Matrix Biology Plus 21 (2024) 100139

Contents lists available at ScienceDirect

Matrix Biology Plus

journal homepage: www.sciencedirect.com/journal/matrix-biology-plus

Designing collagens to shed light on the multi-scale structure–function mapping of matrix disorders

Sonal Gahlawat [a], Vikas Nanda [b,c], David I. Shreiber [a,*]

[a] Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
[b] Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
[c] Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

ARTICLE INFO

Keywords:
Collagen
Collagen Disorders
Synthetic Collagen Peptides
Recombinant Bacterial Collagen
Molecular Dynamics Simulations
Steered Molecular Dynamics Simulations

ABSTRACT

Collagens are the most abundant structural proteins in the extracellular matrix of animals and play crucial roles in maintaining the structural integrity and mechanical properties of tissues and organs while mediating important biological processes. Fibrillar collagens have a unique triple helix structure with a characteristic repeating sequence of $(Gly\text{-}X\text{-}Y)_n$. Variations within the repetitive sequence can cause misfolding of the triple helix, resulting in heritable connective tissue disorders. The most common variations are single-point missense mutations that lead to the substitution of a glycine residue with a bulkier amino acid (Gly → X). In this review, we will first discuss the importance of collagen's triple helix structure and how single Gly substitutions can impact its folding, structure, secretion, assembly into higher-order structures, and biological functions. We will review the role of "designer collagens," i.e., synthetic collagen-mimetic peptides and recombinant bacterial collagen as model systems to include Gly → X substitutions observed in collagen disorders and investigate their impact on structure and function utilizing *in vitro* studies. Lastly, we will explore how computational modeling of collagen peptides, especially molecular and steered molecular dynamics, has been instrumental in probing the effects of Gly substitutions on structure, receptor binding, and mechanical stability across multiple length scales.

Collagens: Introduction

Collagens are the most abundant class of animal proteins, accounting for one-third of the total protein by mass in animals [1,2]. As a primary component of the extracellular matrix (ECM), collagen plays an essential role in defining the shape and architecture of tissues and organs and providing structural integrity [3-5]. Collagens interact with cell surface receptors and various components of the ECM, thereby regulating critical cellular functions, including adhesion, spreading, differentiation, migration, secretion of the ECM and other molecules, and tissue development, repair, and remodeling [6-8]. Collagens assemble into diverse higher-order structures across different tissues, providing various biological functions to these tissues.

In vertebrates, 45 distinct collagen genes have been identified, producing 28 classes of collagens [1,7]. Collagens are divided into five categories based on their function, domain architecture, and supramolecular organization: (1) fibril-forming collagens; (2) fibril-associated collagens with interrupted triple helices (FACIT); (3) multiplexins; (4) network-forming collagens; and (5) transmembrane collagens [9-11]. This review focuses on the fibrillar species for which synthetic model systems have provided significant insight. Fibrillar collagen is composed of three polypeptide chains, called α-chains, that self-assemble to form a rope-like triple helix in a zipper-like fashion [10]. The folding of the triple helix begins with a few monomer units forming a nucleus, followed by a growth phase that results in a large, organized structure [12-15].

Mutations in collagen can impair function by disrupting folding or supramolecular assembly, reducing collagen secretion by cells, and/or disrupting collagen proteostasis *via* the endoplasmic reticulum stress signaling pathway [16-25]. These alterations in collagen structure and

Abbreviations: ECM, extracellular matrix; FACIT, fibril-associated collagens with interrupted triple helices; CMP, collagen-mimetic peptides; CLP, collagen-like proteins; GPC, golgi to plasma membrane carriers; OI, Osteogenesis Imperfecta; EDS, Ehlers-Danlos Syndrome; hMSCs, human mesenchymal stem cells; NMR, nuclear magnetic resonance spectroscopy; Scl1/Scl2, Streptococcal collagen-like proteins; CL, collagen-like domain; Fn, fibronectin; MD, molecular dynamics; SMD, steered molecular dynamics; 3D, three-dimensional.

* Corresponding author.
E-mail address: shreiber@soe.rutgers.edu (D.I. Shreiber).

https://doi.org/10.1016/j.mbplus.2023.100139
Received 31 July 2023; Received in revised form 29 November 2023; Accepted 9 December 2023
Available online 14 December 2023

function ultimately impact the structure and function of tissues and organs. Consequently, mutations in genes coding for α-chains are associated with several inherited connective tissue disorders, including osteogenesis imperfecta (OI), Ehlers-Danlos syndrome (EDS) and its various subtypes, epidermolysis bullosa, many chondrodysplasias, Alport syndrome, and Knobloch syndrome [10,26-28].

In this review, we will briefly summarize the significance of collagen structure and conformation, and how single Gly substitutions affect the triple helix structure, leading to impaired function. We then focus on the applications of "designer collagens," i.e., synthetic collagen-mimetic peptides (CMPs) and recombinant bacterial collagen-like proteins (CLPs), which have been used to investigate the impact of Gly substitutions. Finally, we will discuss how computational modeling mechanics and dynamics are complementing *in vitro* studies of collagen biology.

Collagen's triple helical structure

Fibrillar collagens are initially produced as procollagen, which contains a triple helical domain flanked by two non-helical domains: the N- and C-propeptides. The triple helical domain has an essential $(GXY)_n$ repeating sequence, where X and Y are often proline (Pro, P) and 4-hydroxyproline (Hyp, O), respectively [29-32]. Following the synthesis of α-chains, the C-propeptide domains of all three α-chains associate via specific recognition sequences that are stabilized by inter-chain disulfide bonds [26,27,33-36]. The association of the C-propeptide domains enforces the staggered association of the three chains and nucleate triple helix folding. This nucleation step is followed by the propagation of the helix from the C- to the N-terminus of the molecule in a zipper-like fashion [10,26,37].

Interestingly, animal collagen is thermally unstable at the human body's physiological temperature and prefers to exit as random coils instead of a triple helix [38,39]. Consequently, the triple helix folding does not begin until critical post-translational modifications have occurred: the hydroxylation of at least 100 Pro residues at position Y to Hyp by prolyl-4-hydroxylase [26,40]. This delay in folding ensures the proper propagation and folding of the thermally stable triple helix, which is further stabilized by the Hsp47 chaperone [38,41-43]. Inhibiting prolyl 4-hydroxylation using α, α'-dipyridyl causes the procollagen chains to accumulate within the cell in a random-coiled form instead of a triple helix [44]. Furthermore, the rate of propagation of the triple helix is limited by the *cis* → *trans* isomerization of peptide bonds involving Pro and Hyp residues [45]. While collagen contains all peptide bonds in the *trans* state, a random coiled α-chain contains a high fraction of the peptide bonds in the *cis* state [46,47]. Thus, before an α-chain can fold into a triple helix, all the *cis* peptide bonds need to be isomerized into the *trans* state, a process that requires high activation energy and thus, can be the rate-limiting step for triple helix propagation [48]. The resulting triple helices are transported from the endoplasmic reticulum (ER) to the Golgi and secreted into the ECM via post-Golgi carriers for fibril assembly. Procollagen processing and initial fibril synthesis can occur in closed, elongated, intracellular Golgi-to-plasma membrane carriers (GPC) in both chick and mouse tendons to initiate fibril formation [49-51]. These GPCs are eventually targeted to plasma membrane protrusions projecting from the cell surface, which are called as fibripositors, and are present only during embryonic development [50]. However, in postnatal mouse tendon cells, procollagen processing also occurs in intracellular compartments even in the absence of intracellular fibrillogenesis [52]. The GPCs containing new collagen fibrils can either fuse with the plasma membrane to form a new fibripositor or fuse with the existing fibripositors [49]. In both cases, the collagen fibril grows in two different ways: (1) through the addition of new collagen molecules at the end of the growing fibril, or (2) through end-to-end fusion to generate long collagen fibers [53,54]. The cellular cytoskeleton machinery, in particular actin filaments, has been found to regulate fibripositor-mediated fibril formation and alignment of extracellular collagen fibrils (Fig. 1) [55].

One of the crucial steps in the formation of collagen fibrils is the removal of N- and C- propeptides. N-propeptide removal is catalyzed by members of ADAMTS (a disintegrin and metalloprotease with thrombospondin motifs), including ADAMT-2, ADAMT-3, and ADAMT-14, while the C-propeptide removal is catalyzed by all members of the tolloid family of zinc metalloproteinases, including BMP-1, mTLD, and TLL-1 [56-61]. The removal of both N- and C-propeptides can occur either prior to or in tandem with fibril formation [49,50]. While collagen with retained N-propeptide can still form fibrils that are thin, irregular, and loosely packed, the removal of the C-propeptide is essential for the self-assembly of collagen fibrils [62-64]. Canty-Laird et al. used embryonic and postnatal tendon explants to show that N-propeptides are removed first either within the ER-Golgi compartments or Golgi, while the C-propeptides are removed in post-Golgi compartments, suggesting their individual roles for procollagen secretion and fibril formation, respectively [65]. To further confirm the location of N-propeptide processing, loss of Golgi giantin, an integral component of the Golgi membrane, in cells caused defects in intracellular N-propeptide processing, highlighting an essential role of giantin in procollagen processing [66,67].

The three α-chains adopt a left-handed polyproline II conformation and are supercoiled into a right-handed triple helix, with each chain staggered by one residue relative to the next [37,68]. The triple helix has an essential $(GXY)_n$ repeating sequence, where X and Y are often Pro and Hyp, respectively [29]. The tight packing of the triple helix requires every third residue to be buried in the center of the triple helix. The staggering of the three chains enables the packing of Gly near the center of the helix, with a side chain of only a single hydrogen (H) atom, whereas the bulkier amino acids occupy the outer positions that are exposed to solvents. Unlike most proteins, collagen does not have a hydrophobic core. Instead, structure and stability depend on a network of inter-chain hydrogen bonds between the backbone amide nitrogen of glycine and the backbone carbonyl oxygen of the X position amino acid [28,68,69]. High Pro and Hyp content stabilize the fold. These amino acids have a rigid, five-membered pyrrolidine ring that limits the chain flexibility and enforces the polyproline-II conformation required for a triple helix [29,69,70]. The restricted ring conformation of Pro and Hyp confers stability to collagen folding by reducing the entropy cost [71]. Additionally, Hyp provides stereoelectronic stabilization via $n \rightarrow \pi^*$ interactions, which further supports the *trans* conformation of the peptide bond [72-74]. The occurrence of Hyp residues at the Y position results in faster folding by overcoming one component of the entropic barrier to folding, *cis–trans* isomerization of peptide bonds, and enables intermolecular H-bonding, which dramatically increases the structure's thermal stability [75-78]. The importance of Hyp content in maintaining the thermal stability of collagen is further demonstrated by the direct correlation between the Hyp content at the Y position and the denaturation temperature of collagens in vertebrates [79-83]. Consequently, the triple helix starts to unfold in the regions with low Hyp content, which are more thermolabile than other regions [76,84].

Five triple helices associate both laterally and axially to form a staggered array of microfibrils with distinct periodicity. The staggering process in collagen plays a critical role in producing various triple helix conformations, especially for collagens forming heterotrimers. For collagen-I, consisting of two α1 and one α2 chains, staggering could result in a conformation of three different triple helices with the α2 chain occupying either the leading, middle, or lagging strand position. In contrast, collagen-V consists of three different α-chains, which can produce six different conformations. The staggered microfibrils have an alternating gap-and-overlap region, known as a D-period, every 67 nm or 234 amino acid residues [85-87]. This D-period contains an overlap region of 0.4D and a gap region of 0.6D and is related to the length (L) of the molecule such that $L = 4.46D$ [85]. When visualized under an electron microscope after negative staining, this staggered arrangement appears striated with a characteristic light-and-dark banding pattern

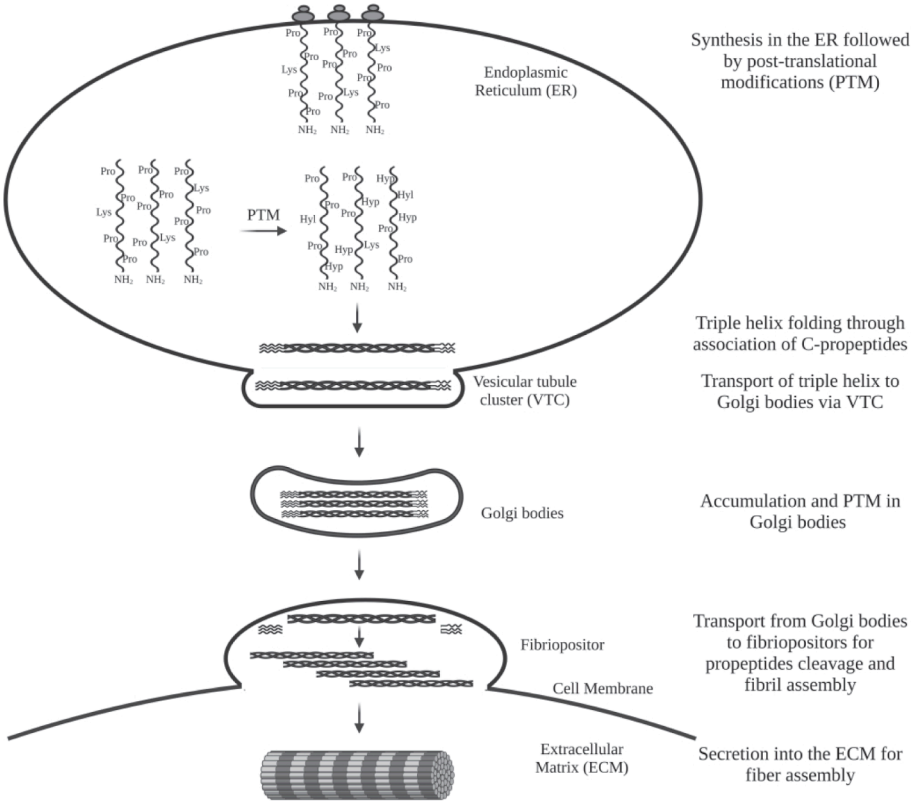

Fig. 1. Collagen synthesis and assembly. Self-assembly in fibrillar collagens from monomers to triple helix to higher-order structures, including fibrils. Created with BioRender.com.

[85,88-90]. The microfibrils further self-assemble into larger fibrils in various architectures, including parallel bundles in tendons and ligaments, orthogonal lattices in cornea, basket weaves in skin, and concentric weaves in long bones [91-95]. The fibrils are stabilized by the formation of inter- and intra-molecular crosslinks at specific locations in the telopeptide domains, which are catalyzed by lysyl oxidase [96]. Lysyl oxidase also catalyzes the intermolecular crosslinking between collagen and elastin [97]. The combination of staggering of α-chains and D-period could have important implications in collagen-cell-ECM interactions, cell behavior, and mechanical properties of fibrils, fibers, and tissues [98-101].

Disruptions of the collagen sequence

Different genetic modifications to collagen have been reported in inherited connective tissue disorders, including deletions, insertions, duplications, splice-site mutations, complex rearrangements, and single-base substitutions. Of the characterized mutations, almost 75% are missense mutations, i.e., single-base substitutions that change the codon of a critical amino acid to another amino acid [102-105]. The vast majority of disease-causing mutations are substitution of a single Gly residue with another bulkier amino acid within the $(GXY)_n$ sequence.

Effect of Gly substitutions on collagen's structure

The repetitive $(GXY)_n$ sequence pattern is not strictly conserved in non-fibrillar collagens that contain naturally occurring interruptions. These interruptions are often associated with regions of lesser stability to provide molecular flexibility and might also be involved in recognition by cellular receptors and other ECM components [113-115]. In human collagen I and III alone, more than 1200 naturally occurring mutations that interrupt the $(GXY)_n$ pattern within fibrillar regions have been noted [106]. The conservation of the repetitive sequence is an absolute necessity in fibrillar collagens, where the replacement of a single Gly residue with another amino acid can lead to severe pathological conditions (Table 1) [10,26-28]. The effects of Gly substitutions are variable and depend on the type of substituted amino acid and its location along the triple helix. The occurrence of high- and low-stability regions within a collagen triple helix further complicates this genotype-phenotype correlation [116]. Fibrillar collagens are more susceptible to glycine substitutions than non-fibrillar collagens. In non-fibrillar collagens, triple helical domains can often accommodate additional interruptions with no or mild effects [10].

In fibrillar collagens, the stoichiometry of α-chains can also impact the disease phenotype. For homotrimers, such as collagen-II and collagen-III, the α-chain from either of the two alleles can assemble to form a homotrimer. In such cases, only 12.5% of the assembled trimers contain three normal chains, while the remaining 87.5% of trimers will contain either one, two, or three pathogenic α-chains. In contrast, in the self-assembly of heterotrimers, such as collagen-I, the ratio of normal-to-pathogenic triple helices will depend on the α-chain (α1(I) or α2(I)) containing the mutation. For collagen-I, mutations in α1(I)-chains will result in 25% of normal and 75% of pathogenic triple helices, whereas mutations in α2(I)-chains will give a 50–50% ratio of normal to pathogenic triple helices.

Effects of Gly substitutions can be grouped into two broad categories: ones that prevent the folding of the triple helix, and ones that produce a structurally abnormal protein (Fig. 2) [10,26,27]. In the former case, substitutions in/near the C-terminal region prevent the C → N terminal zipper-like folding of the triple helix, leading to the accumulation of unfolded, mutant α-chains in the cells, followed by their degradation, possibly through endoplasmic reticulum-mediated or lysosomal degradation pathways [22,117,118]. In contrast, substitutions in the triple helical domain and in/near the N-terminal region permit the association of chains at the C-terminal to form the nucleation domain but partially interfere with the propagation of the triple helix, causing the synthesis of

Table 1
Inherited collagen disorders involving Gly → X substitutions with their underlying collagen source and clinical features [10,26-28,106-112].

Type	Class	Gene(s)	Pathology
I	Fibrillar	COL1A1, COL1A2	Osteogenesis Imperfecta (OI)[a], Classical-like Ehlers-Danlos Syndrome (EDS)[a], Arthrochalasia EDS (Type I and II)[a]
II	Fibrillar	COL2A1	Chondrodysplasias (Achondrogenesis type II or hypochondrogenesis[a], Multiple Epiphyseal Dysplasisa[a], Kniest Dysplasisa[a], Legg-Calve Perthese Disease[a], Otospondylomegaepiphyseal Dysplasia[b], Platyspondylic Lethal Skeletal Dysplasia[a], Spondyloepiphyseal Dysplasia Congenita[a], Stickler Syndrome[a]
III	Fibrillar	COL3A1	Vascular EDS[a]
IV	Network	COL4A1, COL4A2, COL4A3, COL4A4, COL4A5, COL4A6	Alport Syndrome[a,b], Gould syndrome[a], X-linked Deafness[b], Brain Small Vessel Disease[a], Retinal Arterial Tortuosity[a,]
V	Fibrillar	COL5A1, COL5A2, COL5A3	Classical EDS[a]
VI	Network	COL6A1, COL6A2, COL6A3, COL6A5	Bethlem myopathy[a], Ullrich Congenital Muscular Dystrophy[ab]
VII	Anchoring fibrils	COL7A1	Dystrophic Epidermolysis Bullosa[a,b], Nonsyndromic Congenital Nail Disorders[b]
VIII	Network	COL8A1, COL8A2	Fuchs Endothelial Corneal Dystrophy[a]
IX	FACIT	COL9A1, COL9A2, COL9A3	Multiple Epiphyseal Dysplasia[a,b], Stickler Syndrome[a]
X	Network	COL10A1	Schmid metaphyseal chondrodysplasia[a]
XI	Fibrillar	COL11A1, COL11A2	Stickler Syndrome[a], Stospondylomegaepiphyseal Dysplasia[a,b]
XII	FACIT	COL12A1	Bethlem myopathy[a]
XIII	MACIT	COL13A1	Congenital Myasthenic Syndrome[b]
XIV	FACIT	COL14A1	–
XV	MULTIPLEXIN	COL15A1	–
XVI	FACIT	COL16A1	–
XVII	MACIT	COL17A1	Junctional Epidermolysis Bullosa[b]
XVIII	MULTIPLEXIN	COL18A1	Knobloch Syndrome[b]
XIX	FACIT	COL19A1	–
XX	FACIT	COL20A1	–
XXI	FACIT	COL21A1	–
XXII	FACIT	COL22A1	–
XXIII	MACIT	COL23A1	–
XXIV	Fibrillar	COL24A1	–
XXV	MACIT	COL25A1	Congenital Fibrosis of the Extraocular Muscles[b]
XXVI	FACIT	COL26A1	–
XXVII	Fibrillar	COL27A1	Steel Syndrome[b]
XXVIII		COL28A1	–

[a] Represents the autosomal-dominant inheritance.
[b] Represents the autosomal-recessive inheritance of mutations.

a structurally altered protein with impaired functionality. Substitutions in the C-terminal region lead to abnormal α-chains that do not assemble into a triple helix and are ultimately degraded. As a result, there is reduced collagen secretion in the ECM that would directly influence the stiffness of the ECM, since fibrillar collagens are the major contributors to ECM stiffness [119]. These substitutions result in a milder phenotype, which has been supported by the observation of mild OI phenotype in such cases [10,120-122]. However, substitutions that result in the formation of structurally-altered proteins can be more devasting and can show mild to severe phenotypes. Such substitutions permit the association of chains to form a nucleation domain followed by the propagation of the triple helix folding with a possibility of local disruption or a kink at the substitutions site. These abnormally synthesized triple helices can exert a wide spectrum of effects, including overmodification of specific amino acid residues causing delayed and/or reduced collagen secretion, delayed fibril formation, and alteration in the morphology of fibrils and fibers [10,26,118,120,123,124]. Incorporation of abnormal fibers into the ECM can disrupt matrix deposition and their interaction with cell surface receptors and other components of the ECM, and the overall structure, organization, stiffness, and function of the ECM due to the crosslinking between abnormal and normal fibrils to generate structurally altered fibers [10,26,118,120,123,124].

In vitro studies to examine the structure–function relationship in collagen disorders

To better understand the impact of collagen substitutions on triple helix folding and improve our knowledge of the structure–function relationship, several different approaches have been utilized to model single Gly substitutions in inherited connective disorders, including the design and production of synthetic collagen-mimetic peptides (CMPs) and recombinant bacterial collagen.

Synthetic collagen-mimetic peptides (CMPs)

Since the 1960–70's [125-129], short, synthetic CMPs, also known as triple-helical peptides, have been widely used model systems to investigate the structural and biological consequences of collagen mutations. CMPs contain either repeating sequences of $(GXY)_n$ or a host-guest sequence flanked by $(GPO)_n$ repeats (Fig. 3) [130]. The host–guest sequence is defined as the introduction of a "guest" sequence in a "host" sequence, such as introducing an animal collagen sequence in $(GPO)_n$ repeats or bacterial collagen Scl2 sequence [131]. The stability of the triple helix can be improved by adding non-native amino acids, such as 4R-fluoroproline, 4R-hydroxyproline, 4R-fluoroproline, 4R-methylproline, *trans*-4-amino-L-proline, N-methylalanine, etc., within the peptide sequence or covalently crosslinking the three chains. Both of the strategies have been reviewed extensively [130,132,133].

Brodsky et al. established a propensity scale for amino acids in the X and Y positions within the $(GXY)_n$ to relate the amino acid sequence and triple helix stability [134]. Of the 20 amino acids studied, the most stable residues in the X position were Pro > Ala > Gln, whereas for the Y position, Hyp > Arg > Met, i.e., $(GPO)_n$ offers the maximum stability to the triple helix [134,135].

Host-guest CMPs incorporating various types of Gly substitutions observed in collagen disorders have been extensively used to investigate their effects on triple helix structure, stability, and folding to establish structure–function relationships of collagen mutations and relate to disease severity. Bella et al. incorporated a Gly → Ala substitution, which has been identified in several connective tissue diseases, into a $(GPO)_{10}$ peptide [19]. The crystal structure of the peptide showed two interesting features: (1) the three chains formed a triple helix with bulging or "twist relaxation" observed at the substitution site to accommodate three Ala residues within the center of the triple helix, and (2) the loss of three direct interchain H-bonds due to steric hindrance caused by the Ala side chain (Fig. 4) [19]. Four water molecules were also seen inside the molecule, which provided the necessary interchain H-bonding to maintain the overall stability of the triple helix [19]. This indicates that alterations in the $(GPO)_n$ sequence can allow the non-imino acids in the X and Y positions to participate in interchain interactions, including H-bonding, side-chain interactions, or solvent-mediated H-bonding [19,136]. The crystal structure of the $(GPO)_{10}$ peptide with Gly → Ala substitution provided crucial structural information on how a single Gly substitution can cause conformational

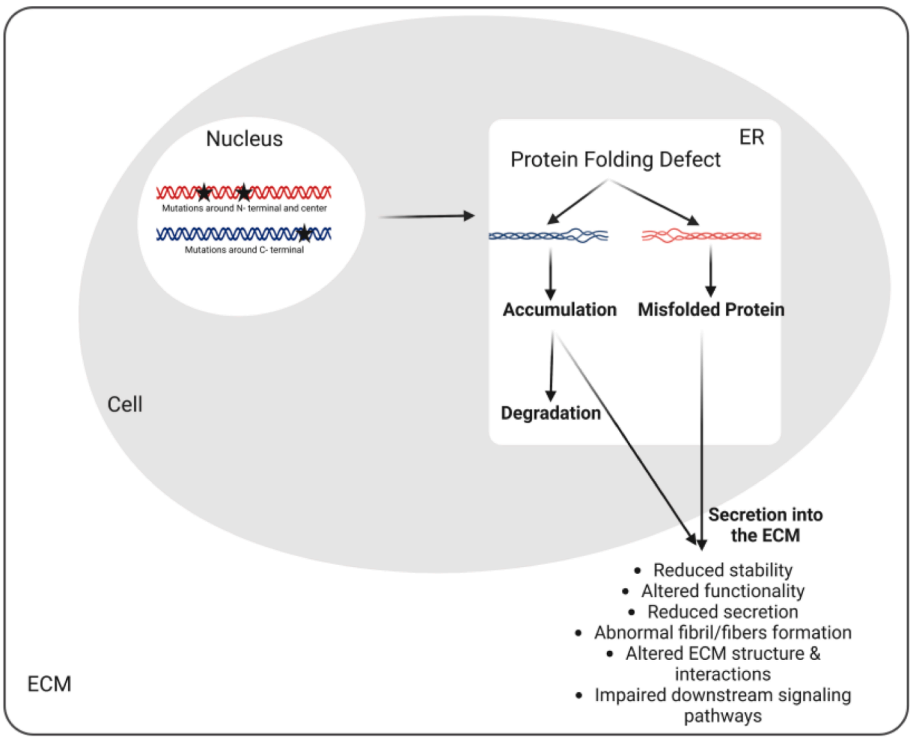

Fig. 2. Fates of defective collagens. Effect of single Gly substitutions in fibrillar collagens on protein folding, secretion, and assembly in the ECM. Created with BioRender.com.

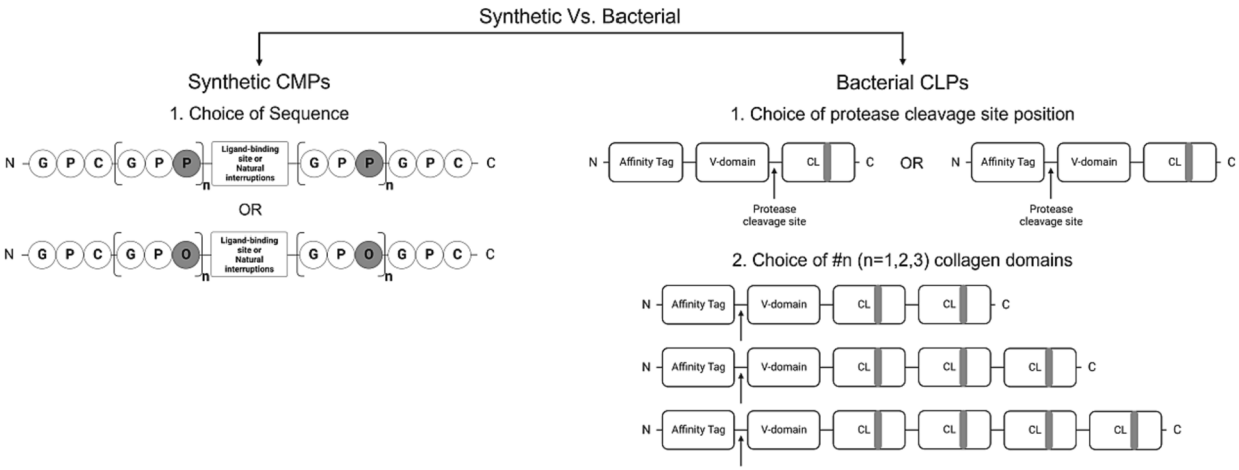

Fig. 3. Schematic structure of designer collagens. CMPs consist of $(GPP)_n$ or $(GPO)_n$ sequences to flank either a ligand-binding site or a natural interruption owing to their high stability. (GPC) sequence can also be added at both the N- and C-terminal to aid in triple helix stability *via* disulfide bonds, for monitoring peptide concentration, and enhancing the peptides attachment to tissue culture treated plastics. In contrast, CLPs (or SCl2 proteins) are designed to contain an affinity tag (such as Strep, His, MBP, GST, etc.) at the N-terminal followed by a non-collagenous sequence-variable (V) domain, which is crucial for trimerization, and collagen-like (CL) domains. A protease cleavage site can be included following either the affinity tag or the V-domain depending on the downstream application. Moreover, multiple CL domains, separated by a linker, can be included in one construct for specific applications, e.g., inclusion of integrin-binding sites (highlighted in grey) for enhanced cell attachment. Created with BioRender.com.

changes and loss of spatial arrangement of the triple helix on both sides of the substitution site. Gly → Ser substitution also confirmed the asymmetric effects of the mutation, i.e., altered H-bonding and increased structural flexibility on the C-terminal side of the substitution site versus the N-terminal side [137].

These structural changes have functional implications. For example, CMPs were designed to include reported Gly → X (X = Ala, Arg, and Val) substitutions occurring within the high-affinity integrin α2β1 binding motif, GROGER, in vascular EDS patients. These mutated sites were flanked with $(GPO)_n$ sequences at both ends [138]. The study again demonstrated that the identity of the residue in a Gly → X substitution could have a range of impacts on triple helix stability, binding with the integrin α2-inserted (α2-I) domain, and the expected severity of a disease phenotype [138]. Cell adhesion assays demonstrated that human mesenchymal stem cells (hMSCs) grown on the Gly → Ala peptide exhibited strong cell adhesion similar to wild-type peptides, while cells cultured on Gly → Val peptide adhered poorly, and those that were adherent displayed altered morphology with a "pancake-like" appearance [138]. While the host–guest peptides approach enabled the investigation of structural consequences of Gly → X substitutions

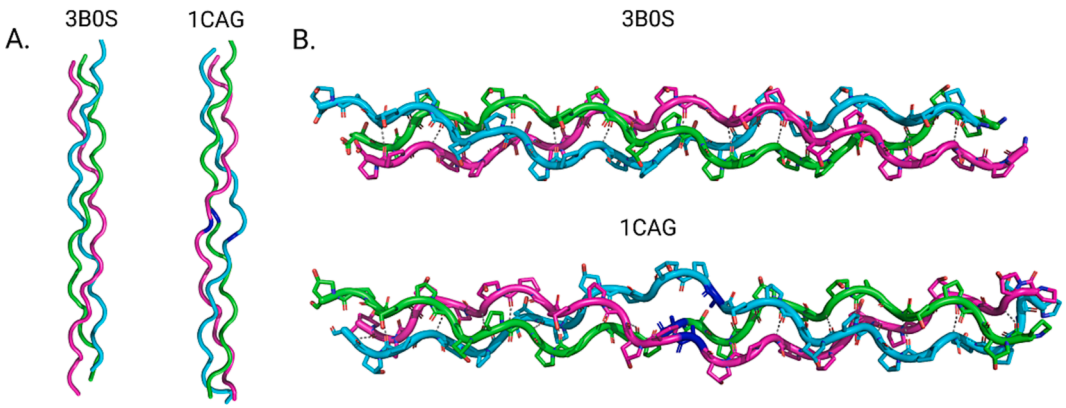

Fig. 4. **(A)**, Triple helical structure of (Gly-Pro-Hyp)^{9} peptide (PDB code 3B0S) and Gly → Ala peptide (PDB code 1CAG). A local unwinding of the triple helix can be seen at the Ala substitution site (highlighted in dark blue). **(B)**, H-bonding is highlighted in dark gray dashes between the three chains. While the normal H-bonding is seen in the native triple helix, Gly → Ala substitution hindered normal H-bonds formation at the Ala site due to bulging of the triple helix to accommodate three Ala residues in the center of the helix. Created with The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

utilizing the highly stable triple helix formed by the $(GPO)_n$ sequence, it failed to consider the role of sequence-dependent variation in the triple helix in determining the impact of Gly substitutions.

Role of CMPs in understanding the impact of local sequence microenvironment on the structure and stability of the triple helix:

Previous studies have highlighted the influence of the local sequence environment surrounding the Gly → X substitution site on the local and global stability of the triple helix. To further demonstrate how the same Gly → X substitution can have variable clinical phenotypes ranging from minimal to severe, Yang et al. designed multiple peptides incorporating a local region of the α1(I) chain of collagen-I, which contain sites for a lethal and a non-lethal Gly → Ser substitution [139]. Although substitution at either location terminated the triple helix propagation upon encountering the Ser site, they differentially impacted the thermal stability of the resulting helix. The lethal mutation caused a 22 °C reduction in the thermal stability of the peptide, which was twice the impact of the non-lethal mutation [18,139,140]. This difference in thermal stability could be attributed to their surrounding sequence, i.e., the sequence surrounding the non-lethal Gly → Ser has a higher imino acid content, whereas the sequence surrounding the lethal one has more ionizable residues [139]. Similar results were obtained when a smaller Gly → Ala substitution was introduced at the same locations [141]. These findings were further corroborated by Bryan et al., where the impact of different Gly → X substitutions (X = Ala, Ser, Arg, and Asp) was compared when surrounded by either a highly stable sequence of $(POG)_{10}$ or comparatively lower stable sequence from natural collagen with less rigidity [142]. As expected, the bulkier amino acid decreased triple helix stability (Ala ≈ Ser > Arg > Asp) [142]. Nonetheless, Gly → X substitutions had a significantly higher impact on structure and stability when surrounded by the stable $(POG)_{10}$ units versus lower stable regions [142].

All these findings highlight that collagen's triple helix is a rodlike structure with some regions more flexible than others. The helical symmetry, in other terms, "helical twist," varies along the collagen length and is determined by the local sequence [116,143,144]. Imino acid-rich regions of the helix demonstrate a tighter 7/2 helical-symmetry (7 amino acids per two helical turns) compared to imino acid-poor regions, which have a 10/3 helical-symmetry (10 amino acids

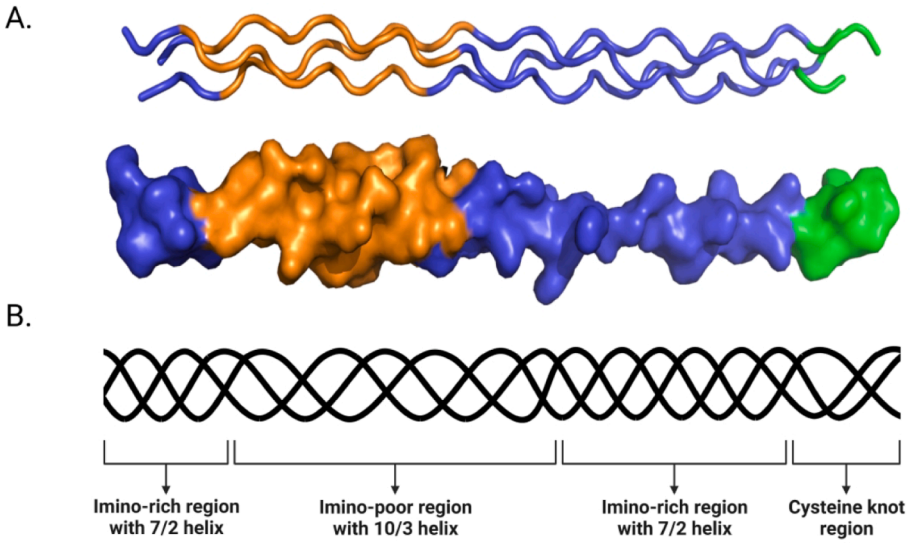

Fig. 5. **(A)**, Triple helical structure of a 42-residue C-terminal sequence from collagen-III (PDB code 3DMW) showing sequence-dependent variation of helical twist in cartoon and surface form. **(B)**, The imino rich regions demonstrate a tighter 7/2 triple helix whereas the imino poor region (OGPRGNRGERGS) is comparatively relaxed. The cysteine knot region (GPCC) demonstrates a different conformation. Created with The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC and Biorender.com.

per three helical turns) (Fig. 5) [116,143-145]. The 7/2 helical-symmetry of the triple helix arises due to the steric constraints (fixed ψ angle and restricted ϕ angle) imposed by the repeating imino acid-rich sequences (such as GPO or GPP) [143]. With an increase in the non-imino acid content in the helix, the symmetry of the triple helix changes towards the 10/3 helical-symmetry. Such variations in the triple helix symmetry have important structural and functional implications, including nucleation, interactions with cellular receptors via ligand binding sites, accessibility of enzymatic cleavage sites, etc. Moreover, there are multiple factors that ultimately affect the triple helix stability, including the identity of substituted amino acid, its location along the length of the triple helix, the local sequence environment surrounding the substitution site, and its proximity to an interaction site, such as integrin receptors, metalloproteases, proteoglycans, etc. [104].

NMR analysis of CMPs to study folding and propagation of the triple helix folding:

Owing to their small size, relatively low molecule weight, and slow folding rates, CMPs are a great tool for nuclear magnetic resonance (NMR) spectroscopy to understand the individual steps involved in the triple helix folding and conformational changes in real-time at the substitution site and N- and C-terminal of the site. Specific residues of CMPs can be labeled with ^{13}C and ^{15}N to study the kinetics of triple helix folding by monitoring the decrease in the intensity of monomer peaks and/or increase in the trimer peaks [146]. Brodsky et al. designed a peptide based on a collagen-III sequence, which was capped by $(POG)_n$ triplets at both ends for stabilization, to probe the influence of amino acid sequence on the dynamics of triple helix folding [147]. Real-time NMR data showed that the folding of peptides into a triple helix occurs in three steps: (i) the dimerization of monomer chains, (ii) the assembly of 'nucleated trimer' species in which a part of the peptide at the C-terminal is in a triple helical environment while the rest is still unfolded, and (iii) the slow propagation of the nucleated species into a triple helix [147]. These synthetic CMPs contain $(GPO)_n$ repeating sequences at both ends that not only stabilize the peptide but also act as the nucleation domain due to high imino acid content [148]. Following the nucleation, the triple helix folding propagates from the C- to N-terminus. Surprisingly, nucleation is possible at the N-terminal $(GPO)_n$ sites when imino-rich sequences are included at both ends [147,149]. On the contrary, nucleation in animal fibrillar collagens occurs at the C-terminal via the initial association of the C-propeptide region of three α-chains. This brings together the C-terminal Hyp residues in close proximity to initiate nucleation, which is followed by the formation of inter-chain disulfide bonds that further stabilize the nucleation domain [150]. Next, the sequences present in the C-propeptide and triple helical domain guide the chain alignment for the propagation of the helix in the N-terminal direction [151].

Bhate et al. further utilized NMR spectroscopy to compare the consequences of Gly → Ala substitution in the nucleation domain versus the propagation domain [141]. Similar to previous results, CMPs with an Ala substitution in the propagation domain formed a triple helix at the C-terminal, which propagated in the C- to N-terminal direction until encountering the Ala substitution site [141]. In contrast, the impact of substitution in the nucleation domain is slightly more complicated. In OI and other connective tissue disorders, Gly substitutions in the nucleation domain can prevent the folding of the triple helix nucleus, resulting in the accumulation of unfolded molecules. However, in synthetic peptides with the Ala substitution in the nucleation domain, the peptides still formed a triple helix around the substitution site due to the incorporation of stable surrounding sequences at both the N- and C-terminal of the peptides, a finding previously reported by Bella et al. [19,141]. This suggests that Gly → Ala substitution in the nucleation domain can initiate triple helix assembly, possibly through a different mode, as more N-terminal residues folded before the substitution site compared to human collagen's folding mechanism [141]. Alternative folding mechanisms have been demonstrated in CMPs in the absence of a strong nucleation domain at the C-terminal [148]. Buevich et al. demonstrated that inserting the nucleation domain at the N-terminal did not change the directionality of triple helix propagation; instead, no directionality was observed [148]. Hyde et al. replaced the imino acid-poor sequence occurring N-terminal to the Gly → Ala substitution site with a high imino acid sequence and improved the refolding of the triple helix around the substitution site as well as the N- and C-terminal [152]. Some of the other triple helix folding mechanisms could involve regions of disulfide bonds, T4-phage foldons, or coiled-coil domains at either end of the terminal, which can act as nucleating sites to initiate triple helix assembly [152-155].

One of the common outcomes of Gly → X substitution is the inability of the triple helix to fold N-terminal of the substitution site. Synthetic peptides with Gly substitutions maintained the triple helical conformation C-terminal of the substitution site but exhibited mobility of monomeric chains both at the substitution site and N-terminal from it [146,156]. It is hypothesized that the short length of the peptides prevents the renucleation of the α-chains N-terminal of the substitution site, unlike full-length collagens obtained from patients with OI and other collagen disorders, where the three α-chains can renucleate N-terminal of the substitution site to form full-length but structurally defective triple helices [139,141]. The introduction of Gly → X substitutions in human collagens causes a significant delay in the triple helix propagation around the substitution site, which ultimately affects their secretion into the ECM [20,23,120,157]. These observations are consistent with the fact that mutations in collagen α-chains result in post-translational over-modifications at the N-terminal of the mutation site [18,139]. These over-modifications occur only in the unfolded chains of the collagen, i.e., N-terminal of the mutation site, since the enzymes (including prolyl hydroxylase, lysyl hydroxylase, and glycosyltransferases) catalyzing these reactions act only on the non-helical chains and do not react with the same chains folded into a triple helix [123]. Consequently, Gly → X substitutions delay the folding of the triple helix N-terminal of the substitution site, which prolongs the exposure time between the unfolded chains and the modifying enzymes that leads to excessive hydroxylation of lysine residues and glycosylation of the hydroxylysine residues on the unfolded chains [18].

Although synthetic peptides are a valuable tool and laid the foundation to investigate the impact of single Gly substitutions on triple helix structure, folding, and stability found in collagen disorders, there are disadvantages to this system. These CMPs are comparatively expensive, and their thermal stability is much greater than that of animal collagen, which can significantly alter the structure stability and binding with active ligand sites [158]. Even with advances in coupling chemistry, yields at each step of synthesis are below 100%, resulting in truncated versions of the peptides that act as impurities. The inclusion of impurities can result in imperfect and less stable triple helix structures [159]. The CMPs are often stabilized by including cysteine residues within the terminal triplets of the peptides. However, these cysteines can undergo random, air-induced oxidation, resulting in inter- and intra-helix disulfide bonds that ultimately give rise to smaller polymers [160]. Lastly, owing to the small size (<10 nm in length) of CMPs and limited sequence diversity within 15–40 amino acid residues, it is challenging to characterize the effects of multiple Gly → X substitutions occurring in different regions of collagen on structure and function [161]. One plausible alternative to synthetic CMPs is recombinant bacterial collagen.

Recombinant bacterial collagen

More than 100 different bacterial species containing collagen-like sequences have been found, including *Streptococcus pyogenes*, *Bacillus anthracis*, *Clostridium taeniosporum*, *Pasteuria ramose*, and more [162-164]. The collagen-like sequences in different bacterial strains can serve two distinct functions: (1) they help the invading pathogen to evade the

immune system of the host, and (2) they interact and bind to multiple cell surface receptors to promote host cell invasion [165]. All of the characterized bacterial collagens vary in amino acid composition and length, yet they all contain the characteristic $(Gly\text{-}X\text{-}Y)_n$ sequence [162,166].

The bacterial collagen from *S. pyogenes* has been widely utilized in biomaterials and as a tool to investigate the sequence-structure–function relationships in normal and pathological collagens. The *S. pyogenes*, a gram-positive bacterium, produces Streptococcal collagen-like proteins 1 and 2 (Scl1 and Scl2), which have been well characterized in terms of their structural and functional properties [162,163,167-170]. Both Scl1 and Scl2 contain an N-terminal signal sequence, a sequence-variable globular domain (V-domain), a collagen-like triple helix domain (CL), which consists of repeating units of $(Gly\text{-}X\text{-}Y)_n$, and a C-terminal cell wall or membrane attachment domain. The V-domain contains a coiled-coil motif crucial for the trimerization and proper folding of the triple helix [169]. This review will mainly focus on the Scl2 protein and its applications for studying collagen-based disorders.

The Scl2 protein is larger than the Scl1 protein [162] and has been widely used to generate various constructs with sequence modifications in recombinant *E. coli* to study the effect of interruptions in the $(Gly\text{-}X\text{-}Y)_n$ repeating sequence on structure, stability, folding, and function (Fig. 3). The recombinant Scl2 protein self-assembles into a stable triple helix despite the lack of Hyp residues, which are crucial for triple helix stability in animal collagen. Compared to animal collagen, the collagenous domain of Scl2 contains a relatively low content of Pro residues (≈12%), comparable to fibril-forming animal collagen (≈20%) [171]. The most remarkable difference between animal collagen and Scl2 is the increased charged residue content, with approximately 15% in animal collagens and 30% in Scl2 [171,172]. These charged and polar amino acid residues in Scl2 stabilize the triple helix *via* electrostatic interactions, forming an ordered hydration network in association with the polar residues [171]. The melting temperature of Scl2 ranges from 35 ∼ 40 °C, which is comparable to animal collagens [162,166]. Without the need for post-translational modifications to maintain structural stability, Scl2 presents an excellent opportunity to produce large quantities of protein with high purity and minimal batch-to-batch variation.

Unlike animal collagen, which contains several ligand-binding sites that interact with cell surface receptors and other ECM components, the Scl2 protein does not inherently bind to human cell surface receptors [162,163,173]. Thus, in terms of bioactivity, Scl2 acts as a "blank-slate" and allows selective integration of specific human collagen ligand-binding sites to study distinct protein-collagen interactions. To date, we and several others have introduced one or more unique biological, functional ligand-binding sites, including integrin, fibronectin, heparin, hyaluronic acid, chondroitin-sulfate, and matrix metalloproteinases, and demonstrated cell-specific adhesion of smooth muscle cells, fibroblasts, myoblasts, and endothelial cells from different species, such as human, bovine, rat, and mouse [163,173-180]. With these modifications, Scl2 can be used to characterize the effects of specific Gly substitution at any location while controlling the surrounding sequence to attribute the impact on cell function to particular interactions with the inserted biological motifs along with the location.

One of the most established ways to investigate the triple helical nature of collagen-based peptides and proteins is through trypsin digestion assays because an intact triple helix is resistant to this enzymatic activity. However, any substitutions that can disrupt the triple helix make the structure prone to trypsin digestion. Recombinant Scl2 was first used to interrogate the effects of single Gly substitutions on the triple helix conformation and folding [181]. Cheng et al. introduced Gly → X substitutions (X = Ser/Arg) either near the middle or the N-terminal of the collagenous domain of Scl2 [181]. While all constructs showed triple helical conformation, the Gly → Arg substitution near the N-terminal showed partial cleavage of the protein upon trypsin digestion, suggesting that the Arg amino acid had a more pronounced effect on the structure of the helix. Whereas the Gly substitutions near the middle of the CL domain resulted in slower folding rates as compared to wild-type constructs, Gly substitutions near the N-terminal significantly impeded the triple helix folding [181]. Unlike animal collagen, where the triple helix folds from the C- to N-terminal, Scl2 contains the trimerization domain at the N-terminus, resulting in the triple helix folding from the N- to C-terminal. This suggests that Gly → X substitutions within or near the nucleation domain (N-terminal in Scl2 and C-terminal in animal collagens) disrupt and/or delay the triple helix folding in both animal collagens and collagen-like peptides/proteins.

The interactions of collagen with cell-surface receptors and other ECM components are dependent on the triple helical conformation of collagen. To better understand how mutations within an integrin-binding site impact this crucial interaction, Yigit et al. incorporated a human α1(I) collagen sequence containing a GFPGER integrin-binding site within the Scl2 domain [182]. From this basic design, Gly → Ser substitutions observed in OI patients were introduced within and near the GFPGER sequence to investigate their impact on structure, integrin binding, and cell adhesion [182]. Similar to the results shown by synthetic peptides, the authors found an asymmetric effect of Gly → Ser substitutions on integrin binding, i.e., substitutions within the GFPGER site and up to nine triplets N-terminal to the sequence significantly altered integrin binding and prevented cell adhesion, whereas substitutions C-terminal to the site had minimal-to-no impact [182]. It is hypothesized that the local sequence N-terminal of the GFPGER sequence could play a role in the initial recognition of the α2-I domain of the integrin receptors [183]. In a similar study, Qui et al. studied the impact of multiple Gly → X (X = Ala/Ser/Val) substitutions within and C-terminal to the GFPGER integrin-binding site in an effort to establish the relationship between the altered biophysical properties (including structural and functional outcomes) and the severity of OI [184]. Although all proteins with Gly substitutions formed a triple helix, the substitutions reduced their thermal stability and sensitivity to trypsin digestion depending on the amino acid and its location, in agreement with previous studies [138,139,182,184]. As expected, the size of the substituting amino acid had a direct effect on the destabilization of the triple helix, which in turn impacted the binding affinity between the peptide and the integrin α2-I domain. Greater degrees of destabilization (Val > Ser > Ala) generated lower binding with integrins (Ala > Ser > Val) [184]. On a similar note, Chumm et al. demonstrated the differential impact of Gly → Ser substitutions within the fibronectin (Fn) binding site of αI(II) chain of collagen-II. As the location of the substitution moved from N- to C-terminus within the Fn binding site, a higher fraction of the less stable and partially-unfolded species in equilibrium with the fully-folded species were found that impacted their thermal stability and sensitivity to trypsin digestion [185]. This is in contrast to Gly substitutions within an integrin-binding site, which had minimal impact on structure and stability [182]. This substitution also interfered with Fn binding [185-187]. These studies confirm how single Gly substitutions can have varying biological consequences, which could affect downstream cellular signaling pathways that rely on the crucial interaction between collagen and cell-surface receptors.

In addition to their potential in investigating the role of single Gly substitutions, the recombinant bacterial collagen also allows the introduction of a short stretch of sequence within the CL domain of Scl2. Hwang and Brodsky introduced natural interruptions from the α5 chain of collagen-IV, where mutations lead to Alport syndrome [188]. Three different interruptions of varying length were included: (1) a 4-residue interruption (VCL-X_4-CL), (2) a 15-residue interruption (VCL-X_{15}-CL), and (3) a Gly → Ser substitution before the 4-residue interruption (VCL(G-S)-X_4-CL) [188]. Although the 4-residue interruption caused small structural perturbations, the peptide remained resistant to trypsin digestion, indicating the formation of a tightly packed triple helix [188]. In comparison, the 15-residue interruption showed varied structural conformations. The central residues of 15-residue interruption were sensitive to trypsin, but not the residues found near the $(GXY)_n$ sequence

boundary, indicating the presence of two conformations of the triple helix with a transition from a flexible, less stable structure to a tightly-packed triple helix [188]. Both of these findings are consistent with previous observations where small interruptions (1 to 4 residues) can be accommodated within the triple helix, as seen in collagen-X, but large-sized interruptions do not pack well within the triple helix, as seen in collagen-VII, and make the triple helix sensitive to enzymatic digestion [189,190]. Furthermore, the introduction of Gly → Ser substitutions before the 4-residue interruption disrupted the folding of the triple helix and made the protein susceptible to trypsin digestion. Similar results were demonstrated by Yeo et al., where the author introduced different Gly → X (X = Glu/Val/Asp) substitutions of Alport syndrome in a much larger interruption sequence (29-residue long) [191]. Consistent with the previous results, all Gly substitutions dramatically altered the structural stability of the triple helix with local bulging at the substitution site, leading to increased sensitivity to trypsin digestion [191]. The disruption of the triple helical structure also led to a significant loss in integrin binding, which was further confirmed by cell adhesion assays [191]. While the $(GXY)_n$ sequence is strictly conserved in fibrillar collagens, the same is not true in non-fibrillar collagens, such as collagen-IV or collagen-X, where natural interruptions are well distributed throughout the molecule and play a crucial role in providing molecular flexibility and supercoiling ability for higher-order structures as well as potential sites for collagen degradation and ligand binding [115,192,193]. Despite the natural occurrence of interruptions in non-fibrillar collagens, destabilization of the triple helix and alteration in the interaction between collagen-integrin receptors indicate that the replacement of a single Gly residue within the $(GXY)_n$ sequence can dramatically influence the structure and folding of the triple helix and its complex interactions with cell-surface receptors, further strengthening the effects of Gly → X substitutions that are in close proximity to natural interruptions and leading to collagen disorders with variable clinical phenotype.

The recombinant Scl2 has proven to be an excellent model to study and replicate the folding of the animal collagen due to the following three parallels: (1) initiation of the folding *via* trimerization of the V-domain (or nucleation in animals) at one end of the helix, (2) propagation of the triple helix formation in one direction (N- to C-terminus in Scl2 versus C- to N-terminus in animal collagens), and (3) a similar temperature dependence (i.e., increased triple helix formation at low temperatures compared to high temperatures) [194,195]. The recombinant Scl2 collagen system has shown tremendous potential as a biomaterial that is used to study the functions of normal and pathological human collagen through the introduction of biologically functional molecules.

Limitations of synthetic peptides and recombinant collagen

Synthetic CMPs and recombinant CLPs have played integral roles in elucidating the structure, assembly, and biochemistry of collagen. However, almost all the studies utilized peptides that self-assemble into homotrimers. Since peptides containing $(GPO)_n$ or $(GPO)_n$ repeating units (n greater than 6) can assemble into a stable triple helix without the need for any crosslinking mechanism, it was significantly easier to synthesize short CMPs that would eventually form homotrimeric triple helices [196,197]. Although these trimers are good models for collagen-II and collagen-III, they do not correctly mimic the effect of Gly substitutions found in heterotrimeric collagen, such as collagen-I. As discussed earlier, the ratio of normal to pathogenic triple helices will depend on whether the mutation is in the α1(I) or α2(I) chain of collagen-I and their staggering to form a triple helix. As such, the homotrimeric triple helix model will potentially exaggerate the effect of Gly substitution compared to the heterotrimeric triple helix. Gauba and Hartgerink developed stable collagen heterotrimers using three different peptide sequences, each with a distinct charge, i.e., $(POG)_{10}$ - neutral, $(PKG)_{10}$ - positive, and $(DOG)_{10}$ - negative [198]. The authors demonstrated that the incorporation of an α-chain containing Gly → Ser substitution had a substantial effect on the thermal stability of the assembled heterotrimers, while the incorporation of additional Ser substitutions had dramatically less impact [198].

While these heterotrimeric triple helices helped examine the effects of one, two, or three Gly substitutions in collagen, the self-assembly of heterotrimers faces two main challenges: lack of control over composition and registration of α-chain for forming a triple helix. First, the mixing of peptides with distinct neutral, positive, and negative charges can lead to the formation of triple helices with ten different combinations: homotrimers (A_3, B_3, and C_3), heterotrimers with two same chains (A_2B, AB_2, A_2C, AC_2, B_2C, C_2B), and a single ABC heterotrimer. With two different peptide sequences, there are still four possibilities: A_3, B_3, A_2B, and AB_2. This problem is further magnified when registration of α-chains is considered. A single ABC heterotrimer can have six different registrations, where each chain can occupy the leading, middle, or lagging strand position. Studies have already shown that the terminal amino acid residue of CMP can have a significant effect on their stability [199]. As such, the use of heterotrimeric triple helices has been limited in studying normal and pathological collagen.

Another significant variation between CMPs and CLPs versus animal collagen is the folding mechanism and stability of the triple helix. Unlike animal collagen, where specific recognition sequences within the C-propeptide region are critical to the initial nucleation of the triple helix, the presence of high imino acid sequences (GPP and GPO) and the V-domain is responsible for the triple helix folding in CMPs and CLPs, respectively. As mentioned earlier, while the triple helix folds from C- to N-terminal in animal collagen, the triple helix propagates from N- to C-terminal in bacterial CLPs. Therefore, Gly substitutions near the N-terminal of the animal collagen and the C-terminal of CLPs are expected to have similar folding effects. In contrast, CMPs can nucleate and propagate from either C- or N-terminal, and thus, such models provide vital information about the impact of Gly substitutions on structure, stability, and interactions with cellular receptors. However, a nucleation domain at both N- and C-terminal can facilitate the folding of the triple helix from both directions, even in peptides containing Gly → X substitution [148].

It is also important to consider the role of sequence microenvironment around the Gly → X substitution site [141]. Nucleation sequences with high imino acid content and stability can also potentially serve as sites for renucleation to occur N-terminal of the Gly → X substitution site [200]. Hyde et al. investigated the influence of different sequences with varied length, stability, and nucleation propensity on triple helix folding N-terminal of the Gly → Ala substitution [152]. Surprisingly, while the incorporation of five triplets from collagen-I N-terminal of the site failed to generate refolding of the triple helix around the Ala mutation site, the inclusion of more imino acid-rich sequence found further downstream of the Ala mutation site, $GPO(GAO)_3$, formed a stable helix [152]. Thus, refolding of the triple helix N-terminal of the Gly → X substitution site is dependent on the sequence and can explain the folding of the triple helix containing Gly → X substitution. However, in the absence of a renucleation domain adjacent to the substitution site, there will be a delay in the triple helix folding until the next renucleation sequence is encountered, resulting in the restart of the folding. The structural perturbation in the triple helix caused by delayed folding can further have severe implications, including the propagation of an altered triple helix with helical phase shifts [201,202].

Insights from *in silico* modeling of collagen disorders

Molecular dynamics simulations

The *in vitro* studies on designer CMPs and CLPs provide critical information about the impact of Gly → X substitutions on structure and function. Combining these with computational modeling and simulations of collagen molecules has allowed scientists to further understand

the influence of mutations on collagen structure, folding, and mechanical strength at multiple levels of hierarchy, ranging from the sub-atomic scale to molecular and mesoscopic scales. Together with the experimental techniques, computational modeling, especially molecular dynamics (MD) and steered MD simulations of collagen molecules, provides critical information about the structure and stability of normal and mutated collagens as well as their interactions with human cell-surface receptors that are not possible to study at the *in vitro* level (Fig. 6).

Identifying the first crystal structure of a collagen peptide with Gly → Ala substitution was a major milestone in showing the impact of Gly → X substitutions on the triple helix and intermolecular H-bonding pattern. However, it has been shown that the crystal structures of protein samples can have different conformations in the crystal form versus in solution [203]. In comparison to the crystal structure, where all three Ala residues had interstitial water bridges instead of direct H-bonds, the MD simulations of $(GPO)_{10}$ peptide with Gly → Ala substitution demonstrated more asymmetric H-bonding [19,204]. Because of the different distribution of water molecules around the different chains of the molecule, the Ala residue in one chain was able to maintain a direct H-bond with the backbone, while the Ala residues in the other two chains could not. Since experimentally-derived crystal structures can exhibit conformational variations because of different crystallization conditions, it is important to consider multiple techniques to obtain structural information to minimize crystallization artifacts that can impact both the structure and the dynamics.

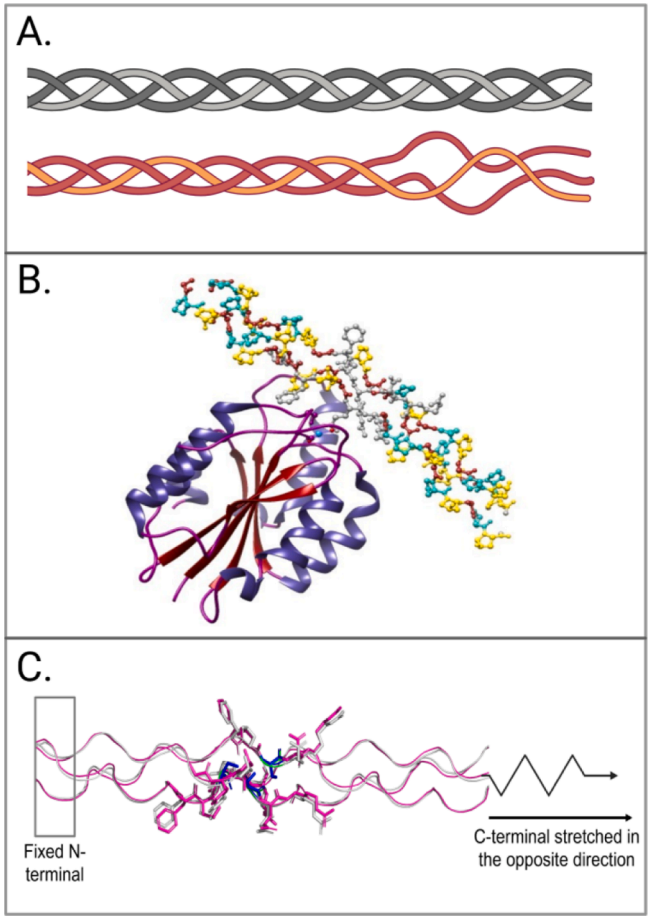

Fig. 6. Computational modeling of collagen sequences have been performed to **(A)** understand the impact of Gly → X mutations on folding and propagation of the triple helix, **(B)** to examine the interaction of collagen with crucial cell surface receptors, including integrin (PDB 1DZ1) and fibronectin, and **(C)** investigate the mechanical properties at different hierarchies of collagen, such as tropocollagen, fibrils, and fibers, using SMDs. Created with BioRender.com.

MD simulations have allowed scientists to gain insight into the folding and propagation of the triple helix and the effect of substitutions on the process. Hartmann et al. utilized unfolded $(GPP)_5$ peptides with an already-formed nucleation site at the C-terminus to study triple helix propagation [15]. The MD simulations matched *in vitro* results and showed that the propagation of the triple helix involves two steps: (1) formation of a transient template, where two α-chains associates together with the third chain still unfolded; and (2) binding of the transient template to the third chain to fold along this template, resulting in the folding of all three chains with a 3-residue stagger [15]. Changing the stability of the nucleus impacted the association of the chains during propagation and resulted in misfolded triple helices [15]. Moreover, the triple helix propagated despite the introduction of Gly → Ala/Thr substitutions in the center of one chain. Substitution in all three chains made it nearly impossible for the triple helix to fold into a stable structure [15]. Lee et al. further extended the MD simulations and applied stochastic boundary conditions to study the influence of Gly → X substitutions on the free energy of the system [205]. The relative free energy changes of mutant collagen molecules versus wild-type were larger in mutants containing two substitutions compared to a single Gly → X substitution [205]. Moreover, the free energy changes correlated well with the size of the substituting amino acid and the clinical severity of OI (Ser < Cys < Val < Glu < Arg < Asp < Trp) [205].

MD simulations have played a central role in describing the role of Gly → X substitutions on triple helix structure and how these substitutions alter the binding between collagen and cell-surface receptors or ECM molecules, including integrins and fibronectin, that play an essential role in regulating cell-signaling pathways. Yigit et al. performed MD simulations on collagen-integrin complexes with a Gly → Ser substitution within or outside of binding sites [182]. The simulations confirmed that Gly → Ser substitutions within the GFPGER integrin-binding site cause disruption in the local interchain H-bonding and alter key interactions between the integrin receptors and integrin-binding site in collagen, thus, explaining the loss of integrin binding in cell adhesion assays [182]. In contrast, no disruptions of key interactions with the integrin receptors were observed when the Gly → Ser substitutions were outside the integrin-binding site [182]. These findings were further supported by Hoop et al., where the effect of Gly → Ala/Arg/Val substitutions on the triple helix conformation and how these structural effects alter the interactions between collagen peptides and α2-I domain of integrin were examined [138]. While Ala, being a smaller amino acid, had no significant impact on the triple helix's diameter, both Arg and Val substantially increased the diameter versus wild-type peptide [138]. These results correlated well with the effects of these substitutions on collagen-α2-I domain interactions that are mediated by Mg^{2+} coordination, where the introduction of Gly → Ala substitutions had no consequences on key interactions with the integrin receptors, but Gly → Arg/Val differentially impacted α2-I-Mg^{2+} coordination [138]. Specifically, a Gly → Arg substitution caused instability in the α2-I-Mg^{2+} coordination and loss in H-bonds and salt bridges, whereas a Gly → Val substitution resulted in the extreme loss of α2-I-Mg^{2+} coordination [138]. In a similar fashion, Gly → Ser substitutions near the N- and C-terminal of the Fn binding site resulted in local unfolding of the triple helix and reduced interaction between the β-strand of collagen peptide and β-sheet of FnI domain, significantly disrupting the formation of interchain H-bonds and salt bridges that can extend up to three-four triplets [185].

MD simulations can be further extended to understand the differential sensitivity of Gly → X substitutions within collagen peptides to digestion by common proteases. Introduction of the Glu/Val or Asp substitution within a natural interruption sequence increased the protein–protein contacts with trypsin by more than 20% and 40%, respectively, in comparison to the wild-type sequence that had the lowest number of contacts between collagen peptide and trypsin [191]. Again, this highlights the impact of local disruptions in the triple helix structure

from Gly substitutions, the magnitude of which is modulated by the type of substituted amino acid, its location along the triple helix, and the local sequence surrounding the substitution.

Steered molecular dynamics (SMD) simulations

Given the rich biomechanical functions of collagens in the body, it is important to understand the mechanical properties of collagenous tissue at different hierarchies, wherein tropocollagen molecules are the building blocks. While uniaxial testing of tissues has provided critical information about their mechanical properties, they fail to consider the information about individual tropocollagen molecules whose arrangement governs the macroscopic properties of tissues. Experimental techniques based on the application of mechanical forces to a single molecule have played a crucial role in studying the interactions between ligands and biomolecules in response to an external force. When combined with molecular modeling, these techniques provide a powerful approach to understand the molecular mechanics of collagen at different hierarchical levels at spatial and temporal scales that are difficult to study experimentally. SMD is one such technique where time-dependent forces can be applied to a single collagen molecule to investigate its mechanical behavior and changes due to alterations in the amino acid sequence [206,207].

While conventional MD simulations can be used to capture conformational changes in proteins and other biomolecules of interest, such biological processes usually occur over a timescale of microseconds to milliseconds and beyond, which makes it significantly challenging for MD simulations to achieve enough conformational sampling from one state to another within the time constraints of a single MD simulation [208]. SMD offers a significant advantage over conventional MD simulations by inducing large conformational changes with enhanced sampling timescales that are otherwise too slow to model. This is achieved by applying a moving harmonic potential (in the form of a spring) to induce motion along certain degrees of freedom (usually a cardinal direction). The application of the external force field drives the system to move from its initial equilibrium state to different metastable states to provide crucial atomistic-level information about the biological process under investigation [209,210]. Moreover, this method can compute free-energy differences between states based on the work applied to the system for an ensemble of trajectories using the Jarzynsky equality [211]. It is important to note that SMDs are non-equilibrium simulation methods, where the harmonic constant and the pulling velocity are vital to understand the mechanical behavior of proteins.

Lorenzo et al. used SMD simulations to relate the mechanical properties of tropocollagen to molecular behavior and factors governing its structure and stability [212]. The repeated $(GPO)_n$ sequence in collagen allows the formation of an intramolecular H-bond network between the NH group of Gly and the CO group of Pro and Hyp as well as with the NH group of other Gly residues located in adjacent chains, which plays a critical role in maintaining the structural stability of tropocollagen molecules [212,213]. During the SMD simulations, stretching of the molecule led to the redistribution of H-bonds, which caused a reduction in the Gly-Pro interactions and an increase in both Gly-Hyp and Gly-Gly interactions with the total amount of H-bonds remaining approximately constant [212]. However, when the central Gly was replaced with Ala, the substituted amino acid could not preserve the interactions [212]. In the same manner, Gautieri et al. introduced seven different OI mutations in a tropocollagen molecule (Ala, Ser, Cys, Arg, Val, Glu, and Asp) to relate the effects on the mechanical properties to the severity of the disorder [214]. Consistent with *in vitro* results, the substitution of Gly with a large, bulky amino acid reduces the Young's modulus of the molecule [214]. This reduction in Young's modulus can be further correlated with the side-chain volume and hydropathy index of the substituting amino acid, i.e., an amino acid with a side-chain volume of $\approx$100 Å^3 and hydropathy index of -3.75 leads to severe phenotype in OI [214]. Given that Gly residue occupies every third position and is buried into the core of the triple helix, substituting Gly with a hydrophobic residue can have minimal impact on the helix disruption because of its fit within the collagen's hydrophobic core, while substitution to a hydrophilic residue can make up for the loss in mechanical properties by forming H-bonds *via* water molecules [214,215]. Thus, the size-chain volume and hydropathy index of an amino acid impact the mechanical response of the triple helix and could be used to predict the severity of the disease.

The local surrounding sequence of Gly → X substitutions also plays a critical role in determining the stability and conformation of the triple helix, which in turn govern its mechanical properties and structural integrity. Mutations within highly stable regions can have drastic effects on triple helix structure and function compared to mutations in less stable regions. Srinivasan et al. applied SMD to study the molecular mechanisms of mutations and their impact on the local sequence environment of collagen type-IV in Alport syndrome [216]. While some Gly → X substitutions caused significant structural changes, including kinking of the triple helix, others showed minimal impact, where perturbations were primarily localized around the substitution site [216]. Moreover, the introduction of the Gly → X substitution severely altered the mechanical properties of tropocollagen, which was demonstrated by following the stress–strain response of the molecule over time and comparing it to the wild-type. As the clinical severity of a mutation increased, a large variation was observed in the elastic modulus. In contrast, mutations associated with less severe clinical presentation exhibited the lowest variability compared to the wild-type [216].

In addition to affecting the mechanical properties of individual tropocollagen molecules, Gly → X may also impact the properties of higher-order structures. To this end, Marlowe et al. built a collagen fibril model utilizing pre-optimized tropocollagen molecules to examine the effect of specific Gly → Ala/Ser/Cys/Asp substitutions on the structure and mechanical properties of fibrils [217]. As expected, the substitution of Gly with a smaller amino acid (Ala, Ser, and Cys) caused local bulging of the triple helix with longer folding times. Substitution with a large, bulky amino acid (Arg) significantly destabilized the triple helix structure of tropocollagen molecules by preventing complete triple helix formation as demonstrated by *in vitro* studies, and in turn, altered the structure, shape, and mechanical properties of mutant fibrils [18,20,217,218]. The results showed that the mechanical strength of mutant fibrils decreased with an increase in the triple helix destabilization [217]. Similar results were observed when the fibrils were sheared by pulling the central tropocollagen, i.e., the interaction between tropocollagen molecules declined as the size of the amino acid increased, which altered the H-bonding pattern between different tropocollagens [217]. This study highlights how Gly → X substitutions not only impact the mechanical properties of tropocollagen at the molecular level, but also alter how these structurally-altered tropocollagen interact with other tropocollagen molecules to form higher-order structures, such as fibrils, fibers, and tissues.

While both MD and SMD simulations have been instrumental in elucidating the impact of Gly → X substitutions on the mechanical properties of individual tropocollagen molecules as well as higher-order fibrils, these studies suffer from a major limitation: the Gly substitutions were introduced into already assembled, energy-minimized, and stabilized tropocollagen molecules. Certain bulky substitutions, such as Asp and Arg, will prevent the formation of a triple helix structure, while others can cause folding defects. The mechanical properties of tropocollagen molecules will be perturbed in both cases, i.e., creating a triple helix structure utilizing the amino acid sequence containing the substitution versus introducing substitution within an already assembled triple helix structure. It would be interesting to explore the time-dependent assembly of tropocollagen molecules containing Gly → X substitutions to obtain triple helix structures that closely resemble the folding of α-chains *in vivo*. Although such studies would be computationally expensive, they would provide essential information about the structural and functional impact of Gly substitutions at different

hierarchies.

Self-assembly of designer collagens into higher-order structures

In animal collagens, five triple helices pack together to form a microfibril, which again self-assembles in axial and lateral directions to create a fiber [219]. This complex and multistep self-assembly is specific to native collagen only; replicating this multi-hierarchical structure has been one of the significant challenges in designer collagens, limiting their use for biomedical applications. The Brodsky group was the first to demonstrate the formation of higher-order aggregates with a filamentous and branched structure formed by self-association of $(POG)_{10}$ repeats. This groundbreaking study laid the foundation for future research to create higher-order structures using synthetic peptides.

Higher-order assembly of CMPs

Various strategies have been employed to create self-assembling CMPs, including electrostatic, hydrophobic, and aromatic interactions, cysteine knots, and chemical crosslinking. Kishimoto et al. used chemical crosslinking agents to generate nanofiber-like structures (10 nm diameter) of poly $(POG)_{10}$ peptides [197]. Similarly, Paramonov et al. utilized a native chemical ligation technique in a $(POG)_9$ sequence containing an N-terminal cysteine and a C-terminal thioester (Fig. 7A) [220]. This resulting native chemical ligation led to the formation of a dense fibril network (10–20 nm in diameter) [220]. While both studies showed the potential of CMPs in generating higher-order structures with a diameter comparable to a collagen fibril, no characteristic light-and-dark banding pattern was observed, which could be explained by the lack of extended sticky ends to generate fibrils in axial and lateral directions.

Another widely used technique to drive the supramolecular assembly of CMPs is a cysteine (Cys) knot that creates disulfide linkages to crosslink the triple helices. Including the Cys knot increases the stability of the resulting triple helices and controls higher-order assembly [221]. In one approach, Koide et al. designed $(POG)_n$-based sequences to include Cys residues within the sequence to provide staggered overhangs that acted as cohesive ends and formed large supramolecular structures, which were confirmed by ultrafiltration and laser diffraction particle size analyzer (Fig. 7B i) [222]. In another approach, Kotch and Raines utilized a cysteine knot to crosslink three triple helices, resulting in $(PYG)_5$ overhanging sticky segments that enabled the intermolecular self-assembly into fibrils of varied lengths (30–400 nm) (Fig. 7B ii) [223]. These very long fibrils are the only collagen-like assemblies produced by synthetic peptides that are longer than the length of a tropocollagen molecule (300 nm). Recently, Krishna and Kiick

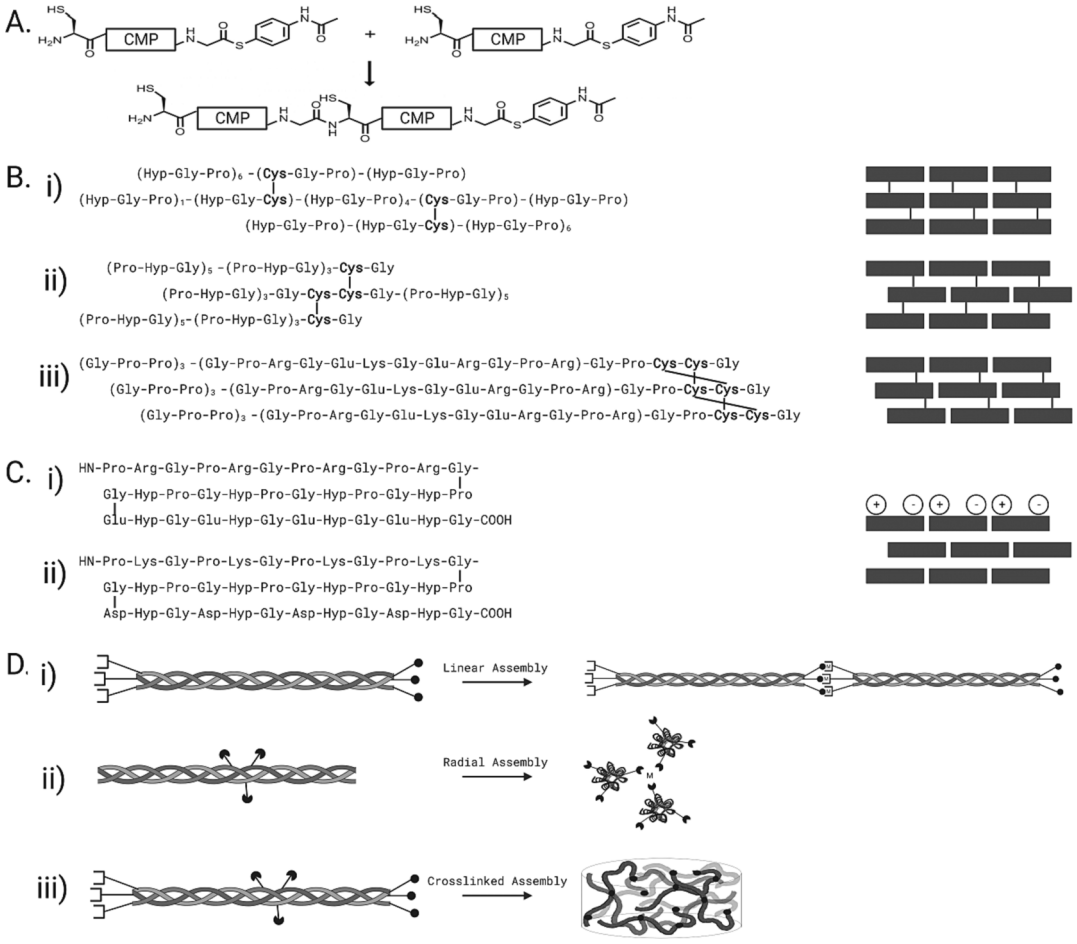

Fig. 7. **(A)**, Assembly of collagen fibrils using native chemical ligation technique. **(B)**, Cysteine knots create disulfide bonds to promote self-assembly of CMPs, where staggering of triple helices will depend on the Cys location. The figure on the right demonstrates the supra-molecule formation with each set of synthetic peptides. **(C)**, Use of electrostatic interactions to drive microfibril formation (as seen on the right) in CMPs with i) Arg and Glu residues, and ii) Lys and Asp, as positively and negative charged, respectively. **(D)**, Three different designs for metal-ion mediated assembly of CMPs along with their self-assembly organization as shown on the right. i) linear design: metal-binding ligands at the N- and C-terminal of the $(POG)_n$ backbone enables head-to-tail assembly of triple helices (side view), ii) radial design: metal-binding ligands at present at the center of the backbones and enables a radial assembly of triple helices (top view), and iii) crosslinking design: metal-binding ligands are present at the N- and C-terminal and the center of the backbone, forming a crosslinked network (side view). The metal-binding ligands are depicted as black solid spheres, bracket, and half-moons. M stands for metal. Created with Biorender.com.

combined cysteine knots and charged amino acids to drive supramolecular assembly into fibrils, ranging from nano- to microscale in length (Fig. 7B iii) [224].

While all the above-listed techniques have successfully created collagen-like fibrils morphology, none of them could recapitulate the light-and-dark banding pattern of native collagen fibrils. Utilizing electrostatic interactions while designing CMPs has successfully driven fibril formation that displayed characteristic D-banding pattern observed in native collagen. Rele et al. designed a self-assembling peptide that included a hydrophobic core domain of $(POG)_n$ repeating units flanked by distinct peptide repeats containing either a positively (Arg) or a negatively (Glu) charged residue (Fig. 7C i) [225]. The presence of charged residues at the terminal facilitated a linear assembly of staggered triple helices, followed by their self-assembly into stable microfibrils with a D-periodicity of approximately 18 nm [225]. While this D-periodicity is significantly lower than native collagen (67 nm), it is an important milestone in creating collagen-mimetic microfibrils with precisely defined periodicity. O'Leary et al. further improvised the peptide sequence by replacing the Arg residue with Lys and the Glu residue with Asp to achieve Lys-Asp salt-bridge hydrogen bond formation (Fig. 7C ii) [226]. The peptides self-assembled into nanofibers, where the triple helix packing was similar to native collagen, and these nanofibers further assembled to form hydrogels [226].

Metal-ion mediated self-assembly of CMPs is another common technique that can direct the assembly of the triple helices into higher-order structures. The Chmielewski group utilized this technology to drive the hierarchical self-assembly in synthetic collagen peptides based on metal–ligand interactions [158,221,227,228]. Moreover, various structures can be produced by selecting a specific metal ion and including the associated metal-coordinating ligands at a distinct location. There are three different design principles based on metal-promoted self-assembly: (1) linear design- where the metal-binding ligands are positioned at the N- and C-terminal of the $(POG)_n$ backbone, enabling head-to-tail assembly of triple helices (Fig. 7D i); (2) radial design- where metal-binding ligands occupy the center positions of the $(POG)_n$ backbone that will allow the radial growth of triple helices (Fig. 7D ii); and (3) crosslink design- is a combination of linear and radial design, where metal-binding ligands are placed both N- and C-termini and the center of the $(POG)_n$ backbone, giving rise to crosslinked network formation (Fig. 7D iii) [227,229-234].

Lastly, incorporating hydrophobic interactions at the termini of the backbone has also been shown to drive the self-assembly of CMPs into fibers and discs. Cejas et al. designed a synthetic $(POG)_{10}$ peptide containing a pentafluorophenylalanine residue at the C-terminal and a Phe residue at the N-terminal, which self-assembled into fibrillar-like structures, similar to collagen fibrils, due to inter-chain aromatic-stacking and ordered hydrophobic interactions [235]. Moreover, the fibril-forming peptide induced platelet aggregation with similar potency as collagen type-I, highlighting the biological activity of self-assembled fibrils. Similarly, McGuinness et al. strategically placed hydrophobic residues (such as Leu or Ile) along the peptide backbone to drive their self-assembly into nanofibers and nanodiscs of varied morphology [236].

Synthetic CMPs have come a long way since their inception. The initial focus was on their usage as collagen surrogates to study the structure and function of collagen with the primary goal of attaining a thermally stable triple helix structure. In recent years, these peptides have demonstrated outstanding potential in mimicking collagen's architecture, including the ability to form fibrils with a D-banding pattern. However, there are still significant challenges that are currently limiting their applications in the biomedical industry. The synthetic fibrils are significantly smaller in length when compared with those of natural collagen with similar diameters between the two, creating the need for designs to improve the self-assembly in the axial direction further is highly desirable. Early work demonstrating cell encapsulation and platelet aggregation is promising and highlights their ability to mimic collagen's biological functions. Still, there is a greater necessity to include essential ligand binding sites and growth factors in these higher-order structures to facilitate cell–cell and cell-ECM interactions for their use as scaffolds for cell and tissue growth.

Higher-order assembly of CLPs

Limitations associated with CMP-derived fibril size and the expense of scaling up peptide production are potentially addressable using designer collagens based on CLPs that can have significantly longer sequences and be expressed recombinantly with high yields and purities. A major goal in CLP design is targeting the D-banding pattern of collagen type-I. Col108 was a synthetic collagen designed by combining fragments from the α1 chain of human collagen type-I. Col108 formed D-periodic bands with a spacing of 35 nm, corresponding to hydrophobic residues on the surface of the helix every 123 residues [237]. This is consistent with the structure of the triple helix, which has a rise of 2.89 Å per residue [32]. Another class of CLPs based on fragments of the bacterial collagen Scl2 also produced banded structures [238]. In these designs, two proline-rich adhesive sequences flanked a 'functional-driver' domain whose sequence could be altered to control morphology and function. The banding pattern could be precisely controlled by sequence design, producing banded fibrils that are 500–800 nm in length. Robust expression and assembly of these proteins allowed structural characterization by cryo-EM tomography, indicating a unique, log-cabin style packing instead of the staggered packing of Col108 and natural collagen. Despite these differences, the inclusion of cell-activating integrin-binding sequences in the functional domains resulted in tissue-culture substrates that encouraged adhesion, growth, and early stages of differentiation of bone progenitor cells. CLPs such as these are a powerful testbed for developing biomaterials with targeted mechanical properties and cellular activities and a useful model system for answering basic questions regarding collagen biology and disease.

Conclusion and future directions

As a research tool, designer collagens have provided a tremendous amount of knowledge to understand the basic characteristics of collagen folding into a stable, triple helix structure and biological interactions with crucial cell-surface receptors and ECM components. By establishing a sequence-structure–function relationship for the triple helix, researchers have determined why some Gly substitutions are more deleterious than others in inherited collagen disorders. The detailed information about the triple helix structure gained through synthetic and recombinant bacterial collagen highlights their potential for a wide range of applications, including probing cell-ECM interactions. Recombinant bacterial collagen can potentially be engineered with defined mechanical properties while presenting matrix cues (such as cell-adhesive ligands) to gain a deeper understanding of how ECM parameters influence cell behavior and function. Future studies should focus on engineering designer collagens that form higher-order structures such as fibrils and fibers to further explore the role of cellular mechanotransduction pathways.

Introducing cells to these designed, higher-order structures can enable investigation of mechanobiology in mimics of normal and pathologic tissues to inform and refine clinical treatments. For example, with the emerging understanding of crosstalk between collagen and the tumor microenvironment contributing to initiation and progression in different types of cancer, there is a need to develop model systems to mimic the complex and diverse cell-tissue microenvironment interactions. Designer collagens provide yet another opportunity to create engineered three-dimensional (3D) tumor models to study the basics of cancer biology in a native-like microenvironment as well as for testing the efficacy of new and available therapeutic agents to improve clinical outcomes.

The field of recombinant bacterial collagen has also inspired the

development of collagen-based biomaterials and synthetic ECM for biomedical applications to avoid animal-derived collagens and growing concerns associated with their use. Animal-derived collagen poses a potential risk of transmitting diseases and other pathogens and can cause immunogenic reactions, thus posing a great safety issue [163,239-242]. Additionally, it is impossible to modify the extracted collagen to improve its functional properties. Other concerns include batch-to-batch variation as well as different species, high extraction costs, difficulty in extracting minor collagen types, as well as religious and ethical issues [243-247]. In combination with the possibilities offered by synthetic biology and protein engineering and advances in 3D printing and biomanufacturing, these tissue-engineered biomaterials will pave the way to create complex systems with multiple hierarchies and different types of collagen and other ECM components to develop designer biomaterials for regenerative medicine. Such systems will provide a native tissue-like microenvironment, allowing precise control over cell behavior and signaling, with the goal of engineering tissue structure and function.

CRediT authorship contribution statement

Sonal Gahlawat: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Vikas Nanda:** Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition. **David I. Shreiber:** Funding acquisition, Visualization, Writing – review and editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We are grateful for funding from the following sources: New Jersey Health Foundation (Grant number: PC101-20) and Marfan Foundation (Grant number: 127611) awarded to D.I.S and National Aeronautics and Space Administration (Grant number: 80NSSC18M0093) awarded to V. N.

References

[1] S. Ricard-Blum, The Collagen Family, Cold Spring Harb Perspect Biol. 3 (2011), a004978, https://doi.org/10.1101/cshperspect.a004978.
[2] K. Tarnutzer, D. Siva Sankar, J. Dengjel, C.Y. Ewald, Collagen constitutes about 12% in females and 17% in males of the total protein in mice, Sci Rep. 13 (2023) 4490, https://doi.org/10.1038/s41598-023-31566-z.
[3] K.A. Jansen, A.J. Licup, A. Sharma, R. Rens, F.C. MacKintosh, G.H. Koenderink, The role of network architecture in collagen mechanics, Biophys J. 114 (2018) 2665–2678, https://doi.org/10.1016/j.bpj.2018.04.043.
[4] L.D. Muiznieks, F.W. Keeley, Molecular assembly and mechanical properties of the extracellular matrix: a fibrous protein perspective, Biochimica et Biophysica Acta (BBA), Molecular Basis of Disease (1832 (2013)) 866–875, https://doi.org/10.1016/j.bbadis.2012.11.022.
[5] T.J. Wess, Collagen Fibril Form and Function, in: Advances in Protein Chemistry, Academic Press, 2005: pp. 341–374. 10.1016/S0065-3233(05)70010-3.
[6] J. Elango, C. Hou, B. Bao, S. Wang, J.E. Maté Sánchez de Val, W. Wenhui,, The molecular interaction of collagen with cell receptors for biological function, Polymers (Basel) 14 (2022) 876, https://doi.org/10.3390/polym14050876.
[7] K.E. Kadler, C. Baldock, J. Bella, R.P. Boot-Handford, Collagens at a glance, J Cell Sci. 120 (2007) 1955–1958, https://doi.org/10.1242/jcs.03453.
[8] F. Kai, H. Laklai, V.M. Weaver, Force matters: biomechanical regulation of cell invasion and migration in disease, Trends Cell Biol. 26 (2016) 486–497, https://doi.org/10.1016/j.tcb.2016.03.007.
[9] M. van der Rest, R. Garrone, Collagen family of proteins, FASEB J. 5 (1991) 2814–2823.
[10] J. Myllyharju, K.I. Kivirikko, Collagens, modifying enzymes and their mutations in humans, flies and worms, Trends Genet. 20 (2004) 33–43, https://doi.org/10.1016/j.tig.2003.11.004.
[11] S. Ricard-Blum, F. Ruggiero, The collagen superfamily: from the extracellular matrix to the cell membrane, Pathol Biol (Paris). 53 (2005) 430–442, https://doi.org/10.1016/j.patbio.2004.12.024.
[12] A. Bachmann, T. Kiefhaber, S. Boudko, J. Engel, H.P. Bächinger, Collagen triple-helix formation in all-trans chains proceeds by a nucleation/growth mechanism with a purely entropic barrier, Proc Natl Acad Sci U S A. 102 (2005) 13897–13902, https://doi.org/10.1073/pnas.0505141102.
[13] D. Silver, J. Miller, R. Harrison, D.J. Prockop, Helical model of nucleation and propagation to account for the growth of type I collagen fibrils from symmetrical pointed tips: a special example of self-assembly of rod-like monomers, Proc Natl Acad Sci U S A. 89 (1992) 9860–9864.
[14] S. Boudko, S. Frank, R.A. Kammerer, J. Stetefeld, T. Schulthess, R. Landwehr, A. Lustig, H.P. Bächinger, J. Engel, Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics, Journal of Molecular Biology. 317 (2002) 459–470, https://doi.org/10.1006/jmbi.2002.5439.
[15] J. Hartmann, M. Zacharias, Mechanism of collagen folding propagation studied by Molecular Dynamics simulations, PLoS Comput Biol. 17 (2021) e1009079.
[16] A.R. Walmsley, M.R. Batten, U. Lad, N.J. Bulleid, Intracellular retention of procollagen within the endoplasmic reticulum is mediated by prolyl 4-hydroxylase, J Biol Chem. 274 (1999) 14884–14892, https://doi.org/10.1074/jbc.274.21.14884.
[17] P.H. Byers, K.A. Holbrook, G.S. Barsh, L.T. Smith, P. Bornstein, Altered secretion of type III procollagen in a form of type IV Ehlers-Danlos syndrome, Biochemical studies in cultured fibroblasts, Lab Invest. 44 (1981) 336–341.
[18] H.P. Bächinger, N.P. Morris, J.M. Davis, Thermal stability and folding of the collagen triple helix and the effects of mutations in osteogenesis imperfecta on the triple helix of type I collagen, Am. J. Med. Genet. 45 (1993) 152–162, https://doi.org/10.1002/ajmg.1320450204.
[19] J. Bella, M. Eaton, B. Brodsky, H.M. Berman, Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution, Science. 266 (1994) 75–81. 10.1126/science.7695699.
[20] M. Raghunath, P. Bruckner, B. Steinmann, Delayed Triple Helix Formation of Mutant Collagen from Patient with Osteogenesis Imperfecta, J. Mol. Biol. 236 (1994) 940–949, https://doi.org/10.1006/jmbi.1994.1199.
[21] L.T. Smith, U. Schwarze, J. Goldstein, P.H. Byers, Mutations in the COL3A1 Gene Result in the Ehlers-Danlos Syndrome Type IV and Alterations in the Size and Distribution of the Major Collagen Fibrils of the Dermis, J, Invest. Dermatol. 108 (1997) 241–247, https://doi.org/10.1111/1523-1747.ep12286441.
[22] A. Forlino, N.V. Kuznetsova, J.C. Marini, S. Leikin, Selective retention and degradation of molecules with a single mutant alpha1(I) chain in the Brtl IV mouse model of OI, Matrix Biol. 26 (2007) 604–614, https://doi.org/10.1016/j.matbio.2007.06.005.
[23] K. Mizuno, S. Boudko, J. Engel, H.P. Bächinger, Vascular Ehlers-Danlos Syndrome Mutations in Type III Collagen Differently Stall the Triple Helical Folding, J Biol Chem. 288 (2013) 19166–19176, https://doi.org/10.1074/jbc.M113.462002.
[24] L.S. Mirigian, E. Makareeva, E.L. Mertz, S. Omari, A.M. Roberts-Pilgrim, A. K. Oestreich, C.L. Phillips, S. Leikin, Osteoblast malfunction caused by cell stress response to procollagen misfolding in α2(I)-G610C mouse model of osteogenesis imperfecta, J Bone Miner Res. 31 (2016) 1608–1616, https://doi.org/10.1002/jbmr.2824.
[25] S. Ishikawa, T. Kosho, T. Kaminaga, M. Miyamoto, Y. Hamasaki, S. Yoshihara, S. Hayashi, K. Igawa, Endoplasmic reticulum stress and collagenous formation anomalies in vascular-type Ehlers-Danlos syndrome via electron microscopy, The J. Dermatol. 48 (2021) 481–485, https://doi.org/10.1111/1346-8138.15766.
[26] H. Kuivaniemi, G. Tromp, D.J. Prockop, Mutations in collagen genes: causes of rare and some common diseases in humans, FASEB J. 5 (1991) 2052–2060, https://doi.org/10.1096/fasebj.5.7.2010058.
[27] J. Myllyharju, K.I. Kivirikko, Collagens and collagen-related diseases, Ann Med. 33 (2001) 7–21, https://doi.org/10.3109/07853890109002055.
[28] M.D. Shoulders, R.T. Raines, Collagen structure and stability, Annu Rev Biochem. 78 (2009) 929–958, https://doi.org/10.1146/annurev.biochem.77.032207.120833.
[29] J.A.M. Ramshaw, N.K. Shah, B. Brodsky, Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host-Guest Triple-Helical Peptides, J. Struct. Biol. 122 (1998) 86–91, https://doi.org/10.1006/jsbi.1998.3977.
[30] A. Rich, F.H.C. Crick, The Structure of Collagen, Nature 176 (1955) 915–916, https://doi.org/10.1038/176915a0.
[31] A. Rich, F.H. Crick, The molecular structure of collagen, J Mol Biol. 3 (1961) 483–506, https://doi.org/10.1016/s0022-2836(61)80016-8.
[32] G.N. Ramachandran, G. Kartha, Structure of collagen, Nature 174 (1954) 269–270, https://doi.org/10.1038/174269c0.
[33] J.D. Schofield, J. Uitto, D.J. Prockop, Formation of interchain disulfide bonds and helical structure during biosynthesis of procollagen by embryonic tendon cells, Biochemistry 13 (1974) 1801–1806, https://doi.org/10.1021/bi00706a004.
[34] P.H. Byers, E.M. Click, E. Harper, P. Bornstein, Interchain disulfide bonds in procollagen are located in a large nontriple-helical COOH-terminal domain, Proc Natl Acad Sci U S A. 72 (1975) 3009–3013.
[35] N.P. Morris, L.I. Fessler, A. Weinstock, J.H. Fessler, Procollagen assembly and secretion in embryonic chick bone, J. Biol. Chem. 250 (1975) 5719–5726, https://doi.org/10.1016/S0021-9258(19)41238-6.

[36] L.N. Lukens, Time of occurrence of disulfide linking between procollagen chains, J. Biol. Chem. 251 (1976) 3530–3538, https://doi.org/10.1016/S0021-9258(17)33376-8.

[37] J.H. Fessler, Self-assembly of collagen, J Supramol Struct. 2 (1974) 99–102, https://doi.org/10.1002/jss.400020204.

[38] E. Leikina, M.V. Mertts, N. Kuznetsova, S. Leikin, Type I collagen is thermally unstable at body temperature, Proc Natl Acad Sci U S A. 99 (2002) 1314–1318, https://doi.org/10.1073/pnas.032307099.

[39] H.P. Bächinger, J. Engel, Thermodynamic vs. kinetic stability of collagen triple helices, Matrix Biol. 20 (2001) 267–269, https://doi.org/10.1016/S0945-053X(01)00138-X.

[40] D.J. Prockop, Mutations that alter the primary structure of type I collagen, The perils of a system for generating large structures by the principle of nucleated growth, J Biol Chem. 265 (1990) 15349–15352.

[41] M. Tasab, M.R. Batten, N.J. Bulleid, Hsp47: a molecular chaperone that interacts with and stabilizes correctly-folded procollagen, EMBO J. 19 (2000) 2204–2211, https://doi.org/10.1093/emboj/19.10.2204.

[42] R.A. Berg, D.J. Prockop, The thermal transition of a non-hydroxylated form of collagen, Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen, Biochemical and Biophysical Research Communications. 52 (1973) 115–120, https://doi.org/10.1016/0006-291X(73)90961-3.

[43] S. Jimenez, M. Harsch, J. Rosenbloom, Hydroxyproline stabilizes the triple helix of chick tendon collagen, Biochem. Biophys. Res. Commun. 52 (1973) 106–114, https://doi.org/10.1016/0006-291X(73)90960-1.

[44] L.I. Fessler, J.H. Fessler, Protein assembly of procollagen and effects of hydroxylation, J Biol Chem. 249 (1974) 7637–7646.

[45] H.P. Bächinger, J. Engel, P. Bruckner, R. Timpl, The Role of Cis-Trans Isomerization of Peptide Bonds in the Coil ⇄ Triple Helix Conversion of Collagen, Eur. J. Biochem. 90 (1978) 605–613, https://doi.org/10.1111/j.1432-1033.1978.tb12641.x.

[46] W.T. Astbury, F.O. Bell, Molecular Structure of the Collagen Fibres, Nature 145 (1940) 421–422, https://doi.org/10.1038/145421a0.

[47] D.A. Torchia, Evidence for cis peptide bonds in copolypeptides of glycine and proline, Biochemistry 11 (1972) 1462–1468, https://doi.org/10.1021/bi00758a021.

[48] U. Reimer, G. Scherer, M. Drewello, S. Kruber, M. Schutkowski, G. Fischer, Side-chain effects on peptidyl-prolyl cis/trans isomerisation11Edited by A. R. Fersht, Journal of Molecular Biology. 279 (1998) 449–460. 10.1006/jmbi.1998.1770.

[49] E.G. Canty, K.E. Kadler, Procollagen trafficking, processing and fibrillogenesis, J. Cell Sci. 118 (2005) 1341–1353, https://doi.org/10.1242/jcs.01731.

[50] E.G. Canty, Y. Lu, R.S. Meadows, M.K. Shaw, D.F. Holmes, K.E. Kadler, Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon, J Cell Biol. 165 (2004) 553–563, https://doi.org/10.1083/jcb.200312071.

[51] R.L. Trelstad, K. Hayashi, Tendon collagen fibrillogenesis: Intracellular subassemblies and cell surface changes associated with fibril growth, Dev. Biol. 71 (1979) 228–242, https://doi.org/10.1016/0012-1606(79)90166-0.

[52] S.M. Humphries, Y. Lu, E.G. Canty, K.E. Kadler, Active Negative Control of Collagen Fibrillogenesis in Vivo: INTRACELLULAR CLEAVAGE OF THE TYPE I PROCOLLAGEN PROPEPTIDES IN TENDON FIBROBLASTS WITHOUT INTRACELLULAR FIBRILS*, J. Biol. Chem. 283 (2008) 12129–12135, https://doi.org/10.1074/jbc.M708198200.

[53] D.F. Holmes, J.A. Chapman, D.J. Prockop, K.E. Kadler, Growing tips of type I collagen fibrils formed in vitro are near-paraboloidal in shape, implying a reciprocal relationship between accretion and diameter, Proc Natl Acad Sci U S A. 89 (1992) 9855–9859, https://doi.org/10.1073/pnas.89.20.9855.

[54] K.E. Kadler, D.F. Holmes, H. Graham, T. Starborg, Tip-mediated fusion involving unipolar collagen fibrils accounts for rapid fibril elongation, the occurrence of fibrillar branched networks in skin and the paucity of collagen fibril ends in vertebrates, Matrix Biol. 19 (2000) 359–365, https://doi.org/10.1016/s0945-053x(00)00082-2.

[55] E.G. Canty, T. Starborg, Y. Lu, S.M. Humphries, D.F. Holmes, R.S. Meadows, A. Huffman, E.T. O'Toole, K.E. Kadler, Actin filaments are required for fibripositor-mediated collagen fibril alignment in tendon, J Biol Chem. 281 (2006) 38592–38598, https://doi.org/10.1074/jbc.M607581200.

[56] A. Colige, S.-W. Li, A.L. Sieron, B.V. Nusgens, D.J. Prockop, C.M. Lapière, cDNA cloning and expression of bovine procollagen I N-proteinase: A new member of the superfamily of zinc-metalloproteinases with binding sites for cells and other matrix components, Proceedings of the National Academy of Sciences. 94 (1997) 2374–2379. 10.1073/pnas.94.6.2374.

[57] A. Colige, I. Vandenberghe, M. Thiry, C.A. Lambert, J. Van Beeumen, S.-W. Li, D. J. Prockop, C.M. Lapiere, B.V. Nusgens, Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying high homology with ADAMTS-2 and ADAMTS-3, J Biol Chem. 277 (2002) 5756–5766, https://doi.org/10.1074/jbc.M105601200.

[58] R.J. Fernandes, S. Hirohata, J.M. Engle, A. Colige, D.H. Cohn, D.R. Eyre, S. S. Apte, Procollagen II amino propeptide processing by ADAMTS-3, Insights on dermatosparaxis, J Biol Chem. 276 (2001) 31502–31509, https://doi.org/10.1074/jbc.M103466200.

[59] E. Kessler, K. Takahara, L. Biniaminov, M. Brusel, D.S. Greenspan, Bone morphogenetic protein-1: the type I procollagen C-proteinase, Science 271 (1996) 360–362, https://doi.org/10.1126/science.271.5247.360.

[60] S.W. Li, A.L. Sieron, A. Fertala, Y. Hojima, W.V. Arnold, D.J. Prockop, The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenic protein-1, Proc Natl Acad Sci U S A. 93 (1996) 5127–5130, https://doi.org/10.1073/pnas.93.10.5127.

[61] I.C. Scott, I.L. Blitz, W.N. Pappano, Y. Imamura, T.G. Clark, B.M. Steiglitz, C. L. Thomas, S.A. Maas, K. Takahara, K.W. Cho, D.S. Greenspan, Mammalian BMP-1/Tolloid-related metalloproteinases, including novel family member mammalian Tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis, Dev Biol. 213 (1999) 283–300, https://doi.org/10.1006/dbio.1999.9383.

[62] M. Miyahara, K. Hayashi, J. Berger, K. Tanzawa, F.K. Njieha, R.L. Trelstad, D. J. Prockop, Formation of collagen fibrils by enzymic cleavage of precursors of type I collagen in vitro, J Biol Chem. 259 (1984) 9891–9898.

[63] D.J.S. Hulmes, K.E. Kadler, A.P. Mould, Y. Hojima, D.F. Holmes, C. Cummings, J. A. Chapman, D.J. Prockop, Pleomorphism in type I collagen fibrils produced by persistence of the procollagen N-propeptide, J. Mol. Biol. 210 (1989) 337–345, https://doi.org/10.1016/0022-2836(89)90335-5.

[64] M. Miyahara, F.K. Njieha, D.J. Prockop, Formation of collagen fibrils in vitro by cleavage of procollagen with procollagen proteinases, J Biol Chem. 257 (1982) 8442–8448.

[65] E.G. Canty-Laird, Y. Lu, K.E. Kadler, Stepwise proteolytic activation of type I procollagen to collagen within the secretory pathway of tendon fibroblasts in situ, Biochem J. 441 (2012) 707–717, https://doi.org/10.1042/BJ20111379.

[66] A.D. Linstedt, H.P. Hauri, Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa, MBoC. 4 (1993) 679–693, https://doi.org/10.1091/mbc.4.7.679.

[67] N.L. Stevenson, D.J.M. Bergen, Y. Lu, M.E. Prada-Sanchez, K.E. Kadler, C. L. Hammond, D.J. Stephens, Giantin is required for intracellular N-terminal processing of type I procollagen, J. Cell Biol. 220 (2021) e202005166.

[68] J. Bella, Collagen structure: new tricks from a very old dog, Biochem J. 473 (2016) 1001–1025, https://doi.org/10.1042/BJ20151169.

[69] A. Bhattacharjee, M. Bansal, Collagen structure: the Madras triple helix and the current scenario, IUBMB Life 57 (2005) 161–172, https://doi.org/10.1080/15216540500090710.

[70] M.P. Williamson, The structure and function of proline-rich regions in proteins, Biochem J. 297 (Pt 2) (1994) 249–260, https://doi.org/10.1042/bj2970249.

[71] J. Josse, W.F. Harrington, Role of pyrrolidine residues in the structure and stabilization of collagen, J. Mol. Biol. 9 (1964) 269–287, https://doi.org/10.1016/S0022-2836(64)80207-2.

[72] J.A. Hodges, R.T. Raines, Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers, J. Am. Chem. Soc. 125 (2003) 9262–9263, https://doi.org/10.1021/ja035881z.

[73] R.W. Newberry, R.T. Raines, The n→π* Interaction, Acc. Chem. Res. 50 (2017) 1838–1846, https://doi.org/10.1021/acs.accounts.7b00121.

[74] H.K. Ganguly, G. Basu, Conformational landscape of substituted prolines, Biophys Rev. 12 (2020) 25–39, https://doi.org/10.1007/s12551-020-00621-8.

[75] R.K. Chopra, V.S. Ananthanarayanan, Conformational implications of enzymatic proline hydroxylation in collagen, Proc Natl Acad Sci U S A. 79 (1982) 7180–7184.

[76] S. Xu, M. Gu, K. Wu, G. Li, Unraveling the Role of Hydroxyproline in Maintaining the Thermal Stability of the Collagen Triple Helix Structure Using Simulation, J. Phys. Chem. B. 123 (2019) 7754–7763, https://doi.org/10.1021/acs.jpcb.9b05006.

[77] M. Bansal, C. Ramakrishnan, G.N. Ramachandran, Stabilization of the collagen structure by hydroxyproline residues, Proc. Indian Acad. Sci. 82 (1975) 152–164, https://doi.org/10.1007/BF03046724.

[78] N. Dai, F.A. Etzkorn, Cis-trans proline isomerization effects on collagen triple-helix stability are limited, J Am Chem Soc. 131 (2009) 13728–13732, https://doi.org/10.1021/ja904177k.

[79] Y. Taga, K. Tanaka, S. Hattori, K. Mizuno, In-depth correlation analysis demonstrates that 4-hydroxyproline at the Yaa position of Gly-Xaa-Yaa repeats dominantly stabilizes collagen triple helix, Matrix Biology Plus. 10 (2021), 100067, https://doi.org/10.1016/j.mbplus.2021.100067.

[80] T.V. Burjanadze, Hydroxyproline content and location in relation to collagen thermal stability, Biopolymers 18 (1979) 931–938, https://doi.org/10.1002/bip.1979.360180413.

[81] X. Zhang, S. Xu, L. Shen, G. Li, Factors affecting thermal stability of collagen from the aspects of extraction, processing and modification, Journal of Leather Science and Engineering. 2 (2020) 19, https://doi.org/10.1186/s42825-020-00033-0.

[82] Y. Taga, M. Kusubata, K. Ogawa-Goto, S. Hattori, Stable Isotope-Labeled Collagen: A Novel and Versatile Tool for Quantitative Collagen Analyses Using Mass Spectrometry, J. Proteome Res. 13 (2014) 3671–3678, https://doi.org/10.1021/pr500213a.

[83] T.V. Burjanadze, Evidence for the role of 4-hydroxyproline localized in the third position of the triplet (Gly-X-Y) in adaptational changes of thermostability of a collagen molecule and collagen fibrils, Biopolymers 21 (1982) 1489–1501, https://doi.org/10.1002/bip.360210803.

[84] C.A. Miles, L. Knott, I.G. Sumner, A.J. Bailey, Differences between the thermal stabilities of the three triple-helical domains of type IX collagen, J Mol Biol. 277 (1998) 135–144, https://doi.org/10.1006/jmbi.1997.1603.

[85] A.J. Hodge, J.A. Petruska, Recent studies with the electron microscope on ordered aggregates of the tropocollagen molecules, Aspects of Protein Structure. (1963) 289–300.

[86] J.P.R.O. Orgel, A. Miller, T.C. Irving, R.F. Fischetti, A.P. Hammersley, T.J. Wess, The In Situ Supermolecular Structure of Type I Collagen, Structure 9 (2001) 1061–1069, https://doi.org/10.1016/S0969-2126(01)00669-4.

[87] J.P.R.O. Orgel, T.C. Irving, A. Miller, T.J. Wess, Microfibrillar structure of type I collagen in situ, Proc Natl Acad Sci U S A. 103 (2006) 9001–9005, https://doi.org/10.1073/pnas.0502718103.

[88] W.J. Tromans, R.W. Horne, G.A. Gresham, A.J. Bailey, Electron microscope studies on the structure of collagen fibrils by negative staining, Z Zellforsch Mikrosk Anat. 58 (1963) 798–802, https://doi.org/10.1007/BF00410661.
[89] B.R. Olsen, Electron microscope studies on collagen. I. Native collagen fibrils, Z Zellforsch Mikrosk Anat. 59 (1963) 184–198, https://doi.org/10.1007/BF00320444.
[90] B.R. Olsen, Electron microscope studies on collagen. II. Mechanism of linear polymerization of tropocollagen molecules, Z Zellforsch Mikrosk Anat. 59 (1963) 199–213, https://doi.org/10.1007/BF00320445.
[91] J. Kastelic, A. Galeski, E. Baer, The Multicomposite Structure of Tendon, Connect. Tissue Res. 6 (1978) 11–23, https://doi.org/10.3109/03008207809152283.
[92] P.P. Provenzano, R. Vanderby, Collagen fibril morphology and organization: Implications for force transmission in ligament and tendon, Matrix Biol. 25 (2006) 71–84, https://doi.org/10.1016/j.matbio.2005.09.005.
[93] R.D. Young, C. Knupp, C. Pinali, K.M.Y. Png, J.R. Ralphs, A.J. Bushby, T. Starborg, K.E. Kadler, A.J. Quantock, Three-dimensional aspects of matrix assembly by cells in the developing cornea, Proc Natl Acad Sci U S A. 111 (2014) 687–692, https://doi.org/10.1073/pnas.1313561110.
[94] I.A. Brown, Scanning electron microscopy of human dermal fibrous tissue, J Anat. 113 (1972) 159–168.
[95] S. Weiner, W. Traub, Bone structure: from angstroms to microns, FASEB J. 6 (1992) 879–885.
[96] R.C. Siegel, J.C. Fu, Collagen cross-linking. Purification and substrate specificity of lysyl oxidase, J. Biol. Chem. 251 (1976) 5779–5785, https://doi.org/10.1016/S0021-9258(17)33123-X.
[97] R.C. Siegel, S.R. Pinnell, G.R. Martin, Cross-linking of collagen and elastin, Properties of lysyl oxidase, Biochemistry. 9 (1970) 4486–4492, https://doi.org/10.1021/bi00825a004.
[98] T.H.C. Brondijk, D. Bihan, R.W. Farndale, E.G. Huizinga, 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 (2012) 5253–5258, https://doi.org/10.1073/pnas.1112388109.
[99] E. Hohenester, T. Sasaki, C. Giudici, R.W. Farndale, H.P. Bächinger, Structural basis of sequence-specific collagen recognition by SPARC, Proceedings of the National Academy of Sciences. 105 (2008) 18273–18277. 10.1073/pnas.0808452105.
[100] K. Poole, K. Khairy, J. Friedrichs, C. Franz, D.A. Cisneros, J. Howard, D. Mueller, Molecular-scale topographic cues induce the orientation and directional movement of fibroblasts on two-dimensional collagen surfaces, J Mol Biol. 349 (2005) 380–386, https://doi.org/10.1016/j.jmb.2005.03.064.
[101] J.M. Wallace, B.G. Orr, J.C. Marini, M.M. Banaszak Holl, Nanoscale Morphology of Type I Collagen is Altered in the Brtl Mouse Model of Osteogenesis Imperfecta, J Struct Biol. 173 (2011) 146–152, https://doi.org/10.1016/j.jsb.2010.08.003.
[102] R. Dalgleish, The human type I collagen mutation database, Nucleic Acids Res. 25 (1997) 181–187, https://doi.org/10.1093/nar/25.1.181.
[103] R. Dalgleish, The Human Collagen Mutation Database 1998, Nucleic Acids Res. 26 (1998) 253–255.
[104] J.C. Marini, A. Forlino, W.A. Cabral, A.M. Barnes, J.D. San Antonio, S. Milgrom, J. C. Hyland, J. Körkkö, D.J. Prockop, A. De Paepe, P. Coucke, S. Symoens, F. H. Glorieux, P.J. Roughley, A.M. Lund, K. Kuurila-Svahn, H. Hartikka, D.H. Cohn, D. Krakow, M. Mottes, U. Schwarze, D. Chen, K. Yang, C. Kuslich, J. Troendle, R. Dalgleish, P.H. Byers, Consortium for Osteogenesis Imperfecta Mutations in the Helical Domain of Type I Collagen: Regions Rich in Lethal Mutations Align With Collagen Binding Sites for Integrins and Proteoglycans, Hum Mutat. 28 (2007) 209–221, https://doi.org/10.1002/humu.20429.
[105] M.G. Pepin, U. Schwarze, K.M. Rice, M. Liu, undefined D. Leistritz, P.H. Byers,, Survival is affected by mutation type and molecular mechanism in vascular Ehlers-Danlos syndrome (EDS type IV), Genet. Med. 16 (2014) 881–888, https://doi.org/10.1038/gim.2014.72.
[106] I.F.A.C. Fokkema, J.T. den Dunnen, P.E.M. Taschner, LOVD: Easy creation of a locus-specific sequence variation database using an “LSDB-in-a-box” approach, Hum. Mutat. 26 (2005) 63–68, https://doi.org/10.1002/humu.20201.
[107] X. Liu, T. Zheng, C. Zhao, Y. Zhang, H. Liu, L. Wang, P. Liu, Genetic mutations and molecular mechanisms of Fuchs endothelial corneal dystrophy, Eye and Vision. 8 (2021) 24, https://doi.org/10.1186/s40662-021-00246-2.
[108] P.M. Rodríguez Cruz, J. Cossins, E. de Paula Estephan, F. Munell, K. Selby, M. Hirano, R. Maroofin, M.Y.V. Mehrjardi, G. Chow, A. Carr, A. Manzur, S. Robb, P. Munot, W. Wei Liu, S. Banka, H. Fraser, C. De Goede, E. Zanoteli, U. Conti Reed, A. Sage, M. Gratacos, A. Macaya, M. Dusl, J. Senderek, A. Töpf, M. Hofer, R. Knight, S. Ramdas, S. Jayawant, H. Lochmüller, J. Palace, D. Beeson, The clinical spectrum of the congenital myasthenic syndrome resulting from COL13A1 mutations, Brain 142 (2019) 1547–1560, https://doi.org/10.1093/brain/awz107.
[109] I. Condrat, Y. He, R. Cosgarea, C. Has, Junctional Epidermolysis Bullosa: Allelic Heterogeneity and Mutation Stratification for Precision Medicine, accessed October 2, 2023, Front. Med. 5 (2019), https://doi.org/10.3389/fmed.2018.00363.
[110] D. Natera-de Benito, J.A. Jurgens, A. Yeung, I.T. Zaharieva, A. Manzur, S. P. DiTroia, S.A. Di Gioia, L. Pais, V. Pini, B.J. Barry, W.-M. Chan, J.E. Elder, J. Christodoulou, E. Hay, E.M. England, P. Munot, D.G. Hunter, L. Feng, D. Ledoux, A. O'Donnell-Luria, R. Phadke, E.C. Engle, A. Sarkozy, F. Muntoni, Recessive variants in COL25A1 gene as novel cause of arthrogryposis multiplex congenita with ocular congenital cranial dysinnervation disorder, Hum Mutat. 43 (2022) 487–498, https://doi.org/10.1002/humu.24333.
[111] J.M.A. Shinwari, A. Khan, S. Awad, Z. Shinwari, A. Alaiya, M. Alanazi, A. Tahir, C. Poizat, N., Al Tassan, Recessive Mutations in COL25A1 Are a Cause of Congenital Cranial Dysinnervation Disorder, Am J Hum Genet. 96 (2015) 147–152, https://doi.org/10.1016/j.ajhg.2014.11.006.
[112] C. Gonzaga-Jauregui, C.N. Gamble, B. Yuan, S. Penney, S. Jhangiani, D.M. Muzny, R.A. Gibbs, J.R. Lupski, J.T. Hecht, Mutations in COL27A1 cause Steel syndrome and suggest a founder mutation effect in the Puerto Rican population, Eur J Hum Genet. 23 (2015) 342–346, https://doi.org/10.1038/ejhg.2014.107.
[113] A. Mohs, M. Popiel, Y. Li, J. Baum, B. Brodsky, Conformational features of a natural break in the type IV collagen Gly-X-Y repeat, J Biol Chem. 281 (2006) 17197–17202, https://doi.org/10.1074/jbc.M601763200.
[114] A.J. Miles, A.P. Skubitz, L.T. Furcht, G.B. Fields, Promotion of cell adhesion by single-stranded and triple-helical peptide models of basement membrane collagen alpha 1(IV)531–543, Evidence for conformationally dependent and conformationally independent type IV collagen cell adhesion sites, J Biol Chem. 269 (1994) 30939–30945.
[115] T. Xu, C.-Z. Zhou, J. Xiao, J. Liu, Unique Conformation in a Natural Interruption Sequence of Type XIX Collagen Revealed by Its High-Resolution Crystal Structure, Biochemistry 57 (2018) 1087–1095, https://doi.org/10.1021/acs.biochem.7b01010.
[116] R.Z. Kramer, J. Bella, P. Mayville, B. Brodsky, H.M. Berman, Sequence dependent conformational variations of collagen triple-helical structure, Nat. Struct. Biol. 6 (1999) 454–457, https://doi.org/10.1038/8259.
[117] Y. Ishikawa, H.P. Bächinger, A molecular ensemble in the rER for procollagen maturation, Biochimica et Biophysica Acta (BBA) - Molecular, Cell Res. (1833 (2013)) 2479–2491, https://doi.org/10.1016/j.bbamcr.2013.04.008.
[118] M.Y. Wong, M.D. Shoulders, Targeting Defective Proteostasis in the Collagenopathies, Curr Opin Chem Biol. 50 (2019) 80–88, https://doi.org/10.1016/j.cbpa.2019.02.021.
[119] Y. Jiang, H. Zhang, J. Wang, Y. Liu, T. Luo, H. Hua, Targeting extracellular matrix stiffness and mechanotransducers to improve cancer therapy, J. Hematol. Oncol. 15 (2022) 34, https://doi.org/10.1186/s13045-022-01252-0.
[120] P.H. Byers, Folding defects in fibrillar collagens, Philos Trans R Soc Lond B Biol Sci. 356 (2001) 151–158, https://doi.org/10.1098/rstb.2000.0760.
[121] A.M. Barnes, A. Ashok, E.N. Makareeva, M. Brusel, W.A. Cabral, M. Weis, C. Moali, E. Bettler, D.R. Eyre, J.P. Cassella, S. Leikin, D.J.S. Hulmes, E. Kessler, J. C. Marini, COL1A1 C-propeptide mutations cause ER mislocalization of procollagen and impair C-terminal procollagen processing, Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. (1865 (2019)) 2210–2223, https://doi.org/10.1016/j.bbadis.2019.04.018.
[122] S. Symoens, D.J.S. Hulmes, J.-M. Bourhis, P.J. Coucke, A. De Paepe, F. Malfait, Type I Procollagen C-Propeptide Defects: Study of Genotype-Phenotype Correlation and Predictive Role of Crystal Structure, Hum. Mutat. 35 (2014) 1330–1341, https://doi.org/10.1002/humu.22677.
[123] J. Engel, D.J. Prockop, The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper, Annu Rev Biophys Biophys Chem. 20 (1991) 137–152, https://doi.org/10.1146/annurev.bb.20.060191.001033.
[124] D.J. Prockop, K.I. Kivirikko, Collagens: molecular biology, diseases, and potentials for therapy, Annu Rev Biochem. 64 (1995) 403–434, https://doi.org/10.1146/annurev.bi.64.070195.002155.
[125] S. Sakakibara, Y. Kishida, K. Okuyama, N. Tanaka, T. Ashida, M. Kakudo, Single crystals of (Pro-Pro-Gly)10, a synthetic polypeptide model of collagen, J. Mol. Biol. 65 (1972) 371–373, https://doi.org/10.1016/0022-2836(72)90288-4.
[126] J. Engel, J. Kurtz, E. Katchalski, A. Berger, Polymers of tripeptides as collagen models: II. Conformational changes of poly(L-prolyl-glycyl-L-prolyl) in solution, J. Mol. Biol. 17 (1966) 255–272, https://doi.org/10.1016/S0022-2836(66)80106-7.
[127] B.B. Doyle, W. Traub, G.P. Lorenzi, E.R. Blout, Conformational investigations on the polypeptide and oligopeptides with the repeating sequence L-alanyl-L-prolylglycine, Biochemistry 10 (1971) 3052–3060, https://doi.org/10.1021/bi00792a011.
[128] F.R. Brown, A. Di Corato, G.P. Lorenzi, E.R. Blout, Synthesis and structural studies of two collagen analogues: poly (L-prolyl-L-seryl-glycyl) and poly (L-prolyl-L-alanyl-glycyl), J Mol Biol. 63 (1972) 85–99, https://doi.org/10.1016/0022-2836(72)90523-2.
[129] B.B. Doyle, E.G. Bendit, E.R. Blout, Infrared spectroscopy of collagen and collagen-like polypeptides, Biopolymers 14 (1975) 937–957, https://doi.org/10.1002/bip.1975.360140505.
[130] G.B. Fields, Synthesis and biological applications of collagen-model triple-helical peptides, Org. Biomol. Chem. 8 (2010) 1237–1258, https://doi.org/10.1039/B920670A.
[131] P.H. von Dreele, N. Lotan, V.S. Ananthanarayanan, R.H. Andreatta, D. Poland, H. A. Scheraga, Helix-Coil Stability Constants for the Naturally Occurring Amino Acids in Water. II. Characterization of the Host Polymers and Application of the Host-Guest Technique to Random Poly(hydroxypropylglutamine-co-hydroxybutylglutamine), Macromolecules 4 (1971) 408–417, https://doi.org/10.1021/ma60022a008.
[132] M. Bhowmick, G.B. Fields, Stabilization of Collagen-Model, Triple-Helical Peptides for In Vitro and In Vivo Applications, Methods Mol Biol. 1081 (2013) 167–194, https://doi.org/10.1007/978-1-62703-652-8_11.
[133] S.A.H. Hulgan, J.D. Hartgerink, Recent Advances in Collagen Mimetic Peptide Structure and Design, Biomacromolecules 23 (2022) 1475–1489, https://doi.org/10.1021/acs.biomac.2c00028.
[134] A.V. Persikov, J.A.M. Ramshaw, A. Kirkpatrick, B. Brodsky, Amino Acid Propensities for the Collagen Triple-Helix, Biochemistry 39 (2000) 14960–14967, https://doi.org/10.1021/bi001560d.

[135] A.V. Persikov, J.A.M. Ramshaw, B. Brodsky, Collagen model peptides: Sequence dependence of triple-helix stability, Pept. Sci. 55 (2000) 436–450, https://doi.org/10.1002/1097-0282(2000)55:6<436::AID-BIP1019>3.0.CO;2-D.
[136] B. Brodsky, Hydrogen bonding in the triple-helix, Proc. Indian Acad. Sci. (Chem. Sci.) 111 (1999) 13–18, https://doi.org/10.1007/BF02869892.
[137] Y. Li, B. Brodsky, J. Baum, NMR Conformational and Dynamic Consequences of a Gly to Ser Substitution in an Osteogenesis Imperfecta Collagen Model Peptide, J Biol Chem. 284 (2009) 20660–20667, https://doi.org/10.1074/jbc.M109.018077.
[138] C.L. Hoop, A.P. Kemraj, B. Wang, S. Gahlawat, M. Godesky, J. Zhu, H.R. Warren, D.A. Case, D.I. Shreiber, J. Baum, Molecular underpinnings of integrin binding to collagen mimetic peptides containing vascular Ehlers-Danlos syndrome-associated substitutions, J. Biol. Chem. (2019) 9685, https://doi.org/10.1074/jbc.RA119.009685.
[139] W. Yang, M.L. Battineni, B. Brodsky, Amino acid sequence environment modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides, Biochemistry 36 (1997) 6930–6935, https://doi.org/10.1021/bi970051h.
[140] A.V. Persikov, J.A.M. Ramshaw, B. Brodsky, Prediction of Collagen Stability from Amino Acid Sequence*, J. Biol. Chem. 280 (2005) 19343–19349, https://doi.org/10.1074/jbc.M501657200.
[141] M. Bhate, X. Wang, J. Baum, B. Brodsky, Folding and conformational consequences of glycine to alanine replacements at different positions in a collagen model peptide, Biochemistry 41 (2002) 6539–6547, https://doi.org/10.1021/bi020070d.
[142] M.A. Bryan, H. Cheng, B. Brodsky, Sequence environment of mutation affects stability and folding in collagen model peptides of osteogenesis imperfecta, Biopolymers 96 (2011) 4–13, https://doi.org/10.1002/bip.21432.
[143] B. Brodsky, A.V. Persikov, Molecular structure of the collagen triple helix, Adv Protein Chem. 70 (2005) 301–339, https://doi.org/10.1016/S0065-3233(05)70009-7.
[144] B. Brodsky, G. Thiagarajan, B. Madhan, K. Kar, Triple-helical peptides: An approach to collagen conformation, stability, and self-association, Biopolymers 89 (2008) 345–353, https://doi.org/10.1002/bip.20958.
[145] K. Okuyama, Revisiting the Molecular Structure of Collagen, Connect. Tissue Res. 49 (2008) 299–310, https://doi.org/10.1080/03008200802325110.
[146] J. Baum, B. Brodsky, Folding of peptide models of collagen and misfolding in disease, Curr Opin Struct Biol. 9 (1999) 122–128, https://doi.org/10.1016/s0959-440x(99)80016-5.
[147] J. Baum, B. Brodsky, Real-time NMR investigations of triple-helix folding and collagen folding diseases, Fold Des. 2 (1997) R53–R60, https://doi.org/10.1016/S1359-0278(97)00028-X.
[148] A.V. Buevich, T. Silva, B. Brodsky, J. Baum, Transformation of the Mechanism of Triple-helix Peptide Folding in the Absence of a C-terminal Nucleation Domain and Its Implications for Mutations in Collagen Disorders*, J. Biol. Chem. 279 (2004) 46890–46895, https://doi.org/10.1074/jbc.M407061200.
[149] X. Liu, D.L. Siegel, P. Fan, B. Brodsky, J. Baum, Direct NMR Measurement of the Folding Kinetics of a Trimeric Peptide, Biochemistry 35 (1996) 4306–4313, https://doi.org/10.1021/bi952270d.
[150] S.H. McLaughlin, N.J. Bulleid, Molecular recognition in procollagen chain assembly, Matrix Biol. 16 (1998) 369–377, https://doi.org/10.1016/s0945-053x(98)90010-5.
[151] N.J. Bulleid, R. Wilson, J.F. Lees, Type-III procollagen assembly in semi-intact cells: chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation, Biochem J. 317 (Pt 1) (1996) 195–202, https://doi.org/10.1042/bj3170195.
[152] T.J. Hyde, M.A. Bryan, B. Brodsky, J. Baum, Sequence dependence of renucleation after a Gly mutation in model collagen peptides, J. Biol. Chem. 281 (2006) 36937–36943, https://doi.org/10.1074/jbc.M605135200.
[153] S. Frank, S. Boudko, K. Mizuno, T. Schulthess, J. Engel, H.P. Bächinger, Collagen Triple Helix Formation Can Be Nucleated at Either End*, J. Biol. Chem. 278 (2003) 7747–7750, https://doi.org/10.1074/jbc.C200698200.
[154] D. Barth, O. Kyrieleis, S. Frank, C. Renner, L. Moroder, The role of cystine knots in collagen folding and stability, part II. Conformational properties of (Pro-Hyp-Gly)n model trimers with N- and C-terminal collagen type III cystine knots, Chemistry 9 (2003) 3703–3714, https://doi.org/10.1002/chem.200304918.
[155] A. McAlinden, T.A. Smith, L.J. Sandell, D. Ficheux, D.A.D. Parry, D.J.S. Hulmes, α-Helical Coiled-coil Oligomerization Domains Are Almost Ubiquitous in the Collagen Superfamily *, J. Biol. Chem. 278 (2003) 42200–42207, https://doi.org/10.1074/jbc.M302429200.
[156] X. Liu, S. Kim, Q.-H. Dai, B. Brodsky, J. Baum, Nuclear Magnetic Resonance Shows Asymmetric Loss of Triple Helix in Peptides Modeling a Collagen Mutation in Brittle Bone Disease, Biochemistry 37 (1998) 15528–15533, https://doi.org/10.1021/bi981147u.
[157] J. Bonadio, P.H. Byers, Subtle structural alterations in the chains of type I procollagen produce osteogenesis imperfecta type II, Nature 316 (1985) 363–366, https://doi.org/10.1038/316363a0.
[158] E.A. Chen, Y.-S. Lin, Using synthetic peptides and recombinant collagen to understand DDR–collagen interactions, Biochimica et Biophysica Acta (BBA) - Molecular, Cell Res. (1866 (2019)), 118458, https://doi.org/10.1016/j.bbamcr.2019.03.005.
[159] D.A. Slatter, D.G. Bihan, R.W. Farndale, The effect of purity upon the triple-helical stability of collagenous peptides, Biomaterials 32 (2011) 6621–6632, https://doi.org/10.1016/j.biomaterials.2011.05.025.
[160] D.A. Slatter, D.G. Bihan, G.E. Jarvis, R. Stone, N. Pugh, S. Giddu, R.W. Farndale, The properties conferred upon triple-helical collagen-mimetic peptides by the presence of cysteine residues, Peptides 36 (2012) 86–93, https://doi.org/10.1016/j.peptides.2012.04.013.
[161] L. He, P. Theato, Collagen and collagen mimetic peptide conjugates in polymer science, Eur. Polym. J. 49 (2013) 2986–2997, https://doi.org/10.1016/j.eurpolymj.2013.05.033.
[162] Z. Yu, B. An, J.A.M. Ramshaw, B. Brodsky, Bacterial collagen-like proteins that form triple-helical structures, J Struct Biol. 186 (2014) 451–461, https://doi.org/10.1016/j.jsb.2014.01.003.
[163] B. An, D.L. Kaplan, B. Brodsky, Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering, Front Chem. 2 (2014), https://doi.org/10.3389/fchem.2014.00040.
[164] M. Rasmussen, M. Jacobsson, L. Björck, Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins *, J. Biol. Chem. 278 (2003) 32313–32316, https://doi.org/10.1074/jbc.M304709200.
[165] J.O. Humtsoe, J.K. Kim, Y. Xu, D.R. Keene, M. Höök, S. Lukomski, K.K. Wary, A Streptococcal Collagen-like Protein Interacts with the α2β1 Integrin and Induces Intracellular Signaling*, J. Biol. Chem. 280 (2005) 13848–13857, https://doi.org/10.1074/jbc.M410605200.
[166] R. Han, A. Zwiefka, C.C. Caswell, Y. Xu, D.R. Keene, E. Lukomska, Z. Zhao, M. Höök, S. Lukomski, Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers, Appl Microbiol Biotechnol. 72 (2006) 109–115, https://doi.org/10.1007/s00253-006-0387-5.
[167] G.L. Rosano, E.A. Ceccarelli, Recombinant protein expression in Escherichia coli: advances and challenges, Front. Microbiol. 5 (2014) 172, https://doi.org/10.3389/fmicb.2014.00172.
[168] A. Yoshizumi, Z. Yu, T. Silva, G. Thiagarajan, J.A.M. Ramshaw, M. Inouye, B. Brodsky, Self-association of streptococcus pyogenes collagen-like constructs into higher order structures, Protein Sci. 18 (2009) 1241–1251, https://doi.org/10.1002/pro.134.
[169] F. Squeglia, B. Bachert, A. De Simone, S. Lukomski, R. Berisio, The Crystal Structure of the Streptococcal Collagen-like Protein 2 Globular Domain from Invasive M3-type Group A Streptococcus Shows Significant Similarity to Immunomodulatory HIV Protein gp41, J Biol Chem. 289 (2014) 5122–5133, https://doi.org/10.1074/jbc.M113.523597.
[170] R. Kananavičiūtė, K. Kvederavičiūtė, D. Dabkevičienė, G. Mackevičius, N. Kuisienė, Collagen-like sequences encoded by extremophilic and extremotolerant bacteria, Genomics 112 (2020) 2271–2281, https://doi.org/10.1016/j.ygeno.2019.12.023.
[171] A. Mohs, T. Silva, T. Yoshida, R. Amin, S. Lukomski, M. Inouye, B. Brodsky, Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline, J Biol Chem. 282 (2007) 29757–29765, https://doi.org/10.1074/jbc.M703991200.
[172] C. Xu, Z. Yu, M. Inouye, B. Brodsky, O. Mirochnitchenko, Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability, Biomacromolecules 11 (2010) 348–356, https://doi.org/10.1021/bm900894b.
[173] E. Cosgriff-Hernandez, M.S. Hahn, B. Russell, T. Wilems, D. Munoz-Pinto, M. B. Browning, J. Rivera, M. Höök, Bioactive hydrogels based on Designer Collagens, Acta Biomater. 6 (2010) 3969–3977, https://doi.org/10.1016/j.actbio.2010.05.002.
[174] N. Seo, B.H. Russell, J.J. Rivera, X. Liang, X. Xu, V. Afshar-Kharghan, M. Höök, An engineered alpha1 integrin-binding collagenous sequence, J Biol Chem. 285 (2010) 31046–31054, https://doi.org/10.1074/jbc.M110.151357.
[175] D.J. Munoz-Pinto, V.R. Guiza-Arguello, S.M. Becerra-Bayona, J. Erndt-Marino, S. Samavedi, S. Malmut, B. Russell, M. Höök, M.S. Hahn, Collagen-mimetic hydrogels promote human endothelial cell adhesion, migration and phenotypic maturation, J Mater Chem B. 3 (2015) 7912–7919, https://doi.org/10.1039/C5TB00990A.
[176] P.A. Parmar, L.W. Chow, J.-P. St-Pierre, C.-M. Horejs, Y.Y. Peng, J. A. Werkmeister, J.A.M. Ramshaw, M.M. Stevens, Collagen-mimetic peptide-modifiable hydrogels for articular cartilage regeneration, Biomaterials 54 (2015) 213–225, https://doi.org/10.1016/j.biomaterials.2015.02.079.
[177] J.F. Diaz Quiroz, P.D. Rodriguez, J.D. Erndt-Marino, V. Guiza, B. Balouch, T. Graf, W.M. Reichert, B. Russell, M. Höök, M.S. Hahn, Collagen-mimetic proteins with tunable integrin binding sites for vascular graft coatings, ACS Biomater Sci Eng. 4 (2018) 2934–2942, https://doi.org/10.1021/acsbiomaterials.8b00070.
[178] Y.Y. Peng, V. Stoichevska, K. Schacht, J.A. Werkmeister, J.A.M. Ramshaw, Engineering multiple biological functional motifs into a blank collagen-like protein template from Streptococcus pyogenes, J Biomed Mater Res A. 102 (2014) 2189–2196, https://doi.org/10.1002/jbm.a.34898.
[179] P.A. Parmar, J.-P. St-Pierre, L.W. Chow, C.D. Spicer, V. Stoichevska, Y.Y. Peng, J. A. Werkmeister, J.A.M. Ramshaw, M.M. Stevens, Enhanced articular cartilage by human mesenchymal stem cells in enzymatically mediated transiently RGDS-functionalized collagen-mimetic hydrogels, Acta Biomater. 51 (2017) 75–88, https://doi.org/10.1016/j.actbio.2017.01.028.
[180] M.B. Browning, D. Dempsey, V. Guiza, S. Becerra, J. Rivera, B. Russell, M. Höök, F. Clubb, M. Miller, T. Fossum, J.F. Dong, A.L. Bergeron, M. Hahn, E. Cosgriff-Hernandez, Multilayer vascular grafts based on collagen-mimetic proteins, Acta Biomater. 8 (2012) 1010–1021, https://doi.org/10.1016/j.actbio.2011.11.015.
[181] H. Cheng, S. Rashid, Z. Yu, A. Yoshizumi, E. Hwang, B. Brodsky, Location of Glycine Mutations within a Bacterial Collagen Protein Affects Degree of

Disruption of Triple-helix Folding and Conformation, J Biol Chem. 286 (2011) 2041–2046, https://doi.org/10.1074/jbc.M110.153965.
[182] S. Yigit, H. Yu, B. An, S. Hamaia, R.W. Farndale, D.L. Kaplan, Y.-S. Lin, B. Brodsky, Mapping the Effect of Gly Mutations in Collagen on α2β1 Integrin Binding, J. Biol. Chem. 291 (2016) 19196–19207, https://doi.org/10.1074/jbc.M116.726182.
[183] F. Carafoli, S.W. Hamaia, D. Bihan, E. Hohenester, R.W. Farndale, An Activating Mutation Reveals a Second Binding Mode of the Integrin α2 I Domain to the GFOGER Motif in Collagens, PLoS One 8 (2013) e69833.
[184] Y. Qiu, A. Mekkat, H. Yu, S. Yigit, S. Hamaia, R.W. Farndale, D.L. Kaplan, Y.-S. Lin, B. Brodsky, Collagen Gly missense mutations: Effect of residue identity on collagen structure and integrin binding, J. Struct. Biol. 203 (2018) 255–262, https://doi.org/10.1016/j.jsb.2018.05.003.
[185] P. Chhum, H. Yu, B. An, B.R. Doyon, Y.-S. Lin, B. Brodsky, Consequences of Glycine Mutations in the Fibronectin-binding Sequence of Collagen, J Biol Chem. 291 (2016) 27073–27086, https://doi.org/10.1074/jbc.M116.753566.
[186] G.A. Hoffmann, M.L. Smith, New Insights into Collagen and Fibronectin Reciprocity during Extracellular Matrix Formation, Chem 5 (2019) 1930–1932, https://doi.org/10.1016/j.chempr.2019.07.012.
[187] K.C. Ingham, S.A. Brew, B.S. Isaacs, Interaction of fibronectin and its gelatin-binding domains with fluorescent-labeled chains of type I collagen, J Biol Chem. 263 (1988) 4624–4628.
[188] E.S. Hwang, B. Brodsky, Folding delay and structural perturbations caused by type IV collagen natural interruptions and nearby Gly missense mutations, J Biol Chem. 287 (2012) 4368–4375, https://doi.org/10.1074/jbc.M111.269084.
[189] T.M. Schmid, T.F. Linsenmayer, Denaturation-renaturation properties of two molecular forms of short-chain cartilage collagen, Biochemistry 23 (1984) 553–558, https://doi.org/10.1021/bi00298a024.
[190] H.P. Bächinger, N.P. Morris, G.P. Lunstrum, D.R. Keene, L.M. Rosenbaum, L. A. Compton, R.E. Burgeson, The relationship of the biophysical and biochemical characteristics of type VII collagen to the function of anchoring fibrils, J Biol Chem. 265 (1990) 10095–10101.
[191] J. Yeo, Y. Qiu, G.S. Jung, Y.-W. Zhang, M.J. Buehler, D.L. Kaplan, Adverse effects of Alport syndrome-related Gly missense mutations on collagen type IV: Insights from molecular simulations and experiments, Biomaterials 240 (2020), 119857, https://doi.org/10.1016/j.biomaterials.2020.119857.
[192] G. Thiagarajan, Y. Li, A. Mohs, C. Strafaci, M. Popiel, J. Baum, B. Brodsky, 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 (2008) 736–748, https://doi.org/10.1016/j.jmb.2007.11.075.
[193] H. Hofmann, T. Voss, K. Kühn, J. Engel, Localization of flexible sites in thread-like molecules from electron micrographs, Comparison of interstitial, basement membrane and intima collagens, J Mol Biol. 172 (1984) 325–343, https://doi.org/10.1016/s0022-2836(84)80029-7.
[194] W.F. Harrington, N.V. Rao, Collagen structure in solution. I. Kinetics of helix regeneration in single-chain gelatins, Biochemistry 9 (1970) 3714–3724, https://doi.org/10.1021/bi00821a010.
[195] W.F. Harrington, G.M. Karr, Collagen structure in solution. II. Analysis of refolding kinetics in terms of nucleation and growth processes, Biochemistry 9 (1970) 3725–3733, https://doi.org/10.1021/bi00821a011.
[196] Y. Feng, G. Melacini, J.P. Taulane, M. Goodman, Acetyl-Terminated and Template-Assembled Collagen-Based Polypeptides Composed of Gly-Pro-Hyp Sequences. 2. Synthesis and Conformational Analysis by Circular Dichroism, Ultraviolet Absorbance, and Optical Rotation, J. Am. Chem. Soc. 118 (1996) 10351–10358, https://doi.org/10.1021/ja961260c.
[197] T. Kishimoto, Y. Morihara, M. Osanai, S. Ogata, M. Kamitakahara, C. Ohtsuki, M. Tanihara, Synthesis of poly(Pro–Hyp–Gly)n by direct polycondensation of (Pro–Hyp–Gly)n, where n = 1, 5, and 10, and stability of the triple-helical structure, Biopolymers 79 (2005) 163–172, https://doi.org/10.1002/bip.20348.
[198] V. Gauba, J.D. Hartgerink, Synthetic Collagen Heterotrimers: Structural Mimics of Wild-Type and Mutant Collagen Type I, J. Am. Chem. Soc. 130 (2008) 7509–7515, https://doi.org/10.1021/ja801670v.
[199] T. Fiala, E.P. Barros, M.-O. Ebert, E. Ruijsenaars, S. Riniker, H. Wennemers, Frame Shifts Affect the Stability of Collagen Triple Helices, J. Am. Chem. Soc. 144 (2022) 18642–18649, https://doi.org/10.1021/jacs.2c08727.
[200] M.S. Ackerman, M. Bhate, N. Shenoy, K. Beck, J.A. Ramshaw, B. Brodsky, Sequence dependence of the folding of collagen-like peptides, Single amino acids affect the rate of triple-helix nucleation, J Biol Chem. 274 (1999) 7668–7673, https://doi.org/10.1074/jbc.274.12.7668.
[201] S.J. Lightfoot, D.F. Holmes, A. Brass, M.E. Grant, P.H. Byers, K.E. Kadler, Type I procollagens containing substitutions of aspartate, arginine, and cysteine for glycine in the pro alpha 1 (I) chain are cleaved slowly by N-proteinase, but only the cysteine substitution introduces a kink in the molecule, J. Biol. Chem. 267 (1992) 25521–25528, https://doi.org/10.1016/S0021-9258(19)74071-X.
[202] B.E. Vogel, R.R. Minor, M. Freund, D.J. Prockop, A point mutation in a type I procollagen gene converts glycine 748 of the alpha 1 chain to cysteine and destabilizes the triple helix in a lethal variant of osteogenesis imperfecta, J. Biol. Chem. 262 (1987) 14737–14744, https://doi.org/10.1016/S0021-9258(18)47857-X.
[203] G.J. Swaminathan, D.E. Holloway, R.A. Colvin, G.K. Campanella, A. C. Papageorgiou, A.D. Luster, K.R. Acharya, Crystal Structures of Oligomeric Forms of the IP-10/CXCL10 Chemokine, Structure 11 (2003) 521–532, https://doi.org/10.1016/S0969-2126(03)00070-4.
[204] I. Fu, D.A. Case, J. Baum, Dynamic water-mediated hydrogen bonding in a collagen model peptide, Biochemistry 54 (2015) 6029–6037, https://doi.org/10.1021/acs.biochem.5b00622.
[205] K.-H. Lee, K. Kuczera, M.M.B. Holl, Effect of osteogenesis imperfecta mutations on free energy of collagen model peptides: a molecular dynamics simulation, Biophys Chem. 156 (2011) 146–152, https://doi.org/10.1016/j.bpc.2011.03.012.
[206] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, K. Schulten, Steered molecular dynamics investigations of protein function, J. Mol. Graph. Model. 19 (2001) 13–25, https://doi.org/10.1016/S1093-3263(00)00133-9.
[207] A. Gautieri, S. Vesentini, F.M. Montevecchi, A. Redaelli, Mechanical properties of physiological and pathological models of collagen peptides investigated via steered molecular dynamics simulations, J. Biomech. 41 (2008) 3073–3077, https://doi.org/10.1016/j.jbiomech.2008.06.028.
[208] K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins, Nature 450 (2007) 964–972, https://doi.org/10.1038/nature06522.
[209] S. Park, K. Schulten, Calculating potentials of mean force from steered molecular dynamics simulations, J Chem Phys. 120 (2004) 5946–5961, https://doi.org/10.1063/1.1651473.
[210] P.-C. Do, E.H. Lee, L. Le, Steered Molecular Dynamics Simulation in Rational Drug Design, J. Chem. Inf. Model. 58 (2018) 1473–1482, https://doi.org/10.1021/acs.jcim.8b00261.
[211] S. Park, F. Khalili-Araghi, E. Tajkhorshid, K. Schulten, Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality, J. Chem. Phys. 119 (2003) 3559–3566, https://doi.org/10.1063/1.1590311.
[212] A.C. Lorenzo, E.R. Caffarena, Elastic properties, Young's modulus determination and structural stability of the tropocollagen molecule: a computational study by steered molecular dynamics, J Biomech. 38 (2005) 1527–1533, https://doi.org/10.1016/j.jbiomech.2004.07.011.
[213] P.J. in't Veld, M.J. Stevens,, Simulation of the Mechanical Strength of a Single Collagen Molecule, Biophys. J . 95 (2008) 33–39, https://doi.org/10.1529/biophysj.107.120659.
[214] A. Gautieri, S. Vesentini, A. Redaelli, M.J. Buehler, Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains, Protein Sci. 18 (2009) 161–168, https://doi.org/10.1002/pro.21.
[215] G. Melacini, A.M. Bonvin, M. Goodman, R. Boelens, R. Kaptein, Hydration dynamics of the collagen triple helix by NMR, J Mol Biol. 300 (2000) 1041–1049, https://doi.org/10.1006/jmbi.2000.3919.
[216] M. Srinivasan, S.G.M. Uzel, A. Gautieri, S. Keten, M.J. Buehler, Alport syndrome mutations in type IV tropocollagen alter molecular structure and nanomechanical properties, J Struct Biol. 168 (2009) 503–510, https://doi.org/10.1016/j.jsb.2009.08.015.
[217] A.E. Marlowe, A. Singh, Y.G. Yingling, The effect of point mutations on structure and mechanical properties of collagen-like fibril: A molecular dynamics study, Mater. Sci. Eng. C 32 (2012) 2583–2588, https://doi.org/10.1016/j.msec.2012.07.044.
[218] J. Xiao, B. Madhan, Y. Li, B. Brodsky, J. Baum, Osteogenesis Imperfecta Model Peptides: Incorporation of Residues Replacing Gly within a Triple Helix Achieved by Renucleation and Local Flexibility, Biophys. J . 101 (2011) 449–458, https://doi.org/10.1016/j.bpj.2011.06.017.
[219] V. Ottani, D. Martini, M. Franchi, A. Ruggeri, M. Raspanti, Hierarchical structures in fibrillar collagens, Micron 33 (2002) 587–596, https://doi.org/10.1016/S0968-4328(02)00033-1.
[220] S.E. Paramonov, V. Gauba, J.D. Hartgerink, Synthesis of Collagen-like Peptide Polymers by Native Chemical Ligation, Macromolecules 38 (2005) 7555–7561, https://doi.org/10.1021/ma0514065.
[221] D.E. Przybyla, J. Chmielewski, Higher-Order Assembly of Collagen Peptides into Nano- and Microscale Materials, Biochemistry 49 (2010) 4411–4419, https://doi.org/10.1021/bi902129p.
[222] T. Koide, D.L. Homma, S. Asada, K. Kitagawa, Self-complementary peptides for the formation of collagen-like triple helical supramolecules, Bioorg. Med. Chem. Lett. 15 (2005) 5230–5233, https://doi.org/10.1016/j.bmcl.2005.08.041.
[223] F.W. Kotch, R.T. Raines, Self-assembly of synthetic collagen triple helices, Proc Natl Acad Sci U S A. 103 (2006) 3028–3033, https://doi.org/10.1073/pnas.0508783103.
[224] O.D. Krishna, K.L. Kiick, Supramolecular Assembly of Electrostatically Stabilized, Hydroxyproline-Lacking Collagen-Mimetic Peptides, Biomacromolecules 10 (2009) 2626–2631, https://doi.org/10.1021/bm900551c.
[225] S. Rele, Y. Song, R.P. Apkarian, Z. Qu, V.P. Conticello, E.L. Chaikof, D-Periodic Collagen-Mimetic Microfibers, J. Am. Chem. Soc. 129 (2007) 14780–14787, https://doi.org/10.1021/ja0758990.
[226] L.E.R. O'Leary, J.A. Fallas, E.L. Bakota, M.K. Kang, J.D. Hartgerink, Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel, Nature Chem. 3 (2011) 821–828, https://doi.org/10.1038/nchem.1123.
[227] D.E. Przybyla, J. Chmielewski, Metal-Triggered Collagen Peptide Disk Formation, J. Am. Chem. Soc. 132 (2010) 7866–7867, https://doi.org/10.1021/ja103148t.
[228] M.M. Pires, D.E. Przybyla, C.M. Rubert Pérez, J. Chmielewski, Metal-Mediated Tandem Coassembly of Collagen Peptides into Banded Microstructures, J. Am. Chem. Soc. 133 (2011) 14469–14471, https://doi.org/10.1021/ja2042645.
[229] M.M. Pires, J. Chmielewski, Self-assembly of Collagen Peptides into Microflorettes via Metal Coordination, J. Am. Chem. Soc. 131 (2009) 2706–2712, https://doi.org/10.1021/ja8088845.
[230] M. He, L. Wang, J. Wu, J. Xiao, Ln3+-Mediated Self-Assembly of a Collagen Peptide into Luminescent Banded Helical Nanoropes, Chemistry – A European Journal. 22 (2016) 1914–1917. 10.1002/chem.201504337.
[231] M.M. Pires, D. Ernenwein, J. Chmielewski, Selective Decoration and Release of His-tagged Proteins from Metal-Assembled Collagen Peptide Microflorettes, Biomacromolecules 12 (2011) 2429–2433, https://doi.org/10.1021/bm2004934.

[232] D.E. Przybyla, J. Chmielewski, Metal-Triggered Radial Self-Assembly of Collagen Peptide Fibers, J. Am. Chem. Soc. 130 (2008) 12610–12611, https://doi.org/10.1021/ja804942w.
[233] V. Hernandez-Gordillo, J. Chmielewski, Mimicking the extracellular matrix with functionalized, metal-assembled collagen peptide scaffolds, Biomaterials 35 (2014) 7363–7373, https://doi.org/10.1016/j.biomaterials.2014.05.019.
[234] M.M. Pires, D.E. Przybyla, J. Chmielewski, A Metal-Collagen Peptide Framework for Three-Dimensional Cell Culture, Angew. Chem. Int. Ed. 48 (2009) 7813–7817, https://doi.org/10.1002/anie.200902375.
[235] M.A. Cejas, W.A. Kinney, C. Chen, G.C. Leo, B.A. Tounge, J.G. Vinter, P.P. Joshi, B.E. Maryanoff, Collagen-Related Peptides: Self-Assembly of Short, Single Strands into a Functional Biomaterial of Micrometer Scale, J. Am. Chem. Soc. 129 (2007) 2202–2203, https://doi.org/10.1021/ja066986f.
[236] K. McGuinness, I.J. Khan, V. Nanda, Morphological Diversity and Polymorphism of Self-Assembling Collagen Peptides Controlled by Length of Hydrophobic Domains, ACS Nano 8 (2014) 12514–12523, https://doi.org/10.1021/nn505369d.
[237] P.J. Kaur, R. Strawn, H. Bai, K. Xu, G. Ordas, H. Matsui, Y. Xu, The Self-assembly of a Mini-fibril with Axial Periodicity from a Designed Collagen-mimetic Triple Helix*, J. Biol. Chem. 290 (2015) 9251–9261, https://doi.org/10.1074/jbc.M113.542241.
[238] J. Hu, J. Li, J. Jiang, L. Wang, J. Roth, K.N. McGuinness, J. Baum, W. Dai, Y. Sun, V. Nanda, F. Xu, Design of synthetic collagens that assemble into supramolecular banded fibers as a functional biomaterial testbed, Nat Commun. 13 (2022) 6761, https://doi.org/10.1038/s41467-022-34127-6.
[239] C. Bonnet, G. Charriere, J. Vaquier, P. Bertin, P. Vergne, R. Treves, Bovine collagen induced systemic symptoms: antibody formation against bovine and human collagen, J Rheumatol. 23 (1996) 545–547.
[240] G. Charriere, M. Bejot, L. Schnitzler, G. Ville, D.J. Hartmann, Reactions to a bovine collagen implant. Clinical and immunologic study in 705 patients, J Am Acad Dermatol. 21 (1989) 1203–1208, https://doi.org/10.1016/s0190-9622(89)70330-3.
[241] R.J. Mullins, C. Richards, T. Walker, Allergic reactions to oral, surgical and topical bovine collagen, Anaphylactic risk for surgeons, Aust N Z J Ophthalmol. 24 (1996) 257–260, https://doi.org/10.1111/j.1442-9071.1996.tb01589.x.
[242] J. Keefe, L. Wauk, S. Chu, F. DeLustro, Clinical use of injectable bovine collagen: A decade of experience, Clin. Mater. 9 (1992) 155–162, https://doi.org/10.1016/0267-6605(92)90095-B.
[243] A. Sorushanova, I. Skoufos, A. Tzora, A.M. Mullen, D.I. Zeugolis, The influence of animal species, gender and tissue on the structural, biophysical, biochemical and biological properties of collagen sponges, J Mater Sci: Mater Med. 32 (2021) 12, https://doi.org/10.1007/s10856-020-06485-4.
[244] Š. Rýglová, M. Braun, T. Suchý, M. Hříbal, M. Žaloudková, L. Vištějnová, The investigation of batch-to-batch variabilities in the composition of isolates from fish and mammalian species using different protocols, Food Res. Int. 169 (2023), 112798, https://doi.org/10.1016/j.foodres.2023.112798.
[245] V. Ferraro, B. Gaillard-Martinie, T. Sayd, C. Chambon, M. Anton, V. Santé-Lhoutellier, Collagen type I from bovine bone, Effect of animal age, bone anatomy and drying methodology on extraction yield, self-assembly, thermal behaviour and electrokinetic potential, Int J Biol Macromol. 97 (2017) 55–66, https://doi.org/10.1016/j.ijbiomac.2016.12.068.
[246] Y.K. Lin, D.C. Liu, Comparison of physical–chemical properties of type I collagen from different species, Food Chem. 99 (2006) 244–251, https://doi.org/10.1016/j.foodchem.2005.06.053.
[247] M.H. Cumming, B. Hall, K. Hofman, Isolation and Characterisation of Major and Minor Collagens from Hyaline Cartilage of Hoki (Macruronus novaezelandiae), Mar. Drugs 17 (2019), https://doi.org/10.3390/md17040223.