8.95 Å	8.61 Å	8.64 Å
Helical pitch†	89.5 Å	60.3 Å	~ 73.4 Å
Triplets per turn (360/ t)	10	7	~ 8.5
Residues per turn	30	21	~ 25.5
Superhelix			
Symmetry (P_0)‡,§	10_7	7_5	$\sim 17_{12}$
Unit twist (κ)	-108°	-102.86°	-106°
Unit height (τ)	2.98 Å	2.87 Å	2.88 Å
Helical repeat	29.8 Å	20.1 Å	~ 49.6 Å
Units per turn (360/ κ)	3.33	3.50	3.40
Reference	[36]	[54]	[66]

containing amino acids other than Pro/Hyp were crystallized and information about side-chain conformation, water-mediated hydrogen bonding or conformational variability was finally obtained.

SEQUENCE DEPENDENCE OF COLLAGEN HELICAL SYMMETRY

Fibre diffraction images of tendons were interpreted with average triple helical collagen structures (all repeating units were considered to have the same conformation and to be related by some helical symmetry operator). One of the helical parameters measured experimentally from fibre diffraction diagrams is the magnitude of the unit twist $|\kappa|$ (Table 1) and values between 107° (3.36 units/turn) and 110° (3.27 units/turn) were reported [36,46,59]. In order to describe the helical symmetry as a ratio of small integers, a rounded-off value of ten repeating units in three turns ($10/3$, 10, or 3.33 units per turn) was adopted, corresponding to a magnitude of the unit twist $|\kappa| = 108^\circ$ [25,36,46]. The different helical parameters for the individual strands and superhelix of the 10_7 (or $10/3$) model are given in Table 1; co-ordinates of a 10_7 molecular model obtained from fibre diffraction are available in PDB format in the Supplementary Online Data (file Fraser1979.pdb) [36].

However, the X-ray diffraction of $(PPG)_{10}$ crystals was clearly indicative of a tighter symmetry, with 7 units in two turns ($7/2$, 7_5 or 3.5 units per turn) [50,54] (helical parameters in Table 1). Both 10_7 and 7_5 symmetries were compatible with the fibre diffraction data and this observation led to Okuyama et al. [60–62] to propose a new structural model for collagen based on the 7_5 (or $7/2$) helix. Many of the crystal structures of collagen model peptides determined to date show 7_5 helical symmetry, particularly those with an amino acid sequence predominantly or entirely made of imino acids (Pro, Hyp and alternative modifications of Pro; collagen peptides with this type of sequence are referred to here as ‘imino-saturated’). However, the crystal structure of a peptide with a sequence of human type III collagen $(POG)_3$ -ITGARGLAG- $(POG)_4$ (PDB code 1BKV, peptide T3-785) shows distinct helical symmetry for the imino-saturated $(POG)_n$ zones (7_5) and the central, imino acid-free ITGARGLAG zone (close to 10_7) (Figure 3) [63,64]. Similarly the crystal

**Figure 3** Helical twist in the structure of collagen

(A) Top view of the crystal structure of the T3-785 peptide, showing the first three POG triplets on each chain. The side chains of the Pro and Hyp amino acids are shown. The unit twist κ in this region averages -102° and the 7-fold symmetry of a 7_5 helix is obvious. (B) Top view of the central region of the T3-785 peptide with sequence ITGARGLAG. All side chains have been truncated to the C β . The unit twist κ in this region averages -107° and the helix approaches locally the 10-fold symmetry of a 10_7 helix. (C) Triple helical model of a $(AAG)_{10}$ sequence with a constant unit twist $\kappa = -106^\circ$. This value results in a helix with approximately 17-fold symmetry where each individual chain is right handed with 17 triplets in two turns (17_2) and the superhelix is left handed with 17 triplets in five turns (17_{12}) (Table 1). Co-ordinates of this model in PDB format are available in the Supplementary Online Data (file Helix1712.pdb). (D) Unit triplets from a 7_5 helix (POG, green), a 10_7 helix (POG, red) and a 17_{12} helix (AAG, blue) superimposed at the central residue. Image prepared with PyMol (<http://www.pymol.org>).

structure of a peptide with a much longer type III collagen sequence GPIGPOGPRGNRGERGSEGSGOHOG σ MO-(GPO) $_2$ -AOGPCCGG (PDB code 3DMW, peptide T3-991, σ M selenomethionine) shows that the central imino acid-free zone and the imino acid-rich flanks are close to 10_7 and 7_5 helices respectively [65].

The differences between a 7_5 and a 10_7 helix are minimal at the level of one individual three-amino-acid-repeating unit (Figure 3D). However, they become important in the context of very long COL \times 3 domains such as those in fibrillar collagens. Thus, it is important to obtain a reasonably accurate figure for the average unit twist of a COL \times 3 domain, so that the conformation of the entire domain is adequately represented. Clarification of the sequence-dependence of collagen helical symmetry has been obtained using a novel method in which the repeating unit of the collagen triple helix is defined as a triplet of residues, each from a different chain, sitting approximately at the same vertical level [66]. The unit twist κ is obtained from the helical operation that relates two consecutive such triplets. Variations of local conformation in the high-resolution crystal structures of collagen peptides were visualized by plotting κ against their trimeric sequence (Figure 4) [66]. This analysis showed that imino-saturated zones have an average κ of -103° (corresponding to a 7_5 helix), whereas imino acid-free zones approach an average κ of -108° (corresponding to a 10_7 helix). Similarly, the average unit height τ increases from 2.84 Å in imino-saturated zones to 2.90 Å in imino acid-free zones. Interestingly, zones with intermediate

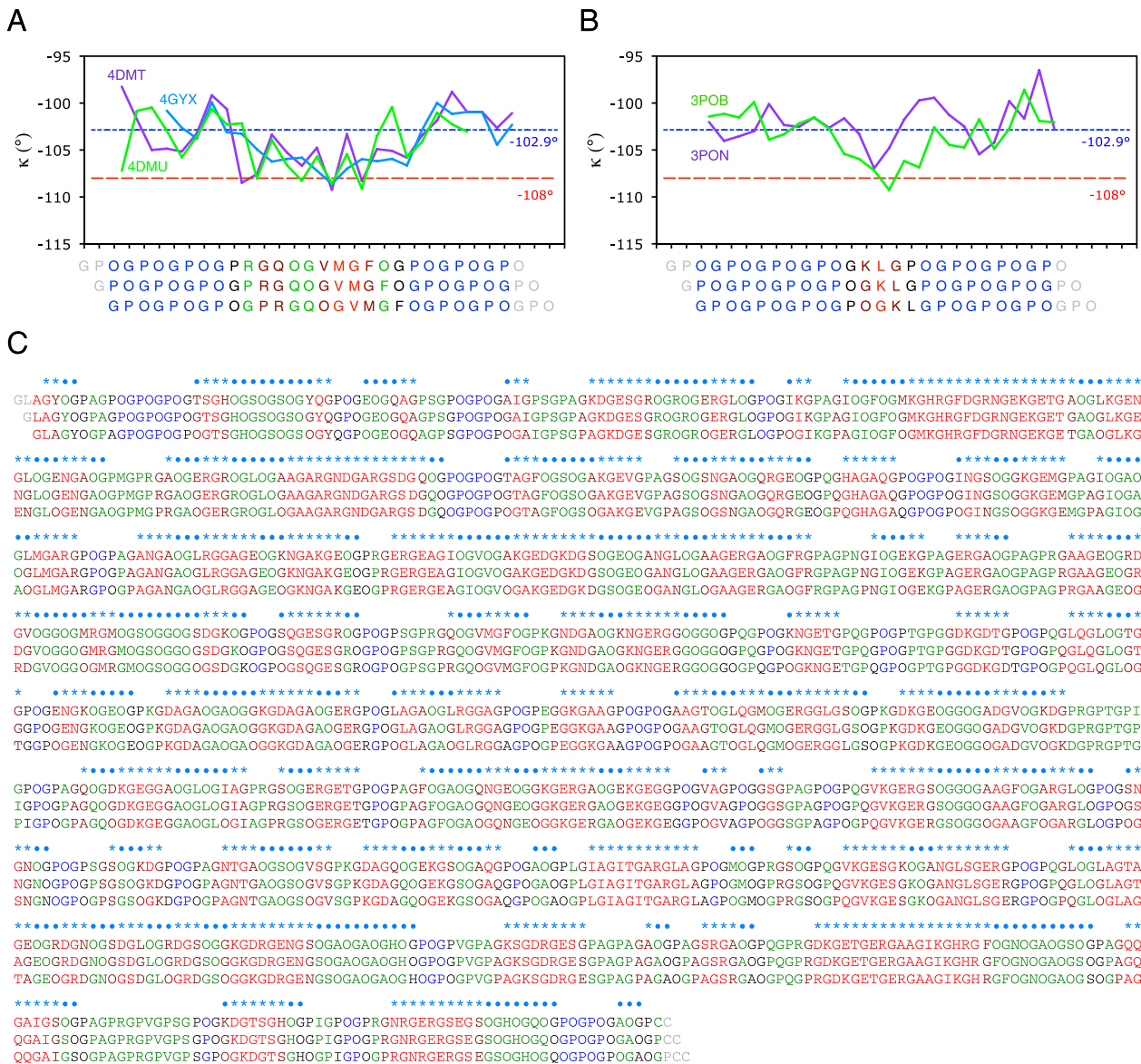


Figure 4 Variation of the unit twist κ as a function of the collagen amino acid sequence

(A) Analysis of unit twist using the method described in [66] for the structures of the peptides $(\text{GPO})_{2/3}\text{-GPRGQGVGMFO-(GPO)}_{2/3}$ as determined by different groups (PDB codes 4DMT and 4GYX), and the complex of one these peptides with von Willebrand factor A3 domain (PDB code 4DMU) [90,181]. For all three structures, the average unit twist, κ , calculated over the $(\text{GPO})_{2/3}$ zones is -103° , and over the GPRGQGVGMFO sequence it is -106° . The average twists and the overall pattern of variation are essentially the same for both peptide structures and the complex, suggesting that there is no distortion over complex formation. (B) Identical analysis for the structure of the peptide $(\text{GPO})_4\text{-GKL-(GPO)}_4$ (PDB code 3PON) and its complex with the CUB2 domain of MASP-1 (PDB code 3POB). The average unit twist, κ , over the central GKL triplets are -105° for the peptide alone and -108° for the complex. These average values and the pattern of variation in both structures suggest a degree of untwisting of the triple helix over complex formation. (C) Predicted variation of the unit helical twist (colour-coded sequence) and possible sites of water-mediated hydrogen bonding (\bullet , \ast) in the COL \times 3 domain of human type III collagen (UniProt P02461). Triplets are named as GP2, GP1 or GP0 depending on their number of imino acids (2, 1 or 0). Individual helical steps between consecutive triplets are predicted to vary their unit twist κ depending on the number of imino acids in both triplets (see [66] for a description of the method). Thus, imino-saturated zones with GP2-GP2 steps (in blue) are expected to average $\kappa = -102.6^\circ$, whereas imino acid-free stretches with GP0-GP0 steps (in red) are expected to average $\kappa = -108^\circ$. Other zones with GP1-GP1 steps (green), GP2-GP1/GP1-GP2 (black) and GP1-GP0/GP0-GP1 (dark red) are expected to show intermediate values of κ . An average value of $\kappa = -106^\circ$ is obtained for the entire COL \times 3 domain of type III collagen. Possible sites of inter-strand ζ 1 (\ast) or ζ 1- γ 1 (\bullet) water bridges (433 and 307 sites respectively) arise when a non-imino acid occupies the X position on any of the three chains (see text). In ζ 1- γ 1 sites (\bullet), Hyp residues are likely to provide additional hydrogen bonding to the bound water molecules (as in Figure 5A). Sites of Pro hydroxylation are indicated as O residues, but the degree of hydroxylation is never 100% and will vary across the different positions.

imino acid content show average κ and τ values in between those corresponding to the 7_5 and 10_7 helices [66].

It has been suggested that the unit twist κ between two consecutive triplets changes gradually through the COL \times 3 domains and is related to the number of imino acids in these two

triplets. Analysis of collagen sequences in terms of the expected values of κ for every pair of consecutive triplets predicts an average value $\kappa = -106^\circ$ for the COL \times 3 domains of fibrillar collagens I, II or III, and similar values can be obtained for uninterrupted COL \times 3 domains in other collagens (Figure 4C)

[66]. This average value of κ corresponds to a helical symmetry intermediate between the 10_7 and 7_5 helices, and reflects the fact that in many collagens the most common pairs of consecutive triplets contain zero to two imino acids (GP0-GP0, GP0-GP1, GP1-GP1), whereas imino-acid rich pairs (three or four imino acids, GP1-GP2, GP2-GP2) are less frequent (Figure 4). There are no simple rational helices corresponding to a collagen triple helix with $\kappa = -106^\circ$. Its average conformation can be approximately described by helices with exotic integer combinations such as 17_{12} : a left-handed superhelix of 17 units in five turns made of three supercoiled right-handed 17_2 helices (Table 1) [66].

There is no question that imino-saturated collagen peptides show a 7_5 structure [58], and there is consensus that the conformational restraints from imino acids favour such tight helix, with 3.5 residues per turn. When the proportion of non-imino acids increases, the conformation of collagen peptides relaxes into a less tightly-wound helix which approaches the 3.33 residues per turn. On average, many COL \times 3 domains will have an intermediate number of residues per turn. Thus, if a monotonous helix is used to represent the entire conformation of a COL \times 3 domain, a helical model matching its average κ should be considered rather than a pure 7_5 or 10_7 model [66]. In the absence of high-resolution structures of collagen fibrils or native long collagen molecules, it is still unknown whether these predictions are accurate. Nevertheless, Orgel et al. [67] conclude from their analysis of fibre X-ray diffraction data that different triple helical symmetries occur along type I and II collagen molecules in their native (fibrillar tissue) environment. They discuss that other factors such as local helix dissociation or molecular packing must have an effect on the local symmetry of the collagen triple helix, in addition to any intrinsic sequence-related preferences towards a 10_7 or 7_5 symmetry.

COLLAGEN HYDROGEN BONDING AND THE WATER-BONDED STRUCTURE

The pattern of hydrogen bonding between collagen strands differs from that of common secondary structure elements, such as α -helices or β -sheets, in that not all peptide bonds from the repetitive –Gly-X-Y– polypeptide chain can form main-chain to main-chain hydrogen bonds. Early fibre diffraction models already deduced the correct interstrand N–H...O=C connectivity where the N–H groups of the Gly residues act as donors and the C=O groups of the residues in the X position of the following strand act as acceptors (so-called Rich and Crick II hydrogen bonding topology) [25,46]. This topology means that only one interstrand hydrogen bond can form per Gly-X-Y triplet. Structurally, these interstrand hydrogen bonds form a ladder along the triple helix where the steps change orientation, following a helical path as they connect different pairs of strands in turn (Figure 2).

Nevertheless, hydrogen exchange experiments on collagen suggested that only one of the three amide hydrogens per triplet exchanged at the high rates expected for a peptide group freely exposed to the surrounding solvent. The remaining two amide hydrogens exchanged much more slowly, although they could be distinguished in two groups. The first group (one amide hydrogen per triplet) showed very low rates of exchange that changed little upon temperature increase. The second group (about 0.7 hydrogens per triplet) showed low rates of exchange that increased rapidly with the increase in temperature [68–70]. These observations suggested that about 1.7 amide groups per every three residues in collagen were involved in hydrogen bonding, but the two interactions were not equivalent. Deformation of the triple helix to account for a second hydrogen bond per triplet is not possible due to stereochemical reasons. Thus, in order to resolve

the dilemma of the two slow amide hydrogens, Ramachandran and Chandrasekharan [71] proposed that when the residue in the X position was not an imino acid, a second ‘hydrogen bond’ could be formed *through* a water molecule between the N–H group of the amino acid in the X position and the C=O group of the Gly residue on the *previous* strand. Therefore, this water-bridged hydrogen bonding would reverse that of the direct interstrand hydrogen bonds. Ramachandran et al. [72] also noted a few years later that the hydroxy groups of Hyp residues could form an additional hydrogen bond to the bridging water molecule.

Confirmation of this ‘water-bonded’ model for collagen was not really possible from fibre diffraction data alone, but water molecules around collagen became clearly visible in the high-resolution crystal structures of collagen model peptides. The first evidence of water-mediated hydrogen bonding between collagen strands was observed in the Gly \rightarrow Ala peptide, where the disruption introduced by the Gly substitution was overcome through water molecules connecting the three strands around the sites of interruption [55,73]. Further structures of the T3-875, EKG (PDB code 1QSU) [74] and other collagen model peptides confirmed the existence of single water molecules connecting the strands of uninterrupted collagen triple helices whenever a free N–H group occurs at the amino acid in the X position, in the manner postulated in the water-bonded model for collagen (Figure 5). These water bridges will be referred to here as $\zeta 1$ bridges, following a naming scheme suggested for the different types of collagen-water hydrogen bonding topology [73]. Water molecules in $\zeta 1$ bridges often appear as particularly well-defined spheres in the electron density maps (Figure 5A) and have crystallographic temperature factors similar to those of the surrounding atoms from the peptide chains. Some $\zeta 1$ bridges are stabilized by additional hydrogen bonding from the hydroxy group of Hyp residues ($\zeta 1-\gamma 1$ bridges, in the naming scheme above). For this to occur the Hyp residues must be placed two positions C-terminal to the C=O acceptor group (Figure 5). In $\zeta 1-\gamma 1$ bridges, the water-binding sites created by the three groups on the collagen triple helix have a remarkably ideal geometry, with the water molecules adopting a tetrahedral co-ordination and with the three hydrogen bonds showing ideal distances and angles. The fourth co-ordination position is usually occupied by another water molecule.

Most high-resolution crystal structures of collagen model peptides to date are disproportionately rich in imino acids, many with imino-saturated stretches at both ends. This is mainly for designer reasons, to ensure the stability of relatively short triple helices. However, a few structures containing non-imino acids in the X positions have been determined (Supplementary Table S1). In peptide structures with resolution 2.0 Å or better, there are 56 observed $\zeta 1$ bridges out of 66 possible (85%). Of the remaining ten cases, there is one unusual two-water $\zeta 2$ bridge and the other positions are disrupted or occupied by other side chains or lattice interactions. Of the water molecules involved in $\zeta 1$ bridges, 38 form hydrogen bonds with hydroxy groups of Hyp residues ($\zeta 1-\gamma 1$ bridges) out of 39 possible (97%). The T3-785 crystal structure also shows two instances of Thr hydroxy groups playing similar roles to the Hyp ones in $\gamma 1$ bridges.

NMR studies of the T3-785 and (POG) $_{10}$ peptides where specific positions had been enriched with ^{15}N -labelled amide groups showed that the Gly and Leu (X position) amide protons exchanged very slowly and slowly respectively, whereas the Ala (Y position) amide protons exchanged rapidly with the solvent. The Gly amide protons at the centre of the (POG) $_{10}$ peptide exchanged even more slowly than the corresponding Gly amide protons in the middle of the T3-785 peptide [75]. The rates of exchange observed on these peptides were consistent with the

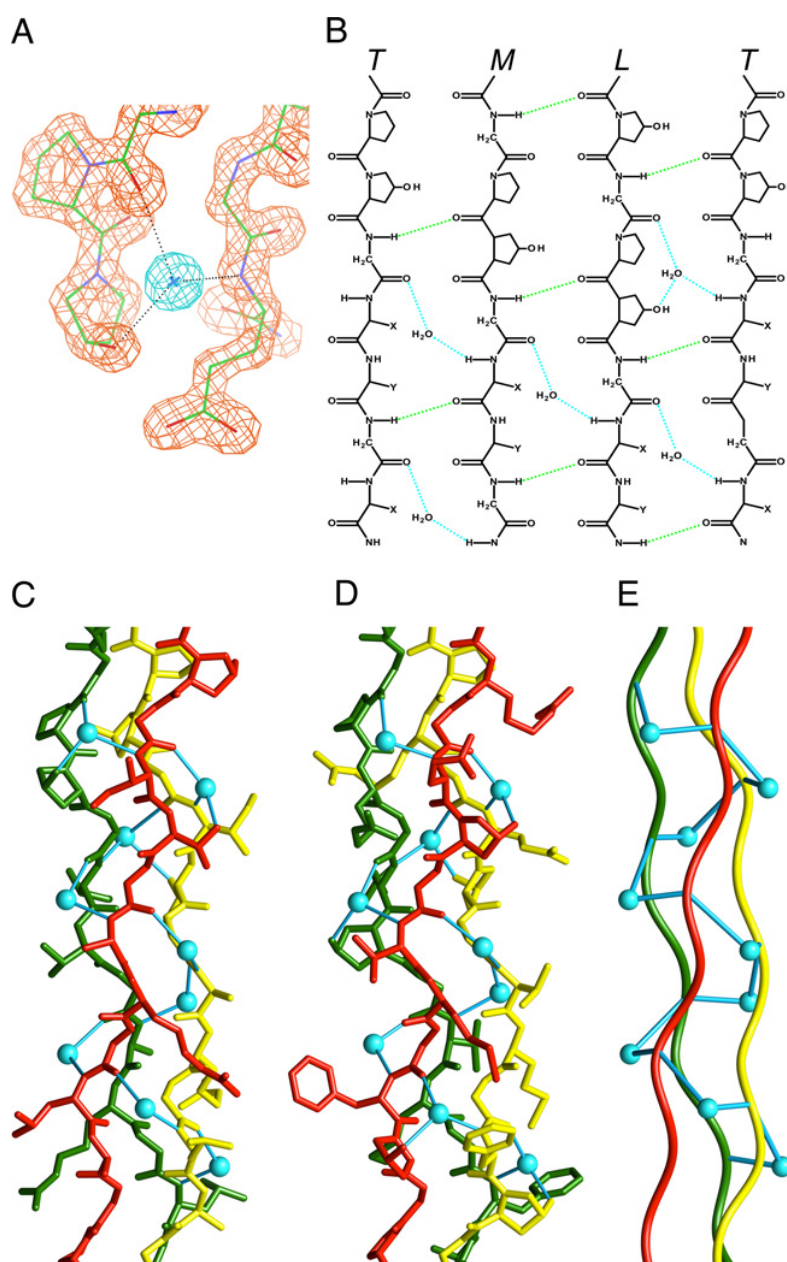


Figure 5 Water-mediated hydrogen bonding in collagen

(A) High-resolution crystal structures of collagen model peptides show clear evidence of water molecules (cyan sphere density) connecting two peptide groups on different strands (orange density) through hydrogen bonding (dotted lines). In this example the hydroxy group of a Hyp residue provides an additional hydrogen bond to the bridging water. The electron density map is from the crystal structure of the EKG peptide (PDB code 1QSU) [74]. (B) Hydrogen bonding topology of the water-bonded collagen model, illustrated for a general sequence with both imino and amino acids. Chains are labelled as *trailing* (T), *middle* (M) and *leading* (L), with the T chain repeated on the right to show the interactions more clearly. Direct interstrand hydrogen bonds are shown in green, water-mediated hydrogen bonds are shown in cyan. The two sets have opposite directionality (TMLT for direct hydrogen bonds, TLMT for water-mediated ones). Due to the one-residue-staggering, the local environments of the water molecules may differ even in a homotrimer structure. (C) Detail of the crystal structure of the T3-785 peptide showing water molecules as cyan spheres connecting two strands via hydrogen bonding (cyan sticks). (D) Detail of the crystal structure of the von Willebrand factor-binding T3-403 peptide (PDB code 4DMT) [90] showing an identical position of the water-bridging molecules despite the obvious differences in amino acid sequence. (E) Ribbon diagram showing schematically the three strands and the water bridges that interconnect them. Interstrand N—H...O=C hydrogen bonds are not shown. The ζ 1 water bridges (see text) build a continuous helix of hydrogen bonds, which runs left-handed around the three strands interconnecting them. Images in (C–E) were prepared with UCSF Chimera [253].

classic experiments of hydrogen exchange on native collagen. The combination of the NMR data and the X-ray structural information of the T3-785 peptide suggests that the water-mediated hydrogen bonding in the X position slows the hydrogen exchange almost in the same manner as a direct hydrogen bond. It is therefore

likely that these water-mediated hydrogen bonds contribute to the local stability of the triple helix and help to maintain the triple helical conformation in regions where there are no imino acids. Even with this second set of hydrogen bonds, regions of triple helix conformation without imino acids will be more

flexible and dynamic than regions where all X and Y positions are occupied by Pro and Hyp residues respectively. Water-mediated hydrogen bonds should be less resistant towards increases in temperature than direct interstrand hydrogen bonds. This would agree with the classic observations of the effect of temperature on collagen hydrogen exchange and the different behaviour of the two populations of slow-exchanging protons [68,70].

Figure 4 also shows all possible positions for water-mediated hydrogen bonding ($\zeta 1$ and $\zeta 1-\gamma 1$ bridges) on the sequence of the COL $\times 3$ domain of type III collagen. There are a total of 740 possible such positions, with some very long stretches where there are no imino acids in the X position. Thus, the triple helical structure with two hydrogen bonding connections per triplet (one mediated through water) is potentially far more common than the structure with strictly one hydrogen bond per triplet (assuming that most of these water bridges do actually form). The mature type III collagen protein is a trimer of three $\alpha_1(\text{III})$ chains, each with 1068 amino acids (after cleavage of the N- and C-terminal propeptides, UniProt P02461). Thus, the number of possible water-mediated hydrogen bonds is equivalent to 0.69 per every three amino acids of mature type III collagen, which would be consistent with the 0.7 amide protons per triplet that have been shown to exchange slowly with the solvent, but not as slowly as the one amide proton per triplet involved in direct interstrand hydrogen bonding.

HIGHLY STRUCTURED HYDRATION NETWORKS AROUND COLLAGEN MODEL PEPTIDES

Water has always been considered an intrinsic component of collagen with a role in maintaining the conformation of the native collagen molecule. Tendons contain tightly bound water, and dehydration increases their mechanical stiffness. Early X-ray fibre diffraction experiments showed evidence of wide-range structural changes when tendons were dehydrated [18,19,76]. These findings have been corroborated through the years. A recent study has shown that water removal from tendons shortens the collagen molecules and fibrils and this shortening translates into tensile forces much larger than these achievable from muscle contraction alone [77]. Several studies on oriented hydrated tendons have been conducted using a variety of techniques (Raman and infrared spectroscopies, calorimetric, dielectric measurements, dynamic mechanical spectroscopy and NMR). They have provided evidence for ordering of water molecules in collagen fibrils that differs from bulk water [16,78–85]. Spectroscopic analyses suggest different groups of water molecules, ranging from strongly bound waters with correlation times ≥ 1 ns to relatively free waters with rotational correlation times on the 10^{-10} s scale [81,82,85]. Before high-resolution structures became available, the groups of water molecules were interpreted on the basis of the water-bonded model proposed from fibre diffraction studies, plus additional hydration layers around the collagen molecules.

The crystal structure of the Gly \rightarrow Ala peptide showed an extensive network of ordered water molecules surrounding the triple helical structure (PDB code 1CGD) [73]. Many of these water molecules were located in positions consistent with hydrogen bonding to the C=O groups of the main chain and the OH groups of the Hyp residue side chains, as indicated by distances and angles between the atoms visible in the crystal structures. Most water molecules showed additional hydrogen bonding (on a stereochemical basis) to other water molecules, which in turn were connected to other waters and eventually back to the appropriate groups on the same peptide or on a different peptide on the crystalline lattice. An elaborate network of bridges

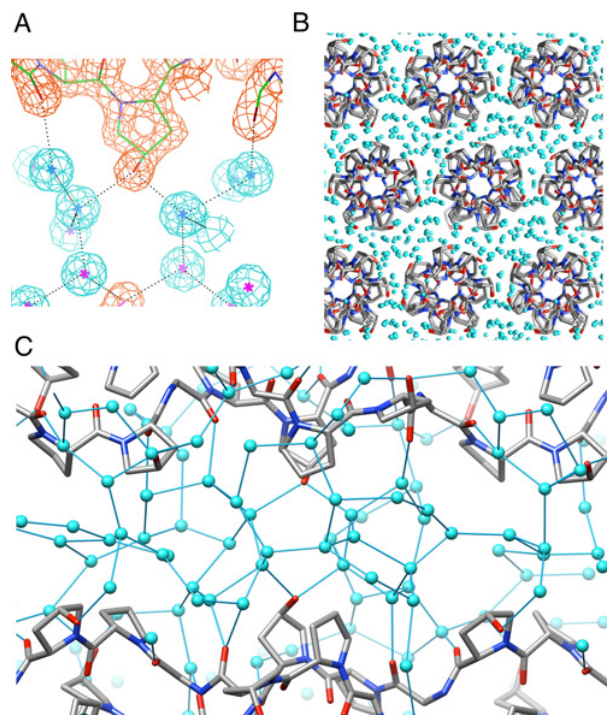


Figure 6 Ordered water networks between neighbouring triple helices in crystal structures of collagen peptides

(A) Water molecules (cyan sphere density) occupy the space between triple helices (orange density) and interconnect them through hydrogen bonding. The electron density map is from the crystal structure of the EKG peptide (PDB code 1QSU) [74]. (B) Lateral packing of triple helices in the crystal structure of the (POG) $_9$ peptide (PDB code 3B0S) [58]. Water molecules occupy the space between triple helices, which have no direct contact among themselves. (C) Extensive hydrogen-bonded networks of water molecules connect neighbouring triple helices in the (POG) $_9$ structure. Main-chain C=O groups and hydroxy groups of Hyp residues provide multiple anchoring points for these water networks. Images in (B) and (C) were prepared with UCSF Chimera [253].

was described and categorized and a predominance of water-bonded motifs with partial pentagonal geometries was noted. The hydroxy groups of the Hyp residues acted as linchpins for the water networks. Later structural determinations have confirmed the existence of these extensive water networks around the triple helical structures, and the same topologies of collagen-water hydrogen bonding have been observed [86,87]. In particular, some of the water bridges involving the side chains of the Hyp residues are ubiquitous in these crystal structures (Figure 6). The extent of water ordering in different structures is dependent on the particular amino acid sequence and the packing of the triple helices in the lattice, and parallel columnar arrangements of collagen model peptides with a high content of imino acids seem particularly effective at inducing large structuring of the surrounding water molecules (Figure 6).

One of the key observations is that the triple helices in the crystal lattices have hardly any direct hydrogen bonding or hydrophobic contact between their side chains (Figure 6). In most crystal structures only a few contacts between side chains extending towards the neighbouring chains are observed (Leu residues in (POG) $_4$ -(LOG) $_2$ -(POG) $_4$, PDB code 2DRX; Asp, Glu, Lys residues in (POG) $_3$ -PKG-E/DOG-(POG) $_3$, PDB code 3T4F, 3U29; Phe, Gln, Arg residues in (GPO) $_3$ -GPRGQOGVMGFO-(GPO) $_3$, PDB code 4DMT; etc.) [88–90]. This situation contrasts with what is usually observed in globular protein crystals, which

build the crystal lattices through clusters of surface residues involved in direct protein–protein interactions. In collagen peptide crystals, the water networks surrounding the triple helices seem to act as cushions or spacers between them. The packing arrangements seen in peptide crystal structures (Figure 6B) are often reminiscent of the expected arrangements of collagen triple helices in fibrils. Furthermore, distances between triple helices in the crystal lattices are similar to the lateral packing distances between collagen molecules in tendon and other tissues examined by fibre X-ray diffraction. Thus, the lateral interactions seen in crystals are probably representative of what occurs in fibrillar collagen assemblies (with the caveat that crystals usually have parallel and antiparallel triple helices on the same lattice), and their interaxial distances are maintained by water molecules that connect adjacent helices. Leikin et al. [91] have reported the existence of *hydration forces* between collagen triple helices that are completely consistent with the observed hydration network. At short interaxial spacings (less than 16.8 Å) the forces are repulsive, whereas at longer spacings the forces are attractive. The repulsive forces start at a distance larger than the diameter of a single triple helix, and result from the compression of the water layer bound to the collagen model. The attractive forces are hydrophilic in nature and result from the dynamic network of water molecules that interconnect laterally the triple helices by hydrogen bonding (Figure 6C) [83,92]. This hydration network seems to have a role in directing assembly of collagen fibrils and maintaining their geometry.

HYDROXYPROLINE AND COLLAGEN STABILITY: RING PROPENSITY AND CHAIN PREORGANIZATION

Collagens are unique in animal proteins for their high content of Hyp, in particular the 4(*R*)-hydroxyproline stereoisomer (Hyp^{4R}, O). This is produced via post-translational modification of Pro residues of individual collagen strands by the enzyme prolyl-4-hydroxylase (P4H) (EC 1.14.11.2). P4H uses Pro, 2-oxoglutarate and O₂ as substrates and produces succinate, Hyp^{4R} and CO₂. The enzyme is very specific: it hydroxylates Pro residues in the Y position of individual polypeptide chains with repeating –Gly-X-Y– sequence, not acting on Pro residues when already on triple helical conformation [93]. Some collagens also contain 3(*S*)-hydroxyproline (Hyp^{3S}, O^{3S}), a rare post-translational modification by a different enzyme, prolyl-3-hydroxylase (P3H) EC 1.14.11.7 [94]. Hyp^{3S} will be discussed separately later.

At a very basic level, the imino acids Pro and Hyp stabilize the PPII conformation via stereochemical restrictions imposed by the imino acid rings. The Ramachandran plot for Pro shows a very limited choice of φ and ψ conformational angles, and one of the main regions corresponds to the PPII conformation [95]. This type of stabilization, often referred to as *preorganization*, is of entropic nature. In this case it decreases the entropic cost of collagen folding by favouring extended PPII-like chain conformations in the unfolded state. These are closer to the final folded state (triple helix) than if they were in a completely random conformation [44]. However, Hyp has an additional stabilizing effect, as demonstrated by the differences in thermal stability between hydroxylated ($T_m = 43^\circ\text{C}$) and non-hydroxylated ($T_m = 27^\circ\text{C}$) human type I collagen [96–98]. Importantly, lack of prolyl hydroxylation in animals results in collagens that are not stable at physiological temperature, and removal of P4H activity is lethal for both vertebrate and invertebrate animal models [99,100]. The very first collagen model peptides provided further evidence for the effect of Hyp on thermal stability: (POG)₁₀ had a temperature of denaturation 30 °C higher than that of the homologous peptide (PPG)₁₀ [52]. The impact of Hyp and many related Pro derivatives

has since been studied extensively using collagen model peptides, and several reviews of these studies have been published [44,101–105]. Biochemical knowledge obtained from these studies has opened a myriad of engineering possibilities, and the chemical biology of proline modifications is now an area of intense activity. A summary of the current state of the field follows.

Crystal structures of collagen model peptides show a distinct preference for the conformation of the imino acid rings, depending on their position in the collagen chain. Proline rings can adopt two states, defined by their φ and χ_1 angles (interdependent in imino acids). The *C γ -endo (down)* conformation (Figure 7) is characterized by positive χ_1 angles close to 25° and values of φ close to –75°. The *C γ -exo (up)* conformation shows negative χ_1 angles close to –20° and values of φ close to –60° [56,106]. These values of φ match closely the mean observed values of the φ conformational angles in the collagen triple helix: $-72 \pm 6^\circ$ for the X position, $-59 \pm 4^\circ$ for the Y position (averages from collagen peptide crystal structures at resolution 1.5 Å or better). This means that imino acids are effectively preorganized to fit into the collagen conformation without any significant strain [106]. The first crystal structure of a collagen model peptide with Hyp residues (PDB code 1CGD) showed that Hyp residues (all in the Y position) had a strong preference for the *C γ -exo* conformation, whereas Pro residues (all in the X position) had a clear preference for the *C γ -endo* conformation [55]. The same trend was observed in several crystal structures of (PPG)_{9/10} peptides (PDB codes 1A3I, 1A3J, 1G9W, 1ITT and 1K6F) where the *C γ -exo* conformation was favoured for Pro residues in the Y position and the *C γ -endo* conformation was preferred for Pro residues in the X position [56,106–108]. These *positional preferences* [87,109] are the consequence of the differences in the φ conformational angle at each position on the collagen triple helix, and have been largely confirmed in all peptide structures determined to date (Supplementary Table S1). On the other hand, crystal structures of amino acids and small peptides show a clear *intrinsic preference* for the *C γ -exo* conformation in Hyp residues, whereas Pro residues show a mixture of *C γ -endo* and *C γ -exo* conformations with a 2:1 preference for the former [106,110]. These observations led to Zagari and co-workers [106] to propose the *propensity-based hypothesis* by which replacing the Pro residues in the Y position with Hyp would stabilize the triple helix by reducing the conformational freedom of the unfolded state and preorganizing the unfolded chain towards the conformation of the folded state. This stabilization, of entropic nature, means that Hyp residues are more effective than Pro residues in the Y position as they are better preorganized to the conformational angles required at that position in the folded state (the triple helix). The propensity-based hypothesis would explain why the stabilizing effect of Hyp is very stereospecific and only the Hyp^{4R} diastereoisomer is found in the Y position of natural collagens. Accordingly, the collagen model peptide with swapped imino acid positions, (OPG)₁₀ does not form stable triple helices [111], and neither does the peptide (P α OG)₁₀ with *allo*-hydroxyproline residues (Hyp^{4S}, α O) in the Y position [112]. Crystal structures of the host–guest peptides (PPG)₄-P α OG-(PPG)₄ (PDB code 1X1K) and (PPG)₄-OPG-(PPG)₄ (PDB code 3A0A) show the Hyp^{4S} and Hyp^{4R} residues on the central triplets following the positional preference for the ring conformation, which is opposite to their own intrinsic preference (Figure 7). Both peptides are thus destabilized with respect to the parent peptide (PPG)₉ in a manner consistent with the propensity-based hypothesis [87,113,114].

Raines and co-workers [44,115–118] have developed the concept further and used it to engineer new hyperstable collagen model peptides with synthetic proline derivatives. The *C γ -exo*

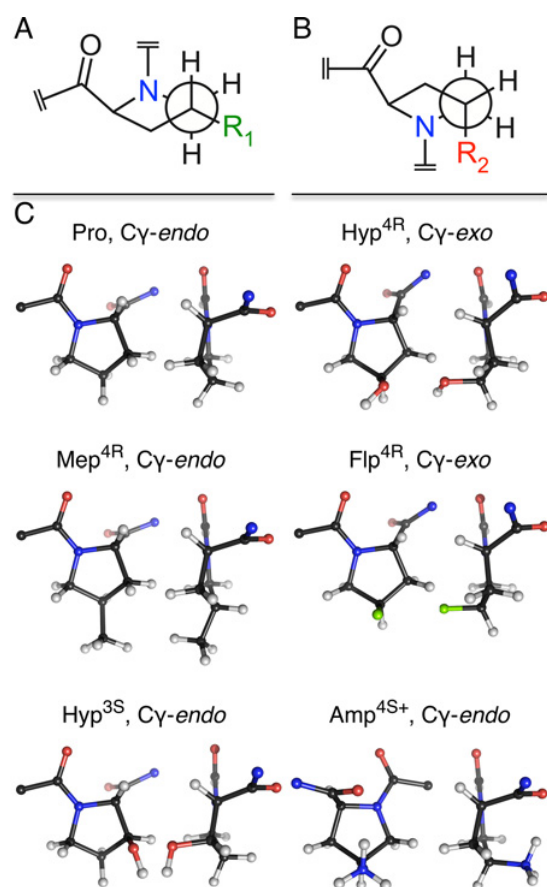


Figure 7 Preferences for ring conformation in proline and several of its derivatives

(A) The $C\gamma$ -endo conformation of 4*R*-proline derivatives places the R_1 group in an equatorial orientation. (B) Electron-withdrawing groups ($-\text{OH}$, $-\text{F}$, $-\text{Cl}$, $-\text{OCH}_3$) in the R_2 position stabilize the $C\gamma$ -exo conformation in 4*R*-proline derivatives through a *gauche* effect, placing the R_2 group in an axial orientation. Adapted from [44]: Shoulders, M.D. and Raines, R.T. (2009) Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958. (C) Stereoelectronic effects favour an axial position for electronegative groups $-\text{OH}$, $-\text{F}$, $-\text{Cl}$ or $-\text{OCH}_3$ in position 4 of the pyrrolidine ring. This results in strong preferences for the $C\gamma$ -exo conformation (*up* puckering) for 4(*R*) stereoisomers and for the $C\gamma$ -endo conformation (*down* puckering) for 4(*S*) stereoisomers. Steric effects favour an equatorial position for $-\text{CH}_3$ and $-\text{SH}$ groups in position 4, resulting in opposite conformation preferences. Proline itself favours the $C\gamma$ -endo conformation in approximately 2:1 ratio. The $-\text{OH}$ group in position 3, as in 3(*S*)-hydroxyproline, also favours an axial orientation which stabilizes the $C\gamma$ -endo conformation. 4(*S*)-aminoproline adopts a $C\gamma$ -endo conformation in an acidic environment but prefers the $C\gamma$ -exo conformation under basic conditions. Theoretical calculations suggest that transannular hydrogen bonds stabilize the $C\gamma$ -endo conformation of 4*S*-proline derivatives with hydrogen bonding donors ($-\text{OH}$, $-\text{NH}_3^+$) in position 4 [119,120]. See [44] for a comprehensive review of the experimental evidence and a detailed explanation of the stereoelectronic effect and its interplay with the steric effect. Abbreviations: Pro, proline; Hyp, hydroxyproline; Flp, fluoroproline; Mep, methylproline; Amp, aminoproline. Only hydrogen atoms on the pyrrolidine rings are shown.

preference for Hyp can be explained by a stereoelectronic *gauche* effect (Figure 7). Similarly other electronegative groups such as fluorine in 4(*R*)-fluoroproline or chlorine in 4(*R*)-chloroproline will favour this conformation (Table 2). Steric effects can also preorganize Pro derivatives to preferred ring conformations. A methyl group will prefer an equatorial position in a Pro ring and thus 4(*R*)-methylproline will favour a $C\gamma$ -endo conformation (Figure 7), whereas 4(*S*)-methylproline will favour a $C\gamma$ -exo conformation (Table 2) (see [44] for a comprehensive review).

Table 2 Intrinsic preferences for the conformation of the pyrrolidine ring in Pro derivatives and abbreviations used throughout the text.

Adapted from [44]: Shoulders, M.D. and Raines, R.T. (2009) Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958 with additional data from [119,121,135,138,139]. *Pro has a slight 2:1 preference for the $C\gamma$ -endo conformation (see the text). †Kep ring conformation is essentially planar due to the $\text{C}=\text{O}$ double bond [138].

Group	Imino acid	$C\gamma$ -endo	$C\gamma$ -exo	
$-\text{H}$	Proline	Pro*	P	
$-\text{OH}$	4-Hydroxyproline	Hyp ^{4S}	O	Hyp ^{4R} αO
	3-Hydroxyproline	Hyp ^{3S}	O ^{3S}	
$-\text{F}$	4-Fluoroproline	Flp ^{4S}	fP ^{4S}	Flp ^{4R} fP ^{4R}
	3-Fluoroproline	Flp ^{3S}	fP ^{3S}	
$-\text{Cl}$	4-Chloroproline	Clp ^{4S}	cP ^{4S}	Clp ^{4R} cP ^{4R}
$-\text{CH}_3$	4-Methylproline	Mep ^{4R}	mP ^{4R}	Mep ^{4S} mP ^{4S}
$-\text{OCH}_3$	4-Methoxyproline		Mop ^{4R}	mO ^{4R}
$-\text{SH}$	4-Mercaptoproline	Mcp ^{4S}		Mcp ^{4R}
$-\text{NH}_3^+$	4-Aminoproline (acid)	Amp ^{4S+}	aP ^{4S+}	Amp ^{4R+} aP ^{4R+}
$-\text{NH}_2$	4-Aminoproline (basic)		Amp ^{4S}	
$-\text{NHCOH}$	4-Formamidoproline	Fmp ^{4S}		Fmp ^{4R}
$-\text{NHCOCH}_3$	4-Acetamidoproline	Acp ^{4S}		Acp ^{4R}
$-\text{N}_3$	4-Azidoproline	Azp ^{4S}		Azp ^{4R}
$=\text{O}$	4-Ketoproline	Kep†		

The torsion angles for the $C\gamma$ -exo conformation match the torsion angles for the Y position in the collagen triple helix. Therefore, replacing a Pro residue in the Y position with a derivative preorganized in the $C\gamma$ -exo conformation (Table 2) introduces an entropic stabilization, and the fluorinated collagen model peptide (Pfp^{4R}G)₁₀ has one of the highest thermal stabilities reported to date ($T_m = 91^\circ\text{C}$, [115]). In contrast, when the electronegative groups are in the 4(*S*) position the favoured conformation is the $C\gamma$ -endo (Table 2). Replacing a Pro residue in the X position with a derivative preorganized in $C\gamma$ -endo conformation will also introduce an entropic stabilization, although other unfavourable factors may overturn this stabilization (discussed below). Generally, however, the effect of Pro substitution in the X position is less pronounced than that of replacing Pro in the Y position. Pro residues have already preference for the $C\gamma$ -endo conformation, whereas the $C\gamma$ -exo substitutes in the Y position reverse that preference and their impact on the entropic stabilization is higher [44]. Opposite effects are expected if $C\gamma$ -exo Pro derivatives are placed in the X position or if $C\gamma$ -endo Pro derivatives are placed in the Y position. In these cases a destabilization of the triple helix would occur. If the electronegative group is also a hydrogen bonding donor the $C\gamma$ -endo conformation could be further stabilized by intramolecular hydrogen bonding to the $\text{C}=\text{O}$ group of the substituted imino acid, as shown in Figure 7 for 4(*S*)-aminoproline under acidic conditions (Amp^{4S+}) (Table 2) [119]. However, such mechanism has been reported as destabilizing for the collagen triple helix when in the X position, due to a weakening interference with the interstrand hydrogen bonds [119–122].

An additional stereoelectronic effect termed $n \rightarrow \pi^*$ interaction [123] has been linked to collagen stabilization by Hyp and other similarly substituted Pro derivatives. This weak interaction occurs between the lone pairs of the oxygen atom of a peptide bond and an empty orbital on the oxygen atom in the next peptide bond. The preferred value for the ψ angle in the PPII conformation (145°) and in the Y position of the collagen triple helix ($151 \pm 4^\circ$) is geometrically ideal for these $n \rightarrow \pi^*$ interactions, and thus this effect will be potentially larger for the imino acids that favour the $C\gamma$ -exo conformation, such as Hyp^{4R} (Table 2) [44,124]. Although the stabilization energy of $n \rightarrow \pi^*$ interactions is modest (less

than 1 kcal/mol, weaker than a 'weak hydrogen bond') [125], their main impact is to displace the equilibrium between the *trans* and *cis* forms of the peptide bond, $K_{trans/cis}$ in imino acid residues [44]. Unfolded collagen strands will show a mixture of Pro residues with *cis* and *trans* peptide bonds, and all of them must be set to *trans* in order to build the triple helical structure. Thus, favouring *trans* peptide bonds in the unfolded state should contribute to an entropic stabilization of collagen. Nevertheless, replacement of Pro-Pro residues from a (PPG)₁₀ peptide with a Pro-*trans*-Pro alkene isostere shows that the *cis*-*trans* isomerization effects on the stability of the triple helix are limited [126].

Combining the conformational preferences of different Pro derivatives allows for the engineering of largely preorganized collagens with improved thermal stability. The rationale for the design is to use derivatives that favour the *C γ -endo* conformation in the X position and derivatives that favour the *C γ -exo* conformation in the Y position (Table 2). Numerous collagen model peptides engineered with Pro residues replaced are consistent with this entropic stabilization (Supplementary Table S2; see [44] for a comprehensive list). For instance, fluorine is more electronegative than oxygen and Flp^{4R} has a stronger preference for the *C γ -exo* conformation than Hyp^{4R}. Thus Flp^{4R} in the Y position should be more effective than Hyp^{4R} in stabilizing collagen, whereas Flp^{4R} in the X position should be destabilizing. Accordingly, the peptide (Pfp^{4R}G)₁₀ is much more stable thermally than the reference peptide (POG)₁₀ (Supplementary Table S2) [127], whereas the peptide (fP^{4R}PG)₁₀ does not form triple helices [128].

These design principles have been demonstrated elegantly with the synthesis, biochemical characterization and crystal structure determination of the collagen model peptide (mP^{4R}fP^{4R}G)₇ (PDB code 3IPN) [118]. This peptide has a large thermal stability compared with the unstable reference peptide (PPG)₇ (Supplementary Table S2). The main component of its stability is entropic, as expected from the preorganization of the chain conformation consequence of the choice of proline derivatives for each position. This preorganization has not altered the structure of the triple helix, which closely resembles that of the peptides (PPG)₁₀ or (POG)₁₀.

For all its conceptual simplicity, the propensity model does not explain all cases and an increasing number of 'exceptions' to the rule have been reported (Supplementary Table S2). For instance Hyp^{4S} does not stabilize the helix at any position. Its preferred *C γ -endo* conformation should make it suitable for incorporation in the X position and yet the peptide (α OPG)₁₀ does not form triple helices [112] and the peptide (α OPG)₁₅ is less stable than its reference peptide (PPG)₁₅ [120,129] (Supplementary Table S2). Incorporation of Flp^{4S} into the X position of peptides (fP^{4S}PG)₇ and (fP^{4S}PG)₁₀ or Flp^{4R} into the Y position of peptides (Pfp^{4R}G)₇ and (Pfp^{4R}G)₁₀ is stabilizing. Preorganization of both X and Y positions should bring extraordinary stability. Yet, peptides with both positions occupied by Flp, (fP^{4S}fP^{4R}G)₇, (fP^{3S}fP^{4R}G)₇ and (fP^{4S}fP^{4R}G)₁₀ either do not form triple helices or are destabilized [130,131] (Supplementary Table S2). Different explanations have been put forward to rationalize these exceptions: unfavourable steric interactions, interference of intramolecular hydrogen bonds with the interstrand N-H...O=C hydrogen bonding, etc. [106,120,131].

The anomaly of Hyp^{4S} is illustrative. When Hyp^{4S} is in the X position its hydroxy group points to the inside of the collagen triple helix. Thus, it was suggested that steric clashes with the neighbouring chains would prevent the formation of a stable triple helix [106]. However, Flp^{4S} is very stabilizing in the X position [128,132] and the size difference between the -OH and -F groups is too small to account for the discrepancy in stability.

An alternative mechanism of destabilization would result from the formation of an intramolecular hydrogen bond between the -OH and C=O groups of Hyp^{4S} (similar to the case shown for Amp^{4S+} in Figure 7). In the X position, the C=O group is already involved in N-H...O=C hydrogen bonding. Accepting a second hydrogen bond would weaken this interstrand interaction [120]. The crystal structure of a host-guest peptide with two triplets Hyp^{4S}-Pro-Gly (PDB code 3B2C) [133] shows conformational diversity for the Hyp^{4S} residues: about 30% adopt the *C γ -exo* conformation, going against both intrinsic and positional preferences of Hyp^{4S} in the X position; the remaining 70% adopts a distorted (shallow) *C γ -endo* conformation with χ_1 values averaging 16° (compared with the average from high-resolution peptide crystal structures, $\chi_1 = 26 \pm 8^\circ$). These distortions probably avoid unfavourable steric interactions with the other chains, and are thought to cause the marked decrease in thermal stability of the peptide (POG)₄-(α OPG)₂-(POG)₄ ($T_m = 49^\circ\text{C}$) with respect to the reference peptide (POG)₁₀ ($T_m = 62^\circ\text{C}$) [133]. Concerning the proposed intramolecular hydrogen bond, the average distance between the -OH group and the carbonyl oxygen is 3.15 (± 0.13) Å, whereas the average angle between the hydrogen bond and the C=O bond is 76 (± 3)°. These geometrical parameters indicate a weak, if any, hydrogen bonding interaction and suggest that, at least for the (POG)₄-(α OPG)₂-(POG)₄ peptide, there will be little interference with the interstrand N-H...O=C hydrogen bonding.

Unexpected results have also been obtained when one Pro-Flp^{4R}-Gly unit is embedded within a Pro-Hyp-Gly context. Thus, (POG)₃-Pfp^{4R}G-(POG)₄ is slightly less stable ($T_m = 44^\circ\text{C}$) than the parent peptide (POG)₈ ($T_m = 47^\circ\text{C}$) despite introducing in the Y position the highly stabilizing Flp^{4R} [134]. On the other hand, adding Hyp to the X and Y positions produces completely the opposite to the expected results. The peptide (OOG)₁₀ with Hyp^{4R} in both X and Y positions goes against the propensity model and yet it is slightly more stable than the reference (POG)₁₀. The related peptide (POG)₃-OOG-(POG)₄ has the same thermal stability ($T_m = 47^\circ\text{C}$) than the parent peptide (POG)₈. On the other hand, the peptide (O^{4S}OG)₁₀ with Hyp^{4S} in the X position and Hyp^{4R} in the Y position follows the propensity model and yet is much less stable than the reference (POG)₁₀ [130] (Supplementary Table S2). Possible reasons for these discrepancies in the case of Hyp-rich peptides are discussed in the next section.

Current knowledge from structural and biochemical analysis of collagen peptides with different proline derivatives has led to relatively high levels of success in predicting the changes in collagen stability and folding kinetics when using these derivatives. This knowledge is guiding ongoing efforts on the design of new collagen peptides with functionalizable groups [135–138] or the engineering of novel collagen peptide mimetics and collagen-based biomaterials [122,139–144]. One example of application is the development of pH-dependent triple helices through synthetic ionizable Pro derivatives such as aminoproline (Amp) residues [119,145–147] or carboxylate-modified prolines [148,149]. Amp-based collagen model peptides have a complex stability profile due to the convergence of several factors like the intrinsic ring preferences for Amp^{4S}/Amp^{4R+}, the possibility of intramolecular hydrogen bonding and the different stereoelectronic properties of the neutral (-NH₂) and protonated (-NH₃⁺) forms of the amino group (Table 2). At neutral and acid pH, the amino group is protonated: Amp^{4S+} prefers to adopt a *C γ -endo* conformation, possibly enforced by an intramolecular transannular hydrogen bond (Figure 7). On the other hand, the uncharged amino group at basic pH will favour the *C γ -exo* conformation for Amp^{4S}, in a similar way to Mep^{4S}. This difference forms the basis of a pH-dependent conformational switch between the two conformations, and thus (POG)₃-aP^{4S}

PG-(POG)₃ is more stable at pH 11 ($T_m = 33^\circ\text{C}$) than at pH 3 ($T_m = 13^\circ\text{C}$) [119]. The related peptide (POG)₃-PaP^{4S}G-(POG)₃ is more stable at pH 11 ($T_m = 44^\circ\text{C}$) due to the uncharged Amp^{4S} preferring the *C γ -exo* conformation and being in the Y position [119]. Partially conflicting results have been obtained using repeating peptides with six triplets of Pro, Amp and Gly residues (Supplementary Table S2). These data suggest that both protonated (pH 3) and neutral Amp^{4R} (pH 12) are stabilizing in the Y position, whereas Amp^{4S+} and Amp^{4R+} are both stabilizing in the X position when protonated (pH 3), but not when neutral [145–147].

HYDROXYPROLINE AND COLLAGEN STABILIZATION: ENTHALPY CONTRIBUTION

The mechanism of stabilization discussed above is essentially of entropic nature: the chain conformation in the unfolded state is primed to a conformation that is closer to the folded state and therefore reduces the overall entropic cost of going from three individual chains to one single triple helix. However, collagen stability has a large enthalpic component, in contrast to what occurs with other proteins, and the stabilization of collagen by Hyp has been correlated with an increase in that enthalpy [70]. An early thermodynamic analysis of the denaturation of the model peptides (POG)₁₀ and (PPG)₁₀ demonstrated that (POG)₁₀ has a much higher denaturation temperature than (PPG)₁₀ and also that it shows larger enthalpy and entropy changes for the coil \rightarrow triple helix transition [53]. The same findings have been obtained in previous studies [128,150].

What is the origin of the large enthalpy of collagen? In the context of protein stability, a high enthalpy of stabilization is usually associated with more (or stronger, or both) hydrogen bonds, more efficient molecular packing leading to stronger van der Waals interactions, stronger electrostatic interactions, etc. Given the non-globular structure of the collagen triple helix and the absence of charge effects between Pro and Hyp, it would follow that the enthalpic component of collagen stabilization by Hyp mainly relates to an increase in hydrogen bonding. However, the structure of the triple helix, which places the hydroxy groups of Hyp at the periphery, does not allow for it. The only obvious alternative is significant hydrogen bonding interaction with water surrounding the triple helix and the formation of water bridges connecting the chains. Different models of hydrogen bonding interactions between the main-chain C=O groups, the Hyp side chains, and water molecules around the triple helix have been proposed in the past, including the water-bonded structure when non-imino acids occupy the X position [71,72,151]. As already discussed, the crystal structure of the Gly \rightarrow Ala peptide revealed an extensive repetitive network of water molecules, interconnecting the triple helical structure through hydrogen bonding with the C=O groups of the peptide bonds and the –OH groups of the Hyp residues [73]. Water networks with the same hydrogen bonding topology have been observed in subsequent, higher-resolution crystal structure determinations of peptides with a significant proportion of POG triplets (Supplementary Table S1). In these structures, the –OH groups of Hyp side chains are central to the connectivity of the hydration networks (Figure 6). However, it is not possible to assess from crystal structures whether the contribution of the hydrogen bonding between Hyp and water is ultimately stabilizing.

There has been a degree of controversy about the physical significance of these water networks with respect to collagen stability. The large body of evidence in support of the propensity-based hypothesis and the stereoelectronic effects (see above) has

offered a convincing case for collagen stability, while apparently dismissing any role of water in the structure and stability of the triple helix. In particular, the entropic cost of engaging many water molecules in hydrogen bonding interactions is thought to largely overcome any stabilizing enthalpic interaction [44]. Nevertheless, neither the propensity-based hypothesis nor stereoelectronic effects offer a reasonable explanation for the enthalpic stabilization of natural collagen. The (OOG)₁₀ peptide is one of the exceptions to the propensity-based predictions. It forms a triple helix that is slightly more stable ($T_m = 65^\circ\text{C}$) than (POG)₁₀, despite placing *C γ -exo* favouring Hyp^{4R} in the X position. Crystal structures of the (OOG)₁₀ peptide (PDB code 1WZB) [152] and the closely related (GOO)₁₀ peptide (PDB code 1YM8) [153] show that essentially all Hyp residues adopt the *C γ -exo* conformation irrespective of their X or Y position, and yet the resulting triple helices have the same helical and hydrogen bonding patterns as other collagen model peptides.

To gain insight into this anomaly, Kobayashi and co-workers [130,150] performed differential scanning calorimetry analyses on (OOG)₁₀ and other peptides. They concluded that the increase in stability with respect to (PPG)₁₀ seen in the different peptides could be enthalpy-dominant or entropy-dominant, and that the mechanisms of collagen stabilization by Flp and Hyp would differ (Figure 8). Thus, the increased stability of (Pfp^{4R}G)₁₀ with respect to (PPG)₁₀ would be entropically driven, consistent with the propensity-based model, stereoelectronic effects, and *cis/trans* equilibrium of the peptide bonds discussed above. In contrast, the increased stability of (POG)₁₀ with respect to (PPG)₁₀ would be enthalpically driven, with a significant contribution from hydrogen bonding to the hydration network surrounding the triple helix, in addition to the entropic factors mentioned above [150]. The (OOG)₁₀ peptide showed smaller enthalpy and entropy changes than those seen in (PPG)₁₀ and (POG)₁₀ (Figure 8). The degree of hydration of (OOG)₁₀ and (POG)₁₀ is very similar in the crystalline state, but (OOG)₁₀ appears to be more hydrated in the unfolded state. This would explain the reduction in enthalpy and entropy differences between folded and unfolded state of (OOG)₁₀ with respect to (PPG)₁₀ or (POG)₁₀ [152].

COLLAGEN AND 3-HYDROXYPROLINE

The other post-translational modification of Pro residues is prolyl-3-hydroxylation, a rare modification that so far has only been reported in collagen proteins. Hyp^{3S} was originally discovered on type I collagen more than 50 years ago [154], but its biological function is still being elucidated [94]. Hyp^{3S} is a quantitatively minor modification, with only a few residues per α chain in collagen types I, II, III, IV and V/XI [155–159]. It occurs only at specific sites, always in the X position of triplets with Hyp^{4R} in the Y position. Hyp^{3S} sites are highly conserved, and show variable occupancy in a given tissue and pronounced specificity across tissues or developmental stages [94]. For example, 3-hydroxylation of a C-terminal (GPP)_n motif in type I collagen is unique to tendon and absent from skin and bone. The levels of Hyp^{3S} in this motif appear to be regulated during development [157,159]. Hyp^{3S} must be important for the self-assembly of collagen supramolecular structures as several genetic variants with defective prolyl-3-hydroxylation have been linked to recessive forms of osteogenesis imperfecta and severe myopia (reviewed in [94,160]).

All of these findings have contributed to a renewed interest in the role of Hyp^{3S} in collagen structure and function. Hyp^{3S} has preference for the *C γ -endo* conformation (Table 2) and, according to the propensity model, it should provide stabilization in the X

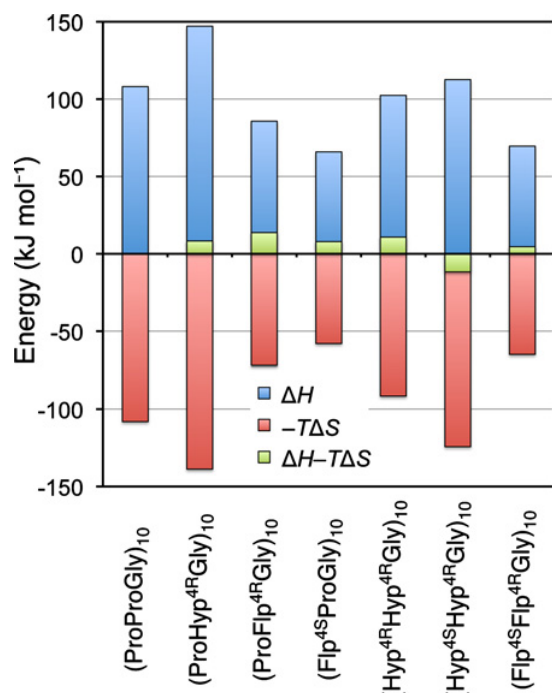


Figure 8 Comparison of thermodynamic parameters for the triple helix \rightarrow coil transition of different collagen peptides normalized at the equilibrium transition temperature of (PPG)₁₀ [130]

Red bars represent the ΔH term, blue bars represent the $-T\Delta S$ term, and yellow-green bars represent ΔG . A positive ΔG bar indicates that the peptide is more stable than (PPG)₁₀ (the transition to coil is not favoured, the triple helix is favoured), whereas a negative ΔG indicates that the peptide is less stable (transition to coil is favoured). Different magnitudes for the red and blue bars indicate the relative differences in the contribution of the enthalpic and entropic term for each peptide. Adapted from [130]: Doi, M., Nishi, Y., Uchiyama, S., Nishiuchi, Y., Nishio, H., Nakazawa, T., Ohkubo, T. and Kobayashi, Y. (2005) Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)₁₀ analogues: (4(S)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)₁₀, (4(R)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)₁₀ and (4(S)-fluoroprolyl-4(R)-fluoroprolyl-Gly)₁₀. *J. Pept. Sci.* **11**, 609–616; and [150] Nishi, Y., Uchiyama, S., Doi, M., Nishiuchi, Y., Nakazawa, T., Ohkubo, T. and Kobayashi, Y. (2005) Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of collagen triple helix. *Biochemistry* **44**, 6034–6042.

position. Initial studies produced conflicting results, as the peptide (GO³⁵O)₁₀ appeared not to form a triple helix at all, whereas the host-guest peptide (GPO)₃-GO³⁵O-(GPO)₄ was slightly more stable than the related (GPO)₈ [161]. On the other hand, the triple helical structure of the peptide (GPO)₃-(GO³⁵O)₂-GPO₄ (2G66) [162] was virtually identical with that of other (PPG)_n or (POG)_n peptides, with all of the Hyp³⁵ residues in the expected *C γ -endo* conformation and no evidence for any destabilizing interaction. Later work by the same group established that Hyp³⁵ does stabilize slightly the collagen triple helix in the X position of Gly-X-Hyp^{4R} triplets [163] and that this stabilization is mainly entropic, as would be expected from the propensity model.

Nevertheless, the very low frequency of Hyp³⁵ residues in collagen (compared with more than 100 Hyp^{4R} positions per collagen chain) argues against a main role for Hyp³⁵ in simply conferring additional stability to the triple helix [94]. Other functions have been proposed where Hyp³⁵ side chains may act as specific points for collagen interaction, with other molecules of collagen or with collagen-binding proteins, and the evidence is slowly emerging. The regular spacing of several Hyp³⁵ sites in collagen types I and II may indicate a possible role in fibril assembly [156]. Also, Hyp³⁵ sites may overlap with fibril surface

binding sites for small leucine-rich proteoglycans [94,160]. The (GPO)₃-(GO³⁵O)₂-GPO₄ structure (PDB code 2G66) shows that the -OH groups from Hyp³⁵ residues point away from the triple helix. This orientation would be consistent with a role in directing specific hydrogen bonding intermolecular interactions to other proteins or to other collagen triple helices [156,162]. A recent report shows that Hyp³⁵ residues in type IV collagen prevent its interaction with platelet-specific glycoprotein VI, thus blocking the initiation of platelet aggregation [164]. All of these studies point towards biological functions for Hyp³⁵ based on molecular interactions. Interestingly, evolution seems to have selected different roles for the two hydroxylated Pro residues: Hyp^{4R} as a widespread non-specific modification to increase thermal stability throughout; and Hyp³⁵ as a specific localized modification to provide sites for molecular recognition.

COLLAGEN STABILIZATION BY OTHER AMINO ACID SIDE CHAINS

Over the last decade, several studies have reported the existence of numerous open reading frames with the hallmark (Gly-X-Y)_n collagen sequence in prokaryotic and viral genomes [165–169]. Some of these bacterial and viral collagen-like proteins have been produced using recombinant techniques, and the formation of triple helical structures confirmed [169–172]. Sequence analyses of non-metazoan collagenomes reveal amino acid compositions quite different from those of vertebrates or invertebrates (Table 3, Figure 9). In particular, the proportion of Pro residues in the Y position is very low in all prokaryotic and viral collagenomes, probably due to the lack of prolyl hydroxylation in these organisms. Absence of Hyp residues appears to be compensated for by an increase in other residues at specific positions, depending on the taxonomic groups (Table 3, Figure 9): Pro residues in the X position (*Escherichia coli*, phages, *Bacillus*); charged residues in the X and Y positions (*Mimivirus*, *E. coli*, *Streptococcus*); Thr residues in the Y position (*Bacillus*); Gln residues in the Y position (phages, *Bacillus*, *Streptococcus*, *E. coli*); or Ala residues in the X position (*Bacillus*). The biochemical characterization of these prokaryotic collagens shows thermal stabilities close to those of vertebrate collagens [169–173], indicating that other mechanisms of stabilization independent of prolyl hydroxylation are at play. Prokaryotic collagens have rapidly generated interest as possible sources of collagen-based biomaterials with designed properties [172,174,175], and thus it is important to elucidate these Hyp-independent mechanisms of collagen stabilization as they may open new avenues for collagen engineering.

Brodsky and co-workers [176–179] have carried out a systematic study of the effect of different amino acids in the X and Y positions on the stability of collagen. Most peptides with a simple repetitive (Gly-X-Y)_n sequence are not stable at room temperature unless X and Y are imino acids. Thus, the usual approach to investigate the contribution of specific triplets to collagen stability is the design of host-guest peptides with sequences such as Ac-(GPO)₃-(GX_AX_B)-(GPO)₄-GG-NH₂ or Ac-(GPO)₃-(GX_AX_B-GX_CX_D)-(GPO)₃-GG-NH₂, where the destabilizing effect of the 'guest' triplets GX_AX_B or GX_AX_B-GX_CX_D is measured by comparison to the reference peptide Ac-(GPO)₈-GG-NH₂ (the most stable of the series). From these stability data, an empirical algorithm for the prediction of collagen stability from its amino acid sequence was developed [178] (Collagen Stability Calculator, <http://compbio.cs.princeton.edu/csc>). This algorithm can be used to predict the melting temperature of short peptides with an architecture similar to the host-guest series above, but also to produce a profile of relative stability for longer collagen sequences

Table 3 Amino acid preferences (%) in collagen triple helical domains of human, mouse, zebrafish, *C. elegans* (worm), *Drosophila melanogaster* (fly), *E. coli*, *Streptococcus*, *Bacillus*, Caudovirales (phages) and Mimiviridae collagenomes

	Human X/Y	Mouse	Fish	Worm	Fly	<i>E. coli</i>	<i>Streptococcus</i>	<i>Bacillus</i>	Phages	Mimiviridae
PO	27/38	27/32	24/42	20/32	46/9	22/7	32/3	37/7	1/2	
ED	19/6	20/6	14/13	25/8	35/1	31/13	6/2	21/10	62/4	
KR	7/22	7/23	9/11	8/27	1/31	13/29	0/1	4/18	0/60	
VILMF	24/10	22/9	21/10	13/5	21/9	4/5	11/4	26/5	16/8	12/15
A	7/6	8/7	13/10	8/4	10/20	14/7	23/2	11/9	1/2	
TS	7/7	8/8	7/10	8/6	5/10	1/19	2/19	8/59	7/18	11/8
QN	6/8	6/9	14/8	6/6	2/15	7/22	3/26	2/28	8/7	
G	1/2	3/2	3/1	1/0	0/0	1/1	3/3			
HYCW	2/1	3/1	2/1	3/2	5/1	0/1	0/0	1/1	1/0	

that can be useful to identify regions of high or low stability along a given collagen sequence [103].

The least destabilizing residues in the host–guest analysis of single $GX_A X_B$ triplets were Pro, Glu, Ala, Lys, Arg, Gln and Asp for the X position and Hyp, Arg, Met, Ile, Gln and Ala for the Y position. Aromatic residues Trp, Tyr and Phe were very destabilizing in either position, and Gly was also very destabilizing, more in the X position. Thus, the least destabilizing triplet without imino acids was GER and the most destabilizing observed experimentally was GGF [178]. There is no clear explanation at the molecular level for many of these effects. For instance, it is noteworthy that Ala residues are relatively favoured at either the X or Y positions of the triple helix, without an obvious reason other than being small and not interfering. The case of Arg in the Y position seems clearer: Arg side chains can form hydrogen bonds to main-chain C=O groups on the following strand, as seen in the crystal structure of the T3–785 peptide (PDB code 1BKV) (Figure 9) [64]. Other peptides with Arg residues in the Y position show the same interaction for some but not all their side chains (PDB code 1Q7D, 3DMW, 4AXY, 4DMT and 4GYX) [65,90,180–182]. Analysis of the side-chain conformation of Arg residues in collagen peptides shows a preferred *tt* conformation for the χ_1 and χ_2 torsion angles that differs from that most commonly observed in globular proteins. In particular, the *tttt* or *ttgg* conformations for χ_1 to χ_4 are observed in Arg side chains where $N\epsilon$ or $N\eta 1$ respectively form a hydrogen bond to the main-chain C=O group on the Y position from the following strand [183].

Extension of the host–guest analysis to the two triplet case $GX_A X_B GX_C X_D$ demonstrated that interstrand electrostatic interactions between residues of opposite charges can be very stabilizing, as shown by host–guest peptides with GPKGEO and GPKGDO sequences [179]. Crystal structures of peptides with these sequences (PDB codes 3T4F and 3U29) confirmed the formation of interstrand ion pairs and additional hydrogen bonding involving the Lys side chains [89]. The KGE and KGD motifs appear to be more common in metazoan collagen sequences than expected from the individual residue frequencies [179], and positional preference analyses of non-metazoan collagens shows a high proportion of charged residues, Glu/Asp in the X position, Lys/Arg in the Y position, for *E. coli*, *Streptococcus* or viral collagenomes (Table 3, Figure 9A). Biochemical analysis of ScL2, a collagen-like protein from *Streptococcus pyogenes*, indicates a high degree of electrostatic stabilization of the triple helix that is a consequence of its relatively high proportion of charged amino acids [184]. Similarly, collagen-like proteins from *E. coli* genomes contain a high proportion of charged amino acids plus a very high preference for Pro in the X position (Table 3,

Figure 9A). The combination of both effects may contribute to the observed stability of the triple helices of these proteins [169]. Some prokaryotic or viral collagen-like sequences show stretches of repetitive sequences alternating charged residues of different sign ($(KGE)_n$ or $(KGD)_n$, while free of imino acids (one predicted collagen-like sequence from white spot shrimp virus contains 33 consecutive KGE repeats, Q8QTH5). Formation of repetitive interstrand ion pairs as shown in Figure 9B could be essential for maintaining a triple helical structure of these atypical collagen-like sequences.

Other mechanisms for collagen stabilization involving Thr residues have been reported. The cuticle collagen of the deep sea hydrothermal vent worm *Riftia pachyptila* has very low proportion of imino acid residues, high content of Thr residues in the Y position (>18%) and thermal stability comparable to that of metazoan collagens [185]. Biochemical analysis and partial sequencing showed that the Thr residues are glycosylated with galactose saccharides [186]. These glycosylated Thr residues are responsible for the thermal stability of the cuticle collagen in place of Hyp. Thus, the synthetic peptide $(GPT)_{10}$ does not form a triple helical structure, but modification of its Thr residues with β -D-galactose induces triple helical formation (Supplementary Table S2) [187]. Interestingly, the peptide $(GOT)_{10}$ does form a stable triple helix ($T_m = 19^\circ\text{C}$), much more stable after Thr glycosylation ($T_m = 55^\circ\text{C}$) [188]. The combination of Hyp residues in the X position and glycosylated Thr residues in the Y position seems to stabilize several invertebrate cuticle collagens [188], and the crystal structures of the peptides $(PPG)_4$ -O(S/T/V)G-(PPG)₄ (PDB codes 3ADM, 3A1H and 3A0M) [109] suggest that this stabilization is due to van der Waals and dipole–dipole interactions between the Hyp residues in the X position and the S/T/V residues in the Y position of the preceding chain. Some prokaryotic collagenomes (*Bacillus* and also *Clostridium* and other firmicutes) also show a clear preference for Thr residues in the Y position (Figure 9A), and glycosylation of the collagen-like protein from *Bacillus anthracis* BclA has been confirmed [189]. However, a recombinant version of BclA produced in *E. coli* and thus unlikely to contain the specific glycosylation of native BclA, still formed a collagen triple helix with a stability of 37°C for the collagen-like region alone [171].

Electrostatic interactions have been used in the design of novel heterotrimer collagens. When charged collagen peptides with repetitive sequences such as $(PRG)_{10}$ or $(EOG)_{10}$ are mixed together with neutral peptides like $(POG)_{10}$, they can self-assemble into heterotrimers with 2:1 or 1:1:1 stoichiometries [190]. Some of these heterotrimers have stabilities comparable to the reference $(POG)_{10}$ homotrimer (Supplementary Table S2) [191]. Interstrand electrostatic interactions between the charged

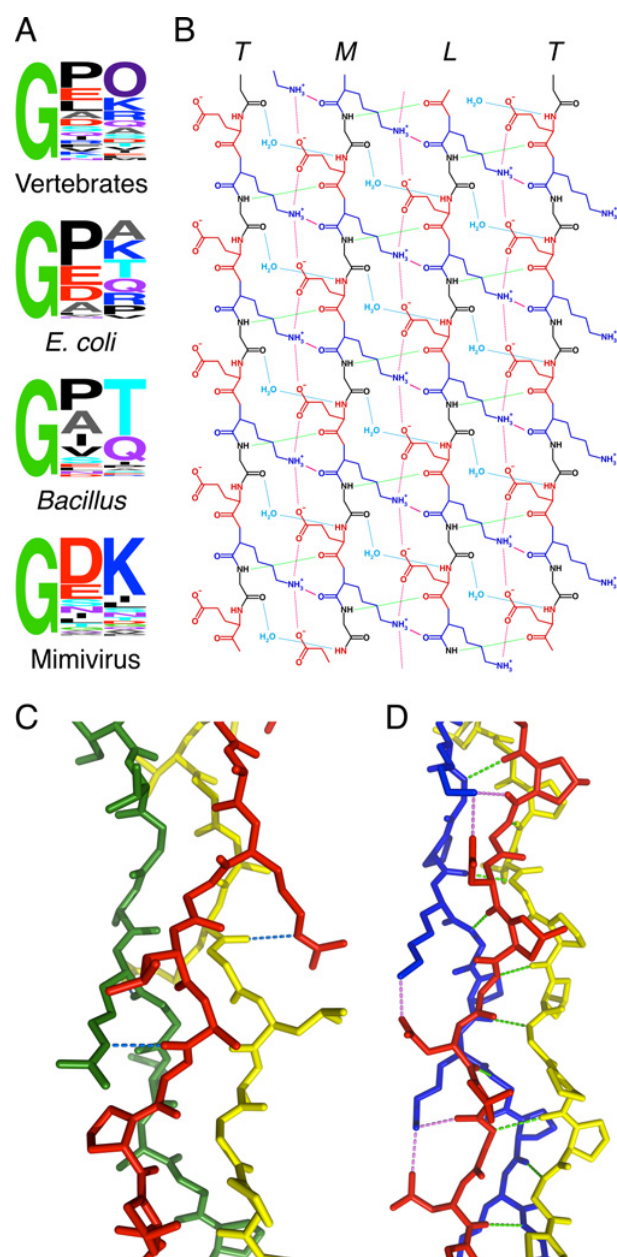


Figure 9 Effect of amino acid side chains other than Pro and Hyp on the stability of the collagen triple helix

(A) Frequency plots showing amino acid residue preferences in the COL \times 3 domains from the collagenomes of several taxonomic groups. The height of each symbol is proportional to its frequency at that position in a repetitive (Gly-X-Y) $_n$ sequence. For simplicity, Pro residues in the Y position of vertebrates are assumed to be all hydroxylated to Hyp. The proportion of Pro residues in the Y position is notoriously low in prokaryotic and viral collagenomes. (B) Collagen triple helices with repetitive (GEK) $_n$ sequences could potentially form a continuous network of salt bridges connecting the three strands (magenta) in a TMLT direction. Interstrand N—H...O=C and water-mediated hydrogen bonding are shown as green and cyan lines respectively. (C) Direct hydrogen bonding of Arg side chains in the Y position to main-chain C=O groups in the following chain, from the structure of the T3-785 peptide (1BKV) [64]. Trailing chain is shown in red, middle chain is shown in yellow, and leading chain is shown in green. The Arg side-chain to main-chain hydrogen bonding follows the TMLT direction. (D) Salt-bridge formation (magenta) in the heterotrimer structure (PKG) $_{10}$:(DOG) $_{10}$:(POG) $_{10}$ (PDB code 2KLW) [192]. The (PKG) $_{10}$ strand (blue) occupies the trailing position, the (DOG) $_{10}$ strand (red) the middle position and the (POG) $_{10}$ strand the leading position. Interstrand N—H...O=C hydrogen bonding is shown in green. Images in (C and D) were prepared with UCSF Chimera [253].

side chains of the individual peptides provide the mechanism of stabilization, and neither the cationic nor anionic peptides are able to form stable homotrimer triple helices. The NMR structure of the heterotrimer triple helix [(PKG) $_{10}$:(DOG) $_{10}$:(POG) $_{10}$] (PDB code 2KLW) shows a single-register triple helix where (PKG) $_{10}$ is the trailing strand, (DOG) $_{10}$ is the middle strand and (POG) $_{10}$ is the leading strand, with multiple interstrand salt-bridge interactions between the Lys and Asp side chains (Figure 9D) [192]. By optimization of the amino acid sequence and the ratio of charged and Hyp residues on each chain, Hartgerink and co-workers [193–196] have produced several collagen peptide designs with better control of the resulting triple helix, favouring heterotrimer assembly and unfavouring the formation of homotrimers.

Similar principles could be applied to the design of heterotrimer collagens based on optimization of steric interactions. Thus, neither the (fP 4S fP 4R G) $_7$ or (PPG) $_7$ peptides can form stable homotrimer helices, yet the heterotrimer [(fP 4S fP 4R G) $_7$] $_2$:(PPG) $_7$ is stable at room temperature (Supplementary Table S2). Molecular models of the homo- and hetero-trimers suggest that the (PPG) $_7$ strand can relax some of the unfavourable steric interactions between fluorine atoms of the neighbouring (fP 4S fP 4R G) $_7$ strands [131]. It can be predicted that favourable combinations of steric and electrostatic interactions will lead to the development of codes for heterotrimer assemblies of collagen peptides with designed sequences [44,131].

COLLAGEN RECOGNITION

The relatively simple 3D structure of the collagen triple helix means that any molecular recognition motif on the COL \times 3 domains of collagens must be essentially linear. Furthermore, such recognition motifs must occur on the framework of a highly repetitive structure imposed by the requirement of the (Gly-X-Y) $_n$ sequence. Use of collagen model peptides has been essential to investigate these questions and in recent years several crystal structure determinations have clarified the mechanisms of collagen recognition by other proteins (Supplementary Table S1). Collagen-binding sites have been mapped on the sequences of COL \times 3 domains and on approximate 2D and 3D models of collagen fibrils based on fibre X-ray diffraction [197–200]. These maps help in the identification of functional domains along the collagen fibril, providing insight into their accessibility to recognition by ligands and the impact of collagen mutations on major ligand-binding sites.

Binding to collagens by cell-surface receptors and extracellular matrix proteins is critical for their biological function. For example, collagens regulate cell behaviour (adhesion, migration and proliferation) through their interaction with specific cellular receptors: collagen-binding integrins, collagen-binding immune receptors and discoidin domain receptors [9,201–203]. The development of collagen ‘toolkits’ by Farndale and co-workers [204,205] has been extremely important for the discovery of the collagen sequence motifs recognized by these receptors and other collagen-binding molecules. The toolkit approach consists of synthesizing a library of collagen peptides with overlapping sequences covering the entire length of long COL \times 3 domains, such as those from type II and type III collagen. This library is then tested for binding against the different collagen-binding proteins or receptors and the collagen recognition motifs are identified from the sequences of the peptides that show binding [204,205]. As a case example, an early version of this method identified the GFOGER sequence as the collagen recognition motif for integrin α 2 β 1 [206,207]. The structure of the peptide (GPO) $_2$ -GFOGER-(GPO) $_3$ co-crystallized with the integrin α 2 I-domain

(PDB code 1DZI) elucidated the molecular basis of the interaction (Figure 10) [41,208]. Additional integrin-binding motifs with sequences similar to GFOGER and with different affinities and specificities have been discovered through a combination of methods and confirmed with synthetic peptides [205,209–211]. Two additional structures of integrin $\alpha 1$ and $\alpha 2$ I-domains in complex with peptides containing GLOGEN or GFOGER motifs have been determined by solution NMR and crystallography (PDB codes 2M32 and 4BJ3) [212,213]. In all of these, structures the Glu residue from the OGE triplet completes the co-ordination of a divalent cation in the integrin I-domain (Figure 10), a mechanism usually seen in integrin interaction with its ligands [41,208].

Collagen toolkits helped to identify the sequence GVMGFO as an interaction motif for three different unrelated proteins: von Willebrand factor (a plasma protein involved in homeostasis), discoidin domain receptor 2 (DDR2, a receptor tyrosine kinase that regulates cell behaviour and extracellular matrix remodelling) and osteonectin (also called SPARC/BM40, an extracellular calcium-binding matrix glycoprotein associated with tissue remodelling and bone mineralization) [214–216]. Crystal structures for each interaction have been reported (PDB codes 4DMU, 2WUH and 2V53) [90,217,218]. In these crystals, the GVMGFO motifs bind to amphiphilic specificity pockets on the surface of the collagen-binding proteins (Figure 10) in addition to other areas of contact specific to each protein. It has been suggested that the convergence of binding mechanisms to the same GVMGFO motif is related to the unique hydrophobic knob created by two large hydrophobic residues separated by Gly, and the relative rarity of such motif on the sequences of COL \times 3 domains of fibrillar collagens [218].

Other structures of collagen peptides in complex with collagen-binding proteins have been recently determined. The collagen-specific chaperone heat-shock protein 47 (Hsp47) is essential for the correct assembly and maturation of collagen triple helices in the endoplasmic reticulum [219]. Crystal structures of Hsp47 in complex with collagen peptides containing a PRG triplet show a 2:1 stoichiometry (Figure 10; PDB codes 4AU2, 4AU3 and 3ZHA) where two Hsp47 molecules bind in a head-to-head fashion to two sites on the homotrimer peptides. Each Hsp47 molecule makes extensive contacts with the leading or trailing strands. The Arg residues from these strands form salt bridges with Asp residues from separate Hsp47 molecules (Figure 10E) [182].

Structural analyses of collagen recognition by other domains include crystal and NMR structures of collagen peptides in complex with Fab fragments from arthritogenic autoantibodies (PDB codes 2Y5T and 4BKL) [220,221], CUB domains from molecules of the innate immune system (PDB codes 3POB and 4LOR) [222,223], metalloproteinase domains (PDB codes 4AUO and 2MQS) [224,225] or the collagen-binding protein CNA from *Staphylococcus aureus* (PDB code 2F6A) [226]. The Fab–collagen complexes show how antibodies recognize epitopes from type II collagen in its native triple helical conformation. In these structures the peptides fill completely the binding clefts with extensive van der Waals, direct and water-mediated hydrogen bonding and salt-bridge interactions between collagen and Fab residues [220,221]. The CUB–collagen complexes show an interesting interaction where one Lys residue from the collagen peptide forms salt bridges with acidic side chains involved in the co-ordination of a Ca^{2+} ion, reminiscent of the interaction between collagen and $\alpha 1/\alpha 2$ integrin I-domains [41], although the Lys residue does not participate in metal co-ordination.

Most complexes between collagen triple helical peptides and their binding partners show the triple helix placed ‘on top’ of the binding domain (Figures 10A and 10B), rather than

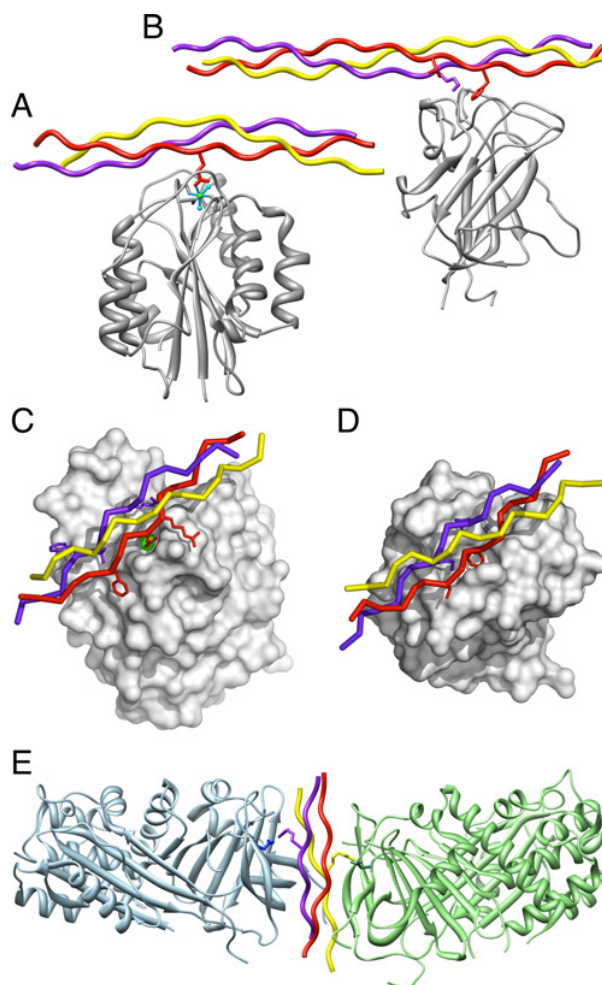


Figure 10 Structures of complexes between collagen model peptides and collagen-binding domains

(A) Ribbon diagram of the integrin $\alpha 2$ I-domain bound to a collagen peptide with the integrin-binding motif GFOGER (PDB code 1DZI) [41]. Trailing chain is shown in purple, middle chain is shown in red, and leading chain is shown in yellow. The Glu side chain from the middle strand completes the co-ordination of a divalent metal ion in the integrin I-domain. (B) Ribbon diagram of the discoidin domain of human DDR2 bound to a collagen peptide with the DDR2-binding motif GVMGFO, where Met has been replaced by the near-isosteric norleucine (Nle) (2WUH) [218]. Strand colours as in (A). Hydrophobic residues Val and Phe from the middle chain and Nle from the trailing chain are shown interacting with the discoidin domain. (C) The Glu side chain of the GFOGER motif binds to a pocket on the $\alpha 2$ I-domain surface containing the divalent cation at the bottom (in green). Other side chains from the collagen peptide interacting with the $\alpha 2$ I-domain are shown. (D) The Phe and Nle side chains from the GVMGFO motif bind to a pocket in the surface of the DDR2 discoidin domain. Other side chains interacting with DDR2 are shown. (E) Ribbon diagram of the complex between two Hsp47 molecules and a collagen peptide (PDB code 4AU2) [182]. Lys residues on the collagen peptide form salt-bridges with an Asp residues on each Hsp47 molecule. Images in (A, B and E) were prepared with UCSF Chimera [253]. Images in (C and D) were prepared with PyMol (<http://www.pymol.org>).

being surrounded by it. Two complexes show 2:1 stoichiometries [182,213], where one triple helical peptide forms a complex with two partners exploiting the repetition of binding sites on homotrimer collagens (Figure 10E). A special case is the structure of a collagen peptide with the collagen-binding protein CNA from *Staphylococcus aureus* where the triple helix is ‘hugged’ by the different subdomains of the bacterial protein [226]. Thus, binding often involves predominantly two of the three strands engaging in van der Waals, hydrogen bonding and salt-bridge

interactions with residues on the surface of the binding partner. Nevertheless, participation of key residues from different chains demonstrates the need for a triple helical structure in these interactions. Some of the collagen triple helices show appreciable bending upon complex formation, whereas others remain quite straight (Figure 10). It is possible that lattice interactions also have an effect on either bending the triple helices or keeping them straight. When the structure of the collagen peptide on its own is also available, it is possible to compare whether there are more subtle structural variations upon binding, as for instance a variation in the internal superhelical twist (Figure 4). However, the resolution of the complexes is often significantly lower than that of the isolated peptides and no general conclusions about changes in superhelical twist can be extracted.

In complexes where there is sufficient resolution to visualize the positions of bound water molecules (PDB codes 1DZI, 2WUH, 2Y5T and 3POB), the water-bonded model is preserved: water molecules bound to the amide groups of amino acids at the X position form additional hydrogen bonds to the carbonyl groups of Gly residues on the preceding chain and, if available, to hydroxy groups from Hyp residues ($\zeta 1$ or $\zeta 1-\gamma 1$ bridges, as in Figure 5) [208]. These water molecules and hydrogen bonding topology are preserved even in the regions of contact between the collagen triple helix and its binding partner, and their network of hydrogen bonds often extends to residues on the surface of the partner (Supplementary Figure S1). Conservation of these water molecules even upon complex formation reinforces the idea that they are an intrinsic feature of the triple helical structure of collagen [208]. Even when water molecules are not visible due to the low resolution of the structures, it is safe to assume that the hydration positions most consistently observed in high-resolution structures may remain occupied upon complex formation. Modelling waters into these positions often suggests additional water-mediated hydrogen bonding between the collagen triple helix and its partner. Molecular dynamics simulations of water molecules around collagen–ligand complexes and their separate components have been used to identify structured water molecules that mediate interactions between the two partners [227]. These water molecules show significant residence times and binding energies, and maintain hydrogen bonding topologies equivalent to those of the isolated collagen triple helix. These molecular dynamics simulations strongly support the notion that hydration on the surface of the triple helix also contributes to collagen molecular recognition [227].

C–H...O=C HYDROGEN BONDING

Conventional views on hydrogen bonding state that both the donor and acceptor atoms must be highly electronegative (O, N, F). There is evidence, however, that hydrogen bonds are possible where one or both of the atoms or groups involved have moderate or low electronegativity. These *weak hydrogen bonds* can occur, for instance, with carbon atoms acting as donors or aromatic rings acting as acceptors [228]. The possibility of hydrogen bonding from $C\alpha$ atoms to carbonyl groups ($C\alpha-H\cdots O=C$) in proteins was already a subject of interest at the time of the fibre diffraction studies on collagen, and it was suggested for the PGII and collagen structures on the basis of model building [39,46,229]. The crystal structure of the Gly→Ala peptide provided experimental evidence for two repetitive patterns of $C\alpha-H\cdots O=C$ hydrogen bonding in the collagen triple helix [230] (Figure 11). One set of hydrogen bonds occurs between the $C\alpha$ of the amino acid in the Y position and the C=O group in the X position of the following strand,

acting as companions to the conventional $N-H\cdots O=C$ interstrand hydrogen bonding. This tandem topology is exactly the same as seen ubiquitously in the β -sheet structure [231]. A second set occurs between the $C\alpha$ of Gly and the C=O groups of Gly and the amino acid in the X position both from the preceding strand, in opposite direction to the $N-H\cdots O=C$ bonds. The geometry of the $Gly-C\alpha-H\cdots O=C$ interactions is indicative of a bifurcated and three-centred hydrogen bonding configuration (Figure 11B), which is a characteristic feature in crystal structures with deficiency of hydrogen bonding donor groups [232].

Weak hydrogen bonds are not so weak after all. The strength of $C\alpha-H\cdots OH_2$ hydrogen bonds has been calculated as about half of that of the water dimer [233], and *ab initio* quantum calculations suggest that $N-H\cdots O$ and $C\alpha-H\cdots O$ hydrogen bonds have a similar contribution to the stability of β -sheets [234]. Thus, given their large numbers in proteins, $C\alpha-H\cdots O$ hydrogen bonds in particular should be considered as important factors in protein structure and stability. They have also been reported to contribute to the stability of protein–protein interfaces and to macromolecular recognition [235–237]. In the collagen case, interstrand $C\alpha-H\cdots O=C$ hydrogen bonds are another characteristic feature of the collagen triple helix. They alleviate the problem of shortage of available hydrogen bonding donors (only one direct $N-H\cdots O=C$ hydrogen bond per triplet) and define an ensemble of stabilizing interactions at the inner core of the triple helix.

Structural analysis of $C\alpha-H\cdots O$ hydrogen bonding offers strategic paths for molecular engineering. The concept of isostructurality exploits the replacement of weaker C–H donor groups with other chemical groups with stronger hydrogen bonding strength, such as N–H. This strategy has been recently applied to collagen successfully. In an elegant study, the introduction of an *aza*-Gly residue in the middle of typical collagen model peptides resulted in hyperstability and faster folding of the triple helix [238]. Molecular dynamics simulations suggest that substitution of the CH_2 group with a NH group maintained the same overall hydrogen bonding, with the C=O groups of *aza*-Gly and Pro residues from the previous chain acting as acceptors (Figure 11C).

COLLAGEN INTERRUPTIONS

The –Gly-X-Y– repetitive pattern is not strictly conserved in the majority of collagens and collagen-like proteins. Interruptions result from substitution of the invariant Gly residues or from removing or adding additional residues in between Gly residues, thus breaking the repetitive sequence. Although these interruptions seem to be an intrinsic part of collagen structure, they are not tolerated in the long COL×3 domains of fibrillar collagens. Missense mutations in these domains where a single Gly residue is replaced with another amino acid result in connective tissue disorders such as osteogenesis imperfecta [1,239], and their clinical severity depends on the substitute residue, the local environment of the substitution and its proximity to collagen interaction sites [101,240,241]. It is thought that interruptions in the COL×3 domains of fibrillar collagens are incompatible with the correct formation of the supramacromolecular fibril assemblies, whereas in non-fibrillar collagens they may have some functional role [242].

Interruptions of the –Gly-X-Y– pattern can range from a few residues to sequences long enough to include entire domains, although in the latter case they are not referred as collagen interruptions. The boundary between these two scenarios is necessarily blurred and could be arbitrarily placed around the

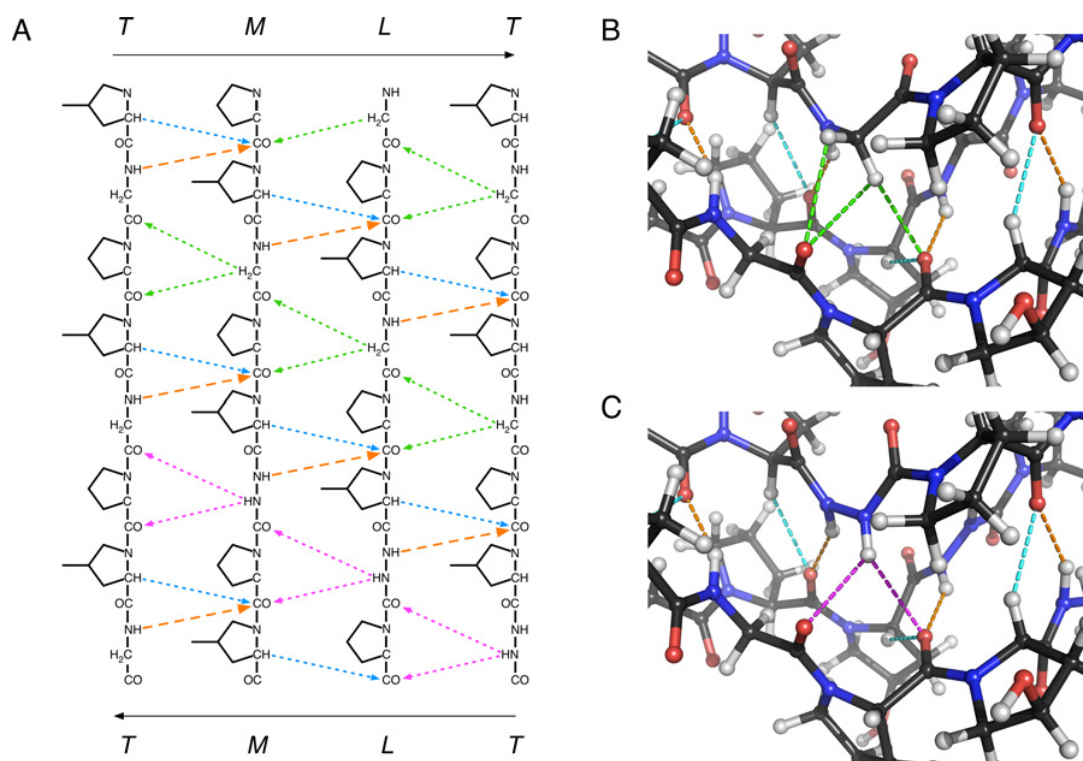


Figure 11 Interstrand hydrogen bonding in the collagen triple helix

(A) Topology diagram, with chains labelled *T*, *M* and *L* as in Figure 5 and with the *T* chain repeated on the right. Conventional N—H...O=C hydrogen bonds (orange) follow a *TMLT* direction. C α —H...O=C hydrogen bonds from the residue in the Y position (cyan) are topologically equivalent to those seen ubiquitously in β -sheets (see the text) and also follow the *TMLT* direction. Gly—C α —H...O=C hydrogen bonds (green) are specific to the collagen triple helix and follow the opposite *TLMT* direction [230]. Engineering of aza-Gly residues in place of Gly [238] probably results in a second set of N—H hydrogen bonds (magenta) that is isostructural with the weak hydrogen bonding from Gly residues and also follows the *TLMT* direction. (B) 3D view of the hydrogen bonds in the collagen triple helix (same colour scheme as in A). Only one set of Gly—C α —H...O=C bonds is shown (green). (C) The same 3D view with an aza-Gly residue modelled in place of the central Gly residue (as in [238]), showing only one set of isostructural hydrogen bonds (magenta). Images in (B) and (C) were prepared with PyMol (<http://www.pymol.org>).

15–20 residues. A nomenclature system (Table 4) has been developed for classification purposes [242,243]. Thus, a typical collagen sequence —Gly-X-Y-Gly-Z-Gly-X-Y-Gly— where a single residue is missing is termed a G1G interruption. Single Gly→Z substitutions in the middle of a collagen sequence result in the altered sequence —Gly-X-Y-Gly-X-Y-Z-X-Y-Gly-X-Y-Gly—, which is equivalent to a G5G interruption. Interruptions in heterotrimer collagens can be classified as commensurate or incommensurate, depending on the relative register of the —Gly-X-Y— pattern at both sides of the interruption (Figure 12). A census of collagen interruptions in the collagenomes from prokaryotes to humans has been recently reported [243]. This census shows a predominance of shorter interruptions in general, but with significant differences in their distribution across different taxonomic groups (Figure 12).

Collagen interruptions must impose a local discontinuity or distortion on the triple helical structure with consequences to its stability at the site of the interruption. Again, collagen model peptides are a useful vehicle to address these questions biochemically [103]. Only two crystal structures of collagen peptides with interruptions have been determined so far. The Gly→Ala peptide (PDB codes 1CAG and 1CGD), with sequence (POG)₃-POGPOAPOG-(POG)₄, is an example of a G5G interruption (underlined). The replacement of a consensus Gly with an Ala residue results in a small bulge at the Ala site, replacement of three N—H...O=C hydrogen bonds with water-

Table 4 G n G nomenclature for the description and classification of the different types of collagen interruptions [243] and selected examples from human type IV collagen sequences

Residues forming the interruption are underlined. Numbering of the first and last residue of each collagen IV sequence corresponds to that of the complete precursor sequence (available from SwissProt).

Type	Motif	Collagen	Sequence
G0G	Gly-Gly	α_5 (IV)	G ²⁷³ IRGPPGPPGGEKGEKGEQG ²⁹²
G1G	Gly-X-Gly	α_3 (IV)	G ⁶⁸⁹ PDGEPGIPGIGFPGPPGPKG ⁷⁰⁹
G3G	Gly-X ₃ -Gly	α_2 (IV)	G ⁴³⁸ LPGPPGPDGFLFLKGAKGAG ⁴⁶⁰
G4G	Gly-X ₄ -Gly	α_5 (IV)	G ³⁸⁰ PPGLPGPPGAAVMGPPGPPGFP ⁴⁰³
G5G	Gly-X ₅ -Gly	α_6 (IV)	G ³⁸⁰ LRGPGVPLPALSGLVPGALGPQ ⁴⁰⁴
G6G	Gly-X ₆ -Gly	α_1 (IV)	G ⁶³⁰ LPGPKGEPGKIVLP ⁶⁵⁵ PPGAEPLG ⁶⁵⁵
G7G	Gly-X ₇ -Gly	α_4 (IV)	G ⁹⁴⁴ EPGLGPPGPMDPNLLGSKGEKGE ⁹⁷⁰
G8G	Gly-X ₈ -Gly	α_5 (IV)	G ¹⁵⁰ IPGMKGEPSIIMSSLLPGKGNPGY ¹⁷⁷
G10G	Gly-X ₁₀ -Gly	α_4 (IV)	G ³⁴⁷ NRGHPGPPGVLTPTPLPLKGG ³⁷⁶ PPGDFPG ³⁷⁶
G12G	Gly-X ₁₂ -Gly	α_2 (IV)	G ³³⁸ EAGDPGPPGLPAYSPHSLAKGARGD ³⁶⁹ PPG ³⁶⁹

mediated hydrogen bonds (ζ 1 bridges) and local untwisting at the replacement site (Figure 13) [55,66,73,244]. The single Gly→Ala replacement brings a dramatic decrease in thermal stability compared with the parent peptide (POG)₁₀ [245]. The Hyp^c peptide (PDB code 1EI8), with sequence (POG)₃-POGPG-(POG)₅ is an example of a G1G interruption where one residue

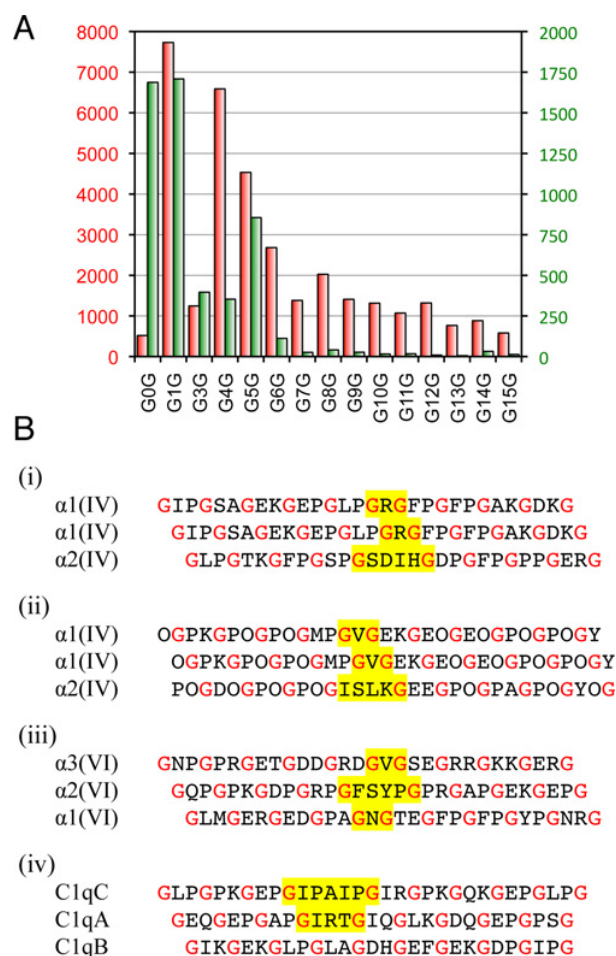


Figure 12 Interruptions in the collagen triple helix

(A) Distribution of different types of interruption in metazoan (red bars, left axis) or prokaryotic and viral (green bars, right axis) sequences of collagens and collagen-like proteins deposited in the UniProt database. (B) Examples of commensurate and incommensurate interruptions in heterotrimer collagens. GnG motifs are highlighted in yellow and Gly residues are shown in red type. (i) Commensurate site in human type IV collagen with two G1G interruptions and one G4G interruption, showing the chain register predicted in [243]. (ii) Observed chain register in a heterotrimer collagen peptide modelling a different commensurate site in human collagen IV with the same types of interruptions [251]. (iii) Commensurate site with three different chains in human collagen VI and the same types of interruptions, showing the chain register predicted in [243]. (iv) Incommensurate site in the human collagen-like protein C1q. Chains C1qC and C1qA have a G5G and a G3G interruption respectively, and chain C1qB is uninterrupted. Chain register shown is as proposed in [254].

has been removed from the consensus collagen triplet. The structure of this peptide shows local disruption of the triple helical structure and unusual Gly–N–H...O=C–Gly hydrogen bonding, with consequences for the spatial relation between the segments at either side of the interruption. The pattern of hydrogen bonding topology is resumed at both sides of the interruption, but the two sets of hydrogen bonds are out of phase [66,243,246]. From the hydrogen bonding topologies observed in these crystal structures it is possible to define ideal cases where the hydrogen bonding connectivity, including water-mediated hydrogen bonds, would be maintained with minimal disruptions to the structure [243]. However, NMR studies combined with molecular dynamics simulations suggest a more complex scenario where other unusual hydrogen bonding arrangements may occur [241,242,247–250].

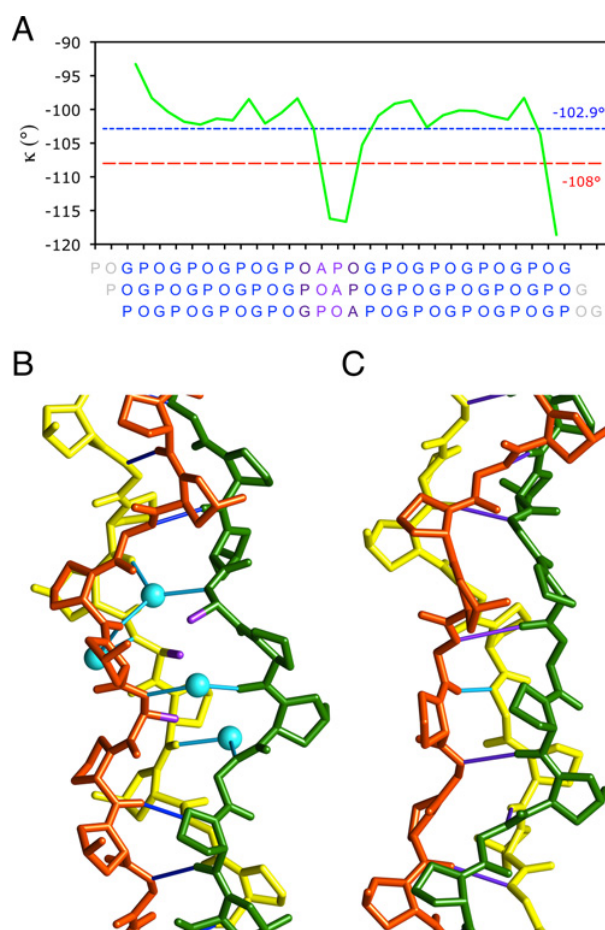


Figure 13 Structural impact of interruptions

(A) Variation of the unit twist (κ) for the collagen model peptide Gly→Ala showing the effect of a G5G interruption (Gly→Ala substitution). A sudden and local untwisting (more negative values of κ) occurs over the helical triplets that contain the Gly→Ala substitution. The two $(POG)_n$ regions flanking the substitution show an average κ of -101° indicating a small overtwisting with respect to the typical 7_5 helix characteristic of imino-saturated stretches. (B) Detail of the crystal structure of the Gly→Ala peptide (PDB code 1CGD) showing how the three side chains of the Ala residues (in purple) point directly towards the centre of the triple helix, causing it to bulge slightly. Four water molecules (cyan spheres) mediate interstrand hydrogen bonding (cyan sticks) around the Ala side chains. Normal hydrogen bonding (blue sticks) is resumed at both sides of the interruption. (C) Detail of the crystal structure of the Hyp⁺ peptide (PDB code 1E18) showing normal interstrand hydrogen bonding (purple sticks) around the interruption. An unusual Gly–N–H...O=C–Gly hydrogen bond occurs at the site of interruption (cyan stick). The helical path of the interstrand hydrogen bonding is locally reversed at the site of interruption and resumed afterwards. Images in (C) and (D) were prepared with UCSF Chimera [253].

These studies also show that the conformation of the triple helix is more flexible at the sites of interruption, and even the crystal structures show evidence of higher flexibility at these specific sites (higher temperature B -factors) [246]. Thus, it is very likely that interrupted collagens in solution have a more dynamic structure and may have local conformations, hydrogen bonding and water-mediated bridges that differ from the ones seen in only two crystal structures of interrupted collagen peptides determined so far. More 3D structures, both from crystallographic and NMR analyses, are needed to achieve a more comprehensive understanding of collagen conformation at the sites of interruption.

The G1G interruption had an even greater negative impact on the stability of the Hyp⁺ peptide than the corresponding G5G

interruption in the Gly→Ala peptide [245]. Yet, the peptide of sequence (GPO)₄-GAAVMGPO-(GPO)₃ (GAAVM peptide), with a G4G interruption, is stable at room temperature even though this G4G interruption is commensurate with the G1G interruption of the Hyp^β peptide [247]. Recently, two peptides with a G5G interruption where the same residues are arranged in different order, (GPO)₅-GPOALOG/GLOAPOG-PO-(GPO)₃, showed significant differences in stability, conformation and flexibility [250]. It follows that not all interruption types have the same impact on collagen stability and that the actual sequences in or around the interruption are also important. Intriguingly, sequence conservation analyses on collagen interruptions show certain patterns of preferences for the residues inside or surrounding the interruptions [242,243]. Understanding the intricate relation between collagen interruptions, their sequence preferences and the impact on stability, conformation and flexibility is critical to understand their biological role, but there are also practical considerations. Interruption sites could be incorporated into engineered collagen-based polymers to introduce flexibility points or sites for interaction with specific receptors. Furthermore, it has been suggested that collagen interruptions may impose some preference on the chain register adopted by heterotrimer collagens [243]. Some support for this hypothesis has been obtained from the recent analysis of a heterotrimer peptide modelling a commensurate interruption in type IV collagen, with two chains with a G1G interruption (GVG) and one chain with a G4G interruption (GISLKG). This study confirms the predicted register although the observed hydrogen bonding topology may be different from the one originally suggested [251]. The study provides yet another example that it may be possible to control the formation of collagen heterotrimers using information embedded in the collagen sequence alone. Thus, collagen interruptions as a whole should not be simply considered as ‘obstacles’ in the sense commonly associated to the pathological mutations of fibrillar collagens, but as elements that can have functional roles in guiding collagen heterotrimer chain register [251] or in influencing the formation of supramolecular arrangements specific to the interrupted collagens.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Our understanding of the structure and stability of the collagen triple helix has improved significantly since the early fibre diffraction models from 50 years ago. High-resolution structural determinations and biochemical characterization of collagen model peptides have largely established the relative contributions to stability of every amino acid type, at least for homotrimer triple helices. Non-natural proline derivatives have led to the synthesis of hyperstable collagens and have improved our understanding of the fundamental mechanisms of collagen stabilization by hydroxyproline.

Collagen triple helical domains outside metazoa have amino acid compositions very different from those in vertebrates. They demonstrate that there are alternative mechanisms for building stable collagen triple helices in the absence of prolyl hydroxylation. These mechanisms have been used for the design of stable collagen mimics and have opened avenues for the engineering of heterotrimer collagens and self-assembling synthetic collagen peptides. Prokaryotic collagens will facilitate the development of bacterial expression systems for cost-effective production of stable collagen-based polymers. These recombinant collagens could be designed with sequence motifs targeting specific collagen receptors or collagen-binding proteins, with the

aim of producing artificial extracellular matrices for biomedical applications.

Use of peptide libraries has advanced enormously our understanding of collagen molecular recognition. Given the largely linear structure of collagen domain, it has become possible to map several ligand recognition sites on 2D representations of the collagen fibril. Furthermore, the atomic details of the interaction between collagen triple helices and several collagen-binding proteins or receptors are now known. This information will be invaluable to improve our understanding of collagen biology.

Challenges remain at the basic molecular level to understand the triple helical conformation at the sites of interruption, which are very varied and numerous in non-fibrillar collagens. The effect of incommensurate interruptions in molecular structure and chain register is largely unknown, and the possible impact of commensurate interruptions in heterotrimer chain selection deserves investigation. Interruptions could also be exploited as sites of local flexibility in recombinant collagen-based biomaterials.

Finally, a main challenge is the development of methods for the synthesis of hierarchical collagen assemblies that mimic collagen fibrils, networks, basement membranes or other supramolecular organizations. This will require a much deeper understanding of the mechanisms of collagen assembly and organization and is critical for future tissue regeneration interventions. Connective tissues differ in composition and organization. Thus, it is essential to elucidate the mechanisms that lead to the generation of different structures with distinct features and functional properties of each tissue. In particular, our understanding of the dynamic interactions of collagen with other biomolecules has to improve. These advances will also result in a better understanding of collagen pathologies and will make possible a new generation of biomimetic biomaterials.

REFERENCES

- Myllyharju, J. and Kivirikko, K.I. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **20**, 33–43 [CrossRef PubMed](#)
- Kadler, K.E., Baldock, C., Bella, J. and Boot-Handford, R.P. (2007) Collagens at a glance. *J. Cell Sci.* **120**, 1955–1958 [CrossRef PubMed](#)
- Gordon, M.K. and Hahn, R.A. (2010) Collagens. *Cell Tissue Res.* **339**, 247–257 [CrossRef PubMed](#)
- Ricard-Blum, S. (2011) The collagen family. *Cold Spring Harb. Perspect. Biol.* **3**, a004978 [CrossRef PubMed](#)
- Mienaltowski, M.J. and Birk, D.E. (2014) Structure, physiology, and biochemistry of collagens. *Adv. Exp. Med. Biol.* **802**, 5–29 [CrossRef PubMed](#)
- Johnstone, I.L. (2000) Cuticle collagen genes. Expression in *Caenorhabditis elegans*. *Trends Genet.* **16**, 21–27 [CrossRef PubMed](#)
- Page, A.P. and Winter, A.D. (2003) Enzymes involved in the biogenesis of the nematode cuticle. *Adv. Parasitol.* **53**, 85–148 [CrossRef PubMed](#)
- Page, A.P. and Johnstone, I.L. (2007) The Cuticle. *WormBook 2007*, 1–15 [PubMed](#)
- Heino, J., Huhtala, M., Kapyla, J. and Johnson, M.S. (2009) Evolution of collagen-based adhesion systems. *Int. J. Biochem. Cell Biol.* **41**, 341–348 [CrossRef PubMed](#)
- Exposito, J.Y., Valcourt, U., Cluzel, C. and Lethias, C. (2010) The fibrillar collagen family. *Int. J. Mol. Sci.* **11**, 407–426 [CrossRef PubMed](#)
- Legay, C. (2000) Why so many forms of acetylcholinesterase? *Microsc. Res. Tech.* **49**, 56–72 [CrossRef PubMed](#)
- Goldstein, B.J., Scalia, R.G. and Ma, X.L. (2009) Protective vascular and myocardial effects of adiponectin. *Nat. Clin. Pract. Cardiovasc. Med.* **6**, 27–35 [CrossRef PubMed](#)
- Gupta, G. and Suroliya, A. (2007) Collectins: sentinels of innate immunity. *BioEssays* **29**, 452–464 [CrossRef PubMed](#)
- Kouser, L., Madhukaran, S.P., Shastri, A., Saraon, A., Ferluga, J., Al-Mozaini, M. and Kishore, U. (2015) Emerging and novel functions of complement protein C1q. *Front. Immunol.* **6**, 317 [CrossRef PubMed](#)
- Bowdish, D.M. and Gordon, S. (2009) Conserved domains of the class A scavenger receptors: evolution and function. *Immunol. Rev.* **227**, 19–31 [CrossRef PubMed](#)

- 16 Fullerton, G.D. and Rahal, A. (2007) Collagen structure: the molecular source of the tendon magic angle effect. *J. Magn. Reson. Imaging* **25**, 345–361 [CrossRef PubMed](#)
- 17 Astbury, W.T. (1938) X-rays adventures among the proteins. *Trans. Faraday Soc.* **34**, 378–388 [CrossRef](#)
- 18 Harrington, W.F. and Von Hippel, P.H. (1961) The structure of collagen and gelatin. *Adv. Protein Chem.* **16**, 1–138 [CrossRef PubMed](#)
- 19 Fraser, R.D.B. and MacRae, T.P. (1973) Collagens. In *Conformation of Fibrous Proteins and Related Synthetic Polypeptides*, pp. 344–402, Academic Press, New York
- 20 Astbury, W.T. and Bell, F.O. (1940) Molecular structure of the collagen fibres. *Nature* **145**, 421–422 [CrossRef](#)
- 21 Pauling, L. and Corey, R.B. (1951) The structure of fibrous proteins of the collagen-gelatin group. *Proc. Natl. Acad. Sci. U.S.A.* **37**, 272–281 [CrossRef PubMed](#)
- 22 Ramachandran, G.N. and Kartha, G. (1954) Structure of collagen. *Nature* **174**, 269–270 [CrossRef PubMed](#)
- 23 Ramachandran, G.N. and Kartha, G. (1955) Structure of collagen. *Nature* **176**, 593–595 [CrossRef PubMed](#)
- 24 Rich, A. and Crick, F.H.C. (1955) The structure of collagen. *Nature* **176**, 915–916 [CrossRef PubMed](#)
- 25 Rich, A. and Crick, F.H. (1961) The molecular structure of collagen. *J. Mol. Biol.* **3**, 483–506 [CrossRef PubMed](#)
- 26 Cowan, P.M., McGavin, S. and North, A.C. (1955) The polypeptide chain configuration of collagen. *Nature* **176**, 1062–1064 [CrossRef PubMed](#)
- 27 Boedtker, H. and Doty, P. (1956) The native and denatured states of soluble collagen. *J. Am. Chem. Soc.* **78**, 4267–4280 [CrossRef](#)
- 28 Rice, R.V. (1960) Reappearance of certain structural features of native collagen after thermal transformation. *Proc. Natl. Acad. Sci. U.S.A.* **46**, 1186–1194 [CrossRef PubMed](#)
- 29 Lewis, M.S. and Piez, K.A. (1964) Sedimentation-equilibrium studies of the molecular weight of single and double chains from rat-skin collagen. *Biochemistry* **3**, 1126–1131 [CrossRef PubMed](#)
- 30 Subramanian, E. (2001) G.N. Ramachandran. *Nat. Struct. Biol.* **8**, 489–491 [CrossRef PubMed](#)
- 31 Balaran, P. (2004) Revisiting an old triumph. *Curr. Sci.* **87**, 549–550
- 32 Bhattacharjee, A. and Bansal, M. (2005) Collagen structure: the Madras triple helix and the current scenario. *IUBMB Life* **57**, 161–172 [CrossRef PubMed](#)
- 33 Sarma, R. (1998) Ramachandran: a Biography of Gopalasamudram Narayana Ramachandran, the Famous Indian Biophysicist. Adenine Press, New York.
- 34 Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 [CrossRef PubMed](#)
- 35 Brodsky, B., Eikenberry, E.F., Belbruno, K.C. and Sterling, K. (1982) Variations in collagen fibril structure in tendons. *Biopolymers* **21**, 935–951 [CrossRef PubMed](#)
- 36 Fraser, R.D., MacRae, T.P. and Suzuki, E. (1979) Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463–481 [CrossRef PubMed](#)
- 37 Cowan, P.M. and McGavin, S. (1955) Structure of poly-L-proline. *Nature* **176**, 501–503 [CrossRef](#)
- 38 Crick, F.H. and Rich, A. (1955) Structure of polyglycine II. *Nature* **176**, 780–781 [CrossRef PubMed](#)
- 39 Ramachandran, G.N., Sasisekharan, V. and Ramakrishnan, C. (1966) Molecular structure of polyglycine II. *Biochim. Biophys. Acta* **112**, 168–170 [CrossRef PubMed](#)
- 40 Adzhubei, A.A., Sternberg, M.J. and Makarov, A.A. (2013) Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **425**, 2100–2132 [CrossRef PubMed](#)
- 41 Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J. and Liddington, R.C. (2000) Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* **101**, 47–56 [CrossRef PubMed](#)
- 42 Tormo, J., Puiggali, J., Vives, J., Fita, I., Lloveras, J., Bella, J., Aymami, J. and Subirana, J.A. (1992) Crystal structure of a helical oligopeptide model of polyglycine II and of other polyamides: acetyl-(glycyl- β -alanyl)₂-NH propyl. *Biopolymers* **32**, 643–648 [CrossRef PubMed](#)
- 43 Bella, J., Puiggali, J. and Subirana, J.A. (1994) Glycine residues induce a helical structure in polyamides. *Polymer* **35**, 1291–1297 [CrossRef](#)
- 44 Shoulders, M.D. and Raines, R.T. (2009) Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958 [CrossRef PubMed](#)
- 45 Parmar, A.S., Nunes, A.M., Baum, J. and Brodsky, B. (2012) A peptide study of the relationship between the collagen triple-helix and amyloid. *Biopolymers* **97**, 795–806 [CrossRef PubMed](#)
- 46 Ramachandran, G.N. (1967) Structure of collagen at the molecular level. In *Treatise on Collagen* (Ramachandran, G.N., ed.), pp. 103–183, Academic Press, New York
- 47 Sakakibara, S., Kishida, Y., Kikuchi, Y., Sakai, R. and Kakiuchi, K. (1968) Synthesis of poly-(L-prolyl-L-prolylglycyl) of defined molecular weights. *Bull. Chem. Soc. Jpn.* **41**, 1273 [CrossRef](#)
- 48 Berg, R.A., Olsen, B.R. and Prockop, D.J. (1970) Titration and melting curves of the collagen-like triple helices formed from (Pro-Pro-Gly) in aqueous solution. *J. Biol. Chem.* **245**, 5759–5763 [PubMed](#)
- 49 Kobayashi, Y., Sakai, R., Kakiuchi, K. and Isemura, T. (1970) Physicochemical analysis of (Pro-Pro-Gly)_n with defined molecular weight—temperature dependence of molecular weight in aqueous solution. *Biopolymers* **9**, 415–425 [CrossRef PubMed](#)
- 50 Okuyama, K., Tanaka, N., Ashida, T., Kakudo, M. and Sakakibara, S. (1972) An X-ray study of the synthetic polypeptide (Pro-Pro-Gly)₁₀. *J. Mol. Biol.* **72**, 571–576 [CrossRef PubMed](#)
- 51 Sakakibara, S., Kishida, Y., Okuyama, K., Tanaka, N., Ashida, T. and Kakudo, M. (1972) Single crystals of (Pro-Pro-Gly)₁₀, a synthetic polypeptide model of collagen. *J. Mol. Biol.* **65**, 371–373 [CrossRef PubMed](#)
- 52 Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y. and Prockop, D.J. (1973) Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **303**, 198–202 [CrossRef PubMed](#)
- 53 Engel, J., Chen, H.T., Prockop, D.J. and Klump, H. (1977) The triple helix in equilibrium with coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)_n and (L-Pro-L-Hyp-Gly)_n. *Biopolymers* **16**, 601–622 [CrossRef PubMed](#)
- 54 Okuyama, K., Arnott, S., Takayanagi, M. and Kakudo, M. (1981) Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)₁₀. *J. Mol. Biol.* **152**, 427–443 [CrossRef PubMed](#)
- 55 Bella, J., Eaton, M., Brodsky, B. and Berman, H.M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75–81 [CrossRef PubMed](#)
- 56 Berisio, R., Vitagliano, L., Mazzarella, L. and Zagari, A. (2002) Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)₁₀]₃. *Protein Sci.* **11**, 262–270 [CrossRef PubMed](#)
- 57 Hongo, C., Noguchi, K., Okuyama, K., Tanaka, Y. and Nishino, N. (2005) Repetitive interactions observed in the crystal structure of a collagen-model peptide, [(Pro-Pro-Gly)₃]₃. *J. Biochem.* **138**, 135–144 [CrossRef PubMed](#)
- 58 Okuyama, K., Miyama, K., Mizuno, K. and Bachinger, H.P. (2012) Crystal structure of (Gly-Pro-Hyp)₃: implications for the collagen molecular model. *Biopolymers* **97**, 607–616 [CrossRef PubMed](#)
- 59 Fraser, R.D., MacRae, T.P., Miller, A. and Suzuki, E. (1983) Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167**, 497–521 [CrossRef PubMed](#)
- 60 Okuyama, K., Takayanagi, M., Ashida, T. and Kakudo, M. (1977) A new structural model for collagen. *Polym. J.* **9**, 341–343 [CrossRef](#)
- 61 Okuyama, K., Xu, X., Iguchi, M. and Noguchi, K. (2006) Revision of collagen molecular structure. *Biopolymers* **84**, 181–191 [CrossRef PubMed](#)
- 62 Okuyama, K. (2008) Revisiting the molecular structure of collagen. *Connect. Tissue Res.* **49**, 299–310 [CrossRef PubMed](#)
- 63 Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B. and Berman, H.M. (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6**, 454–457 [CrossRef PubMed](#)
- 64 Kramer, R.Z., Bella, J., Brodsky, B. and Berman, H.M. (2001) The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol.* **311**, 131–147 [CrossRef PubMed](#)
- 65 Boudko, S.P., Engel, J., Okuyama, K., Mizuno, K., Bachinger, H.P. and Schumacher, M.A. (2008) Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J. Biol. Chem.* **283**, 32580–32589 [CrossRef PubMed](#)
- 66 Bella, J. (2010) A new method for describing the helical conformation of collagen: dependence of the triple helical twist on amino acid sequence. *J. Struct. Biol.* **170**, 377–391 [CrossRef PubMed](#)
- 67 Orgel, J.P., Persikov, A.V. and Antipova, O. (2014) Variation in the helical structure of native collagen. *PLoS One* **9**, e89519 [CrossRef PubMed](#)
- 68 Yee, R.Y., Englander, S.W. and Von Hippel, P.H. (1974) Native collagen has a two-bonded structure. *J. Mol. Biol.* **83**, 1–16 [CrossRef PubMed](#)
- 69 Privalov, P.L., Tiktupulo, E.I. and Tischenko, V.M. (1979) Stability and mobility of the collagen structure. *J. Mol. Biol.* **127**, 203–216 [CrossRef PubMed](#)
- 70 Privalov, P.L. (1982) Stability of proteins. Proteins which do not present a single cooperative system. *Adv. Protein Chem.* **35**, 1–104 [CrossRef PubMed](#)
- 71 Ramachandran, G.N. and Chandrasekharan, R. (1968) Interchain hydrogen bonds via bound water molecules in the collagen triple helix. *Biopolymers* **6**, 1649–1658 [CrossRef PubMed](#)
- 72 Ramachandran, G.N., Bansal, M. and Bhatnagar, R.S. (1973) A hypothesis on the role of hydroxyproline in stabilizing collagen structure. *Biochim. Biophys. Acta* **322**, 166–171 [CrossRef PubMed](#)
- 73 Bella, J., Brodsky, B. and Berman, H.M. (1995) Hydration structure of a collagen peptide. *Structure* **3**, 893–906 [CrossRef PubMed](#)
- 74 Kramer, R.Z., Venugopal, M.G., Bella, J., Mayville, P., Brodsky, B. and Berman, H.M. (2000) Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J. Mol. Biol.* **301**, 1191–1205 [CrossRef PubMed](#)

- 75 Fan, P., Li, M.H., Brodsky, B. and Baum, J. (1993) Backbone dynamics of (Pro-Hyp-Gly)₁₀ and a designed collagen-like triple-helical peptide by ¹⁵N NMR relaxation and hydrogen-exchange measurements. *Biochemistry* **32**, 13299–13309 [CrossRef PubMed](#)
- 76 Traub, W. and Piez, K.A. (1971) The chemistry and structure of collagen. *Adv. Protein Chem.* **25**, 243–352 [CrossRef PubMed](#)
- 77 Masic, A., Bertinetti, L., Schuetz, R., Chang, S.W., Metzger, T.H., Buehler, M.J. and Fratzl, P. (2015) Osmotic pressure induced tensile forces in tendon collagen. *Nat. Commun.* **6**, 5942 [CrossRef PubMed](#)
- 78 Berendsen, H.J.C. and Migchelsen, C. (1965) Hydration structure of fibrous macromolecules. *Ann. N. Y. Acad. Sci.* **125**, 365–379 [CrossRef](#)
- 79 Susi, H., Ard, J.S. and Carroll, R.J. (1971) The infrared spectrum and water binding of collagen as a function of relative humidity. *Biopolymers* **10**, 1597–1604 [CrossRef PubMed](#)
- 80 Nomura, S., Hiltner, A., Lando, J.B. and Baer, E. (1977) Interaction of water with native collagen. *Biopolymers* **16**, 231–246 [CrossRef PubMed](#)
- 81 Grigera, J.R. and Berendsen, H.J.C. (1979) The molecular details of collagen hydration. *Biopolymers* **18**, 47–57 [CrossRef](#)
- 82 Peto, S., Gillis, P. and Henri, V.P. (1990) Structure and dynamics of water in tendon from NMR relaxation measurements. *Biophys. J.* **57**, 71–84 [CrossRef PubMed](#)
- 83 Leikin, S., Parsegian, V.A., Yang, W. and Walrafen, G.E. (1997) Raman spectral evidence for hydration forces between collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11312–11317 [CrossRef PubMed](#)
- 84 Krasnosselskaia, L.V., Fullerton, G.D., Dodd, S.J. and Cameron, I.L. (2005) Water in tendon: orientational analysis of the free induction decay. *Magn. Reson. Med.* **54**, 280–288 [CrossRef PubMed](#)
- 85 Fullerton, G.D., Nes, E., Amurao, M., Rahal, A., Krasnosselskaia, L. and Cameron, I. (2006) An NMR method to characterize multiple water compartments on mammalian collagen. *Cell Biol. Int.* **30**, 66–73 [CrossRef PubMed](#)
- 86 Okuyama, K., Hongo, C., Fukushima, R., Wu, G., Narita, H., Noguchi, K., Tanaka, Y. and Nishino, N. (2004) Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: implications for proline ring puckering. *Biopolymers* **76**, 367–377 [CrossRef PubMed](#)
- 87 Okuyama, K., Hongo, C., Wu, G., Mizuno, K., Noguchi, K., Ebisuzaki, S., Tanaka, Y., Nishino, N. and Bachinger, H.P. (2009) High-resolution structures of collagen-like peptides [(Pro-Pro-Gly)₄-Xaa-Yaa-Gly-(Pro-Pro-Gly)₄]: implications for triple-helix hydration and Hyp(X) puckering. *Biopolymers* **91**, 361–372 [CrossRef PubMed](#)
- 88 Okuyama, K., Narita, H., Kawaguchi, T., Noguchi, K., Tanaka, Y. and Nishino, N. (2007) Unique side chain conformation of a Leu residue in a triple-helical structure. *Biopolymers* **86**, 212–221 [CrossRef PubMed](#)
- 89 Fallas, J.A., Dong, J., Tao, Y.J. and Hartgerink, J.D. (2012) Structural insights into charge pair interactions in triple helical collagen-like proteins. *J. Biol. Chem.* **287**, 8039–8047 [CrossRef PubMed](#)
- 90 Brondijk, T.H., Bihan, D., Farndale, R.W. and Huizinga, E.G. (2012) Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5253–5258 [CrossRef PubMed](#)
- 91 Leikin, S., Rau, D.C. and Parsegian, V.A. (1994) Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 276–280 [CrossRef PubMed](#)
- 92 Leikin, S., Rau, D.C. and Parsegian, V.A. (1995) Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions. *Nat. Struct. Biol.* **2**, 205–210 [CrossRef PubMed](#)
- 93 Gorres, K.L. and Raines, R.T. (2010) Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **45**, 106–124 [CrossRef PubMed](#)
- 94 Hudson, D.M. and Eyre, D.R. (2013) Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification? *Connect. Tissue Res.* **54**, 245–251 [CrossRef PubMed](#)
- 95 Hollingsworth, S.A. and Karplus, P.A. (2010) A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol. Concepts* **1**, 271–283 [CrossRef PubMed](#)
- 96 Berg, R.A. and Prockop, D.J. (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem. Biophys. Res. Commun.* **52**, 115–120 [CrossRef PubMed](#)
- 97 Rosenbloom, J., Harsch, M. and Jimenez, S. (1973) Hydroxyproline content determines the denaturation temperature of chick tendon collagen. *Arch. Biochem. Biophys.* **158**, 478–484 [CrossRef PubMed](#)
- 98 Perret, S., Merle, C., Bernocco, S., Berland, P., Garrone, R., Hulmes, D.J., Theisen, M. and Ruggiero, F. (2001) Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* **276**, 43693–43698 [CrossRef PubMed](#)
- 99 Friedman, L., Higgin, J.J., Moulder, G., Barstead, R., Raines, R.T. and Kimble, J. (2000) Prolyl 4-hydroxylase is required for viability and morphogenesis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4736–4741 [CrossRef PubMed](#)
- 100 Holster, T., Pakkanen, O., Soininen, R., Sormunen, R., Nokelainen, M., Kivirikko, K.I. and Myllyharju, J. (2007) 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.* **282**, 2512–2519 [CrossRef PubMed](#)
- 101 Brodsky, B. and Persikov, A.V. (2005) Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **70**, 301–339 [CrossRef PubMed](#)
- 102 Koide, T. (2007) Designed triple-helical peptides as tools for collagen biochemistry and matrix engineering. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 1281–1291 [CrossRef PubMed](#)
- 103 Brodsky, B., Thiagarajan, G., Madhan, B. and Kar, K. (2008) Triple-helical peptides: an approach to collagen conformation, stability, and self-association. *Biopolymers* **89**, 345–353 [CrossRef PubMed](#)
- 104 Fallas, J.A., O'Leary, L.E. and Hartgerink, J.D. (2010) Synthetic collagen mimics: self-assembly of homotrimers, heterotrimers and higher order structures. *Chem. Soc. Rev.* **39**, 3510–3527 [CrossRef PubMed](#)
- 105 Fields, G.B. (2010) Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* **8**, 1237–1258 [CrossRef PubMed](#)
- 106 Vitagliano, L., Berisio, R., Mazzarella, L. and Zagari, A. (2001) Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **58**, 459–464 [CrossRef PubMed](#)
- 107 Kramer, R.Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A. and Berman, H.M. (1998) X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* **280**, 623–638 [CrossRef PubMed](#)
- 108 Hongo, C., Nagarajan, V., Noguchi, K., Kamitori, S., Okuyama, K., Tanaka, Y. and Nishino, N. (2001) Average crystal structure of (Pro-Pro-Gly)₉ at 1.0 Å resolution. *Polym. J.* **33**, 812–818 [CrossRef](#)
- 109 Okuyama, K., Miyama, K., Morimoto, T., Masakiyo, K., Mizuno, K. and Bachinger, H.P. (2011) Stabilization of triple-helical structures of collagen peptides containing a Hyp-Thr-Gly, Hyp-Val-Gly, or Hyp-Ser-Gly sequence. *Biopolymers* **95**, 628–640 [CrossRef PubMed](#)
- 110 Donohue, J. and Trueblood, K.N. (1952) The crystal structure of hydroxy-L-proline. II. Determination and description of the structure. *Acta Crystallogr.* **5**, 419–431 [CrossRef](#)
- 111 Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S. and Prockop, D.J. (1982) 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.* **219**, 198–203 [CrossRef PubMed](#)
- 112 Inouye, K., Sakakibara, S. and Prockop, D.J. (1976) 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* **420**, 133–141 [CrossRef PubMed](#)
- 113 Jiravanichanun, N., Hongo, C., Wu, G., Noguchi, K., Okuyama, K., Nishino, N. and Silva, T. (2005) Unexpected puckering of hydroxyproline in the guest triplets, Hyp-Pro-Gly and Pro-allo-Hyp-Gly sandwiched between Pro-Pro-Gly sequence. *ChemBioChem* **6**, 1184–1187 [CrossRef PubMed](#)
- 114 Jiravanichanun, N., Nishino, N. and Okuyama, K. (2006) 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* **81**, 225–233 [CrossRef PubMed](#)
- 115 Holmgren, S.K., Bretscher, L.E., Taylor, K.M. and Raines, R.T. (1999) A hyperstable collagen mimic. *Chem. Biol.* **6**, 63–70 [CrossRef PubMed](#)
- 116 Bretscher, L.E., Jenkins, C.L., Taylor, K.M., DeRider, M.L. and Raines, R.T. (2001) Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **123**, 777–778 [CrossRef PubMed](#)
- 117 Kotch, F.W., Guzei, I.A. and Raines, R.T. (2008) Stabilization of the collagen triple helix by O-methylation of hydroxyproline residues. *J. Am. Chem. Soc.* **130**, 2952–2953 [CrossRef PubMed](#)
- 118 Shoulders, M.D., Satyshur, K.A., Forest, K.T. and Raines, R.T. (2010) Stereoelectronic and steric effects in side chains preorganize a protein main chain. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 559–564 [CrossRef PubMed](#)
- 119 Siebler, C., Erdmann, R.S. and Wennemers, H. (2014) Switchable proline derivatives: tuning the conformational stability of the collagen triple helix by pH changes. *Angew. Chem. Int. Ed. Engl.* **53**, 10340–10344 [CrossRef PubMed](#)
- 120 Shoulders, M.D., Kotch, F.W., Choudhary, A., Guzei, I.A. and Raines, R.T. (2010) The aberrance of the 4S diastereomer of 4-hydroxyproline. *J. Am. Chem. Soc.* **132**, 10857–10865 [CrossRef PubMed](#)
- 121 Erdmann, R.S. and Wennemers, H. (2011) Importance of ring puckering versus interstrand hydrogen bonds for the conformational stability of collagen. *Angew. Chem. Int. Ed. Engl.* **50**, 6835–6838 [CrossRef PubMed](#)
- 122 Erdmann, R.S. and Wennemers, H. (2012) Effect of sterically demanding substituents on the conformational stability of the collagen triple helix. *J. Am. Chem. Soc.* **134**, 17117–17124 [CrossRef PubMed](#)

- 123 Choudhary, A., Gandla, D., Krow, G.R. and Raines, R.T. (2009) Nature of amide carbonyl–carbonyl interactions in proteins. *J. Am. Chem. Soc.* **131**, 7244–7246 [CrossRef](#) [PubMed](#)
- 124 Hinderaker, M.P. and Raines, R.T. (2003) An electronic effect on protein structure. *Protein Sci.* **12**, 1188–1194 [CrossRef](#) [PubMed](#)
- 125 Bartlett, G.J., Newberry, R.W., VanVeller, B., Raines, R.T. and Woolfson, D.N. (2013) Interplay of hydrogen bonds and $n \rightarrow \pi^*$ interactions in proteins. *J. Am. Chem. Soc.* **135**, 18682–18688 [CrossRef](#) [PubMed](#)
- 126 Dai, N. and Etzkorn, F.A. (2009) *Cis*–*trans* proline isomerization effects on collagen triple-helix stability are limited. *J. Am. Chem. Soc.* **131**, 13728–13732 [CrossRef](#) [PubMed](#)
- 127 Holmgren, S.K., Taylor, K.M., Bretscher, L.E. and Raines, R.T. (1998) Code for collagen's stability deciphered. *Nature* **392**, 666–667 [CrossRef](#) [PubMed](#)
- 128 Doi, M., Nishi, Y., Uchiyama, S., Nishiuchi, Y., Nakazawa, T., Ohkubo, T. and Kobayashi, Y. (2003) Characterization of collagen model peptides containing 4-fluoroproline; (4*S*)-fluoroproline-Pro-Gly₁₀ forms a triple helix, but (4*R*)-fluoroproline-Pro-Gly₁₀ does not. *J. Am. Chem. Soc.* **125**, 9922–9923 [CrossRef](#) [PubMed](#)
- 129 Motooka, D., Kawahara, K., Sato, N., Nakamura, S., Uchiyama, S., Doi, M., Nishiuchi, Y., Nakazawa, T., Yoshida, T., Ohkubo, T., Nishi, Y. and Kobayashi, Y. (2008) Synthesis and characterization of the collagen model peptides containing 4(*S*)-hydroxyproline. *Proc. Eur. Pept. Symp.* **30**, 612–613
- 130 Doi, M., Nishi, Y., Uchiyama, S., Nishiuchi, Y., Nishio, H., Nakazawa, T., Ohkubo, T. and Kobayashi, Y. (2005) Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)₁₀ analogues: (4(*S*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-Gly)₁₀, (4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-Gly)₁₀ and (4(*S*)-fluoroprolyl-4(*R*)-fluoroprolyl-Gly)₁₀. *J. Pept. Sci.* **11**, 609–616 [CrossRef](#) [PubMed](#)
- 131 Hodges, J.A. and Raines, R.T. (2005) Stereoelectronic and steric effects in the collagen triple helix: toward a code for strand association. *J. Am. Chem. Soc.* **127**, 15923–15932 [CrossRef](#) [PubMed](#)
- 132 Hodges, J.A. and Raines, R.T. (2003) Stereoelectronic effects on collagen stability: the dichotomy of 4-fluoroproline diastereomers. *J. Am. Chem. Soc.* **125**, 9262–9263 [CrossRef](#) [PubMed](#)
- 133 Motooka, D., Kawahara, K., Nakamura, S., Doi, M., Nishi, Y., Nishiuchi, Y., Kang, Y.K., Nakazawa, T., Uchiyama, S., Yoshida, T. et al. (2012) The triple helical structure and stability of collagen model peptide with 4(*S*)-hydroxyprolyl-Pro-Gly units. *Biopolymers* **98**, 111–121 [CrossRef](#) [PubMed](#)
- 134 Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A. and Brodsky, B. (2003) Triple-helix propensity of hydroxyproline and fluoroproline: comparison of host–guest and repeating tripeptide collagen models. *J. Am. Chem. Soc.* **125**, 11500–11501 [CrossRef](#) [PubMed](#)
- 135 Sonntag, L.S., Schweizer, S., Ochsenfeld, C. and Wennemers, H. (2006) The "azido gauche effect"–implications for the conformation of azidoproline. *J. Am. Chem. Soc.* **128**, 14697–14703 [CrossRef](#) [PubMed](#)
- 136 Erdmann, R.S. and Wennemers, H. (2010) Functionalizable collagen model peptides. *J. Am. Chem. Soc.* **132**, 13957–13959 [CrossRef](#) [PubMed](#)
- 137 Tanrikulu, I.C. and Raines, R.T. (2014) Optimal interstrand bridges for collagen-like biomaterials. *J. Am. Chem. Soc.* **136**, 13490–13493 [CrossRef](#) [PubMed](#)
- 138 Choudhary, A., Kamer, K.J., Shoulders, M.D. and Raines, R.T. (2015) 4-ketoproline: an electrophilic proline analog for bioconjugation. *Biopolymers* **104**, 110–115 [CrossRef](#) [PubMed](#)
- 139 Kuemin, M., Nagel, Y.A., Schweizer, S., Monnard, F.W., Ochsenfeld, C. and Wennemers, H. (2010) Tuning the *cis/trans* conformer ratio of Xaa-Pro amide bonds by intramolecular hydrogen bonds: the effect on PPII helix stability. *Angew. Chem. Int. Ed. Engl.* **49**, 6324–6327 [CrossRef](#) [PubMed](#)
- 140 Erdmann, R.S. and Wennemers, H. (2012) Conformational stability of collagen triple helices functionalized in the Yaa position by click chemistry. *Org. Biomol. Chem.* **10**, 1982–1986 [CrossRef](#) [PubMed](#)
- 141 Erdmann, R.S. and Wennemers, H. (2013) Conformational stability of triazolyl functionalized collagen triple helices. *Bioorg. Med. Chem.* **21**, 3565–3568 [CrossRef](#) [PubMed](#)
- 142 Jiang, T., Xu, C., Liu, Y., Liu, Z., Wall, J.S., Zuo, X., Lian, T., Salaita, K., Ni, C., Pochan, D. and Conticello, V.P. (2014) Structurally defined nanoscale sheets from self-assembly of collagen-mimetic peptides. *J. Am. Chem. Soc.* **136**, 4300–4308 [CrossRef](#) [PubMed](#)
- 143 Jiang, T., Xu, C., Zuo, X. and Conticello, V.P. (2014) Structurally homogeneous nanosheets from self-assembly of a collagen-mimetic peptide. *Angew. Chem. Int. Ed. Engl.* **53**, 8367–8371 [CrossRef](#) [PubMed](#)
- 144 Chen, L., Cai, S., Lim, J., Lee, S.S. and Lee, S.G. (2015) Elucidating pH-dependent collagen triple helix formation through interstrand hydroxyproline–glutamic acid interactions. *ChemBioChem* **16**, 407–410 [CrossRef](#) [PubMed](#)
- 145 Babu, I.R. and Ganesh, K.N. (2001) Enhanced triple helix stability of collagen peptides with 4*R*-aminoprolyl (Amp) residues: relative roles of electrostatic and hydrogen bonding effects. *J. Am. Chem. Soc.* **123**, 2079–2080 [CrossRef](#) [PubMed](#)
- 146 Umashankara, M., Babu, I.R. and Ganesh, K.N. (2003) Two prolines with a difference: contrasting stereoelectronic effects of 4*R/S*-aminoproline on triplex stability in collagen peptides [pro(X)-pro(Y)-Gly]_n. *Chem. Commun. (Camb.)*, 2606–2607 [CrossRef](#) [PubMed](#)
- 147 Umashankara, M., Sonar, M.V., Bansode, N.D. and Ganesh, K.N. (2015) Orchestration of structural, stereoelectronic, and hydrogen-bonding effects in stabilizing triplexes from engineered chimeric collagen peptides (Pro^x-Pro^y-Gly)_n incorporating 4(*R/S*)-aminoproline. *J. Org. Chem.* **80**, 8552–8560 [CrossRef](#) [PubMed](#)
- 148 Lee, S.G., Lee, J.Y. and Chmielewski, J. (2008) Investigation of pH-dependent collagen triple-helix formation. *Angew. Chem. Int. Ed. Engl.* **47**, 8429–8432 [CrossRef](#) [PubMed](#)
- 149 Lee, J. and Chmielewski, J. (2010) Folding studies of pH-dependent collagen peptides. *Chem. Biol. Drug. Des.* **75**, 161–168 [CrossRef](#) [PubMed](#)
- 150 Nishi, Y., Uchiyama, S., Doi, M., Nishiuchi, Y., Nakazawa, T., Ohkubo, T. and Kobayashi, Y. (2005) Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of collagen triple helix. *Biochemistry* **44**, 6034–6042 [CrossRef](#) [PubMed](#)
- 151 Suzuki, E., Fraser, R.D.B. and MacRae, T.P. (1980) Role of hydroxyproline in the stabilization of the collagen molecule via water molecules. *Int. J. Biol. Macromol.* **2**, 54–56 [CrossRef](#)
- 152 Kawahara, K., Nishi, Y., Nakamura, S., Uchiyama, S., Nishiuchi, Y., Nakazawa, T., Ohkubo, T. and Kobayashi, Y. (2005) Effect of hydration on the stability of the collagen-like triple-helical structure of [4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-glycine]₁₀. *Biochemistry* **44**, 15812–15822 [CrossRef](#) [PubMed](#)
- 153 Schumacher, M., Mizuno, K. and Bachinger, H.P. (2005) The crystal structure of the collagen-like polypeptide (glycyl-4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl)₉ at 1.55 Å resolution shows up-puckering of the proline ring in the Xaa position. *J. Biol. Chem.* **280**, 20397–20403 [CrossRef](#) [PubMed](#)
- 154 Ogle, J.D., Arlinghaus, R.B. and Lgan, M.A. (1962) 3-Hydroxyproline, a new amino acid of collagen. *J. Biol. Chem.* **237**, 3667–3673 [PubMed](#)
- 155 Dean, D.C., Barr, J.F., Freytag, J.W. and Hudson, B.G. (1983) Isolation of type IV procollagen-like polypeptides from glomerular basement membrane. Characterization of pro- α 1(IV). *J. Biol. Chem.* **258**, 590–596 [PubMed](#)
- 156 Weis, M.A., Hudson, D.M., Kim, L., Scott, M., Wu, J.J. and Eyre, D.R. (2010) Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J. Biol. Chem.* **285**, 2580–2590 [CrossRef](#) [PubMed](#)
- 157 Hudson, D.M., Werther, R., Weis, M., Wu, J.J. and Eyre, D.R. (2014) Evolutionary origins of C-terminal (GPP)_n 3-hydroxyproline formation in vertebrate tendon collagen. *PLoS One* **9**, e93467 [CrossRef](#) [PubMed](#)
- 158 Hudson, D.M., Joeng, K.S., Werther, R., Rajagopal, A., Weis, M., Lee, B.H. and Eyre, D.R. (2015) Post-translationally abnormal collagens of prolyl 3-hydroxylase-2 null mice offer a pathobiological mechanism for the high myopia linked to human LEPREL1 mutations. *J. Biol. Chem.* **290**, 8613–8622 [CrossRef](#) [PubMed](#)
- 159 Taga, Y., Kusubata, M., Ogawa-Goto, K. and Hattori, S. (2016) Developmental stage-dependent regulation of prolyl 3-hydroxylation in tendon type I collagen. *J. Biol. Chem.* **291**, 837–847 [CrossRef](#) [PubMed](#)
- 160 Eyre, D.R. and Weis, M.A. (2013) Bone collagen: new clues to its mineralization mechanism from recessive osteogenesis imperfecta. *Calcif. Tissue Int.* **93**, 338–347 [CrossRef](#) [PubMed](#)
- 161 Mizuno, K., Hayashi, T., Peyton, D.H. and Bachinger, H.P. (2004) The peptides acetyl-(Gly-3(*S*)-Hyp-4(*R*)-Hyp)₁₀-NH₂ and acetyl-(Gly-Pro-3(*S*)-Hyp)₁₀-NH₂ do not form a collagen triple helix. *J. Biol. Chem.* **279**, 282–287 [CrossRef](#) [PubMed](#)
- 162 Schumacher, M.A., Mizuno, K. and Bachinger, H.P. (2006) 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.* **281**, 27566–27574 [CrossRef](#) [PubMed](#)
- 163 Mizuno, K., Peyton, D.H., Hayashi, T., Engel, J. and Bachinger, H.P. (2008) Effect of the -Gly-3(*S*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-tripeptide unit on the stability of collagen model peptides. *FEBS J.* **275**, 5830–5840 [CrossRef](#) [PubMed](#)
- 164 Pokidysheva, E., Boudko, S., Vranka, J., Zientek, K., Maddox, K., Moser, M., Fassler, R., Ware, J. and Bachinger, H.P. (2014) Biological role of prolyl 3-hydroxylation in type IV collagen. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 161–166 [CrossRef](#) [PubMed](#)
- 165 Bamford, J.K. and Bamford, D.H. (1990) Capsomer proteins of bacteriophage PRD1, a bacterial virus with a membrane. *Virology* **177**, 445–451 [CrossRef](#) [PubMed](#)
- 166 Sylvestre, P., Couture-Tosi, E. and Mock, M. (2002) A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol. Microbiol.* **45**, 169–178 [CrossRef](#) [PubMed](#)
- 167 Rasmussen, M., Jacobsson, M. and Björck, L. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278**, 32313–32316 [CrossRef](#) [PubMed](#)
- 168 Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., Scola, B., Suzan, M. and Claverie, J.M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 [CrossRef](#) [PubMed](#)

- 169 Ghosh, N., McKillop, T.J., Jowitz, T.A., Howard, M., Davies, H., Holmes, D.F., Roberts, I.S. and Bella, J. (2012) Collagen-like proteins in pathogenic *E. coli* strains. *PLoS One* **7**, e37872 [CrossRef PubMed](#)
- 170 Xu, Y., Keene, D.R., Bujnicki, J.M., Hook, M. and Lukomski, S. (2002) Streptococcal Sc1 and Sc2 proteins form collagen-like triple helices. *J. Biol. Chem.* **277**, 27312–27318 [CrossRef PubMed](#)
- 171 Boydston, J.A., Chen, P., Steichen, C.T. and Turnbough, Jr, C.L. (2005) Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis*. *J. Bacteriol.* **187**, 5310–5317 [CrossRef PubMed](#)
- 172 Yu, Z., An, B., Ramshaw, J.A. and Brodsky, B. (2014) Bacterial collagen-like proteins that form triple-helical structures. *J. Struct. Biol.* **186**, 451–461 [CrossRef PubMed](#)
- 173 Yu, Z., Brodsky, B. and Inouye, M. (2011) Dissecting a bacterial collagen domain from *Streptococcus pyogenes*: sequence and length-dependent variations in triple helix stability and folding. *J. Biol. Chem.* **286**, 18960–18968 [CrossRef PubMed](#)
- 174 Han, R., Zwiefka, A., Caswell, C.C., Xu, Y., Keene, D.R., Lukomska, E., Zhao, Z., Hook, M. and Lukomski, S. (2006) Assessment of prokaryotic collagen-like sequences derived from streptococcal Sc1 and Sc2 proteins as a source of recombinant GXY polymers. *Appl. Microbiol. Biotechnol.* **72**, 109–115 [CrossRef PubMed](#)
- 175 Ramshaw, J.A., Werkmeister, J.A. and Dumsday, G.J. (2014) Bioengineered collagens: emerging directions for biomedical materials. *Bioengineered* **5**, 227–426 [CrossRef PubMed](#)
- 176 Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A. and Brodsky, B. (2000) Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960–14967 [CrossRef PubMed](#)
- 177 Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A. and Brodsky, B. (2002) Peptide investigations of pairwise interactions in the collagen triple-helix. *J. Mol. Biol.* **316**, 385–394 [CrossRef PubMed](#)
- 178 Persikov, A.V., Ramshaw, J.A. and Brodsky, B. (2005) Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280**, 19343–19349 [CrossRef PubMed](#)
- 179 Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A. and Brodsky, B. (2005) Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry* **44**, 1414–1422 [CrossRef PubMed](#)
- 180 Emsley, J., Knight, C.G., Farndale, R.W. and Barnes, M.J. (2004) Structure of the integrin $\alpha 2\beta 1$ -binding collagen peptide. *J. Mol. Biol.* **335**, 1019–1028 [CrossRef PubMed](#)
- 181 Boudko, S.P. and Bachinger, H.P. (2012) The NC2 domain of type IX collagen determines the chain register of the triple helix. *J. Biol. Chem.* **287**, 44536–44545 [CrossRef PubMed](#)
- 182 Widmer, C., Gebauer, J.M., Brunstein, E., Rosenbaum, S., Zaucke, F., Drogemuller, C., Leeb, T. and Baumann, U. (2012) Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13243–13247 [CrossRef PubMed](#)
- 183 Okuyama, K., Haga, M., Noguchi, K. and Tanaka, T. (2014) Preferred side-chain conformation of arginine residues in a triple-helical structure. *Biopolymers* **101**, 1000–1009 [CrossRef PubMed](#)
- 184 Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M. and Brodsky, B. (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* **282**, 29757–29765 [CrossRef PubMed](#)
- 185 Gaihl, F., Mann, K., Wiedemann, H., Engel, J. and Timpl, R. (1995) Structural comparison of cuticle and interstitial collagens from annelids living in shallow sea-water and at deep-sea hydrothermal vents. *J. Mol. Biol.* **246**, 284–294 [CrossRef PubMed](#)
- 186 Mann, K., Mechling, D.E., Bachinger, H.P., Eckerskorn, C., Gaihl, F. and Timpl, R. (1996) 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.* **261**, 255–266 [CrossRef PubMed](#)
- 187 Bann, J.G., Peyton, D.H. and Bachinger, H.P. (2000) Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett.* **473**, 237–240 [CrossRef PubMed](#)
- 188 Bann, J.G., Bachinger, H.P. and Peyton, D.H. (2003) Role of carbohydrate in stabilizing the triple-helix in a model for a deep-sea hydrothermal vent worm collagen. *Biochemistry* **42**, 4042–4048 [CrossRef PubMed](#)
- 189 Daubenspeck, J.M., Zeng, H., Chen, P., Dong, S., Steichen, C.T., Krishna, N.R., Pritchard, D.G. and Turnbough, Jr, C.L. (2004) Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis* exosporium. *J. Biol. Chem.* **279**, 30945–30953 [CrossRef PubMed](#)
- 190 Gauba, V. and Hartgerink, J.D. (2007) Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.* **129**, 2683–2690 [CrossRef PubMed](#)
- 191 Gauba, V. and Hartgerink, J.D. (2007) Surprisingly high stability of collagen ABC heterotrimer: evaluation of side chain charge pairs. *J. Am. Chem. Soc.* **129**, 15034–15041 [CrossRef PubMed](#)
- 192 Fallas, J.A., Gauba, V. and Hartgerink, J.D. (2009) Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J. Biol. Chem.* **284**, 26851–26859 [CrossRef PubMed](#)
- 193 O'Leary, L.E., Fallas, J.A. and Hartgerink, J.D. (2011) Positive and negative design leads to compositional control in AAB collagen heterotrimers. *J. Am. Chem. Soc.* **133**, 5432–5443 [CrossRef PubMed](#)
- 194 Jalan, A.A., Demeler, B. and Hartgerink, J.D. (2013) Hydroxyproline-free single composition ABC collagen heterotrimer. *J. Am. Chem. Soc.* **135**, 6014–6017 [CrossRef PubMed](#)
- 195 Jalan, A.A. and Hartgerink, J.D. (2013) Pairwise interactions in collagen and the design of heterotrimeric helices. *Curr. Opin. Chem. Biol.* **17**, 960–967 [CrossRef PubMed](#)
- 196 Jalan, A.A., Jochim, K.A. and Hartgerink, J.D. (2014) Rational design of a non-canonical "sticky-ended" collagen triple helix. *J. Am. Chem. Soc.* **136**, 7535–7538 [CrossRef PubMed](#)
- 197 Orgel, J.P., Irving, T.C., Miller, A. and Wess, T.J. (2006) Microfibrillar structure of type I collagen *in situ*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9001–9005 [CrossRef PubMed](#)
- 198 Sweeney, S.M., Orgel, J.P., Fertala, A., McAuliffe, J.D., Turner, K.R., Di Lullo, G.A., Chen, S., Antipova, O., Perumal, S., Ala-Kokko, L. et al. (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.* **283**, 21187–21197 [CrossRef PubMed](#)
- 199 Orgel, J.P., Antonio, J.D. and Antipova, O. (2011) Molecular and structural mapping of collagen fibril interactions. *Connect. Tissue Res.* **52**, 2–17 [CrossRef PubMed](#)
- 200 Parkin, J.D., Antonio, J.D., Pedchenko, V., Hudson, B., Jensen, S.T. and Savige, J. (2011) Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum. Mutat.* **32**, 127–143 [CrossRef PubMed](#)
- 201 Heino, J. (2007) The collagen family members as cell adhesion proteins. *BioEssays* **29**, 1001–1010 [CrossRef PubMed](#)
- 202 Leitinger, B. and Hohenester, E. (2007) Mammalian collagen receptors. *Matrix Biol.* **26**, 146–155 [CrossRef PubMed](#)
- 203 Heino, J. (2014) Cellular signaling by collagen-binding integrins. *Adv. Exp. Med. Biol.* **819**, 143–155 [CrossRef PubMed](#)
- 204 Farndale, R.W., Lisman, T., Bihan, D., Hamaia, S., Smerling, C.S., Pugh, N., Konitsiotis, A., Leitinger, B., de Groot, P.G., Jarvis, G.E. and Raynal, N. (2008) Cell-collagen interactions: the use of peptide Toolkits to investigate collagen–receptor interactions. *Biochem. Soc. Trans.* **36**, 241–250 [CrossRef PubMed](#)
- 205 Hamaia, S. and Farndale, R.W. (2014) Integrin recognition motifs in the human collagens. *Adv. Exp. Med. Biol.* **819**, 127–142 [CrossRef PubMed](#)
- 206 Knight, C.G., Morton, L.F., Onley, D.J., Peachey, A.R., Messent, A.J., Smethurst, P.A., Tuckwell, D.S., Farndale, R.W. and Barnes, M.J. (1998) Identification in collagen type I of an integrin $\alpha 2\beta 1$ -binding site containing an essential GER sequence. *J. Biol. Chem.* **273**, 33287–33294 [CrossRef PubMed](#)
- 207 Knight, C.G., Morton, L.F., Peachey, A.R., Tuckwell, D.S., Farndale, R.W. and Barnes, M.J. (2000) The collagen-binding A-domains of integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$ recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. *J. Biol. Chem.* **275**, 35–40 [CrossRef PubMed](#)
- 208 Bella, J. and Berman, H.M. (2000) Integrin–collagen complex: a metal–glutamate handshake. *Structure* **8**, R121–R126 [CrossRef PubMed](#)
- 209 Siljander, P.R., Hamaia, S., Peachey, A.R., Slatter, D.A., Smethurst, P.A., Ouwehand, W.H., Knight, C.G. and Farndale, R.W. (2004) Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J. Biol. Chem.* **279**, 47763–47772 [CrossRef PubMed](#)
- 210 Raynal, N., Hamaia, S.W., Siljander, P.R., Maddox, B., Peachey, A.R., Fernandez, R., Foley, L.J., Slatter, D.A., Jarvis, G.E. and Farndale, R.W. (2006) Use of synthetic peptides to locate novel integrin $\alpha 2\beta 1$ -binding motifs in human collagen III. *J. Biol. Chem.* **281**, 3821–3831 [CrossRef PubMed](#)
- 211 Hamaia, S.W., Pugh, N., Raynal, N., Nemoz, B., Stone, R., Gullberg, D., Bihan, D. and Farndale, R.W. (2012) Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin $\alpha 1\beta 1$ using collagen toolkits II and III. *J. Biol. Chem.* **287**, 26019–26028 [CrossRef PubMed](#)
- 212 Chin, Y.K., Headey, S.J., Mohanty, B., Patil, R., McEwan, P.A., Swarbrick, J.D., Mulhern, T.D., Emsley, J., Simpson, J.S. and Scanlon, M.J. (2013) The structure of integrin $\alpha 1$ domain in complex with a collagen-mimetic peptide. *J. Biol. Chem.* **288**, 36796–36809 [CrossRef PubMed](#)
- 213 Carafoli, F., Hamaia, S.W., Bihan, D., Hohenester, E. and Farndale, R.W. (2013) An activating mutation reveals a second binding mode of the integrin $\alpha 2$ domain to the GFOGER motif in collagens. *PLoS One* **8**, e69833 [CrossRef PubMed](#)
- 214 Lisman, T., Raynal, N., Groeneveld, D., Maddox, B., Peachey, A.R., Huizinga, E.G., de Groot, P.G. and Farndale, R.W. (2006) A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **108**, 3753–3756 [CrossRef PubMed](#)

- 215 Konitsiotis, A.D., Raynal, N., Bihan, D., Hohenester, E., Fardale, R.W. and Leitingner, B. (2008) Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **283**, 6861–6868 [CrossRef PubMed](#)
- 216 Giudici, C., Raynal, N., Wiedemann, H., Cabral, W.A., Marini, J.C., Timpl, R., Bachinger, H.P., Fardale, R.W., Sasaki, T. and Tenni, R. (2008) Mapping of SPARC/BM-40/osteonection-binding sites on fibrillar collagens. *J. Biol. Chem.* **283**, 19551–19560 [CrossRef PubMed](#)
- 217 Hohenester, E., Sasaki, T., Giudici, C., Fardale, R.W. and Bachinger, H.P. (2008) Structural basis of sequence-specific collagen recognition by SPARC. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18273–18277 [CrossRef PubMed](#)
- 218 Carafoli, F., Bihan, D., Stathopoulos, S., Konitsiotis, A.D., Kvasakul, M., Fardale, R.W., Leitingner, B. and Hohenester, E. (2009) Crystallographic insight into collagen recognition by discoidin domain receptor 2. *Structure* **17**, 1573–1581 [CrossRef PubMed](#)
- 219 Ishida, Y. and Nagata, K. (2011) Hsp47 as a collagen-specific molecular chaperone. *Methods Enzymol.* **499**, 167–182 [CrossRef PubMed](#)
- 220 Dobritzsch, D., Lindh, I., Uysal, H., Nandakumar, K.S., Burkhardt, H., Schneider, G. and Holmdahl, R. (2011) Crystal structure of an arthritogenic anticollagen immune complex. *Arthritis Rheum.* **63**, 3740–3748 [CrossRef PubMed](#)
- 221 Raposo, B., Dobritzsch, D., Ge, C., Ekman, D., Xu, B., Lindh, I., Forster, M., Uysal, H., Nandakumar, K.S., Schneider, G. and Holmdahl, R. (2014) Epitope-specific antibody response is controlled by immunoglobulin V_H polymorphisms. *J. Exp. Med.* **211**, 405–411 [CrossRef PubMed](#)
- 222 Gingras, A.R., Giriya, U.V., Keeble, A.H., Panchal, R., Mitchell, D.A., Moody, P.C. and Wallis, R. (2011) Structural basis of mannan-binding lectin recognition by its associated serine protease MASP-1: implications for complement activation. *Structure* **19**, 1635–1643 [CrossRef PubMed](#)
- 223 Giriya, U., Gingras, A.R., Marshall, J.E., Panchal, R., Sheikh, M.A., Gal, P., Schwaeble, W.J., Mitchell, D.A., Moody, P.C. and Wallis, R. (2013) Structural basis of the C1q/C1s interaction and its central role in assembly of the C1 complex of complement activation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13916–13920 [CrossRef PubMed](#)
- 224 Manka, S.W., Carafoli, F., Visse, R., Bihan, D., Raynal, N., Fardale, R.W., Murphy, G., Enghild, J.J., Hohenester, E. and Nagase, H. (2012) Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12461–12466 [CrossRef PubMed](#)
- 225 Zhao, Y., Marcink, T.C., Gari, R.R., Marsh, B.P., King, G.M., Stawikowska, R., Fields, G.B. and Van Doren, S.R. (2015) Transient collagen triple helix binding to a key metalloproteinase in invasion and development. *Structure* **23**, 257–269 [CrossRef PubMed](#)
- 226 Zong, Y., Xu, Y., Liang, X., Keene, D.R., Hook, A., Gurusiddappa, S., Hook, M. and Narayana, S.V. (2005) A 'Collagen Hug' model for *Staphylococcus aureus* CNA binding to collagen. *EMBO J.* **24**, 4224–4236 [CrossRef PubMed](#)
- 227 Vitagliano, L., Berisio, R. and Simone, A. (2011) Role of hydration in collagen recognition by bacterial adhesins. *Biophys. J.* **100**, 2253–2261 [CrossRef PubMed](#)
- 228 Desiraju, G.R. and Steiner, T. (2001) *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, Oxford
- 229 Krimm, S. and Kuroiwa, K. (1968) Low temperature infrared spectra of polyglycines and C—H—O=C hydrogen bonding in polyglycine II. *Biopolymers* **6**, 401–407 [CrossRef PubMed](#)
- 230 Bella, J. and Berman, H.M. (1996) Crystallographic evidence for C α —H—O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264**, 734–742 [CrossRef PubMed](#)
- 231 Derewenda, Z.S., Lee, L. and Derewenda, U. (1995) The occurrence of C—H...O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248–262 [CrossRef PubMed](#)
- 232 Jeffrey, G.A. and Saenger, W. (1991) *Hydrogen Bonding in Biological Structures*, Springer-Verlag New York [CrossRef](#)
- 233 Scheiner, S., Kar, T. and Gu, Y. (2001) Strength of the C α H...O hydrogen bond of amino acid residues. *J. Biol. Chem.* **276**, 9832–9837 [CrossRef PubMed](#)
- 234 Scheiner, S. (2006) Contributions of NH...O and CH...O hydrogen bonds to the stability of β -sheets in proteins. *J. Phys. Chem. B* **110**, 18670–18679 [CrossRef PubMed](#)
- 235 Jiang, L. and Lai, L. (2002) CH...O hydrogen bonds at protein–protein interfaces. *J. Biol. Chem.* **277**, 37732–37740 [CrossRef PubMed](#)
- 236 Sarkhel, S. and Desiraju, G.R. (2004) N—H...O, O—H...O, and C—H...O hydrogen bonds in protein–ligand complexes: strong and weak interactions in molecular recognition. *Proteins* **54**, 247–259 [CrossRef PubMed](#)
- 237 Bella, J. and Humphries, M.J. (2005) C α —H—O=C hydrogen bonds contribute to the specificity of RGD cell–adhesion interactions. *BMC Struct. Biol.* **5**, 4 [CrossRef PubMed](#)
- 238 Zhang, Y., Malamakal, R.M. and Chenoweth, D.M. (2015) Aza-glycine induces collagen hyperstability. *J. Am. Chem. Soc.* **137**, 12422–12425 [CrossRef PubMed](#)
- 239 Myllyharju, J. and Kivirikko, K.I. (2001) Collagens and collagen-related diseases. *Ann. Med.* **33**, 7–21 [CrossRef PubMed](#)
- 240 Beck, K., Chan, V.C., Shenoy, N., Kirkpatrick, A., Ramshaw, J.A. and Brodsky, B. (2000) Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4273–4278 [CrossRef PubMed](#)
- 241 Xiao, J., Yang, Z., Sun, X., Addabbo, R. and Baum, J. (2015) Local amino acid sequence patterns dominate the heterogeneous phenotype for the collagen connective tissue disease Osteogenesis Imperfecta resulting from Gly mutations. *J. Struct. Biol.* **192**, 127–137 [CrossRef PubMed](#)
- 242 Thiagarajan, G., Li, Y., Mohs, A., Strafaci, C., Popiel, M., Baum, J. and Brodsky, B. (2008) Common interruptions in the repeating tripeptide sequence of non-fibrillar collagens: sequence analysis and structural studies on triple-helix peptide models. *J. Mol. Biol.* **376**, 736–748 [CrossRef PubMed](#)
- 243 Bella, J. (2014) A first census of collagen interruptions: collagen's own stutters and stammers. *J. Struct. Biol.* **186**, 438–450 [CrossRef PubMed](#)
- 244 Bella, J., Brodsky, B. and Berman, H.M. (1996) Disrupted collagen architecture in the crystal structure of a triple-helical peptide with a Gly→Ala substitution. *Connect. Tissue Res.* **35**, 401–406 [CrossRef PubMed](#)
- 245 Long, C.G., Braswell, E., Zhu, D., Apigo, J., Baum, J. and Brodsky, B. (1993) Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. *Biochemistry* **32**, 11688–11695 [CrossRef PubMed](#)
- 246 Bella, J., Liu, J., Kramer, R., Brodsky, B. and Berman, H.M. (2006) Conformational effects of Gly-X-Gly interruptions in the collagen triple helix. *J. Mol. Biol.* **362**, 298–311 [CrossRef PubMed](#)
- 247 Mohs, A., Popiel, M., Li, Y., Baum, J. and Brodsky, B. (2006) Conformational features of a natural break in the type IV collagen Gly-X-Y repeat. *J. Biol. Chem.* **281**, 17197–17202 [CrossRef PubMed](#)
- 248 Li, Y., Brodsky, B. and Baum, J. (2007) NMR shows hydrophobic interactions replace glycine packing in the triple helix at a natural break in the (Gly-X-Y)_n repeat. *J. Biol. Chem.* **282**, 22699–22706 [CrossRef PubMed](#)
- 249 Li, Y., Brodsky, B. and Baum, J. (2009) NMR conformational and dynamic consequences of a Gly to Ser substitution in an osteogenesis imperfecta collagen model peptide. *J. Biol. Chem.* **284**, 20660–20667 [CrossRef PubMed](#)
- 250 Sun, X., Chai, Y., Wang, Q., Liu, H., Wang, S. and Xiao, J. (2015) A natural interruption displays higher global stability and local conformational flexibility than a similar Gly mutation sequence in collagen mimic peptides. *Biochemistry* **54**, 6106–6113 [CrossRef PubMed](#)
- 251 Xiao, J., Sun, X., Balaran, M., Brodsky, B. and Baum, J. (2015) NMR studies demonstrate a unique AAB composition and chain register for a heterotrimeric type IV collagen model peptide containing a natural interruption site. *J. Biol. Chem.* **290**, 24201–24209 [CrossRef PubMed](#)
- 252 Sasisekharan, V. and Yathindra, N. (1999) *Collagen: Ramachandran Triple Helix Revisited*. In *Perspectives in Structural Biology*, pp. 155–168, University Press, Hyderabad
- 253 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 [CrossRef PubMed](#)
- 254 Kilchherr, E., Hofmann, H., Steigemann, W. and Engel, J. (1985) Structural model of the collagen-like region of C1q comprising the kink region and the fibre-like packing of the six triple helices. *J. Mol. Biol.* **186**, 403–415 [CrossRef PubMed](#)