

Review

Revealing Accessibility of Cryptic Protein Binding Sites within the Functional Collagen Fibril

Cody L. Hoop , Jie Zhu, Ana Monica Nunes, David A. Case and Jean Baum *

Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA; cody.hoop@rutgers.edu (C.L.H.); zhujiejz39@gmail.com (J.Z.); anamonicanunes@gmail.com (A.M.N.); david.case@rutgers.edu (D.A.C.)

* Correspondence: jean.baum@rutgers.edu; Tel.: +1-848-445-5284

Academic Editors: Margaret Sunde, Matthew Chapman, Daniel Otzen and Sarah Perrett

Received: 25 August 2017; Accepted: 27 October 2017; Published: 1 November 2017

Abstract: Fibrillar collagens are the most abundant proteins in the extracellular matrix. Not only do they provide structural integrity to all of the connective tissues in the human body, but also their interactions with multiple cell receptors and other matrix molecules are essential to cell functions, such as growth, repair, and cell adhesion. Although specific binding sequences of several receptors have been determined along the collagen monomer, processes by which collagen binding partners recognize their binding sites in the collagen fibril, and the critical driving interactions, are poorly understood. The complex molecular assembly of bundled triple helices within the collagen fibril makes essential ligand binding sites cryptic or hidden from the molecular surface. Yet, critical biological processes that require collagen ligands to have access to interaction sites still occur. In this contribution, we will discuss the molecular packing of the collagen I fibril from the perspective of how collagen ligands access their known binding regions within the fibril, and we will present our analysis of binding site accessibility from the fibril surface. Understanding the basis of these interactions at the atomic level sets the stage for developing drug targets against debilitating collagen diseases and using collagen as drug delivery systems and new biomaterials.

Keywords: collagen; fibrils; ligand; interaction; solvent accessible surface area

1. Introduction

Some of the most fundamental functional fibrils of the human body are formed by collagens, which remarkably play dual structural and biological roles. Of the 28 types of collagens in the extracellular matrix (ECM), seven form higher order fibrils with a complex architecture that function in order to uphold the structural integrity of connective tissues. In addition, their interactions with molecular binding partners, such as cell receptors, enzymes, and other ECM components, are critical for the function and regulation of cellular processes [1–5]. The significance of collagen in cell signaling and in regulation of the ECM is high, along with its role in pathological conditions such as arthritis, cancer, and heart disease [1,2]. Collagens types I–III are the most abundant fibrillar collagens. Collagen I makes up 90% of all collagen, and it is found primarily in bone, dermis, ligaments, and cornea [1,3]. Within these tissues, protein–collagen interactions drive numerous cellular processes, yet how these very specific interactions manage to occur with either collagen monomers or fibril superstructures is not readily evident. Despite all of the incredible advances in understanding different aspects of collagen–ligand interactions through the use of collagen model peptides (CMPs) [6–36], the molecular determinants of the interactions between the full length collagen and its binding partners remains unclear. Here, we review known and theorized structural aspects of collagen I and discuss the binding accessibility of several critical interaction partners based on quantitative solvent accessible surface area (SASA) calculations on the smallest repeating unit (SRU) of the collagen I fibril.

2. The Complex Collagen Architecture

At the foundation of all collagens, there is a collagenous domain composed of repeating Gly-Xaa-Xaa' sequences, where Xaa and Xaa' are often proline (Pro/P) and post-translationally modified hydroxyproline (Hyp/O), respectively. Fibrillar collagens (I–III, V, XI, XXIV, and XXVII) have an uninterrupted GXX' sequence. The linear collagenous sequences are over 1000 amino acids long, termed ' α -chains', and fold into left-handed polyproline type II (PPII) helices. Three α -chains supercoil around a common axis, staggered by one residue, to form the characteristic collagen triple helix. Within the triple helix, the Gly residues are buried on the interior, while the X and X' residues are largely exposed to the solvent and are responsible for the recognition of collagen binding partners [37]. Throughout this review, the triple helix will be considered the monomeric unit. Collagen I is a heterotrimeric triple helix composed of two α 1(I) and one α 2(I) chains. The triple helix presents with a long, rod-like profile, with dimensions of approximately 300 nm long and 1.5 nm in diameter [37–40] (Figure 1A).

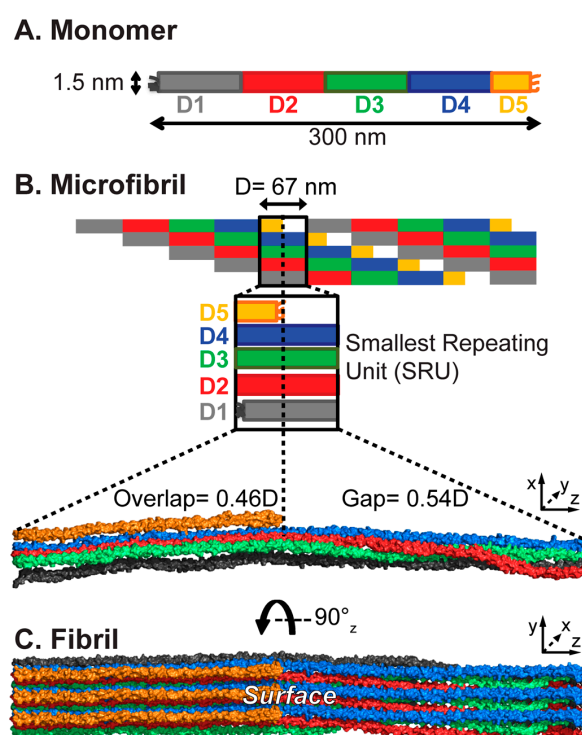


Figure 1. Collagen assembly from triple helical monomers to the collagen fibril architecture. (A) Monomer triple helix: Collagen I is a heterotrimer composed of two α 1 and one α 2 chains that supercoil into a long rod-like triple helix with approximate dimensions of 300×1.5 nm. The monomer consists of 4.46 D-periods, numbered D1–D5; D5 is $0.46D$; (B) Microfibril: Five monomers bundle into pentameric microfibrils, in which one D-period contains the entire collagen sequence from D1 to D5. The short D5 divides the D-period into two distinct regions; the “overlap” region contains segments D1–D5, while the “gap” region is devoid of D5. This creates the characteristic D-banding pattern that can be observed by microscopy. The schematic D-band is rendered in Pymol; (C) Fibril surface: A 90° rotation about the long triple-helical axis reveals the fibril surface. As an example, “surface A” is shown here. For clarity, three D-periods are shown along the y -axis and one along the x -axis.

Collagen is produced as procollagen, containing propeptides in the N and C-termini. Upon secretion into the ECM, the propeptides are cleaved, exposing the telopeptides that are adjacent to the collagenous domain and initiating collagen assembly into large highly organized supermolecular fibril structures [41–43]. These fibrils are not amyloid in conformation, but have differently highly organized hierarchical structures. The buildup of the collagen architecture is presented in Figure 1.

Toward fibril assembly, five monomers bundle into microfibrils. Within the microfibrils, the monomers are staggered axially by 67 nm when hydrated [44]. This consistent 67 nm dimension is called the 'D-period' and sets the periodicity of the fibril assembly. The ≈ 300 nm triple helix spans the length of 4.46 D-periods and is divided into D-segments, D1–5 (distinguished by color in Figure 1), with D5 being a short 0.46D [45]. Since the monomers of the microfibril are offset by one D-period, the microfibril can be cut into units one D-period long, which contain one copy of each D-segment, D1–5, each from a different monomer (Figure 1B). This unit is the SRU, and the smallest building block of the collagen fibril. Inter-microfibrillar crosslinking between D1 and D5 stabilizes the fibril assembly [46]. Due to D5 being less than half of the length of the other D-segments, the SRU is divided into two distinct sections that are defined by their molecular densities: the 'overlap' region, which contains D5, and the 'gap' region, devoid of D5 (Figure 1B). The alternation of these regions creates the characteristic D-banding pattern of dark and light bands observed in microscopy, resulting from differential electron density (in electron microscopy) or heights (in atomic force microscopy) between the two regions. It has been established through several studies that the monomers are not stacked laterally in the microfibril, but are arranged into a quasihexagonal lattice [47–53]. This three-dimensional packing buries some parts of the monomers within the microfibril. From their fiber diffraction study of an intact collagen I fiber from rat tail tendon, Orgel et al. found that while in the overlap region the quasihexagonal packing is largely constant throughout its length, in the gap region, the triple helices take on a right handed twist about the fibril axis, which determines the exposure of particular domains to the accessible surface (Figure 1C) [53]. The fibril SRU contains all of these structural features, which are integrated into the larger fibril assembly, and the entire collagen I sequence for locating the binding domains within the D-period. Thus, the fibril SRU provides the simplest model to assess the accessibility of interaction domains within the collagen fibril.

3. Collagen Interactions and the Fibril Surface

3.1. Identification of Binding Sequences

The complex assembly of monomers into microfibrils and fibrils results in one face being surface exposed and available for binding while other segments are buried within the fibril core [53,54]. Fibrillar collagens have numerous binding partners: interacting with ECM components, binding to cellular receptors, and interacting with enzymes for chemical modification and matrix metalloproteinases (MMPs) during degradation [4,55–59]. Despite its presentation as a long rod with a repeating GXX' sequence, collagen's binding partners recognize specific sequences and motifs across its more than 1000 residue span, and some specifically bind to particular structural conformations. Much progress has been made in determining collagen recognition sequences of cell receptors and interaction partners through the elegant use of triple helical peptides, recombinant and bacterial collagen, and sequence-dependent adhesion assays [5,60–71]. The collagen peptide toolkits developed in the Farndale laboratory are libraries of overlapping peptides that span the entire collagenous sequences of homotrimeric collagens II and III. Each peptide contains 27 residues from natural collagen sequences, including nine unique residues, flanked on each termini by five GPP repeats ((GPP)₅) for stabilization of the triple helix [62]. Although, collagen I is a heterotrimer of two $\alpha 1(I)$ chains and one $\alpha 2(I)$ chain, the collagenous domain of the collagen I $\alpha 1(I)$ chain contains 74.5% and 63.0% sequence identity to the collagenous domains of collagens II and III, respectively. Thus, these peptides provide excellent tools for identifying specific sequence motifs recognized by collagen-binding partners and comparing their relative affinities in the triple helical context.

By evaluating the binding of receptors to the collagen toolkit peptides, specific binding sequences have been determined for collagen binding integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$ [72–76], discoidin domain receptors (DDR) 1 and 2 [64,65,77], von Willebrand factor (VWF) [78,79], glycoprotein VI (GPVI) [80,81], leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) [82,83], osteoclast-associated receptor (OSCAR) [84,85], secreted protein acidic and rich in cysteine (SPARC) [86], *Yersinia* adhesion

A (YadA) [87], procollagen C-proteinases enhancer (PCPE) [43], MMP1 [88], MMP13 [89], and fibromodulin [90]. Known sequences of collagen binding partners have been previously reviewed and mapped onto fibrillar collagen I and III sequences [5,56,57,91]. Due to challenges of acquiring intact monomeric collagen I, only a few studies exist on protein binding to full collagen I monomers, using microscopy techniques to detect ligand binding. The binding of heparin to procollagen type I was localized by rotary shadowing and electron microscopy (EM) [92]. Binding sites of SPARC to procollagens of types I, II, and III have been visualized by rotary shadowing EM [86] and atomic force microscopy (AFM) [93], and to the collagen I monomer lacking propeptides by AFM [93]. These experiments probe primarily protein interactions with collagen when it is in a long triple-helical rod, in which X and X' residues are largely exposed and accessible for binding. The relative affinities of the recognition sequences to the collagen triple helix may not be translatable to the full fibril architecture. In the three-dimensional packing of the fibril, some of the interaction domains become "cryptic" or buried from the accessible surface [4,54,58,94,95], yet cellular functions are still carried out, and it remains difficult to understand how molecular partners recognize and interact with their specific binding motifs when packed in the three-dimensional collagen fibril.

3.2. Fibril Surface Identity

The Orgel laboratory reported the identity of the solvent-exposed surface of the collagen I fibril based on fitting the X-ray fiber diffraction model of the collagen I microfibril from rat tail tendon [53] into the profile of the collagen I fibril that is observed by scanning electron microscopy (SEM) and AFM [96–98]. In the microscopy images, the collagen fibril follows a corrugated pattern with up-ward and downward slopes, the peak of which Orgel et al. suggests can be explained by the C-terminal telopeptide; furthermore they suggest that the C-telopeptide is exposed on the surface with the N-terminus buried within the fibril [98]. This would place segments D5 and D4 on the exterior of the fibril. Segment D5 contains a (GPO)₅ domain adjacent to the C-telopeptide. This imino-acid rich GPO repeat is expected to form the most stable triple helix domain, and may have a protective role to less stable sequences within the fibril core [54]. However, this surface leaves buried and inaccessible high affinity recognition sequences for collagen-binding integrins and the single interaction site for platelet receptor VWF, which is shared with DDRs and SPARC. Herr and Farndale have proposed an alternative orientation with the opposite face exposed that would make these sites more accessible [58]. This raises an interesting question: if critical interaction domains are buried within the collagen fibril interior, how do biological processes that depend on these interactions occur? In order gain a perspective on the residue-specific exposure of identified binding motifs within the collagen fibril, we used computational tools to build a collagen fibril model and calculate SASA from the fibril surface of each residue in our model.

4. Computational Studies of the Collagen Fibril

Streeter and de Leeuw first built an all-atom model of the collagen fibril based on the atomic structure and periodic nature of the fibril, as reported by Orgel (Protein Data Bank (PDB) entry: 3HR2) [53], and performed molecular dynamics (MD) simulations on the collagen model with periodic boundaries to replicate the supermolecular arrangement [99]. The simulation was validated by comparison to experimental observations and theories of the collagen fibril structure. Streeter and de Leeuw also studied the inter-protein interactions within a collagen fibril [100]. Simulations of collagen fibrils and monomers were compared to study the atomic level interactions upon fibrillogenesis. They also compared the process of collagen fibrillogenesis with the process of protein folding and concluded that they both act to shield hydrophobic side-chains from solvent, and they differ in the percentages of major intra-protein hydrogen bond types (e.g., backbone–backbone, backbone–side-chain, and side-chain–side-chain) [100].

Varma and Schieber also performed MD simulations on collagen models with periodic boundaries, and evaluated the effect of several factors on D-band shrinkage [101]. They found that the

large shrinkage that was observed in earlier simulations was not eliminated by optimizations of temperature/pressure coupling algorithms, salt concentration or hydration level, or the cross-linking of monomers, and concluded that the shrinkage was a force-field artifact. A new force field Amber99sb-ildn was found to produce a small D-band shrinkage of <3% [101].

Here, we aimed to create a measure of accessibility for known specific receptor and protein binding sequences within the collagen I fibril. Previously, Orgel et al. created a qualitative ranking of accessibility of collagen binding sites based on their distance from and relative occlusion by the fibril surface [54] and calculated SASA within the fibril to evaluate accessibility of the MMP cleavage site [98]. We have now used a modified SASA method to characterize the ability of a binding partner to access its specific binding site within the collagen fibril SRU that is nearest to the fibril surface.

4.1. Building the Atomic Model of the Collagen Fibril

We have constructed an atomic model of a collagen fibril unit cell based on the low-resolution fiber diffraction structure (PDB entry 3HR2 [53], see Figure 1 for fibril architecture). The fiber diffraction provided structural information about the C α atoms, giving the overall shape and orientation of the collagen monomers. From this structure, we built a complete all-atom monomer model by adding side-chain and backbone atoms, plus water molecules to fill the repeating unit cell, following work by the Leeuw group [99,100]. This was equilibrated by MD simulations with loose tethering of the C α atoms to their positions in the 3HR2 structure [53]. Five collagen triple helix monomers from the equilibrated model were duplicated and packed by D-period to create a microfibril and then cut along the z-axis to the length of one D-period to generate one SRU. A finite, non-periodic fibril SRU model was then built with one complete D-period along the z-axis, and three SRUs along the x- (depth) and y- (height) axes (Figure 2A). Unlike periodic models, this fibril model has a surface on either side, surface layer A (with D5 and D4 on the surface) and B (with D1 on the surface), where proteins or other effectors may bind, and an inner layer that is completely shielded from a fibril surface.

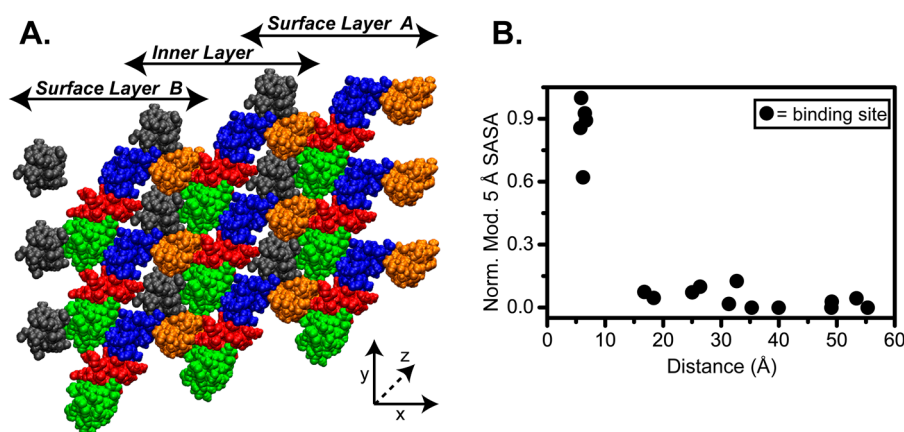


Figure 2. (A) A representative cross-section showing the 3×3 SRU matrix used in the solvent accessible surface area (SASA) calculations. The model contains three SRUs in each the x- and y-dimensions, and the z-dimension extends to one D-period in length. Triple helical monomers are color-coded by D-segments as in Figure 1. Double-headed arrows indicate the three distinct layers in the x-dimension. (B) Correlation of SASA with binding site distance. The modified 5 Å SASA of 17 known binding sites with respect to the orientation proposed by Orgel et al. (“surface A”) is plotted vs. distance of the site from the surface.

4.2. SASA of Collagen I Fibril Reveals Exposure of Partner Binding Sites from the Fibril Surface

In order to assess the site-specific accessibility at the collagen fibril surface, we performed SASA calculations with different probe sizes on all residues of the 3×3 fibril SRU model (Figure 2A) using

the Molecular Surface (MS) program [102,103]. In these calculations, the SASA gives the surface area of an individual residue that is accessible to a probe of a certain radius. We initially calculated SASA with a 1.4 Å probe, the radius of a water molecule, in order to measure the water accessibility within the fibril. Remarkably, there is uniformly high SASA, even into the D-segments of the inner layer, to water molecules throughout the 3×3 fibril SRU model. This indicates that despite appearing densely packed, there is significant space for water within the triple helices, even within the core of the collagen fibril.

In an attempt to model access of ligands to the binding sites in the fibril, we determined the SASA of the different D-segments, using a larger spherical probe with a 5 Å radius, termed the 5 Å SASA. As we are interested in determining binding site exposure relative to the surface A or surface B (see Figure 2A), we used a simple approach in which we subtract the 5 Å SASA of SRUs of the inner layer from the 5 Å SASA of the surface layer. This subtraction gives accessibility only from the fibril surface, when compared to the fibril core. This allows us to probe the accessibility of the residues that are exposed to the fibril surface. We applied the calculation to each residue of the middle SRUs along the *y*-axis of the 3×3 fibril SRU model (Figure 2A) since they are surrounded by all of the neighbors that would be present in the full fibril in order to obtain a baseline for the 5 Å SASA. As the triple helix is composed of three chains, we used the average 5 Å SASA of the three chains at each position for the calculation.

We have plotted the modified 5 Å SASA of 17 known binding sites in collagen I with respect to their distance from the surface A, which exposes D4 and D5 (Figure 2B). We find that of these 17 sites, only five show a high modified 5 Å SASA; the other sites appear to have minimal modified 5 Å SASA suggesting that the majority of the collagen binding sites are not accessible from surface A. This underscores that although access to these binding sites by binding partners is required for a multitude of biological processes, their exposure within the context of the collagen fibril is minimal.

We analyzed the modified 5 Å SASA in residue-specific detail throughout the collagen fibril. For visualization purposes, we present SASA from the perspectives of both surfaces A and B. The residue-specific modified 5 Å SASA for each D-segment relative to a particular surface is plotted in Figure 3A,B. Our SASA analysis reveals non-uniform SASA between the D-segments, as expected due to the quasihexagonal arrangement of the microfibril. In particular, high SASA is observed in the D-segment at the assumed surface (D4 and D5 in surface A and D1 in surface B), and the accessible surface area is significantly lower in D-segments deeper into the fibril. Moreover, the SASA is not uniform within each D-segment along the length of the D-period due to the inherent twisting of monomers within the fibril (see schematic at the top of Figure 3A,B). For example, from surface A, D2 has moderate accessibility in the overlap region and in the N-terminal half of the gap region, but no accessibility in the C-terminal end of the gap region (Figure 3A). The twisting out of the D3 at this end of the gap makes it lowly accessible, and it is non-accessible in the remainder of the D-period. Conversely, surface B gives moderate accessibility to the overlap region and the N-terminal half of the gap region of D3, and moderate accessibility of D2 in the overlap region (Figure 3B). In this orientation, the twisting out of D2 at the C-terminal end of the gap region gives it higher accessibility than even D1, which is at the surface and has the highest exposure for most of the D-period. The variability of SASA both between D-segments and along the D-period makes the exposure of binding sites complex.

The accessibility from the fibril surface is highly determinant of interactions with molecular partners. We therefore analyzed the accessibility of thirteen collagen-binding ligands from both surfaces A and B (Figure 3). The SASA highlights the highly accessible binding sites near the surface in either orientation. Quantification of the residue-specific SASA of the 3×3 collagen fibril SRU model in both possible orientations provides a measure of accessibility of interaction domains when either face is exposed.

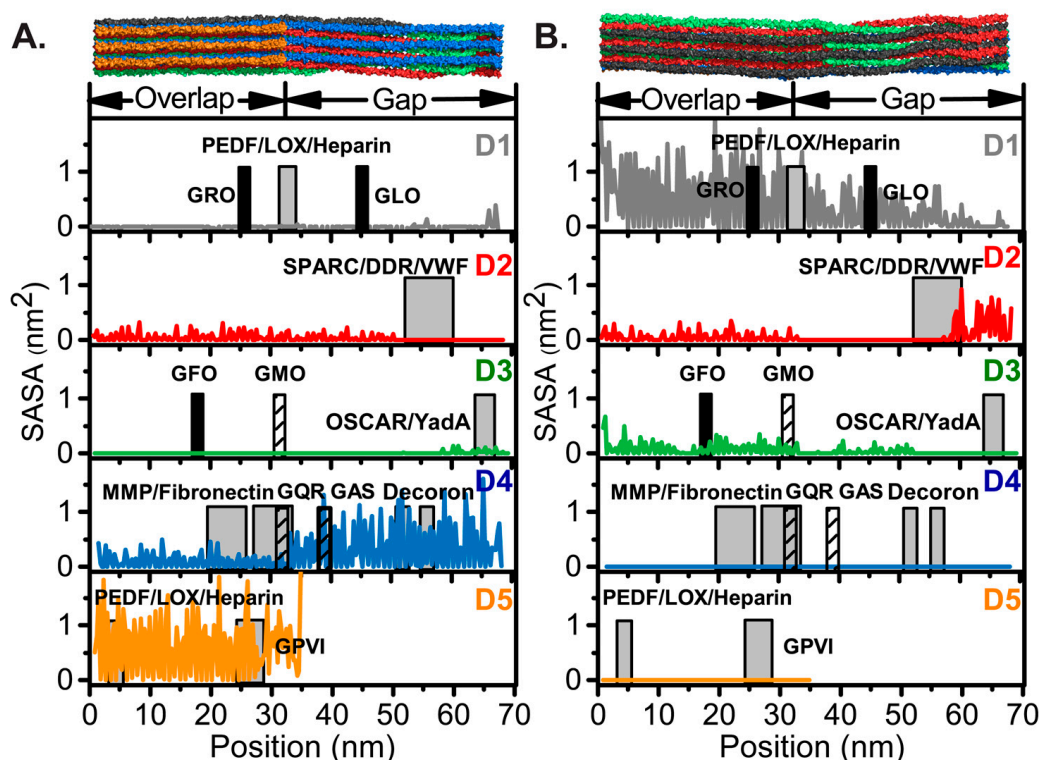


Figure 3. Residue-specific modified 5 Å SASA calculations with respect to surfaces (A) A [98] and (B) B [58]. The D-segments, D1–D5, are indicated and colored as in Figures 1 and 2. The locations of collagen ligands are shown by gray boxes with heparin [104], PEDF [105] and LOX [57] in D1 and D5; SPARC [86], DDRs [64,65] and VWF [78] in D2; OSCAR [84,85] and YadA [87] in D3; MMPs [68], fibronectin [70], and the decorin core protein, decoron [106], in D4; and GPVI [80] in D5. Integrin high and low affinity motifs are represented by black and hashed boxes, respectively, and labeled by the first three residues of the indicated six-residue binding motif. The width of the boxes corresponds to the length of the identified recognition sequence. Abbreviations: PEDF, pigment epithelium-derived factor; LOX, lysyl oxidase; SPARC, secreted protein acidic and rich in cysteine; DDRs, discoidin domain receptors; VWF, von Willebrand factor; OSCAR, osteoclast-associated immunoglobulin-like receptor; Yad A, *Yersinia* adhesin A; MMPs, matrix metalloproteinases; GPVI, glycoprotein VI.

Presentation of the C-terminus to the surface provides direct access to essential fibril interaction sites, including lysyl oxidase (LOX) [57], GPVI [80,107], and decoron, the core protein of decorin [106]. LOX interaction sites are on both D1 and D5 for formation of intermolecular cross-links that regulate the collagen fibril assembly. LOX deaminates lysines and hydroxylysines in preparation of cross-link formation between triple helices to stabilize collagen fibrils [108]. GPVI is a glycoprotein on platelet membranes that interacts with collagen to activate the platelets and induce platelet aggregation [107]. GPVI recognizes and binds to short (GPO)_n tandem repeats on collagen. The only GPO repeat motif longer than (GPO)₂ in the collagen I sequence is in D5 at the very C-terminus of the collagenous domain and contains (GPO)₅ in the α1(I) chain and (GPO)₄ in the α2(I) chain. Decorin is a proteoglycan that interacts with collagen I fibrils to stabilize the suprafibrillar assembly [106,109]. The core protein of decorin, decoron, recognizes two binding motifs in the gap region of D4, KXGDRGE, and AKGDRGE [106,109]. An X-ray crystal structure of decoron has been solved and docked onto the collagen I fibril when D4 is exposed [106]. Access of these sites in the mature fibril of collagen would allow these proteins to carry out their cellular functions.

Furthermore, in this orientation there is a potential means of collagen fibril cleavage by MMPs. MMPs are the only enzymes that are capable of cleaving collagen molecules. Degradation of collagen fibrils is essential for fibril turnover, tissue repair, and development. MMP1, MMP8,

MMP13, and membrane type-MMP1 (MT-MMP1) cleave collagen α -chains at a single, specific Gly-Ile/Leu cleavage site, dissecting the triple helix into three-quarter and one-quarter length fragments. MMPs have been proposed to cleave one α -chain at a time, starting specifically with the α 2 chain in collagen I [110,111]. The MMP cleavage site is located in the overlap region on segment D4, and therefore is accessible upon displacement of the C-telopeptide. In an *in silico* study, Perumal et al. proposed how collagenolysis can occur in the collagen fibril. From the C-terminal fibril surface, the α 2 chain is oriented toward the outside of the fibril and has a greater dissociation from the center of the triple helix, making it more vulnerable to MMP cleavage [98]. Moreover, the protection of the site of collagenolysis by the C-telopeptide increases MMP specificity, as it will only be able to cleave when the C-telopeptide is displaced [98,112]. Degradation of the collagen fibril is known to occur, and thus access of MMPs to their cleavage site is essential.

On the other hand, surface B favors access to high affinity integrin binding motifs and platelet receptor recognition sites. In this orientation, known platelet interaction sites, including those for integrin α 2 β 1 and VWF, are more accessible to their binding partners. Both integrin α 2 β 1 and VWF have been shown to be critical receptors for the platelet adhesion to collagen [55,113]. Many specific binding sequences for platelet and other cell receptors have been determined from adhesion assays to fully exposed triple helical collagen II and III peptides. Moreover, Jokinen et al. showed that through solid-phase cell adhesion assays and immunoelectron microscopy, that integrin α 2 β 1 does in fact adhere to mature collagen I fibrils from bovine skin and facilitates cellular responses, such as cell spreading and formation of cellular projections [114]. This suggests that the integrin α 2I domain must have access to at least one of its recognition domains on the collagen I fibril. In which sequence the integrin α 2I actually binds in collagen I fibrils has yet to be determined.

In Figure 3, black and hashed boxes indicate integrin high and moderate affinity binding motifs, respectively. High affinity integrin α 2 β 1 binding sequences, GROGER and GLOGER [58,62,74,76,115], are on D1 and would be directly on the surface of the collagen I fibril available for α 2I domain binding. High affinity integrin binding site GFOGER and moderate affinity binding site GMOGER, on the N-terminal half of D3 also become more accessible with surface B exposed, allowing accessibility to all high affinity integrin α 2I binding sites. Furthermore, only in this orientation is it possible to gain accessibility to the single VWF binding motif, GVMGFO, on the C-terminal end of D2, which is overlapped with the binding sequences for SPARC and DDRs [64,65,78,86].

Although surface B provides more access to high affinity integrin binding motifs and the GVMGFO motif shared by VWF, DDRs, and SPARC binding partners, this surface does not allow access to the MMP cleavage and binding sites or GPVI and decoron interaction motifs. Yet, integrin binding from surface A is not ruled out, as moderate affinity binding motifs, GQRGER and GASGER, are accessible from this surface, especially if the C-telopeptide is dynamic. The dynamics of the C-telopeptide and proteolysis by MMPs might provide access to functional domains otherwise buried within the fibril, and would therefore support the surface exposure of the C-terminus, as proposed by Perumal et al. [98], rather than the N-terminus. Considering the accessibility of critical fibril binding sites, as assessed by our SASA analysis, and the structural features observed by microscopy, we propose that surface A, that proposed by the Orgel laboratory, provides greater access to critical binding motifs.

5. Perspective

5.1. The Collagen Architecture as a “Smart Fibril”: Regulating Accessibility of Partner Binding Sites

The surface proposed by Orgel et al., placing the C-telopeptide on the fibril surface seems to be more accommodating of essential molecular interactions within the collagen fibril. The placement of the MMP cleavage site beneath the C-terminal D5 allows for the dynamics and motions of D5 to act as a gatekeeper, regulating the accessibility of the MMP binding and cleavage site [94]. MMP cleavage of the collagen fibril is critical for fibril turnover, tissue repair, and development. From the opposite

perspective, this site would be buried deep within the fibril and be inaccessible for collagenolysis. Once degraded by MMPs, new interaction sites would become available in the cleaved collagen.

With the C-telopeptide exposed, high affinity integrin binding sites are not accessible. However, the relative binding affinities for integrins were determined in fully exposed, triple helical peptides, and may not be translatable to the context of the complex assembly of the collagen fibril. Constant strong binding of integrin to the collagen fibril may not be optimal. Integrins aid in cell adhesion and movement along collagen, and if bound too tightly, cell movement may not be possible. It has been hypothesized that this tight regulation is biologically important as it limits platelet aggregation under “healthy” conditions but promotes firm platelet adhesion upon cellular activation caused by disruption of blood vessels [55,58,116,117]. Using a ligand accessibility ranking system, Orgel et al. predicted integrin binding sites to be “cryptic”, buried within the fibril assembly, and only partially accessible in the mature fibril [54]. They proposed that cell-collagen interactions are regulated by binding motifs being exposed only upon structural reorganization or degradation of the fibril [54]. Although the high affinity integrin-binding motif, GFOGER, on D3 is buried deep within the fibril, it may be accessible upon limited proteolysis or during fibril assembly [54,94,98].

Of course, fibrils of different collagen types may have very different microfibrillar arrangements and binding interfaces. Relative affinities of integrin binding motifs and other interaction sequences of collagen binding partners have been largely determined via triple helical peptide binding studies on collagen II and III sequences. While these collagens share a high percentage of sequence identity, the similarities in arrangement into fibrillar structures have yet to be determined. A recent study by Woltersdorf et al. reported on the adhesion of collagen-binding integrins in chondrocytes to a 8:1:1 mixture of collagens II, IX, and XI mimicking cartilage collagen [118]. They found that the chondrocytes adhered to triple helical collagen, but were bound only weakly to collagen in the fibrillar form. By immunoelectron microscopy, they observed recombinant integrin I-domains binding not directly to the collagen fibrils, but to residual non-fibrillar material peripheral to the fibrils [118]. While a previous study had observed the binding of integrins to collagen I fibrils, this recent study suggests that in cartilage collagen fibrils composed of collagens II, IX, and XI, integrins do not have access to their recognition sites.

5.2. Regulation of Protein Binding through Internal Dynamics of the Collagen Fibril

The SASA calculation assumes only a snapshot of a configuration of the fibril in time. The fibril and the monomers within are not static entities, but are constantly undergoing internal motions. These fluctuations could provide a key to understanding how binding partners are able to access their seemingly buried binding sites. For example, from the surface proposed by Orgel, the MMP cleavage site is buried beneath the C-telopeptide. Although from the SASA, this domain is only moderately accessible, backbone motions of the C-telopeptide may create an opening for an MMP to proteolyze at the cleavage site. Likewise, from the orientation, an active binding site on D2 for SPARC, DDRs, and VWF are hidden by the D1. However, the backbone dynamics of D1 in this region may reveal a path of access for these binding receptors. Proteins are dynamic in nature, and therefore, their motions may have large impacts on their capabilities to function. Future studies of internal dynamics of the collagen fibril and ligand binding will deliver new insight about how interaction partners access their specific recognition motifs and the functionality of fundamental collagen fibrils.

Acknowledgments: This work was supported by NIH grant GM 45302 to J.B., American Heart Association (AHA) postdoctoral fellowship 13POST16550007 to A.M.N., AHA postdoctoral fellowship 17POST33410326 to C.L.H., and NIH grant GM 45811 to D.A.C.

Author Contributions: C.L.H., J.Z., A.M.N., D.A.C., and J.B. contributed to the writing of the paper and designing of the figures. J.Z. and D.A.C. performed and analyzed the SASA calculations.

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

References

1. Shoulders, M.D.; Raines, R.T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929–958. [[CrossRef](#)] [[PubMed](#)]
2. Heino, J. The collagen family members as cell adhesion proteins. *Bioessays* **2007**, *29*, 1001–1010. [[CrossRef](#)] [[PubMed](#)]
3. Kadler, K.E.; Baldock, C.; Bella, J.; Boot-Handford, R.P. Collagens at a glance. *J. Cell Sci.* **2007**, *120*, 1955–1958. [[CrossRef](#)] [[PubMed](#)]
4. 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. [[CrossRef](#)] [[PubMed](#)]
5. Leitinger, B. Transmembrane collagen receptors. *Annu. Rev. Cell Dev. Biol.* **2011**, *27*, 265–290. [[CrossRef](#)] [[PubMed](#)]
6. Emsley, J.; Knight, C.G.; Farndale, R.W.; Barnes, M.J. Structure of the integrin $\alpha 2\beta 1$ -binding collagen peptide. *J. Mol. Biol.* **2004**, *335*, 1019–1028. [[CrossRef](#)] [[PubMed](#)]
7. 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. [[CrossRef](#)] [[PubMed](#)]
8. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal structure of a collagen-like polypeptide with repeating sequence Pro-Hyp-Gly at 1.4 Å resolution: Implications for collagen hydration. *Biopolymers* **2000**, *56*, 8–13. [[CrossRef](#)]
9. Fallas, J.A.; Gauba, V.; Hartgerink, J.D. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J. Biol. Chem.* **2009**, *284*, 26851–26859. [[CrossRef](#)] [[PubMed](#)]
10. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)(10)](3). *Protein Sci.* **2002**, *11*, 262–270. [[CrossRef](#)] [[PubMed](#)]
11. Kramer, R.Z.; Bella, J.; Brodsky, B.; Berman, H.M. The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol.* **2001**, *311*, 131–147. [[CrossRef](#)] [[PubMed](#)]
12. Kramer, R.Z.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H.M. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **1999**, *6*, 454–457. [[PubMed](#)]
13. Kramer, R.Z.; Venugopal, M.G.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H.M. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J. Mol. Biol.* **2000**, *301*, 1191–1205. [[CrossRef](#)] [[PubMed](#)]
14. Kramer, R.Z.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H.M. X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* **1998**, *280*, 623–638. [[CrossRef](#)] [[PubMed](#)]
15. Xiao, J.; Madhan, B.; Li, Y.; Brodsky, B.; Baum, J. Osteogenesis imperfecta model peptides: Incorporation of residues replacing Gly within a triple helix achieved by renucleation and local flexibility. *Biophys. J.* **2011**, *101*, 449–458. [[CrossRef](#)] [[PubMed](#)]
16. Xiao, J.; Cheng, H.; Silva, T.; Baum, J.; Brodsky, B.M. Osteogenesis imperfecta missense mutations in collagen: Structural consequences of a glycine to alanine replacement at a highly charged site. *Biochemistry* **2011**, *50*, 10771–10780. [[CrossRef](#)] [[PubMed](#)]
17. Li, Y.; Brodsky, B.; Baum, J. NMR conformational and dynamic consequences of a Gly to ser substitution in an osteogenesis imperfecta collagen model peptide. *J. Biol. Chem.* **2009**, *284*, 20660–20667. [[CrossRef](#)] [[PubMed](#)]
18. Xiao, J.; Addabbo, R.M.; Lauer, J.L.; Fields, G.B.; Baum, J. Local conformation and dynamics of isoleucine in the collagenase cleavage site provide a recognition signal for matrix metalloproteinases. *J. Biol. Chem.* **2010**, *285*, 34181–34190. [[CrossRef](#)] [[PubMed](#)]
19. Xiao, J.; Baum, J. Structural insights from ^{15}N relaxation data for an anisotropic collagen peptide. *J. Am. Chem. Soc.* **2009**, *131*, 18194–18195. [[CrossRef](#)] [[PubMed](#)]
20. Thiagarajan, G.; Li, Y.; Mohs, A.; Strafaci, C.; Popiel, M.; Baum, J.; Brodsky, B. Common interruptions in the repeating tripeptide sequence of non-fibrillar collagens: Sequence analysis and structural studies on triple-helix peptide models. *J. Mol. Biol.* **2008**, *376*, 736–748. [[CrossRef](#)] [[PubMed](#)]

21. Madhan, B.; Xiao, J.; Thiagarajan, G.; Baum, J.; Brodsky, B. NMR monitoring of chain-specific stability in heterotrimeric collagen peptides. *J. Am. Chem. Soc.* **2008**, *130*, 13520–13521. [[CrossRef](#)] [[PubMed](#)]
22. 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. [[CrossRef](#)] [[PubMed](#)]
23. Mohs, A.; Popiel, M.; Li, Y.; Baum, J.; Brodsky, B. Conformational features of a natural break in the type IV collagen Gly-X-Y repeat. *J. Biol. Chem.* **2006**, *281*, 17197–17202. [[CrossRef](#)] [[PubMed](#)]
24. 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. [[CrossRef](#)] [[PubMed](#)]
25. 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. [[CrossRef](#)] [[PubMed](#)]
26. Bhate, M.; Wang, X.; Baum, J.; Brodsky, B. Folding and conformational consequences of glycine to alanine replacements at different positions in a collagen model peptide. *Biochemistry* **2002**, *41*, 6539–6547. [[CrossRef](#)] [[PubMed](#)]
27. Buevich, A.V.; Baum, J. Residue-specific real-time NMR diffusion experiments define the association states of proteins during folding. *J. Am. Chem. Soc.* **2002**, *124*, 7156–7162. [[CrossRef](#)] [[PubMed](#)]
28. Buevich, A.; Baum, J. Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2001**, *356*, 159–168. [[CrossRef](#)] [[PubMed](#)]
29. Buevich, A.V.; Shinde, U.P.; Inouye, M.; Baum, J. Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J. Biomol. NMR* **2001**, *20*, 233–249. [[CrossRef](#)] [[PubMed](#)]
30. 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. [[CrossRef](#)] [[PubMed](#)]
31. Long, C.G.; Braswell, E.; Zhu, D.; Apigo, J.; Baum, J.; Brodsky, B. Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. *Biochemistry* **1993**, *32*, 11688–11695. [[CrossRef](#)] [[PubMed](#)]
32. 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. [[CrossRef](#)]
33. 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. [[CrossRef](#)] [[PubMed](#)]
34. 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. [[CrossRef](#)] [[PubMed](#)]
35. Gauba, V.; Hartgerink, J.D. Synthetic collagen heterotrimers: Structural mimics of wild-type and mutant collagen type I. *J. Am. Chem. Soc.* **2008**, *130*, 7509–7515. [[CrossRef](#)] [[PubMed](#)]
36. 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. [[CrossRef](#)] [[PubMed](#)]
37. Ramachandran, G.N.; Kartha, G. Structure of collagen. *Nature* **1955**, *176*, 593–595. [[CrossRef](#)] [[PubMed](#)]
38. Ramachandran, G.N. *Structure of Collagen at the Molecular Level. Treatise on Collagen*; Ramachandran, G.N., Ed.; Academic Press, Inc.: New York, NY, USA, 1967; Volume 1, Chapter 3.
39. Rich, A.; Crick, F.H. The molecular structure of collagen. *J. Mol. Biol.* **1961**, *3*, 483. [[CrossRef](#)]
40. Rich, A.; Crick, F.H. The structure of collagen. *Nature* **1955**, *176*, 915–916. [[CrossRef](#)] [[PubMed](#)]
41. Colige, A.; Vandenberghe, I.; Thiry, M.; Lambert, C.A.; Van Beeumen, J.; Li, S.W.; Prockop, D.J.; Lapiere, C.M.; Nusgens, B.V. Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying high homology with ADAMTS-2 and ADAMTS-3. *J. Biol. Chem.* **2002**, *277*, 5756–5766. [[CrossRef](#)] [[PubMed](#)]
42. Vadon-Le Goff, S.; Hulmes, D.J.; Moali, C. BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. *Matrix Biol.* **2015**, *44–46*, 14–23. [[CrossRef](#)] [[PubMed](#)]
43. Vadon-Le Goff, S.; Kronenberg, D.; Bourhis, J.M.; Bijakowski, C.; Raynal, N.; Ruggiero, F.; Farndale, R.W.; Stocker, W.; Hulmes, D.J.; Moali, C. Procollagen C-proteinase enhancer stimulates procollagen processing by binding to the C-propeptide region only. *J. Biol. Chem.* **2011**, *286*, 38932–38938. [[CrossRef](#)] [[PubMed](#)]
44. Wess, T.J.; Orgel, J.P. Changes in collagen structure: Drying, dehydrothermal treatment and relation to long term deterioration. *Thermochim. Acta* **2000**, *365*, 119–128. [[CrossRef](#)]

45. Petruska, J.A.; Hodge, A.J. A subunit model for the tropocollagen macromolecule. *Proc. Natl. Acad. Sci. USA* **1964**, *51*, 871–876. [[CrossRef](#)] [[PubMed](#)]
46. Siegel, R.C. Biosynthesis of collagen crosslinks: Increased activity of purified lysyl oxidase with reconstituted collagen fibrils. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4826–4830. [[CrossRef](#)] [[PubMed](#)]
47. Hulmes, D.J.S.; Miller, A. Quasi-hexagonal molecular packing in collagen fibrils. *Nature* **1979**, *282*, 878–880. [[CrossRef](#)] [[PubMed](#)]
48. Fraser, R.D.; MacRae, T.P.; Miller, A.; Suzuki, E. Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **1983**, *167*, 497–521. [[CrossRef](#)]
49. Fraser, R.D.; MacRae, T.P.; Miller, A. Molecular packing in type I collagen fibrils. *J. Mol. Biol.* **1987**, *193*, 115–125. [[CrossRef](#)]
50. Wess, T.J.; Hammersley, A.; Wess, L.; Miller, A. Type I collagen packing, conformation of the triclinic unit cell. *J. Mol. Biol.* **1995**, *248*, 487–493. [[CrossRef](#)]
51. Wess, T.J.; Hammersley, A.P.; Wess, L.; Miller, A. Molecular packing of type I collagen in tendon. *J. Mol. Biol.* **1998**, *275*, 255–267. [[CrossRef](#)] [[PubMed](#)]
52. Orgel, J.P.; Miller, A.; Irving, T.C.; Fischetti, R.F.; Hammersley, A.P.; Wess, T.J. The in situ supermolecular structure of type I collagen. *Structure* **2001**, *9*, 1061–1069. [[CrossRef](#)]
53. 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. [[CrossRef](#)] [[PubMed](#)]
54. 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. [[CrossRef](#)] [[PubMed](#)]
55. Farndale, R.W.; Siljander, P.R.; Onley, D.J.; Sundaresan, P.; Knight, C.G.; Barnes, M.J. Collagen–platelet interactions: Recognition and signalling. *Biochem. Soc. Symp.* **2003**, *70*, 81–94. [[CrossRef](#)]
56. An, B.; Lin, Y.-S.; Brodsky, B. Collagen interactions: Drug design and delivery. *Adv. Drug Deliv. Rev.* **2015**, *97*, 69. [[CrossRef](#)] [[PubMed](#)]
57. 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. [[CrossRef](#)] [[PubMed](#)]
58. Herr, A.B.; Farndale, R.W. Structural insights into the interactions between platelet receptors and fibrillar collagen. *J. Biol. Chem.* **2009**, *284*, 19781–19785. [[CrossRef](#)] [[PubMed](#)]
59. Humphries, S.M.; Lu, Y.; Canty, E.G.; Kadler, K.E. Active negative control of collagen fibrillogenesis in vivo. Intracellular cleavage of the type I procollagen propeptides in tendon fibroblasts without intracellular fibrils. *J. Biol. Chem.* **2008**, *283*, 12129–12135. [[CrossRef](#)] [[PubMed](#)]
60. 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 $\alpha_2\beta_1$ -binding site containing an essential ger sequence. *J. Biol. Chem.* **1998**, *273*, 33287–33294. [[CrossRef](#)] [[PubMed](#)]
61. 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. [[CrossRef](#)] [[PubMed](#)]
62. Farndale, R.W.; Lisman, T.; Bihan, D.; Hamaia, S.; Smerling, C.S.; Pugh, N.; Konitsiotis, A.; Leitinger, B.; de Groot, P.G.; Jarvis, G.E.; et al. Cell-collagen interactions: The use of peptide toolkits to investigate collagen–receptor interactions. *Biochem. Soc. Trans.* **2008**, *36*, 241–250. [[CrossRef](#)] [[PubMed](#)]
63. Koide, T.; Nishikawa, Y.; Asada, S.; Yamazaki, C.M.; Takahara, Y.; Homma, D.L.; Otaka, A.; Ohtani, K.; Wakamiya, N.; Nagata, K.; et al. Specific recognition of the collagen triple helix by chaperone HSP47. II. The HSP47-binding structural motif in collagens and related proteins. *J. Biol. Chem.* **2006**, *281*, 11177–11185. [[CrossRef](#)] [[PubMed](#)]
64. Konitsiotis, A.D.; Raynal, N.; Bihan, D.; Hohenester, E.; Farndale, R.W.; Leitinger, B. Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **2008**, *283*, 6861–6868. [[CrossRef](#)] [[PubMed](#)]
65. Xu, H.; Raynal, N.; Stathopoulos, S.; Myllyharju, J.; Farndale, R.W.; Leitinger, 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.* **2011**, *30*, 16–26. [[CrossRef](#)] [[PubMed](#)]

66. Sekiya, A.; Okano-Kosugi, H.; Yamazaki, C.M.; Koide, T. Pigment epithelium-derived factor (PEDF) shares binding sites in collagen with heparin/heparan sulfate proteoglycans. *J. Biol. Chem.* **2011**, *286*, 26364–26374. [[CrossRef](#)] [[PubMed](#)]
67. Seo, N.; Russell, B.H.; Rivera, J.J.; Liang, X.; Xu, X.; Afshar-Kharghan, V.; Hook, M. An engineered $\alpha 1$ integrin-binding collagenous sequence. *J. Biol. Chem.* **2010**, *285*, 31046–31054. [[CrossRef](#)] [[PubMed](#)]
68. 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. [[CrossRef](#)] [[PubMed](#)]
69. An, B.; DesRochers, T.M.; Qin, G.; Xia, X.; Thiagarajan, G.; Brodsky, B.; Kaplan, D.L. The influence of specific binding of collagen-silk chimeras to silk biomaterials on hMSC behavior. *Biomaterials* **2013**, *34*, 402–412. [[CrossRef](#)] [[PubMed](#)]
70. 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. [[CrossRef](#)] [[PubMed](#)]
71. Peng, Y.Y.; Stoichevska, V.; Schacht, K.; Werkmeister, J.A.; Ramshaw, J.A. Engineering multiple biological functional motifs into a blank collagen-like protein template from streptococcus pyogenes. *J. Biomed. Mater. Res. Part A* **2014**, *102*, 2189–2196. [[CrossRef](#)] [[PubMed](#)]
72. Hamaia, S.W.; Pugh, N.; Raynal, N.; Nemoz, B.; Stone, R.; Gullberg, D.; Bihan, D.; Farndale, R.W. Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin $\alpha 1\beta 1$ using collagen toolkits II and III. *J. Biol. Chem.* **2012**, *287*, 26019–26028. [[CrossRef](#)] [[PubMed](#)]
73. Emsley, J.; Knight, C.G.; Farndale, R.W.; Barnes, M.J.; Liddington, R.C. Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* **2000**, *101*, 47–56. [[CrossRef](#)]
74. Raynal, N.; Hamaia, S.W.; Siljander, P.R.; Maddox, B.; Peachey, A.R.; Fernandez, R.; Foley, L.J.; Slatter, D.A.; Jarvis, G.E.; Farndale, R.W. Use of synthetic peptides to locate novel integrin $\alpha 2\beta 1$ -binding motifs in human collagen III. *J. Biol. Chem.* **2006**, *281*, 3821–3831. [[CrossRef](#)] [[PubMed](#)]
75. Carafoli, F.; Hamaia, S.W.; Bihan, D.; Hohenester, E.; Farndale, R.W. An activating mutation reveals a second binding mode of the integrin $\alpha 2$ i domain to the GFOGER motif in collagens. *PLoS ONE* **2013**, *8*, e69833. [[CrossRef](#)] [[PubMed](#)]
76. Hamaia, S.; Farndale, R.W. Integrin recognition motifs in the human collagens. *Adv. Exp. Med. Biol.* **2014**, *819*, 127–142. [[PubMed](#)]
77. Carafoli, F.; Bihan, D.; Stathopoulos, S.; Konitsiotis, A.D.; Kvensakul, M.; Farndale, R.W.; Leitinger, B.; Hohenester, E. Crystallographic insight into collagen recognition by discoidin domain receptor 2. *Structure* **2009**, *17*, 1573–1581. [[CrossRef](#)] [[PubMed](#)]
78. 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. [[CrossRef](#)] [[PubMed](#)]
79. Brondijk, T.H.; Bihan, D.; Farndale, R.W.; Huizinga, E.G. Implications for collagen I chain registry from the structure of the collagen von willebrand factor A3 domain complex. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5253–5258. [[CrossRef](#)] [[PubMed](#)]
80. 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. [[CrossRef](#)] [[PubMed](#)]
81. Jung, S.M.; Moroi, M.; Soejima, K.; Nakagaki, T.; Miura, Y.; Berndt, M.C.; Gardiner, E.E.; Howes, J.M.; Pugh, N.; Bihan, D.; et al. Constitutive dimerization of glycoprotein VI (GPVI) in resting platelets is essential for binding to collagen and activation in flowing blood. *J. Biol. Chem.* **2012**, *287*, 30000–30013. [[CrossRef](#)] [[PubMed](#)]
82. Lebbink, R.J.; Raynal, N.; de Ruiter, T.; Bihan, D.G.; Farndale, R.W.; Meyaard, L. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor lair on collagens II and III. *Matrix Biol.* **2009**, *28*, 202–210. [[CrossRef](#)] [[PubMed](#)]
83. Brondijk, T.H.; de Ruiter, T.; Ballering, J.; Wienk, H.; Lebbink, R.J.; van Ingen, H.; Boelens, R.; Farndale, R.W.; Meyaard, L.; Huizinga, E.G. Crystal structure and collagen-binding site of immune inhibitory receptor lair-1: Unexpected implications for collagen binding by platelet receptor GPVI. *Blood* **2010**, *115*, 1364–1373. [[CrossRef](#)] [[PubMed](#)]

84. Barrow, A.D.; Raynal, N.; Andersen, T.L.; Slatter, D.A.; Bihan, D.; Pugh, N.; Cella, M.; Kim, T.; Rho, J.; Negishi-Koga, T.; et al. OSCAR is a collagen receptor that costimulates osteoclastogenesis in DAP12-deficient humans and mice. *J. Clin. Investig.* **2011**, *121*, 3505–3516. [[CrossRef](#)] [[PubMed](#)]
85. 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. [[CrossRef](#)] [[PubMed](#)]
86. Giudici, C.; Raynal, N.; Wiedemann, H.; Cabral, W.A.; Marini, J.C.; Timpl, R.; Bachinger, H.P.; Farndale, R.W.; Sasaki, T.; Tenni, R. Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. *J. Biol. Chem.* **2008**, *283*, 19551–19560. [[CrossRef](#)] [[PubMed](#)]
87. Leo, J.C.; Elovaara, H.; Bihan, D.; Pugh, N.; Kilpinen, S.K.; Raynal, N.; Skurnik, M.; Farndale, R.W.; Goldman, A. First analysis of a bacterial collagen-binding protein with collagen toolkits: Promiscuous binding of yada to collagens may explain how yada interferes with host processes. *Infect. Immun.* **2010**, *78*, 3226–3236. [[CrossRef](#)] [[PubMed](#)]
88. Manka, S.W.; Carafoli, F.; Visse, R.; Bihan, D.; Raynal, N.; Farndale, R.W.; Murphy, G.; Enghild, J.J.; Hohenester, E.; Nagase, H. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12461–12466. [[CrossRef](#)] [[PubMed](#)]
89. Howes, J.M.; Bihan, D.; Slatter, D.A.; Hamaia, S.W.; Packman, L.C.; Knauper, V.; Visse, R.; Farndale, R.W. The recognition of collagen and triple-helical toolkit peptides by MMP-13: Sequence specificity for binding and cleavage. *J. Biol. Chem.* **2014**, *289*, 24091–24101. [[CrossRef](#)] [[PubMed](#)]
90. Kalamajski, S.; Bihan, D.; Bonna, A.; Rubin, K.; Farndale, R.W. Fibromodulin interacts with collagen cross-linking sites and activates lysyl oxidase. *J. Biol. Chem.* **2016**, *291*, 7951–7960. [[CrossRef](#)] [[PubMed](#)]
91. Parkin, J.D.; San Antonio, J.D.; Persikov, A.V.; Dagher, H.; Dagleish, R.; Jensen, S.T.; Jeunemaitre, X.; Savige, J. The collagen III fibril has a “flexi-rod” structure of flexible sequences interspersed with rigid bioactive domains including two with hemostatic roles. *PLoS ONE* **2017**, *12*, e0175582. [[CrossRef](#)] [[PubMed](#)]
92. San Antonio, J.D.; Lander, A.D.; Karnovsky, M.J.; Slayter, H.S. Mapping the heparin-binding sites on type I collagen monomers and fibrils. *J. Cell Biol.* **1994**, *125*, 1179–1188. [[CrossRef](#)] [[PubMed](#)]
93. Wang, H.; Fertala, A.; Ratner, B.D.; Sage, E.H.; Jiang, S. Identifying the sparc binding sites on collagen I and procollagen I by atomic force microscopy. *Anal. Chem.* **2005**, *77*, 6765–6771. [[CrossRef](#)] [[PubMed](#)]
94. Orgel, J.P.; San Antonio, J.D.; Antipova, O. Molecular and structural mapping of collagen fibril interactions. *Connect. Tissue Res.* **2011**, *52*, 2–17. [[CrossRef](#)] [[PubMed](#)]
95. Zeltz, C.; Orgel, J.; Gullberg, D. Molecular composition and function of integrin-based collagen glues-introducing colinbris. *Biochim. Biophys. Acta* **2014**, *1840*, 2533–2548. [[CrossRef](#)] [[PubMed](#)]
96. Raspanti, M.; Alessandrini, A.; Gobbi, P.; Ruggeri, A. Collagen fibril surface: TMAFM, FEG-SEM and freeze-etching observations. *Microsc. Res. Tech.* **1996**, *35*, 87–93. [[CrossRef](#)]
97. Hulmes, D.J.; Jesior, J.C.; Miller, A.; Berthet-Colominas, C.; Wolff, C. Electron microscopy shows periodic structure in collagen fibril cross sections. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 3567–3571. [[CrossRef](#)] [[PubMed](#)]
98. 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. [[CrossRef](#)] [[PubMed](#)]
99. Streeter, I.; de Leeuw, N.H. Atomistic modeling of collagen proteins in their fibrillar environment. *J. Phys. Chem. B* **2010**, *114*, 13263–13270. [[CrossRef](#)] [[PubMed](#)]
100. Streeter, I.; de Leeuw, N.H. A molecular dynamics study of the interprotein interactions in collagen fibrils. *Soft Matter* **2011**, *7*, 3373–3382. [[CrossRef](#)] [[PubMed](#)]
101. Varma, S.; Botlani, M.; Hammond, J.R.; Scott, H.L.; Orgel, J.P.; Schieber, J.D. Effect of intrinsic and extrinsic factors on the simulated d-band length of type I collagen. *Proteins* **2015**, *83*, 1800–1812. [[CrossRef](#)] [[PubMed](#)]
102. Connolly, M. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558. [[CrossRef](#)]
103. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709–713. [[CrossRef](#)] [[PubMed](#)]
104. 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. [[CrossRef](#)] [[PubMed](#)]
105. Broadhead, M.L.; Becerra, S.P.; Choong, P.F.; Dass, C.R. The applied biochemistry of pedf and implications for tissue homeostasis. *Growth Factors* **2010**, *28*, 280–285. [[CrossRef](#)] [[PubMed](#)]

106. Orgel, J.P.; Eid, A.; Antipova, O.; Bella, J.; Scott, J.E. Decorin core protein (decoron) shape complements collagen fibril surface structure and mediates its binding. *PLoS ONE* **2009**, *4*, e7028. [[CrossRef](#)] [[PubMed](#)]
107. Nakamura, T.; Jamieson, G.A.; Okuma, M.; Kambayashi, J.; Tandon, N.N. Platelet adhesion to native type I collagen fibrils. Role of GPVI in divalent cation-dependent and -independent adhesion and thromboxane A2 generation. *J. Biol. Chem.* **1998**, *273*, 4338–4344. [[CrossRef](#)] [[PubMed](#)]
108. Csiszar, K. Lysyl oxidases: A novel multifunctional amine oxidase family. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *70*, 1–32. [[PubMed](#)]
109. Keene, D.R.; San Antonio, J.D.; Mayne, R.; McQuillan, D.J.; Sarris, G.; Santoro, S.A.; Iozzo, R.V. Decorin binds near the C terminus of type I collagen. *J. Biol. Chem.* **2000**, *275*, 21801–21804. [[CrossRef](#)] [[PubMed](#)]
110. Chung, L.; Dinakarandian, D.; Yoshida, N.; Lauer-Fields, J.L.; Fields, G.B.; Visse, R.; Nagase, H. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J.* **2004**, *23*, 3020–3030. [[CrossRef](#)] [[PubMed](#)]
111. Lu, K.G.; Stultz, C.M. Insight into the degradation of type-I collagen fibrils by MMP-8. *J. Mol. Biol.* **2013**, *425*, 1815–1825. [[CrossRef](#)] [[PubMed](#)]
112. Sternlicht, M.D.; Werb, Z. How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* **2001**, *17*, 463–516. [[CrossRef](#)] [[PubMed](#)]
113. Moroi, M.; Jung, S.M. Platelet receptors for collagen. *Thromb. Haemost.* **1997**, *78*, 439–444. [[PubMed](#)]
114. Jokinen, J.; Dadu, E.; Nykvist, P.; Kapyla, J.; White, D.J.; Ivaska, J.; Vehvilainen, P.; Reunanen, H.; Larjava, H.; Hakkinen, L.; et al. Integrin-mediated cell adhesion to type I collagen fibrils. *J. Biol. Chem.* **2004**, *279*, 31956–31963. [[CrossRef](#)] [[PubMed](#)]
115. Siljander, P.R.; Hamaia, S.; Peachey, A.R.; Slatter, D.A.; Smethurst, P.A.; Ouwehand, W.H.; Knight, C.G.; Farndale, R.W. Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J. Biol. Chem.* **2004**, *279*, 47763–47772. [[CrossRef](#)] [[PubMed](#)]
116. Farndale, R.W.; Slatter, D.A.; Siljander, P.R.; Jarvis, G.E. Platelet receptor recognition and cross-talk in collagen-induced activation of platelets. *J. Thromb. Haemost.* **2007**, *5* (Suppl 1), 220–229. [[CrossRef](#)] [[PubMed](#)]
117. Farndale, R.W. Collagen-induced platelet activation. *Blood Cells Mol. Dis.* **2006**, *36*, 162–165. [[CrossRef](#)] [[PubMed](#)]
118. Woltersdorf, C.; Bonk, M.; Leitinger, B.; Huhtala, M.; Kapyla, J.; Heino, J.; Gil Girol, C.; Niland, S.; Eble, J.A.; Bruckner, P.; et al. The binding capacity of $\alpha 1\beta 1$ -, $\alpha 2\beta 1$ - and $\alpha 10\beta 1$ -integrins depends on non-collagenous surface macromolecules rather than the collagens in cartilage fibrils. *Matrix Biol.* **2017**, *63*, 91–105. [[CrossRef](#)] [[PubMed](#)]



© 2017 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/>).