

Review

Collagen Mimetic Peptides

Yujia Xu * and Michele Kirchner

Department of Chemistry, Hunter College of the City University of New York, 695 Park Ave., New York, NY 10065, USA; mkirchne@hunter.cuny.edu

* Correspondence: Yujia.xu@hunter.cuny.edu; Tel.: +1-(212)-772-4310

Abstract: Since their first synthesis in the late 1960s, collagen mimetic peptides (CMPs) have been used as a molecular tool to study collagen, and as an approach to develop novel collagen mimetic biomaterials. Collagen, a major extracellular matrix (ECM) protein, plays vital roles in many physiological and pathogenic processes. Applications of CMPs have advanced our understanding of the structure and molecular properties of a collagen triple helix—the building block of collagen—and the interactions of collagen with important molecular ligands. The accumulating knowledge is also paving the way for developing novel CMPs for biomedical applications. Indeed, for the past 50 years, CMP research has been a fast-growing, far-reaching interdisciplinary field. The major development and achievement of CMPs were documented in a few detailed reviews around 2010. Here, we provided a brief overview of what we have learned about CMPs—their potential and their limitations. We focused on more recent developments in producing heterotrimeric CMPs, and CMPs that can form collagen-like higher order molecular assemblies. We also expanded the traditional view of CMPs to include larger designed peptides produced using recombinant systems. Studies using recombinant peptides have provided new insights on collagens and promoted progress in the development of collagen mimetic fibrillar self-assemblies.

Keywords: collagen mimetic peptides; fibril-forming collagen peptide; homotrimer triple helix; heterotrimeric triple helix; recombinant collagen peptides; design of collagen mimetic peptides; collagen receptors; collagen-based biomaterials; extracellular matrix; synthetic collagen

Citation: Xu, Y.; Kirchner, M. Collagen Mimetic Peptides. *Bioengineering* **2021**, *8*, 5. <https://doi.org/10.3390/bioengineering8010005>

Received: 23 November 2020

Accepted: 31 December 2020

Published: 5 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The term collagen *mimetic* often conjures up two different ideas: Those that intend to capture the biological functions of collagen by mimicking the structural hierarchy of collagen building up from the triple helix, and those “inspired” by the properties of collagen and trying to mimic its nano-scale structure and function using non-biological polymers. Examples of the latter include the molecular scaffold made of electrospun polymers with a similar diameter and morphology as collagen fibrils, or nano-scale tubes self-assembled from non-peptide building blocks but decorated with certain amino acid residues on the surface mimicking the functions of collagen [1–5]. Applications of collagen mimetic peptides (CMPs) belong to the former. Peptides are developed to resemble collagens in their amino acid sequence, in their structure, and in their bioactivity. The principle of such an approach falls within the general premise of structural biology that at the foundation of the biological functions of a biomolecule is its molecular structure.

1.1. The Macromolecular Assembly of Collagen

Collagen is a family of extracellular matrix proteins with considerable diversity both in structure and in function. A total of 28 different types of collagen have been identified in this super family, among which the fibrillar collagens are the most abundant and are also the best characterized [6–8]. The major fibrillar collagens include collagen types I, II, and III. Collagen type I is the major collagen in bones, skin, and tendon. Collagen type II

presents primarily in cartilage. Type III collagen often coexists with type I in skin, and in blood vessel walls. Other types of fibrillar collagens are present at a lower amount and are often found coexisting with the three major types.

The structural hierarchy of all collagens starts from the building block: The collagen triple helix [8,9]. A collagen triple helix consists of three polypeptide chains (often referred to as the α chains) coming together in parallel with a precise one residue staggering at the ends [9,10]. The three chains tightly wrap around each other about a common axis in a right-handed helical twist to form a rod-like helical conformation (Figure 1A). The tight packing of the triple helix requires a Gly residue at every third position, giving rise to the characteristic (Gly-X-Y)_n repeating sequence. The obligatory Gly residues are buried at the center of the helix, the side chains of X and Y residues are largely exposed to solvent. The triple helix is often considered a “linear molecule” because of its uniform backbone conformation characterized by an ~0.86 nm helical rise per Gly-X-Y tripeptide [11–14]. The side chains of the X and Y residues can be described as a linear sequential array in an N-to-C directionality spiraling around the surface of the molecule (Figure 1A).

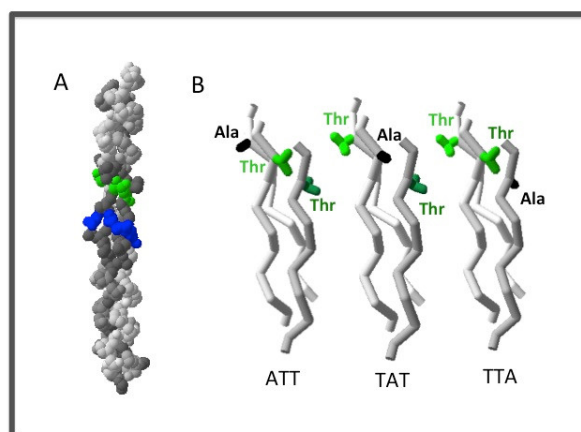


Figure 1. The rod-shaped conformation of the triple helix. **(A)** The structure of the homotrimer triple helix T73–785 [15] was generated using DeepView–Swiss-PdbViewer (PDB: 1bkv). The helix is shown with the N-terminus on top; the Thr residues are shown in green, Arg in blue, hydrophobic residues in dark gray. **(B)** In order to show the asymmetric structure of a heterotrimer associated with different chain registers, a type I collagen like the AAB heterotrimer model was created using DeepView by replacing the Thr of one of the three chains of 1bkv to Ala (side chain shown in black). The amino acid residues included in this heterotrimer model are ITGARGLAG for the two identical strands, and IAGARGLAG for the third, mutated strand. The three structures are shown in the identical view of the backbone, with the N-terminus on top.

The three polypeptide chains of a triple helix can be identical in the form of a homotrimer, or they can be different in amino acid sequences forming a heterotrimeric triple helix [8]. Collagen type II and collagen type III are homotrimers, while collagen type I is a heterotrimer consisting of two identical α_1 chains, and one α_2 chain. There is about 72% sequence similarity between the two α chains in the triple helix domain of type I collagen. Because of the one residue stagger between the adjacent strands in the triple helix, the analogous residues in each strand are unique even in a homotrimer environment (Figure 1A) [10,16–18]. The three strands are usually called leading, middle, and trailing, as viewed from their N-termini. The chain-stagger-related asymmetry in structure is particularly pronounced in a heterotrimer (Figure 1B). Thus, for type I collagen the surface features of the triple helix can be very different depending on which chain is in the leading, middle, or trailing position. There are three possible chain registers for type I collagen: $\alpha_1\alpha_1\alpha_2$, $\alpha_1\alpha_2\alpha_1$, and $\alpha_2\alpha_1\alpha_1$, which are often referred to as α_2 trailing, α_2 middle, and

$\alpha 2$ leading, respectively. Unfortunately, determining the chain register is not at all easy. The correct chain register of type I was accepted to be $\alpha 1\alpha 2\alpha 1$ [19,20]. Emerging data from studies using CMPs, however, are challenging this chain alignment in favor of an $\alpha 1\alpha 1\alpha 2$ register with the $\alpha 2$ chain in the trailing position (details below) [21]. Inside the cell, the C-terminal globular domain, the C-propeptide, was believed to be responsible for both chain selection and chain registration [22]. Structural studies of type I collagen C-propeptide have provided a mechanism for heterotrimerization of the C-propeptide. How the structure of the C-propeptide determines the chain alignment of the triple helix domain, however, remains a mystery [23,24].

Collagens in tissues are higher order, supramolecular assemblies of triple helices. Fibrillar collagens self-associate laterally with a specific 67 nm staggering at the ends to form fibrils (Figure 2A,C) [25–30]. Fibrillar collagens are large molecules consisting of more than 1000 residues per single polypeptide chain in uninterrupted (Gly-X-Y) repeating sequences, forming a long triple helix about 300 nm long and ~1.5 nm in diameter. Each triple helix comprises about 4.4×67 nm in its total length. The staggered arrangement would thus generate long, smooth fibrils with alternating gap-and-overlap regions every 67 nm. This 67-nm structure is termed a *D*-period, which consists of a $0.4D$ overlap zone and a $0.6D$ gap region. The overlap and the gap zones appear as light and dark bands, respectively, when examined using an electron microscope, giving rise to the characteristic striation appearance of collagen fibrils (Figure 2A). In the fibrils, the triple helix further adopts a right-handed super-twist around the microfibrils [31–35]. Because of this super-twist, there is an uneven exposure of different parts of the triple helix on the fibril surface (Figure 2B); as to which specific sections of the triple helix might be exposed on the surface of the fibrils is still under debate [32,36,37].

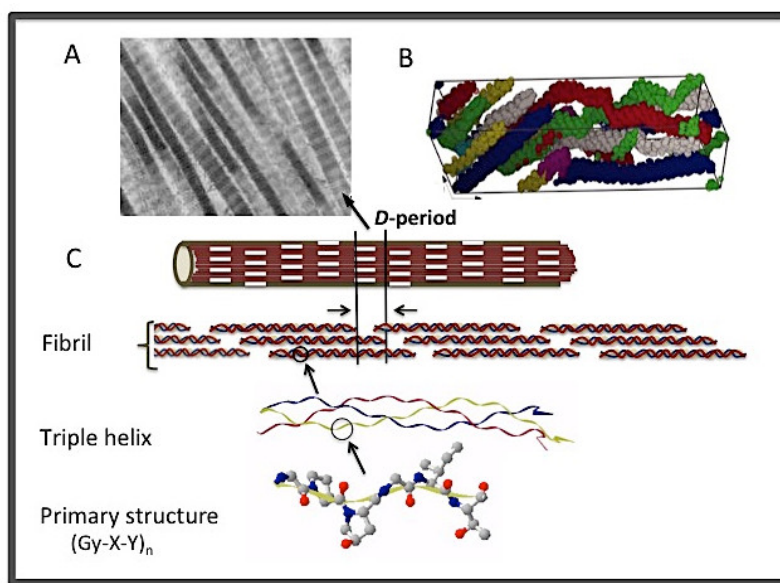


Figure 2. The structural hierarchy of fibrillar collagen. (A) Electron micrograph of collagen fibrils showing the characteristic striation pattern of the *D*-period and the tipped ends. (B) The unit cell of collagen fibril showing the staggered and intertwined arrangement of five triple helices (in different colors) due to the super-twist of the triple helices in fibrils [33]. (C) The different stages of fibrillogenesis from the primary structure to fibrils.

The *D*-period of collagen fibrils is an important feature that has been linked to the tensile strength of bones, the stiffness of the extracellular matrix, and other biomechanical properties of tissues [38–43]. It is the holy grail of all development of collagen mimetic biomaterials to capture the structure and function of this unique, yet ubiquitous molecular

scaffold of all connective tissues. Some of the recent progress in achieving collagen mimetic fibrils is mentioned later in this review. Fibrillogenesis in tissues is a complex process and remains poorly understood. The fibril assembly process per se, however, is a self-assembly process that can be reproduced in vitro from acid-dissolved fibrils [44,45]. It is generally considered that the axial repeating *D*-period of fibril assembly is determined by the molecular interactions contained in each triple helix, although the exact molecular mechanism is not resolved [11–13]. Alternatively, some studies have attributed the deterministic factor of fibrillogenesis to the involvement of the telopeptides: Two short stretches of peptides at the N- and C-termini of the triple helix domain that do not confer to the Gly-X-Y sequence pattern and do not adopt to a triple helix conformation [46]. Later studies have shown that triple helices without the telopeptides can form fibrils, albeit with a slower kinetics [47–49].

Collagen plays much more than a structural role in tissues. It is a dynamic molecular scaffold that supports cell adhesion, cell migration, and cell differentiation [36,50–54]. Cell receptors recognize specific regions of the triple helix. Studies using CMPs to identify these recognition sites are described in Section 2 below, and a comprehensive review of the distribution and organization of these epitopes is given by San Antonio et al. in this special issue. One major family of collagen-binding proteins is integrin. Integrins are the major cell adhesion proteins that bind to the extracellular matrix (ECM) and function as signal transducers of various signaling pathways that can induce global cell responses and affect gene expression [55]. Epithelial and endothelial cells will undergo programmed cell death, or apoptosis, when they lose contact with the ECM. Cells can also control their affinity to collagen through inside-out signaling. The intricate interplay between cells and the ECM is integral to the development of all tissues and organs. Other important collagen receptors include discoidin domain receptor (DDR1 and DDR2), platelet glycoprotein GPIIb/IIIa, immune receptors, the plasma protein von Willebrand factor (vWF), and other macromolecules and proteins [21,36,51,56].

Collagen catabolism is another critical interaction of collagen for maintaining the ECM homeostasis [57]. The hydrolysis of fibrillar collagens is due to the action of matrix metalloproteinases (MMPs) MMP-1, MMP-8, and MMP-13. All three enzymes cut collagens into $\frac{3}{4}$ and $\frac{1}{4}$ length fragments, but with different kinetics to different collagens. MMP-8 preferentially cleaves type I collagen, while MMP-1 has greater catalytic activity on the type III collagen. MMP-13 cleaves type II collagen 5 and 6 times faster than collagens type I and type III, respectively. MMP-2 and MMP-9, which are broadly categorized as gelatinases, also participate in the homeostasis of the ECM by hydrolyzing a partially unfolded, or gelatin form of collagens [58,59]. The well-regulated digestion activity of the MMPs controls the turnover of collagen in normal growth and in tissue remodeling. The unregulated proteolysis by MMP is also the cause of pathological conditions such as arthritis, periodontal diseases, and cancer metastasis [60–62].

Another type of collagen which has often been studied using CMPs is the basement membrane collagen—collagen type IV. Unlike fibrillar collagen, type IV collagen self-assembles into a chicken wire-like molecular network joined at the ends through the N- and C-globular domains. The molecular composition of type IV collagen is also quite complicated [8]. There are six different α chains of type IV collagen that form at least three distinct triple helices with the stoichiometries of $2\alpha_1\alpha_2$, $\alpha_3\alpha_4\alpha_5$, and $2\alpha_5\alpha_6$, respectively. Different isoforms are found in different tissues, and in different developmental stages. While the molecular assembly of the type IV collagen may not have the biomechanical properties of fibrillar collagen, it is the critical molecular scaffold of the basement membrane supporting angiogenesis, and has many cell recognition sites [63]. Molecular interactions with type IV collagens are involved in cancer metastasis; these interactions are potential targets of cancer drugs [64–66].

We are just starting to understand the structural and the molecular details of collagen, and the interactions involving collagen. As is described in more detail, CMPs have been indispensable in these studies. Many reactions take place in an overcrowded ECM

environment where collagen triple helices organize into different molecular scaffolds or networks. How the supramolecular structures of collagen modulate the cell-ECM interactions is yet to be fully elucidated.

1.2. Collagen-Based Biomaterials

Collagen-based biomaterials in the form of sutures and films for wound healing have been used for more than a century [67]. The field has grown tremendously since then to include a broad range of potential medical products based on collagen. The clear advantage of collagen devices is their ability to interact with the host. These products can act as a scaffold for new tissue formation prior to resorption, or be used for soft tissue augmentation. Collagens from animals are often the source for manufacturing these materials. However, the purification process of collagen from animal tissues is a difficult and expensive process. Because of the constant tissue remodeling and modification of collagen, collagen preparations isolated from tissues are often heterogeneous with a high degree of variations in terms of covalent modifications and composition. While collagens per se are considered poor immunogens, the impurities in collagen preparations are known to elicit immunogenic responses. In recent years, there have been increasing concerns of the pathogenicity of collagen devices made from animal collagens or from collagen of human tissues.

Collagen mimetic peptides, thus, represent a desirable alternative for being safer and potentially less expensive. The one other great advantage of mimetic materials, whether produced by chemical synthesis or by genetic engineering, is the ability to modify the amino acid sequences for a better control of the properties and activities of the material. The goal of the collagen mimetic materials should mimic both the physical and biochemical properties of native collagen. Additionally, control of the turnover rate is another important factor for the proper tissue remodeling process, and also for the life span of the material. CMP-based biomaterials bearing epitopes for platelet activation have already demonstrated their potential for application [68,69]. The effectiveness and value of the collagen mimetic biomaterials will largely depend on the ability to modulate and to fine-tune the bioactivities of the material for specific applications.

2. Collagen Mimetic Peptides by Chemical Synthesis

2.1. The Homotrimeric CMPs

Synthetic peptides have been an integral part of protein science. The collagen field is no exception. What is unique to collagen is the requirement to have three peptide chains come together and fold into a triple helix conformation with a specific one-residue stagger at the ends. An early approach was to use chemical cross-links at the C-terminus to bring about the correct association of the three polypeptide chains [70,71]. Later, it was found that peptides with repeating sequences of (Gly-Pro-Hyp)_n with $n > 6$ can self-assemble into a stable triple helix without the need of cross-links, and the thermal stability increases with an increase in the number of tripeptide units [72]. Thus, while the triple helix (Gly-Pro-Hyp)₆ was only marginally stable with a melting temperature (T_m) of 10 °C, that of a (Gly-Pro-Hyp)₁₀ can reach 68 °C. Peptides with (Gly-Pro-Pro)_n repeating sequences also form a triple helix, but have a much lower thermal stability; in comparison with (Gly-Pro-Hyp)₁₀, the thermal stability of (Gly-Pro-Pro)₁₀ is only about 27 °C [73,74]. Since these pioneering studies, it has become quite feasible to synthesize CMP 21–50 residues in sizes with more sequencing variety. The only requirement is to follow the (Gly-Xxx-Yyy)_n sequence pattern with Gly at every third position, although repeating (Gly-Pro-Hyp)_n or (Gly-Pro-Pro)_n, with $n = 2-4$, are frequently included at the N- and/or C-termini for added stability [10,75]. Later studies showed that the multiple (Gly-Pro-Hyp)_n or (Gly-Pro-Pro)_n at the end of the peptide can also function as the nucleation domain to facilitate the asso-

ciation of the three chains in the desired mutual one residue staggering [76–79]. The spontaneous folding of such CMPs inevitably leads to homotrimeric triple helices consisting of three identical polypeptide chains.

2.1.1. The Sequence–Structure Relationship

Studies using the CMP have clearly demonstrated that the rod-shaped triple helix is not uniform in structure or stability [10,80,81]. The three polypeptide chains of a triple helix are connected by a set of hydrogen bonds between the backbone NH of Gly and the backbone CO (X-position) of the neighboring chain [15,82–84]. These main-chain H-bonds present rather uniformly throughout the helix. Additional water-mediated H-bonds can be found between the CO of the Gly with the NH of an X-residue if the X-residue is not a Pro. Crystal structures of CMPs also revealed a sequence-dependent variation in the helical twist. Regions with a high content of Gly-Pro-Hyp or Gly-Pro-Pro form a tighter 7/2 helix (3.5 residues/turn), while the “imino acid-poor” regions adapt to a more relaxed 10/3 symmetry (3.3 residues/turn) [9,10]. This sequence-dependent variation can potentially play a role in the molecular recognition of collagen.

The stability of the CPMs is sensitive to the residues in the X and/or Y positions. Using a set of host–guest peptides, Brodsky and colleagues examined the effects of all 20 amino acid residues plus hydroxyproline in X and/or Y positions(s), and in different combinations [80,81,85–88]. The Gly-Pro-Hyp was the most stable tripeptide; the host peptide (Gly-Pro-Hyp)₃ had a T_m of 47.3 °C in neutral buffer. Replacing the Pro in the *guest* site positioned in the middle of the peptides often resulted in a decrease of T_m ranging from 2–15 °C depending on the identity of the substituted residue, while that caused by replacing Hyp in the Y position ranged from 0 to 21 °C. Having a charged residue like Glu, Asp, Arg, or Lys generally had an unfavorable effect with the exception of Arg in the Y position, which appeared to have a similar stabilizing effect as a Hyp. However, when the charged residues were present in pairs in the sequences of Lys-Gly-Glu (KGE) or Lys-Gly-Asp (KGD), significant stabilizing effects were reported: The T_m of a host–guest peptide with a sequence G_KGE_ or G_KGD_ can increase by 15.4 to 17.5 °C, respectively. This significant stabilizing effect was attributed to a set of inter-chain salt bridges between a pair of oppositely charged residues. In the extended backbone conformation of the triple helix, little interactions could take place between the charged groups in the same chain. Furthermore, the effects of different residues on the overall stability of the triple helix appeared to follow a simple additive rule. A stability calculator was developed based on these studies that could provide reasonably accurate estimations of the thermal stability of CMP 18–50 residues in size, and has been used broadly in the sequence design of homotrimeric CMPs [80].

As the peptide becomes longer, the triple helix is more stable until it reaches a plateau at about 50 residues or so. An empirical curve was used to predict the length dependence of the CMPs [80]. For longer chains, the decrease in entropy of the polypeptide chains in the more constrained folded structure could off-balance the enthalpy contribution in the form of H-bond formation and other interactions. This entropy penalty is more significant for the folding of a longer triple helix. A more quantitative interpretation of the length dependence on the thermal stability would require thermodynamic studies under an equilibrium condition. The thermal unfolding process of CMPs generally does not satisfy this condition [9,81]. How to extrapolate the sequence–stability relationship of CMPs to natural collagen also remains an intriguing question, since natural collagens are not only much longer, but also have much more diverse amino acid sequences. This subject will be revisited in Section 3 on the discussion of recombinant peptides. The findings of the CMPs are generally applicable to larger triple helices, at least on the qualitative level.

2.1.2. The Binding Sites of Collagen Receptors

Defining the binding sites of cell receptors and other macromolecules on collagen is considered one of the crown accomplishments of CMPs [21,50]. Given the complex supra-molecular structure of collagen in the ECM, defining the site where a collagen-binding ligand may interact is difficult. The collagen-binding proteins themselves are also frequently membrane-bound molecular complexes, or in the case of vWF, a complex multi-domain protein. CMPs carrying 6–27 residues modeling a selected section of the $\alpha 1$ chain of collagen type I were used to study the interaction of collagen with integrin [89,90]. The I-domain (or A-domain) of the α -subunit of integrin $\alpha 2\beta 1$ and $\alpha 1\beta 1$ was identified as the domain to interact with both collagen types I and IV in a metal-dependent fashion [91–93]. Peptides containing the sequence GFOGER were first selected as potential binding sites because of their high affinity to the isolated I-domain and ability to support $\alpha 2\beta 1$ -mediated cell adhesion. Subsequent studies using mutations indicated that the recognition of the I-domain is entirely contained in the six-residue sequence, with the residues Arg (R) and Glu (E) being the most critical for the binding [90].

The molecular recognition mechanism of collagen binding by integrin was revealed by the crystal structure of the complex of the I-domain with a 27 mer CMP containing GFOGER at the guest site [94]. The Glu from the middle strand of the triple helix formed a critical interaction with the required divalent ion, and the Arg residue from the middle strand formed a salt bridge with Asp219 of the I-domain to stabilize the complex. The fact that both the Glu and Arg involved in specific interactions with the I-domain all came from the middle strand of the triple helix appeared to offer an explanation as to why the homotrimeric peptide binds the I-domain with similar or even higher affinity as that of heterotrimeric collagen type I. The sequence of the alpha 2 chain of type I collagen in the equivalent location is GPOGES. In the molecular complex, the binding site of the I-domain had close contacts with only one out of the three chains; the interactions with the residues of the other two chains may have facilitated the binding with less specific interactions. The Phe in the middle, and the trailing strands made hydrophobic contacts with the I-domain, while the one in the leading strand was exposed on the surface of the complex. The hydrophobic residues Ala and Leu in the other high-affinity sequence GAOGER and GLOGER, respectively, were expected to provide similar interactions during binding [95]. A specific involvement of Hyp in the six-residue binding site was suspected because of its common presence in the identified high affinity sites. The Hyp of the trailing strand was buried in the interface of the complex, a replacement of residues with larger side chains are likely to create steric clashes and decrease the affinity. No other, more specific interactions involving this Hyp were identified. Later studies found replacing the Hyp with Pro reduced binding affinity but did not abolish the integrin activation [96]. Studies using cross-linked heterotrimeric peptides have indicated the binding affinity of the three different chain registers showed similar affinity to the I-domain. Interestingly, the binding affinity of the heterotrimers was significantly lower than that of the homotrimer.

The peptide approach was later developed into a system known as the ToolKit III [50,95,97]. The ToolKit III consists of 57 peptides with overlapping sequences at the 27-residue guest site that cover the entire sequence of type III collagen. Repeating GPO or GPP sequences flanking the guest site were included to facilitate the triple helix formation. Since sequences of $(GPO)_n$ can interact with the GPVI of platelet, the $(GPP)_n$ sequence is preferred especially for studies with platelets [98]. The ToolKit approach is particularly good when studying the interactions of collagen receptors with collagen type III and collagen type II, due to their homotrimeric nature. Applications using the ToolKit peptides led to the identification of the epitopes of integrin $\alpha 1\beta 1$, $\alpha 2\beta 1$, the vWF on type III collagen, and the DDR2, DDR1, and the immune receptors on type III collagen and type II collagen [97,99–103]. Ideally, these peptides should be covalently crosslinked as well, since the trimerization process of triple helix folding is sensitive to peptide concentration. For this reason, the peptides of the ToolKit III have a Cys at the N- and the C-termini. Given the triple helix conformation, two adjacent Cys residues are often needed in order

to cross-link all three chains in a set of disulfide bonds; a single Cys in a peptide can only cross-link one other neighboring chain and leaves the –SH on the third chain unpaired. Peptides with free –SH groups are often prone to non-specific aggregations.

In tissues, these molecular interactions with collagen receptors take place with collagens in fibril form or in other supramolecular structures. Some epitopes showing a high affinity in a CPM study may not be fully exposed on the surface of the packed fibrils (Figure 1B) [32,36,37,104]. Even if the specific molecular interactions are available, the binding kinetics and binding affinity are likely to be affected by the structural context of the epitopes. Molecular modeling indicates that an I-domain should be able to bind to a triple helix without steric clashes with neighboring helices on the fibrils based on the center-to-center distance of the closely packed triple helices being 15 Å in fibrils [94]. However, the I-domain is also just a part of the integrin complex and does not work in isolation. It remains to be tested in studies of the I-domain and other collagen-binding proteins and/or domains with the triple helix in a higher molecular complex.

2.2. The Heterotrimeric CMPs

In the past decade or so, progress has been made in designing and creating heterotrimeric collagen peptides. Two of the major collagens, collagen type I and collagen type IV, are heterotrimers, and are known to interact extensively with cell receptors. Heterotrimeric peptides are used as model systems to understand these interactions [19,21,96]. Making heterotrimeric triple helices with the correct register faces many challenges; not the least of them is the lack of clear knowledge of the actual chain registration of the native collagens. Because of the one residue staggered arrangement, a triple helix formed from two or three different peptide chains will create a different structure environment for the interaction with cell receptors (Figure 1B).

2.2.1. The Chain Register Affects both the Stability and the Binding Affinity of the Triple Helix

As it was in the case for the homotrimer CMPs, terminal cross-links were first used to circumvent the problems of the chain selection and chain registration. To gain more detailed characterization of the mechanism of proteolysis of collagen type I by the metalloproteinase, a hetero CMP was synthesized to mimic the MMP-1 digestion site between residues 772–783 of the $\alpha 1$ and $\alpha 2$ chains of type I collagen [19,105]. Altogether, 4 Cys residues were included at the C-terminus of the peptide chains which were designed to cross-link the three chains through two disulfide bonds. The chain register was fixed to $\alpha 1\alpha 2\alpha 1$ ($\alpha 2$ as the middle chain). An additional (GPO)₅ at the N-terminus was also included to give the peptide a thermal stability of ~33 °C. The enzyme digest assays performed at room temperature indicated the triple-helical substrates were cleaved with a single cut through the three chains in a manner similar to those observed in studies using collagens isolated from tissues. This similar enzyme activity was also used to argue that the chain register of type I collagen must be $\alpha 1\alpha 2\alpha 1$. However, an NMR study of the peptide showed the C-terminal half of the peptide appeared to be disordered. It is not clear if the disordered conformation was due to the low helix propensity of the sequences or the structural constraint imposed by the C-terminal cross-link. Later studies by crystallography also indicated that the regions surrounding the interchain disulfide bonds are often flexible and the chain register may not be fixed by their use. Additionally, there is a chance the disulfide bond may reshuffle, and lead to unexpected cross-links [106,107].

Another set of similar cross-linked hetero-CMPs were used to model the $\alpha 1\beta 1$ binding sites of collagen type IV. The study revealed the chain register affected both the stability and the folding kinetics of the triple helix [108,109]. The trimer of the $\alpha 1\alpha 2\alpha 1$ register was more stable (T_m of 42 °C) but had lower affinity to integrin, while the trimer with a $\alpha 2\alpha 1\alpha 1$ was less stable (T_m of 30 °C), but showed higher affinity [108,109]. The high sta-

bility was postulated to affect the binding because collagen needs to undergo conformational changes in the integrin adhesion region upon binding, and a less rigid conformation may favor such a conformational adjustment.

2.2.2. The Self-Assembled Heterotrimeric Triple Helix

A heterotrimeric triple helix formed by self-association of three polypeptide chains without cross-linking is not just an interesting, fundamental problem to solve for protein design, but one that would offer more flexibility in chemical synthesis and more versatility for applications. The challenge is the control of both the chain composition and the chain register. The control of composition is a tractable problem if different compositions lead to different stability. By simply mixing two peptides A and B, there will be 8 possible combinations of trimers: AAA, BBB, ABB, BAB, BBA, AAB, ABA, and BAA; where the AAB and ABA have the same chain composition but represent two different triple helical structures due to the different chain register. The population of each of the 8 possible configurations is partitioned into a Boltzmann distribution based on their stability; the desired configuration can be the dominant species if its stability is significantly higher than any of the other competing species. The control of the chain register turns out to be a closely related problem (see below), since the chain registers also affect the stability of the triple helix.

One stabilizing factor that can be exploited for the design of a stable heterotrimeric triple helix is the inter-chain salt bridges [80,86,88]. By strategically placing a pair of oppositely charged residues on neighboring chains, the interactions between the charged pair lead to the formation of a salt bridge (an ionic H-bond) that can significantly stabilize the triple helix. In contrast, unpaired charged residues in the X or Y position are known to cause destabilizing effects on the triple helix due to charge repulsion, and can be used to discourage unwanted conformations. Nanda and coworkers implemented this idea and computationally generated three peptide sequences A, B and C, which by mixing in a 1:1:1 ratio formed an ABC heterotrimer (here ABC stands for chain composition not chain registration) [110,111]. By several rounds of optimization to increase the stability of the ABC trimer while destabilizing other competing configurations, the ABC trimer emerged as the only triple helix in the solution. However, the ABC heterotrimer could have 6 possible chain alignments. The most likely alignment was the one designated as *abc* with A, B, C chains in the leading, middle, and trailing positions, respectively. Based on the computer simulation, *abc* had the most favorable charge-paired interactions and fewest charge repulsions and thus, should have been the one that had the highest stability. Later, a crystal structure study of the designed ABC heterotrimer confirmed the *abc* chain alignment [112]. The amino acid sequences of the A, B, and C peptides, however, were nothing close to that of natural collagen. It will be interesting to find out if some of the sequence features can be utilized to generate heterotrimers with more sequence diversity.

Using an intuitive experimental approach, Hartgerink and coworkers exploited the geometric and sequence specificity of electrostatic interactions of charged residues nearly exhaustively in a series of 30-residue peptides [113–121]. Among the major findings are that peptides carrying 5 or more like charges such as (DOG)₁₀ or (PKG)₁₀ will not form homotrimers because of a strong charge repulsion. However, when such decapositive and decanegative peptides are mixed with (POG)₁₀ in a 1:1:1 molar ratio, they form stable ABC heterotrimers. Most of all, the NMR study shows that such heterotrimers are stabilized by a set of axially oriented charge-pair interactions between a Lys residue in the leading and middle chains with a Glu or Asp residue in the middle and trailing chains, respectively; with the interactions between the Lys-Asp being particularly strong. Furthermore, the orientation bias of the Lys-Asp interaction means only in the alignment of (PKG)₁₀(EOG)₁₀(POG)₁₀ (i.e., with (PKG)₁₀, (EOG)₁₀, and (POG)₁₀ in the leading, middle, and trailing positions, respectively) can all 10 sets of salt bridges be satisfied, and this maximized stability predetermined that only heterotrimer in this chain register will dominate in the solution. The same type of the axial Lys-Asp salt bridge was also found in the

crystal structure of the heterotrimer mentioned above by Zheng et al. [112]. This directional interchain charge-pair interaction has since been used to develop self-assembled heterotrimers or a long super-helix [21,122,123].

Another equally important consideration in designing a stable heterotrimer is to destabilize the competing compositions. While the original ABC heterotrimer based on the inter-chain charge-charge interaction was remarkably strong, with a T_m above 56 °C, they cannot outcompete the homotrimer (POG)₁₀, which has a T_m of 58 °C [115]. Thus, mixing of the three peptides generated a mixture of heterotrimer and the (POG)₁₀ homotrimer. In another effort, Hartgerink and coworkers developed a stable ABC heterotrimer using a decapositive (PKG)₁₀ and a decanegative (EPG)₁₀ mixed with a zwitterionic peptide (DKG)₁₀. Since the zwitterionic peptide does not form a stable homotrimer, the ternary mixture produced a single component ABC heterotrimer with a single chain registration (117). The peptide has a T_m of 38 °C which is quite remarkable considering this trimer has no Hyp.

By maximizing the axial Lys-Asp interactions while reducing the possibility of like-charge repulsions at the same time, a stable heterotrimer in the AAB arrangement was achieved in the composition of 2(GPKGEO)₅(POGDOG)₅ [119,120]. The heterotrimer has a unique AAB chain register because only in this chain alignment are the interactions of the directional, axial Lys-Asp salt bridge maximized, while unfavorable, unpaired charge residues are minimized. This study is a clear demonstration that, by strategically placing the charged residues, both the chain composition and the chain register can be controlled. At the same time, however, the need to have several of the charged residues in fixed locations can restrict the sequence diversity of the peptides and limit their applications.

In a new development, it was found that short stretches of peptides of five (Gly-X-Y) tripeptide units consisting of the charge-pairing residues in the optimal positions can be used as a hetero-nucleation site to develop peptides carrying sequences from natural collagens [21]. A set of heterotrimeric peptides in the AAB composition modeling the cell adhesion epitopes or the vWF binding site of type I collagen in all three possible chain alignments were developed. These heterotrimeric peptides have the hetero-nucleation sites at both the N- and the C-termini, flanking 12 residue-specific CBP (collagen-binding protein)-binding epitopes in the center (Figure 3). The specific amino acid sequences of the nucleation sites were computationally generated using a genetic algorithm; a group of Lys and Asp residues were strategically positioned to form a unique set of optimally oriented Lys-Asp salt bridges which stabilize the triple helix having the desired chain alignment. The folded structure was further “covalently captured” by converting the salt bridges into an amide bond between the ammonium group of the Lys side chain and the carboxylate group of the Asp. Based on the structural, stability, and binding studies, the triple helix in the $\alpha 1\alpha 1\alpha 2$ alignment was found to be less thermally stable than those in the other two possible chain registers, but had the highest binding affinity to DDR1 and vWF. The $\alpha 1\alpha 1\alpha 2$ heterotrimer was also found to induce higher levels of cellular DDR1 and DDR2 kinase activation. These findings appeared to indicate that the $\alpha 1\alpha 1\alpha 2$, rather than the $\alpha 1\alpha 2\alpha 1$ as proposed by previous studies, was most likely to be the correct chain alignment of type I collagen. In this conclusion, it is implied that the correct chain register of collagen was not necessarily the one with the highest stability, but the one with highest bioactivity.

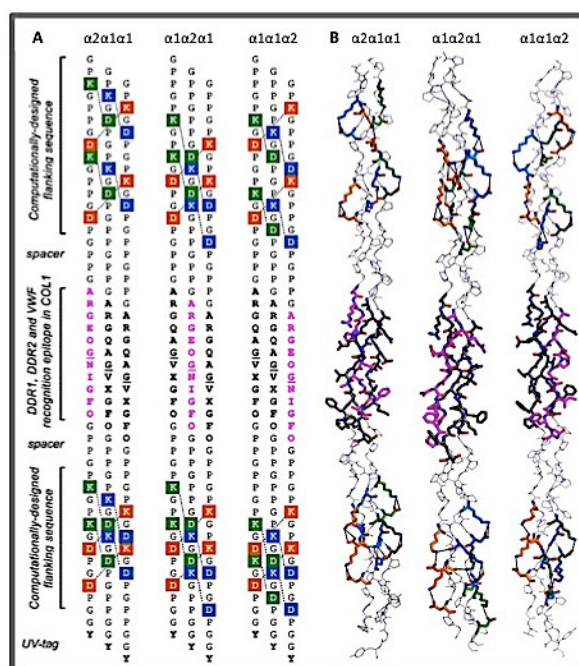


Figure 3. The heterotrimeric peptides modeling the collagen-binding epitopes on type I collagen (21). **(A)** The design and the amino acid sequences of the heterotrimers mimicking type I collagen in three different registers: $\alpha 2\alpha 1\alpha 1$, $\alpha 1\alpha 2\alpha 1$, and $\alpha 1\alpha 1\alpha 2$; here X is norleucine, a methionine bioisostere. The N- and C-termini flanking sequences are generated by a computer program utilizing genetic algorithm to optimize the location of the Lys-Asp charge-pairings in the helix. **(B)** The corresponding crystal structure of the heterotrimers in three registers.

The covalently captured heterotrimeric triple helix can extend the ToolKit approach to study the binding and MMP specificity of heterotrimeric collagens. The covalent-capture can potentially be a more effective alternative than Cys-based cross-linking strategies to generate cross-linked heterotrimeric peptides that are more homogeneous, and more stable. Such systems can find a wide range of applications for collagen research.

2.3. Applications of CMPs

Collagen rarely functions as an individual triple helix in tissues. To truly “mimic” collagen, CMPs need to further associate into fibril-like supramolecular structures. Developing higher order molecular assemblies through the self-association of CMPs, however, has proven to be challenging: Small sizes and limited sequence diversity are two likely limiting factors.

2.3.1. Self-Assembled Fibrillar Structures.

A common approach to form a long triple helix through self-assembly is to use the so called “sticky-end” strategy. In the approach by the Raines’ group, the sticky-end CMPs were created using a cross-linked “core” that brought together polypeptide chains with different lengths extending from the C- and N-termini of the core as branches (Figure 4A,B) [124]. The branches were peptides with $(POG)_n$ or $(PPG)_n$ sequences that had a high propensity to trimerize to form a triple helix. The three strands at the N-terminal end of the core formed an intramolecular triple helix but with an overhang that was complementary to the segment sticking out from the C-terminal end of the core, thus the “sticky ends”. The intermolecular assembly between the complementary sticky ends then produced helices that could reach a length of 200 nm or longer. The functionality of these long triple helices has yet to be fully evaluated. In another approach using disulfide bond cross-

are assembled through lateral, staggered association of triple helices. While these self-assembled triple helices are wonderfully long, they lack certain structural elements of the native collagen fibrils, chief among them the axial structure of the *D*-period. Similarly, the mechanical support provided by the CMP hydrogels are likely to be different from the support of the collagen molecular scaffold in the ECM, despite similar morphology as shown by TEM images. None of the supramolecular CMPs have the *D*-period-like structure seen in collagen fibrils. In one report, it was proposed that a *D*-periodic microfibril was formed from blunt-end self-assembly of a 36-residue triple helix [128]. However, the proposed 17.9 nm *D*-period of the microfibrils is nearly twice the length of the constituent triple helix. The microfibrils must have formed through a very different molecular arrangement than the *D*-periodic collagen fibrils.

2.3.2. Interaction of CMPs with Damaged Collagens

There have been many interesting studies using CMPs to produce nano-structures in different shapes, sizes, and compositions. Some of these molecular assemblies have shown promise for biomedical and other applications [129–139]. There is a good account of some of these works in a recent review [56]. We would, however, like to end this section on synthetic CMPs with the application of CMPs that covers the other end of the spectrum—by going small. This application uses CMPs in an unfolded, single chain conformation to study damaged collagen in tissues.

As indicated in some of the studies above, multiple (Gly-Pro-Hyp) peptide repeats have a high tendency to trimerize with two other peptides to form a triple helix. This tendency makes it a high affinity ligand which can bind to unfolded or partially unfolded collagens. Based on this property, Yu and colleagues have developed several effective peptide probes to detect damaged collagen using a (POG)₉ peptide with a fluorophore conjugated to the N-terminus [140,141]. These collagen hybridizing peptides (CHPs) were found to have a high tendency to form a hybridized triple helix with unfolded chains in damaged collagen, both under in vitro and in vivo conditions. The CHPs are also remarkably stable and resistant to proteolysis in serum. However, the high propensity of (GPO)₉ for triple helix formation has its downside in this application, since the affinity to damaged collagen diminishes once the CHPs trimerize themselves to form a triple helix. To prevent the self-trimerization, the peptides had to be heated above 80 °C before injection for in vivo applications; the *T_m* of a (GPO)₉-based triple helix can reach above 50 °C. To overcome this problem, new features were included in the CHPs. In one clever approach, a nitrobenzyl (NB) group was attached to the backbone nitrogen of a Gly residue located at the center of the peptides [142]. This “NB-caged” CHP cannot trimerize to form a triple helix due to steric clashes of the bulky NB group. Once delivered to the site of detection, the NB group can be removed by a brief radiation of UV light and free the CHPs so they can hybridize with damaged collagen. In a more recent approach, 2*S*,4*S*-fluoroproline (flp, or f) was used to replace Pro [143,144]. The homotrimer (GfO)₉ has a low stability at room temperature, but its strands can effectively bind damaged collagen [145]. The range of applications of the CHP probe is remarkable. It has been used to visualize matrix turnover caused by proteolytic migration of cancer cells in a 3D collagen gel, and to detect the ECM changes associated with mechanical stress, different types of cancers, and connective tissue diseases in mouse models [143,146,147].

Works in the Raines' lab utilized a similar concept to develop CHP probes that can hybridize to the cell surface collagenous protein of *Streptococcus pyogenes*, which is a bacterium responsible for serious infections of wounds and connective tissues [148]. These CHPs utilized (GPP)₇ peptides, which have a low tendency for trimerization (thermal stability of (GPP)₇ triple helix < 27 °C), but maintain the ability to hybridize with damaged collagen at room temperature. Such CHPs can potentially be used to detect bacterial infection at the wound bed. Another modified CHP with the sequence Cy5-G(SG)₂-(fOG)₇ was used to assess the wound healing and tissue remodeling process in injured human

skin, the burned damages in tissues, and the abnormal ECM in bone associated with developmental defects [149–152].

3. The Recombinant Collagen Peptides

There has been an increase in the use of recombinant peptides produced by a recombinant system using designed genes for collagen research. The CMPs produced by chemical synthesis are often limited to <50 residues; the need to include multiple GPO or GPP sequences for stability and folding further limited their ability to model natural collagens, which frequently have more than 1000 amino acid residues per a single polypeptide chain. Because of the feasibility of including large stretches of amino acid sequences to model more extended ranges of collagen, researchers have turned to the recombinant peptide to gain, almost literally, a broader view of collagen and of the mechanisms of its functions. The expression system of *Escherichia coli* (*E. coli*) is often the system of choice. The simpler prokaryotic genome is easier to manipulate and to achieve a high yield [153]. The obvious limitation is the lack of post-translational modifications, and for collagen, a lack of Hyp in particular. However, emerging studies have showed that collagen mimetic peptides without Hyp can still interact with integrin, support cell adhesion, be recognized by MMP and be cleaved on the same site, and self-assemble into fibrils [96,154–157]. Another limitation is the ability to generate heterotrimers, although exciting new progress has been made in this account (details are given in the next section). In general, the bacterial expression system of collagen peptides is considered a valuable tool for collagen research, albeit not perfect.

We would like to clarify that the recombinant peptides are not recombinant collagens. There have been many works in the recent decades devoted to producing full-chain human collagen molecules using an expression system. In such cases, an entire gene(s) of a human collagen (or of another organism) was cloned into a host expression system. The collagen molecules so produced were meant to be an exact replicate of human collagen. Such recombinant collagens have great potential for many applications; a full chapter in this special issue is devoted to such recombinant collagens. Here, we focus on collagen-like peptides with designed amino acid sequences ranging from ~30–300 amino acid residues in size, which can be produced either by a eukaryotic expression system or, more frequently, by a prokaryotic expression system.

3.1. The Sequence–Stability Relationship Revisited

Further expanding their work on the triple helix propensity of amino acid sequences, Brodsky and coworkers analyzed the different factors in the overall thermal stability of a series of recombinant peptides derived from bacterial collagen Sc12 [155,158–160]. The size of the triple helix domain in the peptides ranged from 75 to 237 residues. Despite lacking any hydroxyproline, these recombinant collagens were stable, with a T_m between 23.5 and 35.6 °C depending on the specific sequence and the length of the peptide. The unexpected stability was attributed to the high content of Pro in X positions, and the high level of charge-based interactions. As a result, the T_m was found to be sensitive to pH and the ionic strength of the buffer; a 3–14 °C decrease was reported for some peptides as the pH was reduced from 7 to 2.8 [158]. The stability was found to increase with chain length and reach a plateau when the size reached about 150 residues. For peptides longer than 150 residues, the value of the T_m was more or less stabilized at 36–39 °C with less dependence on the amino acid sequence. Replacing the stabilizing Pro residues with residues with a lower propensity reduced the thermal stability of the triple helix, although the exact extent of change in the T_m did not quantitatively follow the simple additive rules of short CMPs [155].

3.1.1. Defining the Sequence Requirements of Fibronectin Binding, and of the Proteolysis of MMPs

The ability to include longer stretches of a native sequence of collagen makes it possible to study interactions of collagen with larger molecular complexes using the recombinant peptides. Peptides derived from Scl2 are particularly good as a model for such studies. Because of its prokaryotic origin, Scl2 has a low affinity to the human collagen receptors and low susceptibility to MMPs. For this reason, Scl2 was often considered a collagen “blank slate” [160].

By including an amino acid sequence of 3 to 8 Gly-X-Y triplets (9–24 residues) taken from type II collagen in the center *guest* site of Scl2, an 18-residue amino acid sequence was identified as the binding site of fibronectin on type II collagen [161]. Fibronectin is a large, dimeric glycoprotein that interacts with collagen to maintain the integrity of the ECM. The 18-residue binding site of fibronectin is nearly three times the size of the footprints of the collagen-binding domain of integrin or of vWF on collagen, and would be very difficult to characterize using short CMPs. Similar studies of the recombinant peptides with other collagen receptors further expanded our understanding of the molecular recognition process of collagen [154,162].

The Scl2-based peptides were also used to define the sequence selectivity of MMPs [163]. Peptides with an insertion of 12–18 amino acid residues from type III collagen at the *guest* site of Scl2 were produced as the substrate of MMP-1 and MMP-13. A 5-tripeptide sequence was identified as the minimum requirement for MMP digestion, including 4 residues before and 11 residues after the cleavage site. The ratio of k_{cat}/K_m of the reaction was close to that observed using human collagen as a substrate, but the values of both K_m and k_{cat} were about 10-fold higher. This discrepancy was partially attributed to the lack of Hyp in the recombinant peptides, which can affect the affinity of the peptide as a substrate. The same cleavage site and the high k_{cat} indicate the mechanism of the enzymatic reaction is similar for the two substrates.

3.1.2. The Impact of Gly Substitution Mutations

CMP 30–45 residues in size have been used to elucidate the structural and stability effects on collagen by mutations linked with brittle bone disease (osteogenesis imperfecta or OI). The majority of OI mutations are missense mutations that lead to the replacement of the obligatory Gly to a different amino acid. One unique feature of the OI mutations is that the same type of Gly substitution often results in phenotypes of the disease with very different severity depending on the location of the Gly. Because of their limited sizes, studies using CMPs could not fully resolve why the location of a mutation has such a profound impact. In one study, a recombinant peptide containing the 63-residue Hyp-free region of the $\alpha 1$ chain of type I collagen was produced using a bacterial expression system [164]. Several Gly substitution mutations in this region were linked to OI with very different phenotypes. For the same Gly to Ser substitution, the OI was mild when it takes place at Gly⁹⁰¹, but just 12 residues away at Gly⁹¹³ it causes a severe type of OI characterized by prenatal death. In this 90-residue model peptide F877, the authors were able to demonstrate that the structural impact of the Gly replacing mutation is modulated by the local stability and sequence context of the mutation site. Located next to a relatively unstable region consisting of no imino acids, the conformational alteration related to the Ser substitution at Gly⁹¹³ triggered a large scale unfolding of the triple helix, while the effects at Gly⁹⁰¹ were better confined to the close vicinity of the mutation site. A region of more than 20 residues in size was found to be unfolded because of the substitution of Gly⁹¹³. Conformational change of this scale is difficult to study using short CMPs. In a study using the Scl2 consisting of about 300 residues, Brodsky and coworkers were also able to demonstrate that the relative location of the Gly substitution site to the folding nucleation domain had a different overall impact on the structure and stability of the triple helix [165,166].

3.2. The Heterotrimeric Recombinant Peptides

The recombinant peptides offer a different strategy to make heterotrimers by utilizing a heterotrimeric nucleation site appended to the triple helical domain. The nucleation site functions as the C-propeptide of type I or type IV collagen to bring together three different polypeptide chains. Recent studies of the non-collagenous domains of collagen type IX indicate that the NC2 domain of collagen type IX can function as a nucleation site of heterotrimers [167]. Collagen type IX is a heterotrimer consisting of three different chains, and is a member of the fibril-associated collagens with interrupted triple helices (FACIT). The isolated NC2 domain itself consists of 3 different polypeptide chains, each about 38 residues long, and forms a stable α -helix coiled coil. It was further demonstrated that the NC2 domain can be used to direct the folding of a hetero-trimeric triple helix containing the vWF binding site of type I collagen [106,168]. For each peptide chain of the NC2 domain (IX α 1, IX α 2, and IX α 3), two fusion proteins were expressed by connecting to either the peptide mimicking the α 1 chain of type I collagen (I α 1) or the one mimicking the α 2 chain (I α 2). For instance, by mixing I α 1-IX α 1, I α 1-IX α 2, and I α 2-IX α 3 in a 1:1:1 ratio, a molecular chimera forms with the triple helix domain with a specific register referred to as 112. It should be noted that, since the chain register of the NC2 domain is not known, the register 112 is not the same as α 1 α 1 α 2, which generally denotes the chain alignment with the two α 1 chains in the leading and middle positions, and the α 2 in the trailing. However, by simply mixing the group of 6 different fusion proteins, the vWF domain in all three different chain registers can be produced. After stability and binding studies, the heterotrimer with the 112 arrangement showed the highest thermostability and the highest binding affinity. It was, therefore, concluded that by connecting the peptide I α 1 to the IX α 1 and IX α 2 chains, and peptide I α 2 to the IX α 3 chain, the NC2 domain will lead to the formation of a heterotrimer in the same chain register as that of native collagen type I [167]. This approach to heterotrimers is still in its early stages; further studies are needed to fully establish its effectiveness and applicability.

3.3. The Collagen Mimetic Fibrils

The longer triple helices with more diverse amino acid side chains on its surface also facilitated the lateral self-association of the triple helix. Using bacterial collagen, Brodsky and coworkers showed that peptides with a 79 GXY tripeptide can form fibrous bundles at neutral pH (Figure 5) [159]. When the size of the peptide was doubled to include two identical triple helix units, in a construct of CL-CL where CL = 79 tripeptides, the peptide forms a more discrete structure characterized by an in register, end-on-end self-assembly (Figure 5). The length of the end-on-end assembly is ~140 nm, which is in a good agreement with that of a CL-CL molecule; the diameter of the assemblies is 4–5 nm which is about 4 \times that of a single triple helix. It is not clear if the triple helices are in a parallel arrangement as in fibrillar collagen. This work demonstrated that triple helices in the range of ~100 residues in size can self-associate to form stable fibril structures even in the absence of Hyp. Previous work seemed to indicate that interactions involving the hydroxyl of Hyp are necessary for the stabilization of the fibril assembly [131].

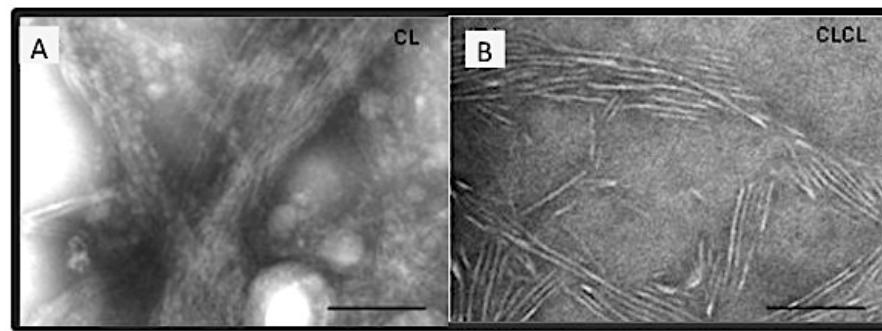


Figure 5. The self-assembled fibrils of CL and CL-CL peptides. **(A)** The mesh-like aggregates of CL. **(B)** The discrete, end-on-end assembly of CL-CL [159]. The scale bars are 100 nm in both pictures.

The self-assembly of the axial repeating structure of collagen fibrils indicates the specific *D*-staggering arrangement representing the most stable conformation during the self-association of the triple helix. Thus, in addition to providing sufficient stability for mutual association, there ought to be a built-in mechanism for conformational bias for the specific staggered structure. This bias was achieved by using recombinant peptides with repeating sequence units (SUs) in its primary structure [13,169]; each SU consists of 123 amino acid residues in an uninterrupted (Gly-X-Y)_n repeating sequence. Two collagen mimetic peptides: Col108 and 2U108 with, respectively, 3 and 2 identical SUs arranged in tandem plus a C-terminal overhang region were shown to self-assemble into mini-fibrils with an axial repeating structure of 35 nm as examined using electron microscopy (Figure 6A–D) and AFM [13]. This 35-nm repeating structure of the mini-fibrils consists of a ~15-nm overlap region and 20-nm gap, and is designated the *d*-period to distinguish it from the 64 nm *D*-period of natural fibrillar collagen. These mini-fibrils can reach 1 µm in length with a diameter about 50–70 nm. In a sense it is a self-assembly at an entirely different scale compared to the self-assembled long triple helix of CMPs. Structural analysis and biophysical studies have further indicated that the 35 nm *d*-period results from the self-assembly of the triple helices with a mutual staggering of 1 SU at the ends, in a way reminiscent to that of the *D*-staggering of collagen fibrils [13].

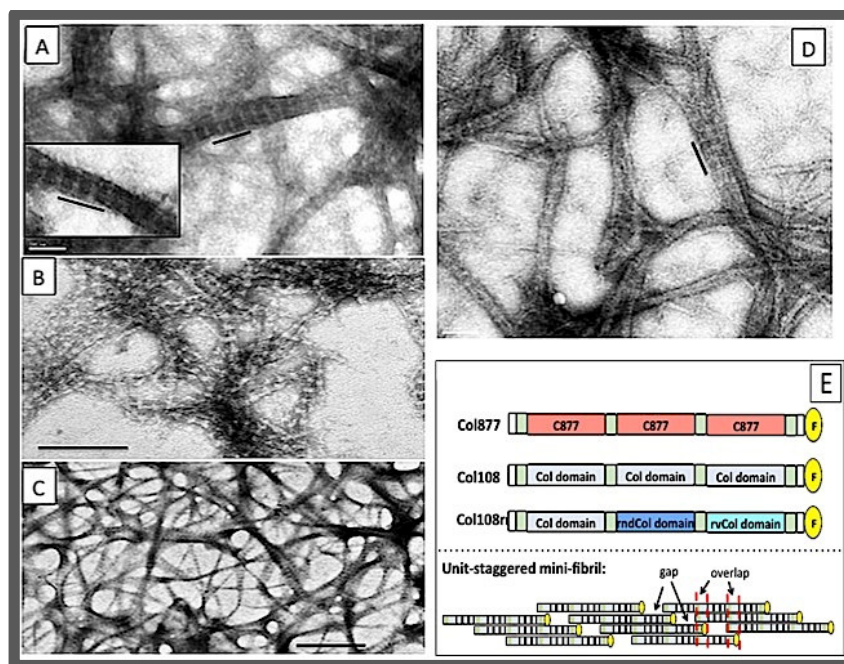


Figure 6. The unit-staggered mini-fibrils with a 35-nm *d*-period. (A–D) TEM images of Col108 mini-fibrils (13). The scale bars are 100 nm in (A,D), 200 nm in (B), and 500 nm in (C). (E) The schematic depictions of the peptides and the unit-staggering of the mini-fibril. The identical sequence units (SUs) are shown as rectangles in the same color: Col877 (3 red), Col108 (3 light blue), Col108rr (3 different SUs in three different colors); the yellow circles represent the foldon domain; the (GPP)₄ sequence that is present in the N-terminus of each SU is shown in green blocks (13). The unit-staggered mini-fibrils are drawn to highlight the alternating gaps and overlap zones, and the in-register alignment of the interacting residues (black vertical bands).

The design of the Col108 and 2U108 was based on the idea that, in the linear conformation of the triple helix, the structural periodicity of the fibrils should come directly from sequence periodicity. The tandem placement of the identical SUs, each consisting of 123 residues, determines a 123-residue sequence periodicity in the primary structure of Col108 and 2U108: Residues in the first sequence unit are repeated three and two times, respectively, every 123 residues. Several lines of observations support the sequence periodicity based *d*-period of the mini-fibrils:

- (1) The size of the *d*-period was in a good agreement with the size of the section of triple helix formed by one SU (123 amino acid residues, or 41 GXY triplets) based on the average helical rise of 0.86 nm per GXY tripeptide. Furthermore, the overlap region was in good agreement with the size of the C-terminal overhang;
- (2) Another peptide, the 1U108, contained only one SU. This peptide did not have a 123-residue sequence periodicity and did not form fibrils [169];
- (3) The pattern of sequence periodicity was the determining factor. In another peptide, peptide Col108rr, the sequences in each of the three SUs of Col108 were shuffled such that while the amino acid composition of this peptide was the same as Col108, the sequence periodicity was lost. As expected, the Col108rr only formed non-specific aggregates [170];
- (4) In a new peptide, peptide Col877, the SU of Col108 was replaced by another 123-residues containing residues 877–986 of the $\alpha 1$ chain of human type I collagen [170]. Thus, in contrast to Col108rr, Col877 had a very different amino acid composition from that of Col108, but had the same 123-residue sequence periodicity in its primary structure. Remarkably, Col877 can form mini-fibrils with the same 35-nm *d*-period as Col108. The Col877 mini-fibrils are a clear demonstration that the unit-staggering arrangement is at the foundation of the *d*-period axial structure.

- (5) The Col877 mini-fibril, and the lack of the fibril assembly of Col108rr and 1U108 indicated the foldon domain was not the determining factor of the fibril assembly, since all peptides, including Col108 and 2U108, contained the foldon domain.

The important finding of the unit-staggered mechanism of the mini-fibrils is that the *D*-staggered fibril assembly of collagen depends on the properties of the triple helix per se. Although factors such as the Hyp and telopeptides can facilitate the fibrillogenesis process, they may not be the determining factors [47,156,157]. The large stretch of sequence of natural collagen in Col877 also means the Col877 mini-fibrils can potentially be a model system to study binding and the effects of OI mutations on type I collagen at the fibril level.

4. Conclusions

The knowledge of the structure, and the sequence–function relationships of the triple helix have led to a wide range of applications for CMPs. One particularly productive area of research is the use of CMPs to probe cell–ECM interactions. The next immediate challenge is to understand and to replicate the modulation of these interactions by the supramolecular structure of collagen in the ECM. For this purpose, a model of a collagen fibril is needed to investigate (1) the relationship between the supramolecular structures of collagen and the biomechanical properties of the ECM, and (2) the determining factors of the molecular assemblies of collagen. Fibrillar collagen has been the focus of the effort to understand the molecular properties of the ECM because of its abundance, and because collagen fibrillogenesis is largely a self-association process of the triple helix without the involvement of any other globular domains. To fully capture the mechanical properties and the biological activity of collagen fibrils, the CMP-based systems most likely need to have a *D*-period like axial structure.

The research of CMPs is often inspired by the goal of developing synthetic collagen, or synthetic ECM. In addition to their far-reaching impact on biomedical applications and on applications in bioengineering, the ability to create complex systems like collagen stands as the ultimate test of our understanding of protein, of protein design, and of chemical synthesis. These are complicated systems. Collagen fibrils, for example, are heterotypical assemblies in tissues consisting of many different kinds of collagen and other macromolecules. The achievements we have gained so far in creating fibril-forming CMPs, or even creating a *D*-like structure by design, are rather modest; they are exceedingly simplified structures compared to natural collagen. But it is often such simplified model systems that aid us in peeling away the complexity of the tissues and the ECM one layer at a time. The allure of a collagen mimetic, after all, is functionality without the complexity.

Author Contributions: Writing—original draft preparation, Y.X.; writing—review and editing, Y.X. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NIH, SC1 GM121273.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hartgerink, J.D.; Beniash, E.; Stupp, S.I. Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* **2001**, *294*, 1684–1688, doi:10.1126/science.1063187.
2. Hartgerink, J.D.; Beniash, E.; Stupp, S.I. Peptide-amphiphile nanofibers: A versatile scaffold for the preparation of self-assembling materials. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5133–5138, doi:10.1073/pnas.0726999999.

3. Wojtowicz, A.M.; Shekaran, A.; Oest, M.E.; Dupont, K.M.; Templeman, K.L.; Hutmacher, D.W.; Guldberg, R.E.; Garcia, A.J. Coating of biomaterial scaffolds with the collagen-mimetic peptide GFOGER for bone defect repair. *Biomaterials* **2010**, *31*, 2574–2582.
4. Lowery, J.L.; Datta, N.; Rutledge, G.C. Effect of fiber diameter, pore size and seeding method on growth of human dermal fibroblasts in electrospun poly(epsilon-caprolactone) fibrous mats. *Biomaterials* **2010**, *31*, 491–504, doi:10.1016/j.biomaterials.2009.09.072.
5. Xie, J.; Li, X.; Lipner, J.; Manning, C.N.; Schwartz, A.G.; Thomopoulos, S.; Xia, Y. “Aligned-to-random” nanofiber scaffolds for mimicking the structure of the tendon-to-bone insertion site. *Nanoscale* **2010**, *2*, 923–926, doi:10.1039/c0nr00192a.
6. Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect Biol.* **2011**, *3*, a004978, doi:10.1101/cshperspect.a004978.
7. Ricard-Blum, S.; Ruggiero, F. The collagen superfamily: From the extracellular matrix to the cell membrane. *Pathol. Biol.* **2005**, *53*, 430–442, doi:10.1016/j.patbio.2004.12.024.
8. Birk, D.E.; Bruckner, P. Collagen suprastructures. *Top. Curr. Chem.* **2005**, *247*, 185–205, doi:10.1007/b103823.
9. Engel, J.; Bachinger, H.P. Structure, stability and folding of the collagen triple helix. *Top. Curr. Chem.* **2005**, *247*, 7–33, doi:10.1007/b103818.
10. Brodsky, B.; Thiagarajan, G.; Madhan, B.; Kar, K. Triple-helical peptides: An approach to collagen conformation, stability, and self-association. *Biopolymers* **2008**, *89*, 345–353, doi:10.1002/bip.20958.
11. Hulmes, D.J.; Miller, A.; Parry, D.A.; Piez, K.A.; Woodhead-Galloway, J. Analysis of the primary structure of collagen for the origins of molecular packing. *J. Mol. Biol.* **1973**, *79*, 137–148.
12. Hulmes, D.J.; Miller, A.; Parry, D.A.; Woodhead-Galloway, J. Fundamental periodicities in the amino acid sequence of the collagen alpha1 chain. *Biochem. Biophys. Res. Commun.* **1977**, *77*, 574–580.
13. Kaur, P.J.; Strawn, R.; Bai, H.; Xu, K.; Ordas, G.; Matsui, H.; Xu, Y. The self-assembly of a mini-fibril with axial periodicity from a designed collagen-mimetic triple helix. *J. Biol. Chem.* **2015**, *290*, 9251–9261, doi:10.1074/jbc.M113.542241.
14. Hulmes, D.J. Building collagen molecules, fibrils, and suprafibrillar structures. *J. Struct. Biol.* **2002**, *137*, 2–10, doi:10.1006/jsbi.2002.4450.
15. Kramer, R.Z.; Bella, J.; Brodsky, B.; HM., B. The crystal and molecular structure of a collagen-like peptide with a Biologically relevant sequence. *J. Mol. Biol.* **2001**, *311*, 131–147.
16. Li, Y.; Brodsky, B.; Baum, J. NMR shows hydrophobic interactions replace glycine packing in the triple helix at a natural break in the (Gly-X-Y)_n repeat. *J. Biol. Chem.* **2007**, *282*, 22699–22706, doi:10.1074/jbc.M702910200.
17. Li, Y.; Kim, S.; Brodsky, B.; Baum, J. Identification of partially disordered peptide intermediates through residue-specific NMR diffusion measurements. *J. Am. Chem. Soc.* **2005**, *127*, 10490–10491, doi:10.1021/ja052801d.
18. Fan, P.; Li, M.-H.; Brodsky, B.; Baum, J. Backbone Dynamics of (Pro-Hyp-Gly)₁₀ and a Designed Collagen-like Triple-Helical Peptide by ¹⁵N NMR Relaxation and Hydrogen-Exchange Measurements. *Biochemistry* **1993**, *32*, 13299–13309.
19. Ottl, J.; Battistuta, R.; Pieper, M.; Tschesche, H.; Bode, W.; Kuhn, K.; Moroder, L. Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases. *FEBS Lett.* **1996**, *398*, 31–36.
20. Piez, K.A.; Trus, B.L. Sequence regularities and packing of collagen molecules. *J. Mol. Biol.* **1978**, *122*, 419–432, doi:10.1016/0022-2836(78)90419-9.
21. Jalan, A.A.; Sammon, D.; Hartgerink, J.D.; Brear, P.; Stott, K.; Hamaia, S.W.; Hunter, E.J.; Walker, D.R.; Leitingner, B.; Farndale, R.W. Chain alignment of collagen I deciphered using computationally designed heterotrimers. *Nat. Chem. Biol.* **2020**, *16*, 423–429, doi:10.1038/s41589-019-0435-y.
22. Prockop, D.J.; Kivirikko, K.I.; Tuderman, L.; Guzman, N.A. The biosynthesis of collagen and its disorders (first of two parts). *N. Engl. J. Med.* **1979**, *301*, 13–23, doi:10.1056/nejm197907053010104.
23. Sharma, U.; Carrique, L.; Vadon-Le Goff, S.; Mariano, N.; Georges, R.-N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N.; et al. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671.
24. DiChiara, A.S.; Li, R.C.; Suen, P.H.; Hosseini, A.S.; Taylor, R.J.; Weickhardt, A.F.; Malhotra, D.; McCaslin, D.R. A cysteine-based molecular code informs collagen C-propeptide assembly. *Nat Commun.* **2018**, *9*, 4206.
25. Wood, G.C.; Keech, M.K. The formation of fibrils from collagen solutions. 1. The effect of experimental conditions: Kinetic and electron-microscope studies. *Biochem. J.* **1960**, *75*, 588–598.
26. Piez, K.A. Structure and assembly of the native collagen fibril. *Connect. Tissue Res.* **1982**, *10*, 25–36.
27. Piez, K.A. Molecular and Aggregate Structure of the Collagens. In *Extracellular Matrix Biochemistry*; Piez, K.A., Reddi, A.H., Eds.; Elsevier: New York, NY, USA; Amsterdam, The Netherlands; Oxford, UK, 1984.
28. Piez, K.A.; Miller, A. The structure of collagen fibrils. *J. Supramol. Struct.* **1974**, *2*, 121–137, doi:10.1002/jss.400020207.
29. Piez, K.A.; Trus, B.L. Microfibrillar structure and packing of collagen: Hydrophobic interactions. *J. Mol. Biol.* **1977**, *110*, 701–704.
30. Piez, K.A.; Trus, B.L. A new model for packing of type-I collagen molecules in the native fibril. *Biosci. Rep.* **1981**, *1*, 801–810.
31. Orgel, J.; Sella, I.; Madhurapantula, R.S.; Antipova, O.; Mandelberg, Y.; Kashman, Y.; Benayahu, D.; Benayahu, Y. Molecular and ultrastructural studies of a fibrillar collagen from octocoral (Cnidaria). *J. Exp. Biol.* **2017**, *220*, 3327–3335, doi:10.1242/jeb.163824.
32. Orgel, J.P.; Antipova, O.; Sagi, I.; Bitler, A.; Qiu, D.; Wang, R.; Xu, Y.; San Antonio, J.D. Collagen fibril surface displays a constellation of sites capable of promoting fibril assembly, stability, and hemostasis. *Connect Tissue Res.* **2011**, *52*, 18–24, doi:10.3109/03008207.2010.511354.

33. Orgel, J.P.; Irving, T.C.; Miller, A.; Wess, T.J. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9001–9005, doi:10.1073/pnas.0502718103.
34. Orgel, J.P.; Persikov, A.V.; Antipova, O. Variation in the helical structure of native collagen. *PLoS ONE* **2014**, *9*, e89519, doi:10.1371/journal.pone.0089519.
35. Orgel, J.P.; San Antonio, J.D.; Antipova, O. Molecular and structural mapping of collagen fibril interactions. *Connect Tissue Res.* **2011**, *52*, 2–17, doi:10.3109/03008207.2010.511353.
36. Sweeney, S.M.; Orgel, J.P.; Fertala, A.; McAuliffe, J.D.; Turner, K.R.; Di Lullo, G.A.; Chen, S.; Antipova, O.; Perumal, S.; Ala-Kokko, L.; et al. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.* **2008**, *283*, 21187–21197, doi:10.1074/jbc.M709319200.
37. Herr, A.B.; Farndale, R.W. Structural insights into the interactions between platelet receptors and fibrillar collagen. *J. Biol. Chem.* **2009**, *284*, 19781–19785, doi:10.1074/jbc.R109.013219.
38. Chang, S.W.; Buehler, M.J. Molecular biomechanics of collagen molecules. *Mater. Today* **2014**, *17*, 70–76, doi:10.1016/j.mattod.2014.01.019.
39. Buehler, M.J. Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12285–12290, doi:10.1073/pnas.0603216103.
40. Davidenko, N.; Schuster, C.F.; Bax, D.V.; Farndale, R.W.; Hamaia, S.; Best, S.M.; Cameron, R.E. Evaluation of cell binding to collagen and gelatin: A study of the effect of 2D and 3D architecture and surface chemistry. *J. Mater. Sci. Mater. Med.* **2016**, *27*, 148.
41. Malcor, J.-D.; Hunter, E.J.; Davidenko, N.; Bax, D.V.; Cameron, R.; Best, S.; Sinha, S.; Farndale, R.W. Collagen scaffolds functionalized with triple-helical peptides support 3D HUVEC culture. *Regen. Biomater.* **2020**, *7*, 471–482.
42. Plant, A.L.; Bhadriraju, K.; Spurlin, T.A.; Elliott, J.T. Cell response to matrix mechanics: Focus on collagen. *Biochim. Et Biophys. Acta* **2009**, *1793*, 893–902.
43. Sherman, V.R.; Yang, W.; Meyers, M.A. The materials science of collagen. *J. Mech. Behav. Biomed.* **2015**, *52*, 22–50, doi:10.1016/j.jmbbm.2015.05.023.
44. Kadler, K.E.; Hill, A.; Canty-Laird, E.G. Collagen fibrillogenesis: Fibronectin, integrins, and minor collagens as organizers and nucleators. *Curr. Opin. Cell Biol.* **2008**, *20*, 495–501, doi:10.1016/j.ceb.2008.06.008.
45. Kadler, K.E.; Hojima, Y.; Prockop, D.J. Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process. *J. Biol. Chem.* **1987**, *262*, 15696–15701.
46. Prockop, D.J.; Fertala, A. Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides. Demonstration that assembly is driven by specific binding sites on the monomers. *J. Biol. Chem.* **1998**, *273*, 15598–15604.
47. Kuznetsova, N.; Leikin, S. Does the triple helical domain of type I collagen encode molecular recognition and fiber assembly while telopeptides serve as catalytic domains? Effect of proteolytic cleavage on fibrillogenesis and on collagen-collagen interaction in fibers. *J. Biol. Chem.* **1999**, *274*, 36083–36088.
48. Comper, W.D.; Veis, A. The mechanism of nucleation for in vitro collagen fibril formation. *Biopolymers* **1977**, *16*, 2113–2131, doi:10.1002/bip.1977.360161004.
49. Helseth, D.L., Jr.; Veis, A. Collagen self-assembly in vitro. Differentiating specific telopeptide-dependent interactions using selective enzyme modification and the addition of free amino telopeptide. *J. Biol. Chem.* **1981**, *256*, 7118–7128.
50. Farndale, R.W.; Lisan, T.; Bihan, D.; Hamaia, S.; Smerling, C.S.; Pugh, N.; Konitsiotis, A.; Leiting, B.; de Groot, P.G.; Jarvis, G.E.; et al. Cell-collagen interactions: The use of peptide Toolkits to investigate collagen-receptor interactions. *Biochem. Soc. Trans.* **2008**, *36*, 241–250, doi:10.1042/BST0360241.
51. Di Lullo, G.A.; Sweeney, S.M.; Korkko, J.; Ala-Kokko, L.; San Antonio, J.D. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* **2002**, *277*, 4223–4231, doi:10.1074/jbc.M110709200.
52. Sweeney, S.M.; Guy, C.A.; Fields, G.B.; San Antonio, J.D. Defining the domains of type I collagen involved in heparin-binding and endothelial tube formation. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7275–7280.
53. Vogel, W.F.; Abdulhussein, R.; Ford, C.E. Sensing extracellular matrix: An update on discoidin domain receptor function. *Cell. Signal.* **2006**, *18*, 1108–1116, doi:10.1016/j.cellsig.2006.02.012.
54. Leiting, B.; Hohenester, E. Mammalian collagen receptors. *Matrix Biol. J. Int. Soc. Matrix Biol.* **2007**, *26*, 146–155, doi:10.1016/j.matbio.2006.10.007.
55. Barczyk, M.; Carracedo, S.; Gullberg, D. Integrins. *Cell Tissue Res.* **2009**, *339*, 269, doi:10.1007/s00441-009-0834-6.
56. Banerjee, J.; Azevedo, H.S. Crafting of functional biomaterials by directed molecular self-assembly of triple helical peptide building blocks. *Interface Focus* **2017**, *7*, 20160138.
57. Amar, S.; Smith, L.; Fields, G.B. Matrix metalloproteinase collagenolysis in health and disease. *Biochim. Et Biophys. Acta. Mol. Cell Res.* **2017**, *1864*, 1940–1951, doi:10.1016/j.bbamcr.2017.04.015.
58. Barillari, G. The Impact of Matrix Metalloproteinase-9 on the Sequential Steps of the Metastatic Process. *Int. J. Mol. Sci.* **2020**, *21*, doi:10.3390/ijms21124526.
59. Cui, N.; Hu, M.; Khalil, R.A. Biochemical and Biological Attributes of Matrix Metalloproteinases. *Prog. Mol. Biol. Transl. Sci.* **2017**, *147*, 1–73, doi:10.1016/bs.pmbts.2017.02.005.

60. Fallas, J.A.; O'Leary, L.E.; Hartgerink, J.D. Synthetic collagen mimics: Self-assembly of homotrimers, heterotrimers and higher order structures. *Chem. Soc. Rev.* **2010**, *39*, 3510–3527, doi:10.1039/b919455j.
61. Lauer-Fields, J.L.; Juska, D.; Fields, G.B. Matrix metalloproteinases and collagen catabolism. *Biopolymers* **2002**, *66*, 19–32, doi:10.1002/bip.10201.
62. Lauer-Fields, J.L.; Tuzinski, K.A.; Shimokawa, K.; Nagase, H.; Fields, G.B. Hydrolysis of triple-helical collagen peptide models by matrix metalloproteinases. *J. Biol. Chem.* **2000**, *275*, 13282–13290, doi:10.1074/jbc.275.18.13282.
63. Pedchenko, V.; Kitching, A.R.; Hudson, B.G. Goodpasture's autoimmune disease—A collagen IV disorder. *Matrix Biol. J. Int. Soc. Matrix Biol.* **2018**, *71–72*, 240–249, doi:10.1016/j.matbio.2018.05.004.
64. Miyake, M.; Hori, S.; Morizawa, Y.; Tatsumi, Y.; Toritsuka, M.; Ohnishi, S.; Shimada, K.; Furuya, H.; Khadka, V.S.; Deng, Y.; et al. Collagen type IV alpha 1 (COL4A1) and collagen type XIII alpha 1 (COL13A1) produced in cancer cells promote tumor budding at the invasion front in human urothelial carcinoma of the bladder. *Oncotarget* **2017**, *8*, 36099–36114, doi:10.18632/oncotarget.16432.
65. Vaniotis, G.; Rayes, R.F.; Qi, S.; Milette, S.; Wang, N.; Perrino, S.; Bourdeau, F.; Nyström, H.; He, Y.; Lamarche-Vane, N.; et al. Collagen IV-conveyed signals can regulate chemokine production and promote liver metastasis. *Oncogene* **2018**, *37*, 3790–3805, doi:10.1038/s41388-018-0242-z.
66. Jayadev, R.; Chi, Q.; Keeley, D.P.; Hastie, E.L.; Kelley, L.C.; Sherwood, D.R. α -Integrins dictate distinct modes of type IV collagen recruitment to basement membranes. *J. Cell Biol.* **2019**, *218*, 3098–3116, doi:10.1083/jcb.201903124.
67. Ramshaw, J.A.; Werkmeister, J.A.; Glattau, V. Collagen-based biomaterials. *Biotechnol. Genet. Eng. Rev.* **1996**, *13*, 335–382.
68. Kumar, V.A.; Shi, S.; Wang, B.K.; Li, I.C.; Jalan, A.A.; Sarkar, B.; Wickremasinghe, N.C.; Hartgerink, J.D. Drug-triggered and cross-linked self-assembling nanofibrous hydrogels. *J. Am. Chem. Soc.* **2015**, *137*, 4823–4830, doi:10.1021/jacs.5b01549.
69. Kumar, V.A.; Taylor, N.L.; Jalan, A.A.; Hwang, L.K.; Wang, B.K.; Hartgerink, J.D. A nanostructured synthetic collagen mimic for hemostasis. *Biomacromolecules* **2014**, *15*, 1484–1490, doi:10.1021/bm500091e.
70. Fields, C.G.; Lovdahl, C.M.; Miles, A.J.; Hagen, V.L.; Fields, G.B. Solid-phase synthesis and stability of triple-helical peptides incorporating native collagen sequences. *Biopolymers* **1993**, *33*, 1695–1707, doi:10.1002/bip.360331107.
71. Heidemann, E.; Roth, W. Synthesis and Investigation of Collagen Model Peptides. *Adv. Polym. Sci.* **1982**, *43*, 143–203.
72. Sakakibara, S.; Inouye, K.; Shudo, K.; Kishida, Y.; Kobayashi, Y.; Prockop, D.J. Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **1973**, *303*, 198–202, doi:10.1016/0005-279590164-5.
73. Suto, K.; Noda, H. Conformational change of the triple-helical structure. IV. Kinetics of the helix-folding of (Pro-Pro-Gly)_n (n equals 10, 12, and 15). *Biopolymers* **1974**, *13*, 2477–2488, doi:10.1002/bip.1974.360131207.
74. Fields, G.D.; Prockop, D.J. Perspectives on Synthesis and Applications of Triple-Helical Collagen Model Peptides. *Biopolymers* **1996**, *40*, 345.
75. Brodsky, B.; Persikov, A.V. Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **2005**, *70*, 301–339, doi:10.1016/S0065-323370009-7.
76. Baum, J.; Brodsky, B. Folding of peptide models of collagen and misfolding diseases. *Curre. Opin. Struct. Biol.* **1999**, *9*, 122–128.
77. Buevich, A.V.; Silva, T.; Brodsky, B.; Baum, J. 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.* **2004**, *279*, 46890–46895, doi:10.1074/jbc.M407061200.
78. Hyde, T.J.; Bryan, M.A.; Brodsky, B.; Baum, J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J. Biol. Chem.* **2006**, *281*, 36937–36943, doi:10.1074/jbc.M605135200.
79. Xu, Y.; Bhate, M.; Brodsky, B. Characterization of the Nucleation Step and Folding of a Collagen Triple-Helix Peptide. *Biochemistry* **2002**, *41*, 8143–8151.
80. Persikov, A.V.; Xu, Y.; Brodsky, B. Equilibrium thermal transitions of collagen model peptides. *Protein Sci.* **2004**, *13*, 893–902.
81. Persikov, A.V.; Ramshaw, J.A.M.; Brodsky, B. Prediction of Collagen Stability from Amino Acid Sequence. *J. Biol. Chem.* **2005**, *280*, 19343–19349.
82. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H.M. Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. *Science* **1994**, *266*, 75–81.
83. Okuyama, K.; Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)₁₀. *J. Mol. Biol.* **1981**, *152*, 427–443, doi:10.1016/0022-283690252-7.
84. Nagarajan, V.; Kamitori, S.; Okuyama, K. Structure analysis of a collagen-model peptide with a (Pro-Hyp-Gly) sequence repeat. *J. Biochem.* **1999**, *125*, 310–318, doi:10.1093/oxfordjournals.jbchem.a022288.
85. Persikov, A.V.; Ramshaw, J.A.; Kirkpatrick, A.; Brodsky, B. Triple-helix propensity of hydroxyproline and fluoroproline: Comparison of host-guest and repeating tripeptide collagen models. *J. Am. Chem. Soc.* **2003**, *125*, 11500–11501, doi:10.1021/ja036673+.
86. Persikov, A.V.; Ramshaw, J.A.; Kirkpatrick, A.; Brodsky, B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry* **2005**, *44*, 1414–1422, doi:10.1021/bi048216r.
87. Persikov, A.V.; Ramshaw, J.A.M.; Kirkpatrick, A.; Brodsky, B. Amino Acid Propensities for the Collagen Triple-Helix. *Biochemistry* **2000**, *39*, 14960–14967.
88. Persikov, A.V.; Ramshaw, J.A.M.; Kirkpatrick, A.; Brodsky, B. Peptide investigations of pairwise interactions in the collagen triple-helix. *J. Mol. Biol.* **2002**, *316*, 385–394.

89. Knight, C.G.; Morton, L.F.; Onley, D.J.; Peachey, A.R.; Messent, A.J.; Smethurst, P.A.; Tuckwell, D.S.; Farndale, R.W.; Barnes, M.J. Identification in collagen type I of an integrin alpha2 beta1-binding site containing an essential GER sequence. *J. Biol. Chem.* **1998**, *273*, 33287–33294, doi:10.1074/jbc.273.50.33287.
90. Knight, C.G.; Morton, L.F.; Peachey, A.R.; Tuckwell, D.S.; Farndale, R.W.; Barnes, M.J. The collagen-binding A-domains of integrins alpha(1)beta(1) and alpha(2)beta(1) recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. *J. Biol. Chem.* **2000**, *275*, 35–40.
91. Tuckwell, D.; Calderwood, D.A.; Green, L.J.; Humphries, M.J. Integrin alpha 2 I-domain is a binding site for collagens. *J. Cell Sci.* **1995**, *108*, 1629–1637.
92. Michishita, M.; Videm, V.; Arnaout, M.A. A novel divalent cation-binding site in the A domain of the beta 2 integrin CR3 (CD11b/CD18) is essential for ligand binding. *Cell* **1993**, *72*, 857–867, doi:10.1016/0092-867490575-b.
93. Kamata, T.; Liddington, R.C.; Takada, Y. Interaction between collagen and the alpha(2) I-domain of integrin alpha(2)beta(1). Critical role of conserved residues in the metal ion-dependent adhesion site (MIDAS) region. *J. Biol. Chem.* **1999**, *274*, 32108–32111, doi:10.1074/jbc.274.45.32108.
94. Emsley, J.; Knight, C.G.; Farndale, R.W.; Barnes, M.J. Structural basis of collagen recognition by integrin alpha2beta1. *Cell* **2000**, *101*, 47–56.
95. Raynal, N.; Hamaia, S.W.; Siljander, P.R.; Maddox, B.; Peachey, A.R.; Fernandez, R.; Foley, L.J.; Slatter, D.A.; Jarvis, G.E. Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *J. Biol. Chem.* **2006**, *281*, 3821–3831.
96. Slatter, D.A.; Foley, L.A.; Peachey, A.R.; Nietlispach, D.; Farndale, R.W. Rapid synthesis of a register-specific heterotrimeric type I collagen helix encompassing the integrin alpha2beta1 binding site. *J. Mol. Biol.* **2006**, *359*, 289–298, doi:10.1016/j.jmb.2006.02.071.
97. Lisman, T.; Raynal, N.; Groeneveld, D.; Maddox, B.; Peachey, A.R.; Huizinga, E.G.; de Groot, P.G.; Farndale, R.W. A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **2006**, *108*, 3753–3756.
98. Knight, C.G.; Morton, L.F.; Onley, D.J.; Peachey, A.R.; Ichinohe, T.; Okuma, M.; Farndale, R.W.; Barnes, M.J. Collagen-platelet interaction: Gly-Pro-Hyp is uniquely specific for platelet Gp VI and mediates platelet activation by collagen. *Cardiovasc Res.* **1999**, *41*, 450–457, doi:10.1016/s0008-636300306-x.
99. Xu, H.; Raynal, N.; Stathopoulos, S.; Myllyharju, J.; Farndale, R.W.; Leitingner, B. Collagen binding specificity of the discoidin domain receptors: Binding sites on collagens II and III and molecular determinants for collagen IV recognition by DDR1. *Matrix Biol. J. Int. Soc. Matrix Biol.* **2011**, *30*, 16–26, doi:10.1016/j.matbio.2010.10.004.
100. Konitsiotis, A.D.; Raynal, N.; Bihan, D.; Hohenester, E.; Farndale, R.W.; Leitingner, B. Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **2008**, *283*, 6861–6868, doi:10.1074/jbc.M709290200.
101. Jarvis, G.E.; Raynal, N.; Langford, J.P.; Onley, D.J.; Andrews, A.; Smethurst, P.A.; Farndale, R.W. Identification of a major GpVI-binding locus in human type III collagen. *Blood* **2008**, *111*, 4986–4996, doi:10.1182/blood-2007-08-108472.
102. Lebbink, R.J.; Raynal, N.; de Ruiter, T.; Bihan, D.G.; Farndale, R.W.; Meysaard, L. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol. J. Int. Soc. Matrix Biol.* **2009**, *28*, 202–210, doi:10.1016/j.matbio.2009.03.005.
103. Zhou, L.; Hinerman, J.M.; Blaszczyk, M.; Miller, J.L.; Conrady, D.G.; Barrow, A.D.; Chirgadze, D.Y.; Bihan, D.; Farndale, R.W.; Herr, A.B. Structural basis for collagen recognition by the immune receptor OSCAR. *Blood* **2016**, *127*, 529–537, doi:10.1182/blood-2015-08-667055.
104. Perumal, S.; Antipova, O.; Orgel, J.P. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2824–2829, doi:10.1073/pnas.0710588105.
105. Otl, J.; Gabriel, D.; Murphy, G.; Knauper, V.; Tominaga, Y.; Nagase, H.; Kroger, M.; Tschesche, H.; Bode, W.; Moroder, L. Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases. *Chem. Biol.* **2000**, *7*, 119–132.
106. Boudko, S.P.; Bachinger, H.P. The NC2 domain of type IX collagen determines the chain register of the triple helix. *J. Biol. Chem.* **2012**, *287*, 44536–44545, doi:10.1074/jbc.M112.417543.
107. Boudko, S.P.; Engel, J.; Okuyama, K.; Mizuno, K.; Bachinger, H.P.; Schumacher, M.A. Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J. Biol. Chem.* **2008**, *283*, 32580–32589, doi:10.1074/jbc.M805394200.
108. Sacca, B.; Renner, C.; Moroder, L. The Chain Register in Heterotrimeric Collagen Peptides Affects Triple Helix Stability and Folding Kinetics. *J. Mol. Biol.* **2002**, *324*, 309–318.
109. Sacca, B.; Sinner, E.K.; Kaiser, J.; Lübken, C.; Eble, J.A.; Moroder, L. Binding and docking of synthetic heterotrimeric collagen type IV peptides with alpha1beta1 integrin. *Chembiochem. Eur. J. Chem. Biol.* **2002**, *3*, 904–907, doi:10.1002/1439-76333:9<904::aid-cbic904>3.0.co;2-i.
110. Xu, F.; Zahid, S.; Silva, T.; Nanda, V. Computational design of a collagen A:B:C-type heterotrimer. *J. Am. Chem. Soc.* **2011**, *133*, 15260–15263, doi:10.1021/ja205597g.
111. Xu, F.; Zhang, L.; Koder, R.L.; Nanda, V. De novo self-assembling collagen heterotrimers using explicit positive and negative design. *Biochemistry* **2010**, *49*, 2307–2316, doi:10.1021/bi902077d.
112. Zheng, H.; Lu, C.; Lan, J.; Fan, S.; Nanda, V.; Xu, F. How electrostatic networks modulate specificity and stability of collagen. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6207–6212, doi:10.1073/pnas.1802171115.

113. Fallas, J.A.; Hartgerink, J.D. Computational design of self-assembling register-specific collagen heterotrimers. *Nat. Commun.* **2012**, *3*, 1087, doi:10.1038/ncomms2084.
114. Fallas, J.A.; Lee, M.A.; Jalan, A.A.; Hartgerink, J.D. Rational design of single-composition ABC collagen heterotrimers. *J. Am. Chem. Soc.* **2012**, *134*, 1430–1433, doi:10.1021/ja209669u.
115. Gauba, V.; Hartgerink, J.D. Surprisingly high stability of collagen ABC heterotrimer: Evaluation of side chain charge pairs. *J. Am. Chem. Soc.* **2007**, *129*, 15034–15041, doi:10.1021/ja075854z.
116. Gauba, V.; Hartgerink, J.D. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.* **2007**, *129*, 2683–2690, doi:10.1021/ja0683640.
117. Jalan, A.A.; Demeler, B.; Hartgerink, J.D. Hydroxyproline-free single composition ABC collagen heterotrimer. *J. Am. Chem. Soc.* **2013**, *135*, 6014–6017.
118. Jalan, A.A.; Hartgerink, J.D. Pairwise interactions in collagen and the design of heterotrimeric helices. *Curr. Opin. Chem. Biol.* **2013**, *17*, 960–967, doi:10.1016/j.cbpa.2013.10.019.
119. Jalan, A.A.; Hartgerink, J.D. Simultaneous control of composition and register of an AAB-type collagen heterotrimer. *Biomacromolecules* **2013**, *14*, 179–185, doi:10.1021/bm3015818.
120. O'Leary, L.E.; Fallas, J.A.; Hartgerink, J.D. Positive and negative design leads to compositional control in AAB collagen heterotrimers. *J. Am. Chem. Soc.* **2011**, *133*, 5432–5443, doi:10.1021/ja111239r.
121. Russell, L.E.; Fallas, J.A.; Hartgerink, J.D. Selective assembly of a high stability AAB collagen heterotrimer. *J. Am. Chem. Soc.* **2010**, *132*, 3242–3243, doi:10.1021/ja909720g.
122. Tanrikulu, I.C.; Forticaux, A.; Jin, S.; Raines, R.T. Peptide tessellation yields micrometre-scale collagen triple helices. *Nat. Chem.* **2016**, *8*, 1008–1014, doi:10.1038/nchem.2556.
123. O'Leary, L.E.; Fallas, J.A.; Bakota, E.L.; Kang, M.K.; Hartgerink, J.D. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nat. Chem.* **2011**, *3*, 821–828, doi:10.1038/nchem.1123.
124. Kotch, F.W.; Raines, R.T. Self-assembly of synthetic collagen triple helices. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3028–3033, doi:10.1073/pnas.0508783103.
125. Yamazaki, C.M.; Kadoya, Y.; Hozumi, K.; Okano-Kosugi, H.; Asada, S.; Kitagawa, K.; Nomizu, M.; Koide, T. A collagen-mimetic triple helical supramolecule that evokes integrin-dependent cell responses. *Biomaterials* **2010**, *31*, 1925–1934, doi:10.1016/j.biomaterials.2009.10.014.
126. Koide, T.; Homma, D.L.; Asada, S.; Kitagawa, K. Self-complementary peptides for the formation of collagen-like triple helical supramolecules. *Bioorg. Med. Chem. Lett* **2005**, *15*, 5230–5233, doi:10.1016/j.bmcl.2005.08.041.
127. Sarkar, B.; O'Leary, L.E.; Hartgerink, J.D. Self-assembly of fiber-forming collagen mimetic peptides controlled by triple-helical nucleation. *J. Am. Chem. Soc.* **2014**, *136*, 14417–14424, doi:10.1021/ja504377s.
128. Rele, S.; Song, Y.; Apkarian, R.P.; Qu, Z.; Conticello, V.P.; Chaikof, E.L. D-periodic collagen-mimetic microfibers. *J. Am. Chem. Soc.* **2007**, *129*, 14780–14787, doi:10.1021/ja0758990.
129. Bai, H.; Xu, K.; Xu, Y.; Matsui, H. Fabrication of Au nanowire in uniform length and diameter using a new monodisperse and rigid biomolecular template, collagen triple helix. *Angewante Chem.* **2007**, *46*, 3319–3322.
130. Kaur, P.; Maeda, Y.; Mutter, A.C.; Matsunaga, T.; Xu, Y.; Matsui, H. 3D Self-Assembly of Triple Helix Peptide Nanowires into Micron-Sized Crystalline Cubes with Joint Hubs of Ligand-Conjugated Au Nanoparticles. *Nature* **2009**, *459*, 8375–8378.
131. Kar, K.; Amin, P.; Bryan, M.A.; Persikov, A.V.; Mohs, A.; Wang, Y.H.; Brodsky, B. Self-association of collagen triple helix peptides into higher order structures. *J. Biol. Chem.* **2006**, *281*, 33283–33290, doi:10.1074/jbc.M605747200.
132. Kar, K.; Ibrar, S.; Nanda, V.; Getz, T.M.; Kunapuli, S.P.; Brodsky, B. Aromatic interactions promote self-association of collagen triple-helical peptides to higher-order structures. *Biochemistry* **2009**, *48*, 7959–7968, doi:10.1021/bi900496m.
133. Kar, K.; Wang, Y.H.; Brodsky, B. Sequence dependence of kinetics and morphology of collagen model peptide self-assembly into higher order structures. *Protein Sci. Publ. Protein Soc.* **2008**, *17*, 1086–1095, doi:10.1110/ps.083441308.
134. McGuinness, K.; Khan, I.J.; Nanda, V. Morphological diversity and polymorphism of self-assembling collagen peptides controlled by length of hydrophobic domains. *ACS Nano* **2014**, *8*, 12514–12523, doi:10.1021/nn505369d.
135. Parmar, A.S.; James, J.K.; Grisham, D.R.; Pike, D.H.; Nanda, V. Dissecting Electrostatic Contributions to Folding and Self-Assembly Using Designed Multicomponent Peptide Systems. *J. Am. Chem. Soc.* **2016**, *138*, 4362–4367, doi:10.1021/jacs.5b10304.
136. Cepas, M.A.; Kinney, W.A.; Chen, C.; Vinter, J.G.; Almond, H.R., Jr.; Balss, K.M.; Maryanoff, C.A.; Schmidt, U.; Breslav, M.; Mahan, A.; et al. Thrombogenic collagen-mimetic peptides: Self-assembly of triple helix-based fibrils driven by hydrophobic interactions. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8513–8518, doi:10.1073/pnas.0800291105.
137. Cepas, M.A.; Kinney, W.A.; Chen, C.; Leo, G.C.; Tounge, B.A.; Vinter, J.G.; Joshi, P.P.; Maryanoff, B.E. Collagen-related peptides: Self-assembly of short, single strands into a functional biomaterial of micrometer scale. *J. Am. Chem. Soc.* **2007**, *129*, 2202–2203, doi:10.1021/ja066986f.
138. Przybyla, D.E.; Chmielewski, J. Metal-triggered radial self-assembly of collagen peptide fibers. *J. Am. Chem. Soc.* **2008**, *130*, 12610–12611, doi:10.1021/ja804942w.
139. Przybyla, D.E.; Chmielewski, J. Metal-triggered collagen peptide disk formation. *J. Am. Chem. Soc.* **2010**, *132*, 7866–7867, doi:10.1021/ja103148t.
140. Li, Y.; Yu, S.M. Targeting and mimicking collagens via triple helical peptide assembly. *Curr. Opin. Chem. Biol.* **2013**, *17*, 968–975, doi:10.1016/j.cbpa.2013.10.018.

141. Li, Y.; Ho, D.; Meng, H.; Chan, T.R.; An, B.; Yu, H.; Brodsky, B.; Jun, A.S.; Michael Yu, S. Direct detection of collagenous proteins by fluorescently labeled collagen mimetic peptides. *Bioconjug. Chem.* **2013**, *24*, 9–16, doi:10.1021/bc3005842.
142. Li, Y.; Foss, C.A.; Summerfield, D.D.; Doyle, J.J.; Torok, C.M.; Dietz, H.C.; Pomper, M.G.; Yu, S.M. Targeting collagen strands by photo-triggered triple-helix hybridization. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14767–14772, doi:10.1073/pnas.1209721109.
143. Hwang, J.; San, B.H.; Turner, N.J.; White, L.J.; Faulk, D.M.; Badylak, S.F.; Li, Y.; Yu, S.M. Molecular assessment of collagen denaturation in decellularized tissues using a collagen hybridizing peptide. *Acta Biomater.* **2017**, *53*, 268–278, doi:10.1016/j.actbio.2017.01.079.
144. Hodges, J.A.; Raines, R.T. Stereoelectronic and steric effects in the collagen triple helix: Toward a code for strand association. *J. Am. Chem. Soc.* **2005**, *127*, 15923–15932, doi:10.1021/ja054674r.
145. Barth, D.; Milbradt, A.G.; Renner, C.; Moroder, L. A (4R)- or a (4S)-fluoroproline residue in position Xaa of the (Xaa-Yaa-Gly) collagen repeat severely affects triple-helix formation. *Chembiochem. Eur. J. Chem. Biol.* **2004**, *5*, 79–86, doi:10.1002/cbic.200300702.
146. Zitnay, J.L.; Li, Y.; Qin, Z.; San, B.H.; Depalle, B.; Reese, S.P.; Buehler, M.J.; Yu, S.M.; Weiss, J.A. Molecular level detection and localization of mechanical damage in collagen enabled by collagen hybridizing peptides. *Nat. Commun.* **2017**, *8*, 14913, doi:10.1038/ncomms14913.
147. Wahyudi, H.; Reynolds, A.A.; Li, Y.; Owen, S.C.; Yu, S.M. Targeting collagen for diagnostic imaging and therapeutic delivery. *J. Control. Release Off. J. Control. Release Soc.* **2016**, *240*, 323–331, doi:10.1016/j.jconrel.2016.01.007.
148. Ellison, A.J.; Dempwolff, F.; Kearns, D.B.; Raines, R.T. Role for Cell-Surface Collagen of *Streptococcus pyogenes* in Infections. *ACS Infect. Dis.* **2020**, *6*, 1836–1843, doi:10.1021/acsinfecdis.0c00073.
149. Schroeder, A.B.; Karim, A.; Ocotl, E.; Dones, J.M.; Chacko, J.V.; Liu, A.; Raines, R.T.; Gibson, A.L.F.; Eliceiri, K.W. Optical imaging of collagen fiber damage to assess thermally injured human skin. *Wound Repair Regen. Off. Publ. Wound Health Soc. Eur. Tissue Repair Soc.* **2020**, *28*, 848–855, doi:10.1111/wrr.12849.
150. Ellison, A.J.; Tanrikulu, I.C.; Dones, J.M.; Raines, R.T. Cyclic Peptide Mimetic of Damaged Collagen. *Biomacromolecules* **2020**, *21*, 1539–1547, doi:10.1021/acs.biomac.0c00103.
151. Dones, J.M.; Tanrikulu, I.C.; Chacko, J.V.; Schroeder, A.B.; Hoang, T.T.; Gibson, A.L.F.; Eliceiri, K.W.; Raines, R.T. Optimization of interstrand interactions enables burn detection with a collagen-mimetic peptide. *Org. Biomol. Chem.* **2019**, *17*, 9906–9912, doi:10.1039/c9ob01839e.
152. Song, J.Y.; Pineault, K.M.; Dones, J.M.; Raines, R.T.; Welik, D.M. Hox genes maintain critical roles in the adult skeleton. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7296–7304, doi:10.1073/pnas.1920860117.
153. Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci. Publ. Protein Soc.* **2009**, *18*, 936–948, doi:10.1002/pro.102.
154. An, B.; Brodsky, B. Collagen binding to OSCAR: The odd couple. *Blood* **2016**, *127*, 521–522, doi:10.1182/blood-2015-12-682476.
155. Yu, Z.; Brodsky, B.; Inouye, M. Dissecting a bacterial collagen domain from *Streptococcus pyogenes*: Sequence and length-dependent variations in triple helix stability and folding. *J. Biol. Chem.* **2011**, *286*, 18960–18968, doi:10.1074/jbc.M110.217422.
156. Perret, S.; Merle, C.; Bernocco, S.; Berland, P.; Garrone, R.; Hulmes, D.J.; Theisen, M.; Ruggiero, F. Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* **2001**, *276*, 43693–43698, doi:10.1074/jbc.M105507200.
157. Olsen, D.R.; Leigh, S.D.; Chang, R.; McMullin, H.; Ong, W.; Tai, E.; Chisholm, G.; Birk, D.E.; Berg, R.A.; Hitzeman, R.A.; et al. Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. *J. Biol. Chem.* **2001**, *276*, 24038–24043.
158. Mohs, A.; Silva, T.; Yoshida, T.; Amin, R.; Lukomski, S.; Inouye, M.; Brodsky, B. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* **2007**, *282*, 29757–29765, doi:10.1074/jbc.M703991200.
159. Yoshizumi, A.; Yu, Z.; Silva, T.; Thiagarajan, G.; Ramshaw, J.A.; Inouye, M.; Brodsky, B. Self-association of streptococcus pyogenes collagen-like constructs into higher order structures. *Protein Sci. Publ. Protein Soc.* **2009**, *18*, 1241–1251, doi:10.1002/pro.134.
160. Yu, Z.; An, B.; Ramshaw, J.A.; Brodsky, B. Bacterial collagen-like proteins that form triple-helical structures. *J. Struct. Biol.* **2014**, *186*, 451–461, doi:10.1016/j.jsb.2014.01.003.
161. An, B.; Abbonante, V.; Yigit, S.; Balduini, A.; Kaplan, D.L.; Brodsky, B. Definition of the native and denatured type II collagen binding site for fibronectin using a recombinant collagen system. *J. Biol. Chem.* **2014**, *289*, 4941–4951, doi:10.1074/jbc.M113.530808.
162. An, B.; Abbonante, V.; Xu, H.; Gavrilidou, D.; Yoshizumi, A.; Bihan, D.; Farndale, R.W.; Kaplan, D.L.; Balduini, A.; Leiting, B.; et al. Recombinant Collagen Engineered to Bind to Discoidin Domain Receptor Functions as a Receptor Inhibitor. *J. Biol. Chem.* **2016**, *291*, 4343–4355, doi:10.1074/jbc.M115.674507.
163. Yu, Z.; Visse, R.; Inouye, M.; Nagase, H.; Brodsky, B. Defining requirements for collagenase cleavage in collagen type III using a bacterial collagen system. *J. Biol. Chem.* **2012**, *287*, 22988–22997, doi:10.1074/jbc.M112.348979.
164. Xu, K.; Nowak, I.; Kirchner, M.; Xu, Y. Recombinant collagen studies link the severe conformational changes induced by osteogenesis imperfecta mutations to the disruption of a set of interchain salt bridges. *J. Biol. Chem.* **2008**, *283*, 34337–34344, doi:10.1074/jbc.M805485200.
165. Cheng, H.; Rashid, S.; Yu, Z.; Yoshizumi, A.; Hwang, E.; Brodsky, B. Location of glycine mutations within a bacterial collagen protein affects degree of disruption of triple-helix folding and conformation. *J. Biol. Chem.* **2011**, *286*, 2041–2046, doi:10.1074/jbc.M110.153965.

166. Xiao, J.; Cheng, H.; Silva, T.; Baum, J.; Brodsky, B. Osteogenesis imperfecta missense mutations in collagen: Structural consequences of a glycine to alanine replacement at a highly charged site. *Biochemistry* **2011**, *50*, 10771–10780, doi:10.1021/bi201476a.
167. Boudko, S.P.; Zientek, K.D.; Vance, J.; Hacker, J.L.; Engel, J.; Bachinger, H.P. The NC2 domain of collagen IX provides chain selection and heterotrimerization. *J. Biol. Chem.* **2010**, *285*, 23721–23731, doi:10.1074/jbc.M110.128405.
168. Boudko, S.P.; Engel, J.; Bachinger, H.P. The crucial role of trimerization domains in collagen folding. *Int J. Biochem Cell Biol.* **2012**, *44*, 21–32, doi:10.1016/j.biocel.2011.09.009.
169. Strawn, R.; Chen, F.; Jeet Haven, P.; Wong, S.; Park-Arias, A.; De Leeuw, M.; Xu, Y. To achieve self-assembled collagen mimetic fibrils using designed peptides. *Biopolymers* **2018**, *109*, e23226, doi:10.1002/bip.23226.
170. Chen, F.; Strawn, R.; Xu, Y. The predominant roles of the sequence periodicity in the self-assembly of collagen-mimetic mini-fibrils. *Protein Sci. Publ. Protein Soc.* **2019**, *28*, 1640–1651, doi:10.1002/pro.3679.