

RESEARCH ARTICLE

# The collagen III fibril has a “flexi-rod” structure of flexible sequences interspersed with rigid bioactive domains including two with hemostatic roles

J. Des Parkin<sup>1</sup>, James D. San Antonio<sup>2</sup>, Anton V. Persikov<sup>3</sup>, Hayat Dagher<sup>1</sup>, Raymond Dalgleish<sup>4</sup>, Shane T. Jensen<sup>5</sup>, Xavier Jeunemaitre<sup>6,7</sup>, Judy Savage<sup>1\*</sup>

**1** From the University of Melbourne Department of Medicine (Northern Health), Melbourne, VIC, Australia, **2** Operations, Stryker Global Quality and Operations, Malvern, PA, United States of America, **3** Lewis-Sigler Institute for Integrative Genomics, Princeton University, Carl Icahn Lab, Princeton, NJ, United States of America, **4** Department of Genetics, University of Leicester, Leicester, United Kingdom, **5** Wharton Business School, University of Pennsylvania, Philadelphia, PA, United States of America, **6** INSERM U970 Paris Cardiovascular Research Centre, Paris France, **7** University Paris Descartes, Paris Sorbonne Cite, Paris, France

\* [jasavage@unimelb.edu.au](mailto:jasavage@unimelb.edu.au)



**OPEN ACCESS**

**Citation:** Parkin JD, San Antonio JD, Persikov AV, Dagher H, Dalgleish R, Jensen ST, et al. (2017) The collagen III fibril has a “flexi-rod” structure of flexible sequences interspersed with rigid bioactive domains including two with hemostatic roles. PLoS ONE 12(7): e0175582. <https://doi.org/10.1371/journal.pone.0175582>

**Editor:** Jie Zheng, University of Akron, UNITED STATES

**Received:** January 9, 2017

**Accepted:** March 20, 2017

**Published:** July 13, 2017

**Copyright:** ©2017 Parkin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial or not-for profit sectors. The second author of this manuscript (JSA) is an employee of private industry, Stryker, Inc. Stryker provided salary support for JSA, but did not have any additional role in the study design, data collection

## Abstract

Collagen III is critical to the integrity of blood vessels and distensible organs, and in hemostasis. Examination of the human collagen III interactome reveals a nearly identical structural arrangement and charge distribution pattern as for collagen I, with cell interaction domains, fibrillogenesis and enzyme cleavage domains, several major ligand-binding regions, and intermolecular crosslink sites at the same sites. These similarities allow heterotypic fibril formation with, and substitution by, collagen I in embryonic development and wound healing. The collagen III fibril assumes a “flexi-rod” structure with flexible zones interspersed with rod-like domains, which is consistent with the molecule’s prominence in young, pliable tissues and distensible organs. Collagen III has two major hemostasis domains, with binding motifs for von Willebrand factor,  $\alpha 2\beta 1$  integrin, platelet binding octapeptide and glycoprotein VI, consistent with the bleeding tendency observed with *COL3A1* disease-causing sequence variants.

## Introduction

The collagens are the major proteins of the extracellular matrix. Each molecule is a homo- or heterotrimer of three  $\alpha$  chains with G-X-Y repeats. Twenty-eight different collagens have been identified with 46 distinct chains [1], that serve as scaffolds for the attachment of cells and matrix proteins, but are also biologically active.

Collagen I is the most abundant collagen, and a heterotrimer of two  $\alpha 1(I)$  and one  $\alpha 2(I)$  chains [2]. Its fibrils have a characteristic banding pattern on heavy metal staining because of their charged residues [3, 4]. The basic repeating structure of the fibril is the D-period, 67 nm long, composed of one region of complete molecular overlap (overlap zone) and one of

and analysis, decision to publish, or preparation of the manuscript and has no financial or proprietary interest in the research. The specific roles of JSA in this publication are articulated in the “author contributions” section.

**Competing interests:** The employment of JSA in private industry (Stryker, Inc.) does not alter the authors’ adherence to PLOS ONE policies on sharing data and materials.

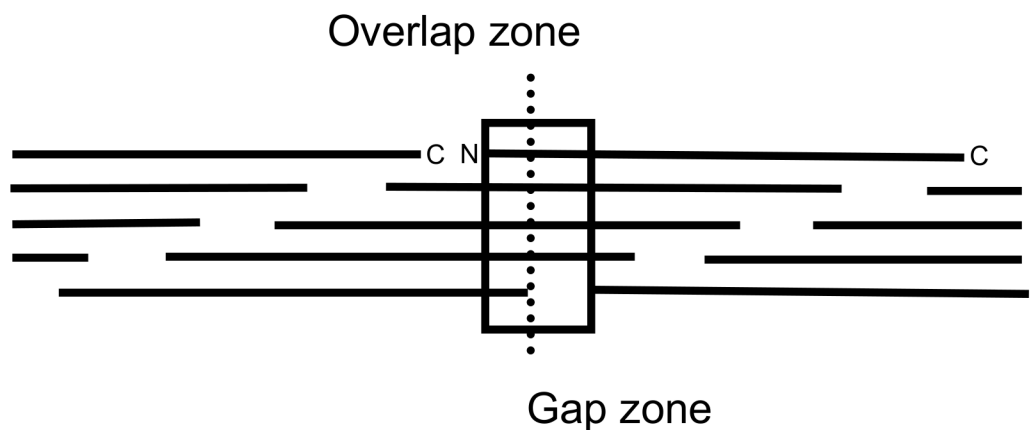
incomplete overlap (gap zone). Each D-period contains the complete collagen monomer (M) sequence derived from overlapping consecutive elements designated as M 1–5, where M1 is the most N-terminal and M5 the most C-terminal segment of the collagen sequence (Fig 1).

Collagen III is another fibrillar collagen. It can form heterofibrils with collagen type I, and is often replaced by collagen I in embryonic development and wound healing [5]. It is a homotrimer of three identical  $\alpha 1(\text{III})$  chains, that together with the collagen I  $\alpha 1(\text{I})$  and  $\alpha 2(\text{I})$  chains are members of the A clade [6], and whose corresponding genes (*COL3A1*, *COL1A1* and *COL1A2*) have arisen from a common progenitor.

The three collagen III  $\alpha 1$  chains are cross-linked in a staggered array through lysine residues at four sites on each molecule to form microfibrils [7]. These fibrils demonstrate the same periodic banding as seen with collagen I. Fibril size varies with tissue type and developmental stage, but overall, the fibrils are finer than for collagen I and recognized on staining as reticulin [8].

Collagen III co-localizes with collagen I in many tissues including the vasculature, bowel and skin [9]. It is generally less abundant than collagen I, except in the walls of distensible organs and blood vessels [10], but its fibril cross-links and proteoglycan-rich matrix contribute mechanical strength and distensibility [11]. Collagen III also participates in a variety of biological functions. It has critical roles in cell-binding, hemostasis and angiogenesis [12], tissue remodeling, fetal development [13], and wound healing [14, 15]. It is affected by microbial infections, ageing, diabetes, and inherited disease. Disease-causing *COL3A1* variants result predominantly in vascular Ehlers-Danlos syndrome (EDS type IV) (MIM 130050). This affects one in 150,000 individuals, and is autosomal-dominantly inherited [16]. Clinical features include easy bruising, and arterial, intestinal or uterine rupture [15]. The atypical form, acrogeria, is characterized by translucent skin and premature aging of the face and peripheries [16,17], but these features are present to some extent in all affected individuals. *COL3A1* variants have also been associated with cerebral aneurysms [18], gastroesophageal reflux disease [19], and pelvic organ prolapse [20].

Interactomes are protein maps which indicate structural features and the sites of interactions with other molecules. The unique molecular structure of collagens, with their predominantly rigid, triple helical conformation, allows the construction of maps wherein triple helical domains can be represented as 2D linear arrays of three polypeptide chains. The collagen I interactome has been used to deduce functional and disease-associated domains from ligand-



**Fig 1. Arrangement of five overlapping collagen III  $\alpha 1$  chains in the fibril D-period with overlap and gap zone indicated.**

<https://doi.org/10.1371/journal.pone.0175582.g001>

binding data [21,22,23]. Here we have constructed and analyzed the collagen III interactome. Note that databases exist which archive disease variants mapping to the collagens (<http://www.le.ac.uk/genetics/collagen/>), or through which researchers can access and plot data of interest onto maps of fibrillar collagens, including collagen III [24]. However, to date a comprehensive analysis of all known ligand binding sites and disease variants for collagen III has not been reported.

## Results

The amino acid numbering systems for procollagen III and the mature collagen III protein that are used here are shown in [S1 Table](#).

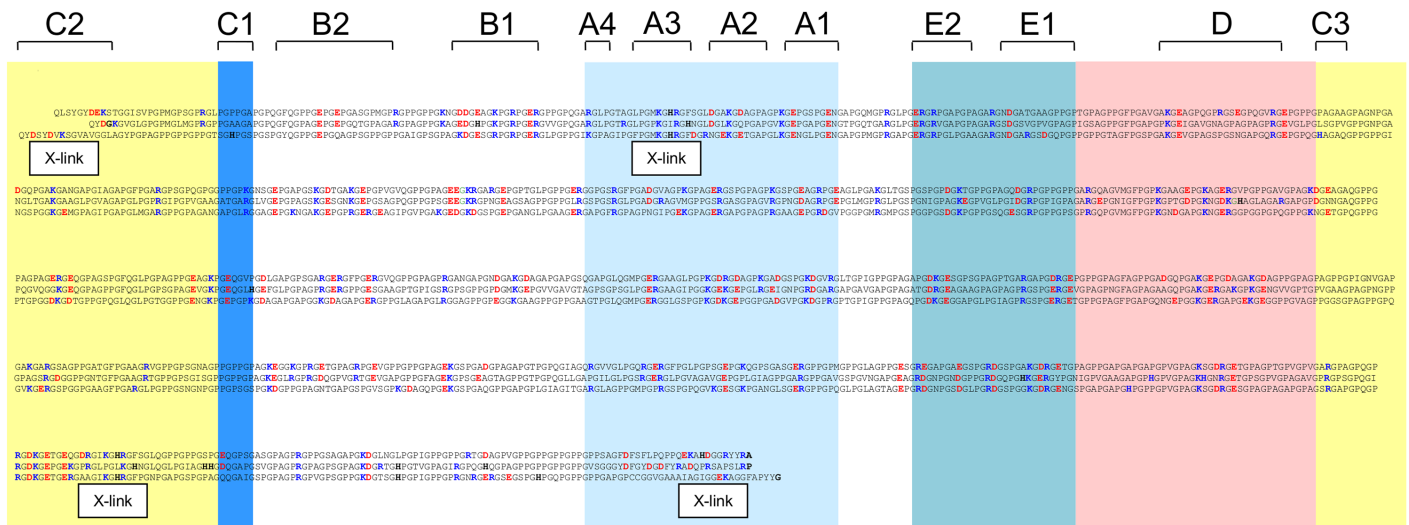
### D-period and banding pattern

When the cross-linking K residues in the collagen  $\alpha 1$ (III) chain were aligned with those in the collagen  $\alpha 1$ (I) and  $\alpha 2$ (I) chains, collagen III had a nearly identical D-period arrangement to collagen I (Figs 1 and 2), with the major crosslink pairs being separated by 843 residues in both collagen types [7,25].

When the D-periods of the  $\alpha 1$ (III) chain were examined for positive- (R, K) and negatively-charged (E, D) residues, the charge, although not necessarily the residue type, was typically in register across the D-period, and also with the  $\alpha 1$ (I) and  $\alpha 2$ (I) chains, and hence with the collagen I D-period (Fig 2). The charge was not however symmetrical about the molecules' mid-points which implies unidirectional alignment of collagen III and I during their assembly into heterotypic fibrils. Charge alignment meant that binding motifs for some ligands, such as proteoglycans and HSP47, are also in the same locations for both collagens.

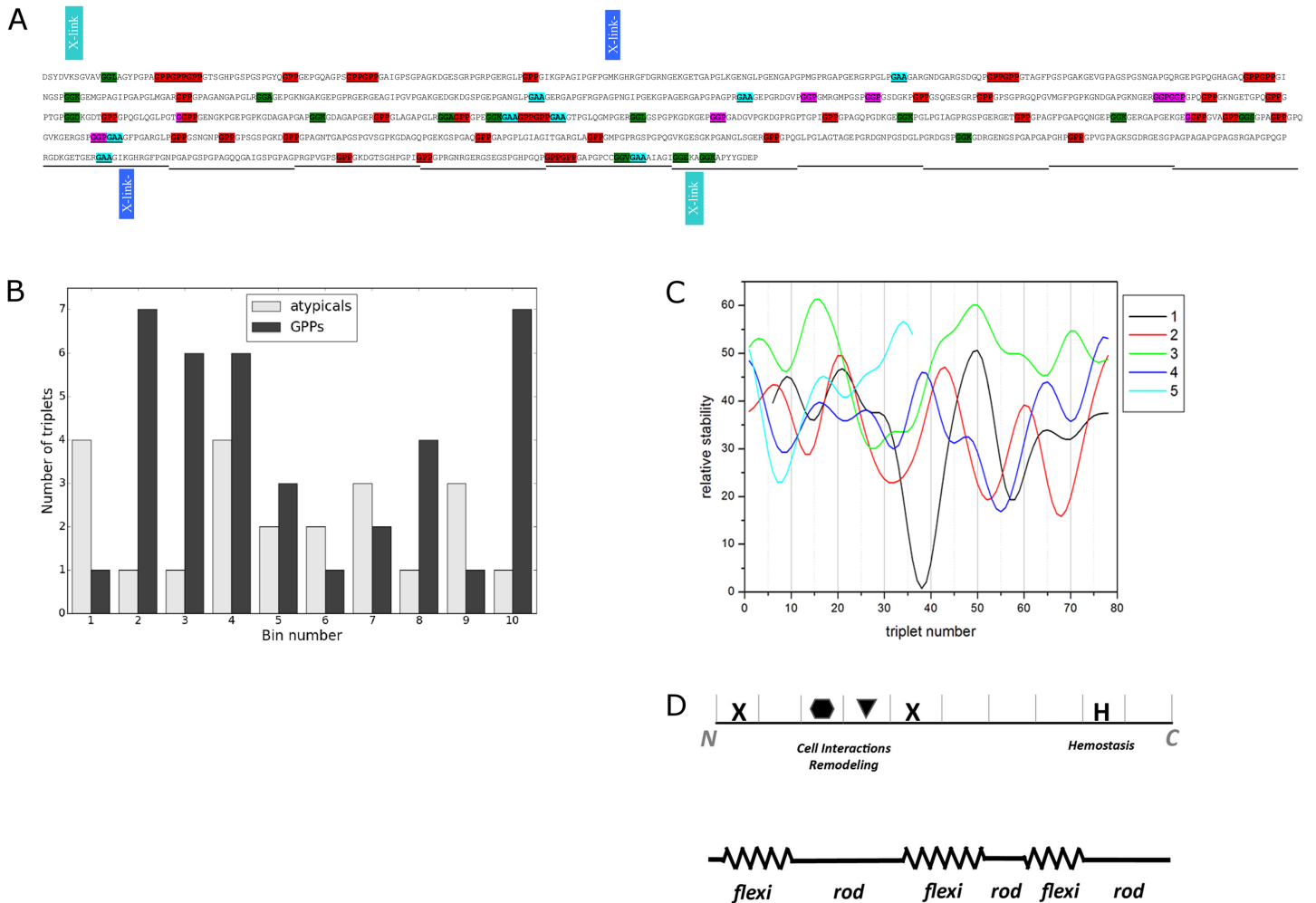
### Analysis of fibril stability and structure

Collagen III has a higher G, and lower P content than the other fibrillar collagens [26], and nearly twice the number of atypical triplets as type I collagen (Fig 3). GPP-rich regions promote triple-helical rigidity with more compact 7/2 symmetry, while non-P ('atypical') residues



**Fig 2. Arrangement of collagen I  $\alpha 1$  (top),  $\alpha 2$  (middle) and collagen III  $\alpha 1$  (bottom) chains demonstrates approximate alignment of charge across all three molecules in the D-period.**

<https://doi.org/10.1371/journal.pone.0175582.g002>



**Fig 3.** a. Collagen III D-period demonstrating location of GPP residues (in red) proposed to confer stability and atypical triplets (GAA, GGP, etc, in other colors) proposed to contribute flexibility.; b. Each monomer of the collagen III D-period was divided into ten bins of equal size (about 23 residues) starting at the N-terminus. The number of GPP triplets and atypical triplets (GAA, GGP, etc) were added for all five monomers and plotted against bin number; c. Relative triple helix stability for amino acid sequences of each monomer within the collagen III D-period was predicted using the collagen stability calculator ([http://semimajor.net/collagen\\_calculator/](http://semimajor.net/collagen_calculator/)). Monomers 1–5 represent five segments of the collagen III sequence, beginning at the N-terminus. Three comprise approximately 234 amino acid residues but the two that included the gap zone were shorter. Examining these stability curves together suggests that some regions of the D-period are more stable, and others less so, suggesting a “flexi-rod” model for collagen III; d. Top: Schematic of type III collagen monomer with sites for cell interactions and remodeling flanked by crosslinks (X) and hemostasis domain (H). Bottom: Stability modeling indicates clusters of atypical collagen sequences of lower stability (springs) are interspersed with rigid zones (rods) hosting sequences that carry out crucial biologic functions.

<https://doi.org/10.1371/journal.pone.0175582.g003>

alter the triple helical twist [27]. Thermodynamic studies on collagen peptides show that the GPP triplets are the major contributors to triple-helix stability, and that atypical triplets are destabilizing [28,29]. Atypical and GPP triplets tended to predominate in different bins of the collagen III molecule (Fig 3). While atypical triplets occurred more often in bins 1, 6, 7, and 9, GPP triplets were more common in bins 2, 3, 4, 8, and 10. There was a strong inverse correlation between atypical and GPP triplets with a Pearson correlation coefficient of -0.52. This suggested the presence of alternating rigid and flexible regions within the D-period, along the length of the collagen III molecule.

The collagen stability calculator demonstrated co-localization of local stability variations (Fig 3), with three major regions of decreased stability, including triplets 5–15, 30–40, and 50–60. We thus propose collagen III functions as a “flexi-rod” in which a confluence of atypical

triplets creates flexible domains, allowing focal expansion or of several discrete fibril regions (Fig 3). Yet, two of the three flexible domains incorporated hydroxyl (OH)-K substrates for intermolecular crosslinking between triple helices. Thus the degree of flexibility of the collagen III fibril afforded by the confluence of atypical amino acid triplets may be modulated by the fibril's extent of crosslinking and may vary, from highly flexible in young organisms or early in wound healing, to less flexible with the more highly cross-linked protein with ageing or later in scar formation. One of the flexible zones lacked the potential for intermolecular crosslink formation and fell in the middle of the gap zone, a thinner, more pliable fibril region. This suggested that the flexi-rod structure of collagen III, regardless of crosslinking, still confers some flexibility to the polymer.

The rod-like domains containing the major cell- and ligand- binding sites were located between the relatively flexible zones on collagen III (Fig 3). This may allow the molecule to preserve the triple helical conformations of the regions critical for biologic function.

A stability analysis of the type I collagen fibril also revealed a "flexi-rod" structure (data not shown). However, the more abundant atypical triplets in collagen III are consistent with a more flexible polymer, and with collagen III's predominance in pliable tissues such as the skin of young animals, wounds, and in distensible organs.

## Sites related to structure, assembly, turnover, modification, cleavage and ligand-binding

These include N- and C-propeptides, A-rich sequences, F residues, C cross-links, K cross-links, O and N glycosylation sites, P glycosylation sites, and cleavage sites (N and C propeptidases, matrix metalloproteinases (MMP), serine proteases, chaperones, Secreted Protein Acidic and Rich in Cysteine (SPARC), and Discoidin domain receptors (DDR1 and DDR2) (Fig 4).

**N- and C-propeptides.** Collagen  $\alpha 1$ (III) has N- and C-propeptides that are normally cleaved. Retention of the quasi-rigid N-terminal propeptide potentially restricts fiber growth [5,7]. The collagen  $\alpha 1$ (III) C-propeptide includes the NC1 recognition sequence, GNPEL-PEDVLDVSSR [30], critical for triple helix nucleation [31], as well as a HtrA1 serine protease cleavage site [32].

**A-rich sequences.** Collagen III has two A-rich sequences, GAAGARGNDGAR and GPA-GERGAPGPAGPRGAA that may contribute to elasticity [33]. Similar sequences occur at these locations in collagen I.

**F residues.** Collagen III has three major cross-fibril clusters of F residues in the same locations as in collagens II and III [34] where they are proposed to promote collagen monomer assembly into fibrils. For the collagen III sequence RGQPGVMGF, F is a crucial component of the DDR2, SPARC and vWF-collagen binding complexes [35,36,37].

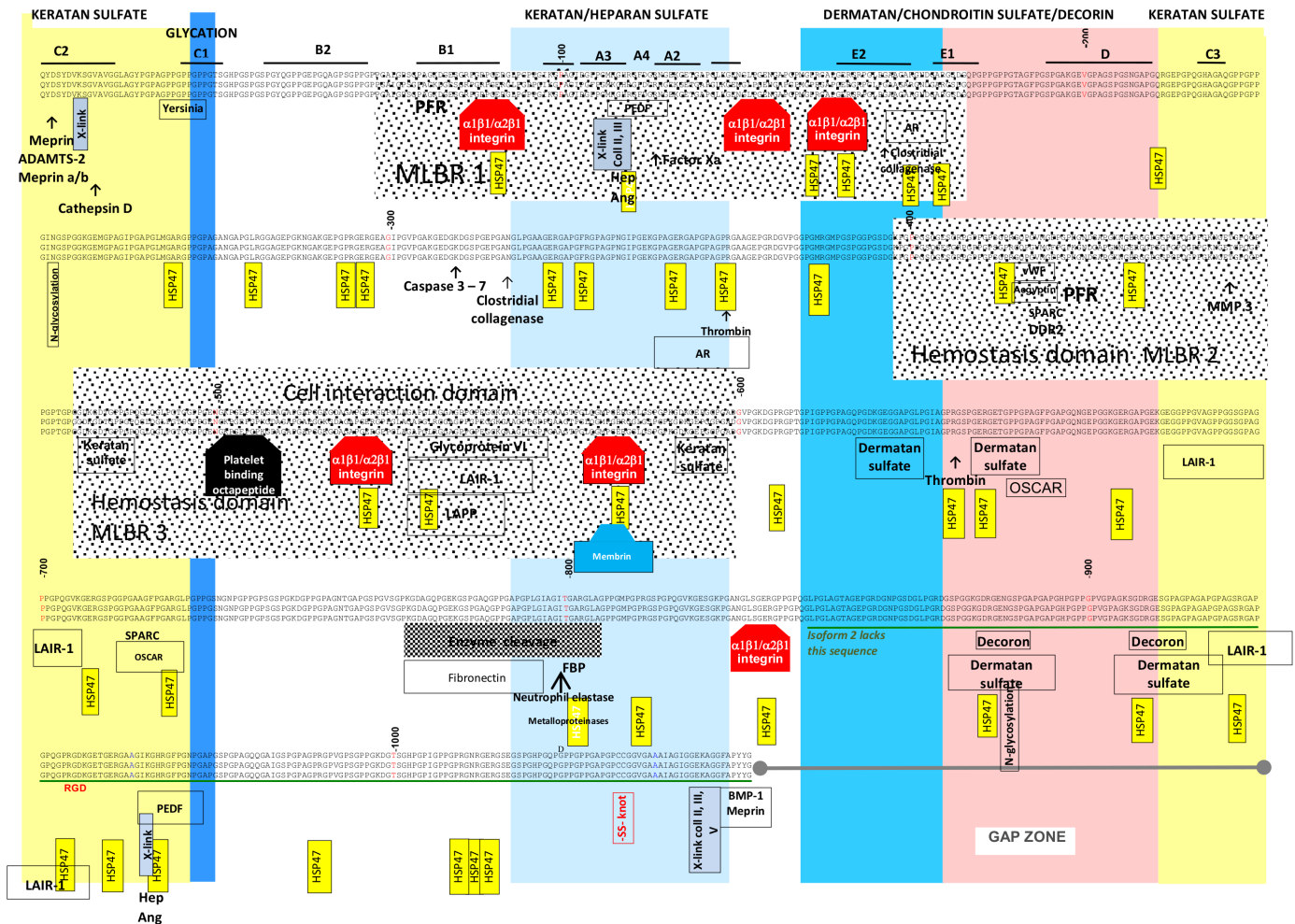
**C cross-links.** The C residues at 1196–1197 in the collagen III  $\alpha 1$  chains form a disulfide knot [38], that facilitates molecular realignment after denaturation. The mature collagen I protein does not have C residues which selects against trimerization with collagen III. Instead collagen I uses the C-terminal (GPP)<sub>5</sub> motif.

**K cross-links.** Divalent and trivalent crosslinks covalently join collagens III and I [77]. The K covalent cross-links at residues 110 and 953 stabilize collagen III fibril formation. The other cross-link sites are in the collagen III N- and C-telopeptides [77].

**O- and N-glycosylation sites.** Collagen III has one O-glycosylation site [39], and N-glycosylation sites at the GXN motifs.

**P hydroxylation sites.** Most P residues in the Yaa position of the GXY triplet in collagen III are 4-hydroxylated. The GHPGPIGPPGPR motif that is 3-hydroxylated in collagen I and II





**Fig 4. Human collagen III interactome demonstrating the positions of putative Cell interaction domain, Fibrillogenesis and enzyme cleavage domain, Major ligand-binding regions (MLBR1, 2 and 3), Hemostasis domains, and miscellaneous ligand binding sites and structural features on the collagen triple helix in the fibril's D-period.** Sequences or residues comprising binding sites or structural features are indicated by labels that appear beneath the relevant sequence or residue(s) and approximate their spans. Some enzyme cleavage sites are indicated by arrows pointed to the approximate center of the enzyme's recognition sequence. Abbreviations used in the interactome for the first time include: ang, angiogenesis; AR, A-rich; coll, collagen; hep, heparin/heparan sulfates; FBP, fibronectin-binding protein; and PFR, platelet fibrinogen receptor.

<https://doi.org/10.1371/journal.pone.0175582.g004>

is not hydroxylated in collagen III [77]. The absence of this cross-link may contribute to collagen III's inability to form homotypic fibrils [77].

**N and C propeptidases.** Meprin a/b cleaves both the N and C-propeptides [40] of collagen III. The N-propeptide is also removed by ADAMTS-2, and the C-propeptide by procollagen C-proteinase, also known as bone morphogenetic protein 1 (BMP1) [40].

**MMP.** MMPs are important in extracellular matrix degradation and tissue remodeling, and activating other MMPs. Collagen III is cleaved by MMPs 1, 2, 3, 9 and 13 at residues 791–806, located about three quarters of the length of the molecule, near the fibrils gap zone [41,42]. There are also cleavage sites nearby for elastase, trypsin and thermolysin thought to act on the soluble, denatured protein [43].

**Serine proteases and other enzymes.** Collagen III  $\alpha 1$  chains have consensus sequences for thrombin and factor Xa cleavage, which are potentially important for collagen III's hemostatic role but may only be relevant to the denatured molecule [44].

**Chaperones.** HSP47 is a procollagen-specific chaperone that recognizes GXR where the R residue is critical [45]. HSP47 stabilizes the collagen triple helix, clamping the three chains in register. Each collagen III  $\alpha 1$  chain binding site may have one or two chaperones bound to it, which prevents lateral aggregation of collagen molecules [45]. Not all of the predicted HSP47 binding sites may be functional.

**SPARC.** SPARC acts as a chaperone but also modulates collagen fibril assembly [46] and tissue remodeling. The minimal binding motif in several collagens is GVMGFO [36,46]. Collagen III and I have a common high affinity SPARC binding site but collagen I has an additional low affinity site [46].

**Discoidin domain receptors (DDR1 and DDR2).** DDRs are tyrosine kinase receptors that induce the expression of MMPs and BMP, and potentially regulate collagen production and organization. Their main binding site in collagen III includes the GVMGFO motif that also binds SPARC, and contributes to the binding of vWF, overlaps with a conserved F residue, and is in the same location in collagen I [47].

### Isoform (P02461-2)

Expression of the collagen III isoform lacking C-terminal residues 694–996 (UniProt KB) has not been confirmed *in vivo*. It lacks the final monomer of the D-period, which includes the fibrillogenesis domain and sites for the disulfide knot, cross-links, and some ligands, but retains the K residues required for cross-linking.

### Binding sites for extracellular matrix molecules, integrins, cells and other ligands

These include the extracellular matrix molecules (proteoglycans, collagens, fibronectin, thrombospondin 4, integrins, RGD motif, leucocyte-associated immunoglobulin-like receptors (LAIR) 1 and 2, Osteoclast-associated Receptor (OSCAR) and other binding motifs.

**Proteoglycans.** The binding motifs for proteoglycans in collagen III are GE/DR/KGE/DXGXXGX [48] (Table 1). Binding sites for decoron (the core protein of decorin) and dermatan sulfate proteoglycan overlap with the motif KXGDRGE at the e1-d band junction [49]. These are at the same locations as in collagen I [22].

The heparan sulfate binding motif is KGHRGF in collagen III and there are two sites at the N and C-termini in the same locations as for collagen I. The heparan sulfate proteoglycans act as co-receptors for growth factors and are pro-angiogenic [50].

Pigment epithelium-derived factor (PEDF), also known as Serpin F1, has a variety of functions including being anti-angiogenic and anti-tumorigenic. It inhibits VEGF expression but upregulates that of thrombospondin [51]. It binds to the same residues in collagen III as heparin and heparan sulfate [52].

Decoron and probably biglycan bind at two sites in bands d and e in collagens I and III [49, 53]. Decorin also binds to fibronectin, complement component C1q, epidermal growth factor receptor (EGFR) and transforming growth factor  $\beta$  (TGF $\beta$ ). Decorin binding to collagen III inhibits fibrillogenesis, but is pro-angiogenic [49, 54].

**Collagens.** Collagen III binds covalently to collagens I and II through cross-links at their N- and C-termini to form heterotypic fibrils [77], that are thinner than those comprising collagen I alone [55]. Collagen III lacks 3OH-P residues [25] which may impair its ability to extend laterally and determines its peripheral location in the fibril [55,77]. Notably, the collagen III triple helix is fifteen residues longer than that of collagen I or II but it is not known how this feature may affect the assembly, structure or function of heterotypic fibrils. In cartilage, collagen III copolymerizes via a trivalent cross-link with collagen II [56] through the same sites

**Table 1. Major structural features and ligand binding sites for collagen III.**

Name	Motif	Sequence	Residues	Comments	Detection method	Reference
Structural features						
C cross-links	CC	CC	1043–1044	Forms bonds between collagen III $\alpha 1$ chains	X Ray crystallography	[38]
K cross-links	GMXGHR (bovine) IAGIGGEYA	GMKGRHRIGKGR IAGIGGEKA	108–113951–956 1052–60	Cross links collagen III and collagen I or II	Enzyme digestion and sequencing	[7,99]
Modification sites and sites related to assembly, trafficking and turnover						
4 P-hydroxylation	G-Xaa-P	N/A	Multiple sites	Stabilizes collagen triple helix	Protein sequence	[100]
O-linked glycosylation	PGMKGRHG	PGMKGRHG	107–114	Regulates cross-link maturation	Cyanogen bromide fragments, protein sequence	[99, 101]
N-linked glycosylation	NXaaS/T	NGS NGS	236–238 884–886	Unlikely in native collagen	Protein sequence	NetNGly 1.0
NC1 recognition sequence	GNPELPEVDVSSR	C-terminal propeptide	C-terminal to onset of triple helix	Highly conserved sequence, critical to triple helix nucleation	Protein sequence	[31]
HSP47	GXR	GXR	Multiple sites, in the same locations as collagen I $\alpha 1$ and $\alpha 2$	Procollagen-specific chaperone that facilitates folding in the ER	Protein sequence	[102]
Secreted Protein Acidic and Rich in C (SPARC)	GYMGFO GAAGFO	GYMGFO GAAGFP	423–428 717–722	Matricellular protein with roles in cell differentiation and wound healing	Rotary shadowing, cyanogen bromide, synthetic peptides	[46]
Membrin	MxxE	MPGE	574–577	Transports proteins within the Golgi	Protein sequence	Minimotif
Glycation sites	Select K residues	K	Undefined	Advanced glycation end-products may form at select residues	Sites in collagen I $\alpha 1$ (434) and $\alpha 2$ chains (453, 479, 924); cross-fibril glycation zone for type I/III fibrils, blue stripe, Fig 4.	[88]
Cleavage sites						
Procollagen III N-proteinase		NYSP/QYDS	.../1-4	Cleaves native collagen III	Protein sequence	[103,104]
ADAMTS-2		NYSP/QYDS	.../1-4	N-proteinase that cleaves collagens I, II, III, probably distinct from procollagen III N-proteinase	Protein sequence	[104]
Mepirin a/b	G/DEPMDF	NYSP/QYDS PYG/DEPM	.../1-4 1065-1068/...	Cleaves both N- and C-propeptides	SDS-PAGE	[40]
Bone morphogenetic protein 1 (BMP1); Procollagen C-proteinase, PCP)	G/DEPMDF	C-propeptide	1065-1068/...	Cleaves C-propeptide	SDS-PAGE	[40]
Metalloproteinase-1 (MMP1, 3 and 9)	GFLGIAGITGARGLA	GFLGIAGITGARGLA	792–803	Cleaves collagen III along the molecule; activates other MMPs	Agarose gels and sequencing	[105]
MMP3 (stromelysin 1)	KNGETGPQGP	KNGETGPQGP	457–466	Cleaves collagen III along the molecule; activates other MMPs	Agarose gels and sequencing	[105]
High temperature requirement A 1 (HtrA 1) serine protease	GPVCEL/C	C-propeptide	C-terminal to triple helix	Serine protease, binds to C-terminus	Solid phase binding assay	[32]

(Continued)



Table 1. (Continued)

Name	Motif	Sequence	Residues	Comments	Detection method	Reference
Clostridium collagenase	LGPA	LGPA	160–163 325–328	Cleavage	Protein sequence	[83,84,106]
Caspase 1	WTD/SS	WTD/SS in C- propeptide	C-terminal to triple helix	Cysteine-aspartic protease that activates proteins by cleavage, digests ECM	Protein sequence	[107]
Caspase 3–7	DGKDG	DGKDG	311–315	Cysteine-aspartic protease, involved in apoptosis and digests extracellular matrix	Protein sequence	[107]
Cathepsin D	KSGVAGGI (bovine)	KSGVAGGI	8–16	Non-specific tissue proteinase that	Protein sequence	[108]
Caspase 3–7	DGKDG	DGKDG	311–315	Cysteine-aspartic protease, involved in apoptosis and digests ECM	Protein sequence	[107]
Neutrophil elastase	I/T	IT	799/800 and 800/801	Cleaves in MMP enzyme cleavage domain	Protein sequence	[69]
Trypsin	R/G	RG	Multiple sites	Cleaves in MMP enzyme cleavage domain	Protein sequence	[109]
Thermolysin	G/L	GL	804/805	Cleaves in MMP enzyme cleavage domain	Protein sequence	[110]
Thrombin	AGPR/GAA AGPR/GSP	AGPR/GAA AGPR/GSP	365/366 641/642	Cleavage	Protein sequence	[44]
Factor Xa	DGR/NG	DGR/NG	118/119	Cleavage	Protein sequence	[44]
Binding sites for extracellular matrix molecules, cells and other ligands						
Extracellular molecules						
Collagen I, II	N-terminal site unresolved KGHR KGHR IAGIGEXA	IAGIGGEKA GPAGMPGFPMKGRH (C-propeptide)	99–119, 1052–1068	Form heterotypic fibrils in tissue-specific manner	CNBr fragments and synthetic peptides with Western blots	[77]
Keratan sulfate proteoglycan (KSPG)	GERGEGPAGS on collagen I GDRGDAGPKGA on collagen 1	GDKGDTGPPGP GDRGEPGGGA	474–484 588–598	Proposed to regulate spacing between collagen fibrils	Protein sequence	[48]
Dermatan sulfate proteoglycan (DSPG)	GDKGESGFSGP on collagen I GDRGEPGGFP on collagen I	GDKGEGGAPGL GERGETGGP	624–634 645–655	Proposed to regulate spacing between collagen fibrils	Protein sequence	[48]
	GDRGETGPAGP on collagen I	GDRGENGSPA	879–889	Proposed to regulate spacing between collagen fibrils	Protein sequence	[48]
	GDRGETGPAGP on collagen I	GDRGESGPA	909–919	Proposed to regulate spacing between collagen fibrils	Protein sequence	[48]
Heparin	KGHR	KGHR	110–113 953–956	Inhibits angiogenesis	Solid phase binding assay, endothelial tube formation assays	[111,112]
Pigment Epithelium-derived Factor (PEDF, SERPIN F1)	MKGRGFDRNG IKGRGFPGNOG	MKGRGFDRNG IKGRGFPGNPG	109–120 952–963	Complexes with collagen in cornea and vitreous	Solid phase binding assay	[52,113]
Decoron (decorin core protein)	KXDGRGE on collagen I; AKGDRGE on collagen I	GKGRGE KSGDRE	877–883 907–914	Inferred from sequence, position and site on A clade map	Rotary shadowing, X Ray crystallography	[49]

(Continued)

Table 1. (Continued)

Name	Motif	Sequence	Residues	Comments	Detection method	Reference
Fibronectin-binding protein	Binds about 122 nm from C-terminus of collagen I	Undefined binding sequence	about 800	Links fibronectin and collagen III. Binds near MMP cleavage site and near decorin with which it interacts	Rotary shadowing, X Ray crystallography	[57]
Secreted protein Acidic and Rich in Cysteine (SPARC, BM40, Osteogenin)		GVMGFO GAAAGFP	423–428 717–722	Secreted glycoprotein important in mineralization	CNBr fragments, synthetic peptides and rotary shadowing	[46]
Thrombospondin	Binds close to both N- and C-termini	Undefined binding sequence(s)	Unknown	Platelet aggregation and antiangiogenic, mediates cell-cell and cell-ECM binding	Rotary shadowing	[58]
Tenascin X	Unknown	Unknown	Unknown	Binds to collagen III and decorin	Solid phase binding assay	[14,115]
Cells						
Integrin $\alpha 1\beta 1$ , $\alpha 2\beta 1$	GRGER (GROGER)	GROGER	84–89	High affinity, expressed on platelets	Rotary shadowing and synthetic peptides, collagen III toolkit	[60]
Integrin $\alpha 2\beta 1$	GLPGER (GLOGER)	GLOGER	135–140, 150–155	High affinity	Rotary shadowing and synthetic peptides, collagen III toolkit	[61]
Integrin $\alpha 2\beta 1$	GAPGER (GAOGER)	GAOGER	525–530	Low affinity	Rotary shadowing and synthetic peptides, collagen III toolkit	[61]
Integrin $\alpha 2\beta 1$	GMPGER (GMOGER)	GMOGER	573–578	High affinity	Rotary shadowing and synthetic peptides, collagen III toolkit	[61]
Integrin $\alpha 2\beta 1$	GLSGER	GLSGER	834–839	High affinity	Rotary shadowing and synthetic peptides, collagen III toolkit	[61]
RGD	RGD	RGD	938–940	Integrin- binding site, not usually relevant in native collagen	Protein sequence	[64]
DDR2 (Discoidin domain receptor-2)	GVMGFO	GVMGFP	423–428	Cell surface collagen receptor, tyrosine kinase receptor that activates fibroblasts	Synthetic peptides	[116]
LAIR-1 (Leukocyte associated Ig-like receptor-1)	(GPO) 10	GAPLRRGGAGPPPEEGKGAAGPPGPP GEGGPOGVAGPOGSSPAGPOGPGVK GPAGPAGAPGFPAGSRGAOGPPGPRGDK	537–563 681–707 915–941	Inhibits immune responses, found on mononuclear cells and thymocytes	Solid phase binding to collagen III and (GPO) <sub>10</sub> peptide	[65]
OSCAR (Osteoclast Associated Receptor)	GPPFPAGFPAGP GGPAAAGFPGAR	GPPFPAGFPAGP GGPAAAGFPGAR	651–662 714–725	Tyrosine kinase-coupled receptor that costimulates osteoclastogenesis	Solid phase binding to synthetic peptides and type II toolkit	[67]
nCAM1 (Neural cell adhesion molecule1); ICAM1 (Intercellular adhesion molecule 1)	Motif unknown	Unknown	One binding site	Membrane glycoprotein but also part of the ECM	Solid phase binding assays of ligand to native proteins	[117]
MAG (Myelin-associated glycoprotein)	Motif unknown	Unknown	Two binding sites	Membrane glycoprotein, but also part of the ECM	Solid phase radio-ligand binding assay	[117,118]
Hemostasis						
Von Willebrand Factor (VWF)	RGQPGVMGF	RGQPGVMGF	422–430	Multimeric glycoprotein, adheres to collagen and platelet surface receptors at high shear rate flow	Synthetic triple helical peptides and solid phase binding assay	[71]

(Continued)

Table 1. (Continued)

Name	Motif	Sequence	Residues	Comments	Detection method	Reference
Platelet fibrinogen receptor (Ib/IIa)	PXXXD	PAGKD	72–76	Controversial, biological significance unknown	Protein sequence	[76]
Glycoprotein VI	GAOGLRGGAGPQPEGGKGAAGPOGPO	GAPGLRGGAGPFGPEGGKGAAGPEGPP	537–563	Platelet membrane protein mediating collagen-induced aggregation	Platelet and GPVI receptor binding to collagens and triple helical peptides in solid phase assays	[78]
Platelet-binding octapeptide, kindlin-3	KOGEQGPK	KOGEQGPK	502–509	Mediates platelet adhesion independent of GpVI and GpIIa	Platelet binding to collagen peptides in solid phase binding assay	[79]
Platelet-derived growth factor (PDGF)	Motif not known	Unknown	C-terminal part of collagen III	Binds to a tyrosine kinase receptor to stimulate cell growth	Solid phase binding assay	[119]
Proteins of infectious agents						
Langerin	ASQNTIYHCKNS	ASQNTIYHCKNS in pro-peptide	N-terminal to triple helix	C-type lectin	Mass spectrometry	[82]
YadA (Yersinia adhesion A)	GPO to (GPO) 6	(GPP) 3	24–32	Outer membrane protein important in adhesion	Solid phase binding to synthetic triple-helical peptides	[85]
AAEL010235 Aegyptin	RGQPGVMGF (high affinity) (GPO) 10, GFOGER (lower affinity)	Several	419–427	Anticoagulant from saliva of blood sucking insect, Aedes aegypti; inhibits binding of vWF, glycoprotein VI and $\alpha 2 \beta 1$ integrin	Solid phase binding assay	[72,73]
Calin	Motif unknown	Unknown	About 419–427	Anticoagulant in leech saliva (Hirudo medicinalis)	Inhibits VWF binding to collagen III	[74]
Leech antiplatelet agent (LAPP)	Bovine GPFPRGGAGPPEGGK	GAPGLRGGAGPPEGGK	537–554	Anticoagulant in Leech saliva	Inhibits VWF binding to collagen III	[75]
MIP (Macrophage infectivity potentiator protein)	Motif unknown	Unknown	Unknown	Virulence factor found in Legionella	Solid phase binding assay	[86]
Miscellaneous ligands						
G protein-coupled receptor 56 (GPR56)	Motif unknown	Unknown	Unknown	Orphan G-protein coupled receptor, important in cortical development	Solution phase "pull down" binding assay	[120]
PR47 (47 Kd platelet receptor for collagen III)	Motif unknown	Unknown	Unknown	Membrane receptor for collagen III similar to collagen I receptor	Solid phase binding assay	[121,122,123]

<https://doi.org/10.1371/journal.pone.0175582.t001>

[77]. Collagen V often occurs together with collagen III but there is no evidence currently for intermolecular crosslinking.

**Fibronectin.** There is no evidence for direct binding of collagen III to fibronectin. However a fibronectin-binding protein interacts with collagen III at about residue 800, after taking into account the propeptide [57], which is the same location that fibronectin binds in collagen I and II. This means that collagen III potentially binds to fibronectin and decorin on the same monomer, near the sites for MMP cleavage and heparan sulfate binding. The proximity to the fibrillogenesis domain suggests a regulatory role of fibronectin in collagen III chain assembly and degradation [21], and in particular, MMP cleavage potentially separates the fibronectin- and decorin- binding sites.

**Thrombospondin 4.** Thrombospondin is found in extracellular matrix, but also in platelets, and it contributes to platelet aggregation and inhibits angiogenesis. Thrombospondin binds to the N- and C-termini of collagen III [58,59].

**Integrins.** Collagen III binds to the  $\alpha 1\beta 1$  and  $\alpha 2\beta 1$  integrins, and potentially a variety of cells. Most P residues in the collagen III  $\alpha 1$  chain are hydroxylated, which enhances integrin binding. The collagen III  $\alpha 1$  chain has three integrin binding sites including GRPGER [60], GLPGEN, a high affinity site for  $\alpha 1\beta 1$  [61], and GMPGER, of rather lower affinity for  $\alpha 2\beta 1$  than GRPGERV [62]. Two lower affinity sites, GAPGER [60,63] and GLSGER have also been described. The GAPGER motif at residues 525–530 in the cell interaction domain [60], is at the same location as the collagen I high affinity site, GFPGER.

The RGD archetypal integrin- binding motif is thought to be less important in collagen where the triple helical conformation renders it inaccessible [64]. However the single RGD site in collagen III is in the same location as in collagen I, and may be exposed after enzymatic cleavage and thus become available for ligation by some integrin receptors [64].

**LAIR 1 and 2.** LAIR-1 is an immune regulatory receptor expressed on mononuclear cells [65]. LAIR-2 blocks the binding of LAIR-1 to collagen III [66]. Both LAIR-1 and -2 bind to  $(GPO)_3$ , and there are at least two sites for interaction on collagen III.

**OSCAR.** This co-stimulates monocytes through the Fc receptor and activates osteoclast formation, resulting in bone resorption. OSCAR binds to GPOGPAGFOGAO with GPOGPXGFX as the minimum motif, where P can be substituted by A [67]. This sequence overlaps with a dermatan sulfate binding site. There is a second possible OSCAR binding site, GGPGAAGFPGAR.

**Other cell-binding motifs.** Collagen III also binds to ICAM1 on endothelial and immune cells, and NCAM1 on neurons, glia, skeletal muscle and natural killer cells but the binding motifs are unknown.

## Major structural domains

These include the major ligand-binding regions 1, 2 and 3 (MLBR1, 2 and 3), a cell-interaction domain, a Gap, an MMP/enzyme cleavage domain and the fibrillogenesis domain.

**MLBR 1, 2 and 3.** MLBR1 is found in approximately the same locations in collagen III and collagen I, and both sites share many binding motifs and ligands. Both MLBR1 have motifs for integrin-binding, intermolecular crosslinking, angiogenesis, and heparin-binding.

However, MLBR2 on collagen III includes the hemostasis- binding sites on Monomer 2, and MLBR3 includes the cell interaction domain and several neighboring functional domains on Monomer 3. This contrasts with collagen I where MLBR2 (with the greatest number of ligand binding sites) is located on Monomer 4 and MLBR3 on Monomer 5.

Notably, only collagen I has binding sites for the ligands involved in mineralization: cartilage oligomeric matrix protein (COMP or thrombospondin 5) and phosphophoryn. Where

collagen III is increased in diseased bone or dentin, the tissues are looser, less-mineralized, and heal poorly after fractures [68].

**Cell interaction domain.** The cell interaction domain of collagen III is relatively exposed and allows access to cell surface integrins [21,22], e.g.,  $\alpha 2\beta 1$  integrin ligation of GAPGER [61], as well as sites for platelets and LAIR-collagen ligation (8,9). Collagen I has an  $\alpha 1\beta 1/\alpha 2\beta 1$  integrin binding site, GFPGER, at this location.

**Gap.** The gap zone of collagen III is near the MMP/enzyme cleavage domain, and access to binding sites and functional domains therein may be prevented or modulated to a great extent by collagen III's C-terminus, whereas its removal could allow access.

**MMP/enzyme cleavage domain.** This domain falls near the overlap/gap zone border. Generally, enzyme access to the collagen III fibril is limited except at this location, where a paucity of imino acids is proposed to yield a looser fibril structure to facilitate enzyme access [69].

**Fibrillogenesis domain.** Sites involved in protein folding, fibrillogenesis and fibril stabilization in both collagen III and I include the C-proteinase cleavage motif, disulfide knot, and cross-link sites.

## Functional and disease-associated sites

The interactome had sites that reflected the normal biological functions of collagen III (hemostasis, angiogenesis, cell-binding) and that were associated with disease (infection, ageing and diabetes, and genetic variants).

**Hemostasis regulatory domains.** The Hemostasis domain 1 is located on M2 in band d, and binds von Willebrand factor (vWF), and salivary proteins from biting insects.

vWF mediates platelet adhesion in damaged vessels, and has been implicated in angiogenesis, and cancer spread [70]. It may normally be inaccessible because of proteoglycan binding, but trauma may expose the subendothelial collagen III binding sites. vWF binds to a highly conserved motif (RGQPGVMGF) in collagen III and I [71]. After circulating platelets attach to collagen III through the vWF binding site, they may then engage the glycoprotein VI and integrin  $\alpha 2\beta 1$  sites.

This domain also contains motifs for the mosquito salivary protein aegyptin [72, 73], and salivary products from ticks and leeches (aegyptin, calin and rLAPP) [74]. Some of these anti-coagulants bind directly to collagen III, blocking vWF binding [74,75]. Aegyptin from the salivary gland of *Aedes aegypti* mosquito binds to RGQPGVMGF in collagen III.

There is a predicted motif (PKGND) similar to the platelet fibrinogen receptor binding site (Glycoprotein IIb/IIIa) (PXXXD) near the vWF binding site and a PAGKD motif near the integrin  $\alpha 2\beta 1$  motif [76].

The Hemostasis domain 2 is located at the cell interaction domain, with binding sites for platelet receptors, glycoprotein VI, Type III collagen binding protein (TIIICBP; kindlin) and integrin  $\alpha 2\beta 1$  [75]. Glycoprotein VI and kindlin-3 are involved in platelet glycoprotein IIb/IIIa activation.

Glycoprotein VI binding may be mediated by (GPO)<sub>4</sub> at the N-terminus of collagen III where P hydroxylation is required for binding [77] but fewer repeats and interrupted sequences, such as GPOGPEGGKGAAGPOGPO, are also effective [78]. Both sites are recognized by LAIR1.

Collagen III binding protein (TIIICBP; kindlin-3) may recognize a linear platelet-binding octapeptide (KOGEOGPK), located between the vWF and integrin-binding motifs [79], but its biological relevance is controversial.

The integrin  $\alpha 2\beta 1$  (Gp Ia/IIa) binding site facilitates platelet binding and activation through other receptors, notably Glycoprotein VI.



**Angiogenesis.** Major binding sites involved in angiogenesis are the same for collagen III and collagen I, and include sites for heparin,  $\alpha 2\beta 1$  integrin, vWF, fibronectin and decorin [80, 81]. All these except for vWF are located at the C-terminus, and are absent from the terminal fragment after MMP cleavage.

**Infection.** Collagen III has binding sites for a number of mainly adhesive proteins produced by infectious organisms. These include langerin, a C-type lectin, which binds to ASQ-NITYHCKNS, a motif found in collagen III and I propeptides [82]. Collagen III is also susceptible to cleavage by *Clostridium* collagenase at the LGPA motif [83, 84]. *Yersinia* adhesion A (YadA) binds to (GPO)<sub>3</sub> [85] and several other sites, especially those rich in imino acids, or hydrophobic or uncharged residues. Aegyptin, calin and rLAPP, from mosquitoes, ticks and leeches, all bind at different collagen III sites.

Collagen III also binds to macrophage infectivity potentiator protein (MIP) which is a *Legionella* virulence factor [86]. Its binding site in collagen III site is unknown, and the motif in collagen IV (CPSGWS) is not present in collagen III [87].

**Ageing and diabetes.** Collagens are long-lived proteins, and glycation in diabetes and ageing results in advanced glycation end-products, stiffness and interference with ligand binding [11]. The main glycation sites in collagen I localize to K residues within bands c and d [22,88], which are conserved in collagen III. Glycation at residue 431 potentially interferes with the binding of vWF, SPARC and DDRs.

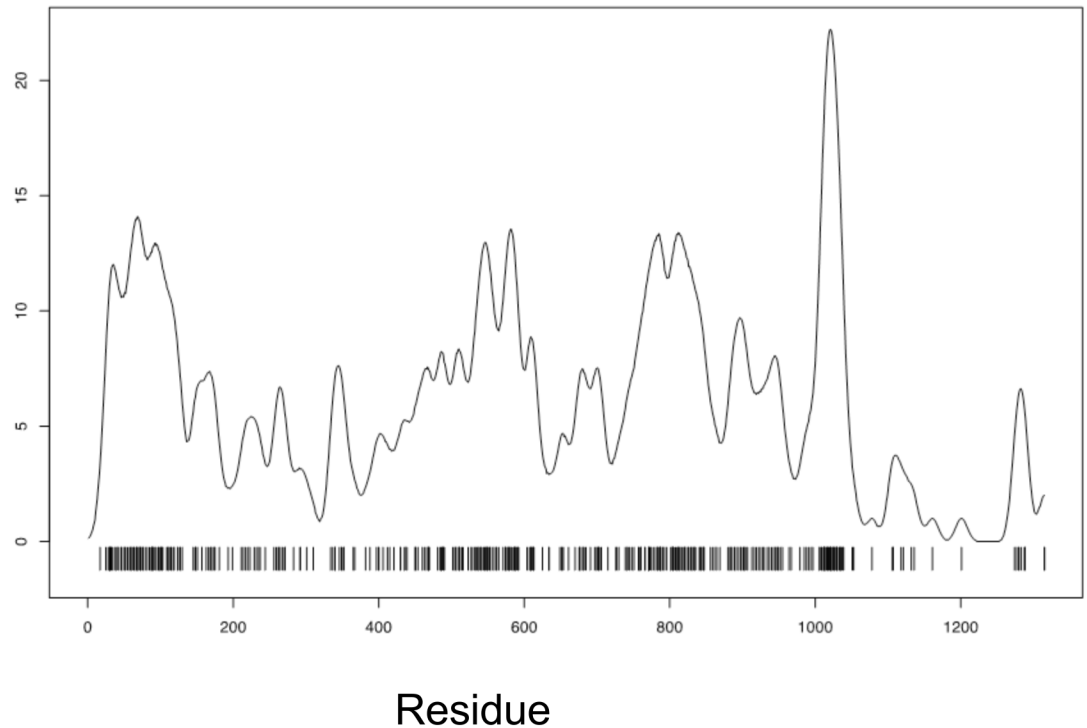
**Genetic variants.** There were 396 different missense variants in the *COL3A1* variant database affecting residues in the collagen III D-period. There were also three nonsense variants and a two-base deletion all in the terminal exon. This corresponded to a total of 400 variants at 254 unique locations affecting the D-period monomers.

Ten thousand random variant maps were generated each with 400 variants, randomly distributed throughout the collagen III  $\alpha 1$  chain. The highest variant frequency peak was then calculated from each random map, and compared with the highest mutation frequency peak in the observed variant map (22.2). There were more frequent variants in residues 1000–1050 when repeated variants at the same residues were considered (Fig 5). This was also true for collagen I. None of the 1000 random maps had such a high peak variant frequency, and the observed variant peak was very significant, with a p value of < 0.0001. This increase was not significant when only unique locations were considered.

There were also fewer variants in the C-propeptide residues 1068 to 1315 compared with the randomly-generated variant maps ( $p < 0.0001$ ). None of the other 10,000 randomly-generated maps had a 250 residue window with 16 or fewer variants. There are fewer sequence variants in the C-propeptides than expected by chance ( $p < 0.0001$ ). This is also true for collagen I. The collagen III C-propeptide is critical in chain aggregation prior to triple-helix formation. Two chain-recognition sequences of 12 and 3 amino acids ensure that only procollagen III participates in triple helix formation. The sequences contribute to a complex three-dimensional structure comprising helices,  $\beta$ -strands and turns [30]. The C-propeptide includes a single N-linked glycosylation site, eight conserved C residues essential for the inter-chain disulfide bonds, and six residues for  $\text{Ca}^{2+}$  binding. Together, these represent 157 of the 245 C-propeptide residues. Counterintuitively, most of the few disease-causing variants in the C-propeptide do not affect residues in identifiable structural or functional motifs (S2 Table). However, amino acid substitutions, for example, at propeptide positions 92, 211 or 219, might disrupt the three-dimensional structure [89] but their pathogenicity is still unproven. Thus, substitutions at some residues in the C-propeptide may result in EDS, some in perinatal lethality, and others in a minimal or undetectable phenotype.

For both collagen III and I, the increased numbers of missense variants in the Cell interaction domain and C-terminus were congruent [21,22], but there was no obvious correlation between the apparently non-random distributions of variants and functionally important sites.

Mutation  
frequency  
(Gaussian  
smoother)



**Fig 5. Missense mutation distribution in collagen III.** Missense mutation frequency along the residues of the collagen III chain. This demonstrates the non-random distribution of variation.

<https://doi.org/10.1371/journal.pone.0175582.g005>

This study examined the effect of variant location, rather than variant type, on clinical phenotype. Already *COL3A1* variant type (G substitutions, splice site mutations, indels) are known to affect disease severity, being associated with an earlier age at onset of the first major complication [90].

Bruising is common with *COL3A1* missense variants, but massive hemorrhage is rare except with viscus rupture with vascular EDS, despite the two hemostasis domains found in collagen III. Collagen I also has a site for vWF, but binding is low affinity and bleeding does not occur with *COL1A1* variants.

Some *COL3A1* variants are associated with particular clinical phenotypes. *COL3A1* missense variants in residues 652–925 have been reported more often with acrogeria [91]. The p. A698T [18,20] near the glycoprotein VI binding motif and the cell interaction domain has been associated with bleeding cerebral aneurysms or pelvic organ prolapse [92]. Gastroesophageal reflux and hiatus hernia may both be linked to *COL3A1* [19] although no allele has yet been identified.

Other mutant genes that produce a vascular EDS-like phenotype often encode binding partners of collagen III. These include the collagen I  $\alpha 1$  and  $\alpha 2$  chains, collagen V  $\alpha 1$  and  $\alpha 2$  chains, tenascin X, lysyl hydroxylase and ADAMTS2.

## Discussion

The collagen III interactome demonstrates an almost identical arrangement of structural and functional domains as collagen I. Both D-periods have the same number and spacing of intermolecular crosslink sites, and the location of charged residues, and hence fibril banding, is largely congruent. In addition, many of the major structural and functional domains (cell

interaction domains, fibrillogenesis and enzyme cleavage sites, and major ligand-binding regions) were found in the same locations. These similarities enable collagen III and I to assemble in register to produce heterofibrils and carry out common biologic functions. They also enable collagen III to be replaced by collagen I in embryogenesis and wound healing. Yet, a major difference between collagen I and III identified here relates to the potential for greater flexibility of collagen III, which makes it an ideal structural component of embryonic tissues, early wound healing, and distensible organs in the adult (vasculature, uterus, small bowel).

In some tissues the retention of the N-propeptide by collagen III is consistent with its inability to extend laterally, and its fibrils being smaller and more peripherally located in the heterotypic fibril than those of collagen I.

Clusters of disease-causing missense variants that may reflect residues necessary for molecular folding, fibril assembly, or ligand interactions, as well as variant-free regions map to the same locations in both collagens III and I [22]. Yet, the clinical phenotypes associated with variants in the collagen III and I genes are very different. Osteogenesis imperfecta is characterized by bone fragility, and vascular EDS by distensible organ rupture, acrogeria and bleeding. This suggests different tissue expression patterns and binding partners. Thus, for example, collagen I binds to proteins that are required for biomineralization, with the overlap zone of collagen I incorporating binding sites for COMP and phosphophoryn that are not present in collagen III.

The organ rupture seen in vascular EDS may result from increased collagen III in large arteries and hollow organs rather than altered elastic properties since collagen III does not bind to elastin [16]. The clinical features of acrogeria probably relate to reduced collagen III in subcutaneous tissues [93]. An association with uterine prolapse, hiatus hernia and gastroesophageal reflux, if substantiated, may reflect less abundant collagen III and increased tissue laxity [19,20]. Cumulative 'wear- and- tear' in an organ structurally dependent on collagen III may contribute to rupture in adult life.

Although both collagens III and I bind vWF, only *COL3A1* variants commonly produce bleeding. This is probably because collagen III has more hemostatic ligand binding sites and a higher affinity for vWF [71], and is more abundant in the vascular sub-endothelium and more accessible on the outer fibril surface. The collagen III  $\alpha 1$  chain has several closely- related binding motifs for platelet proteins that are critical in platelet immobilization and located in the Cell interaction domain. These sites have not been demonstrated in type I collagen. In addition, the collagen III  $\alpha 1$  hemostasis domain is about 100 residues from the vWF binding motif, and platelets bound to the Cell interaction domain may bind simultaneously to the large vWF multimer.

In conclusion, interactomes summarize our understanding of a molecule's structure and function, suggest further interactions and roles, and help explain how missense variants produce clinical phenotypes. The collagen III interactome emphasizes its resemblance to collagen I, indicates how its architecture confers flexibility, and explains its role in hemostasis. These results may also have practical applications in the design of bioactive yet flexible extracellular matrix scaffolds for a variety of uses in medical devices.

## Experimental procedures

### Construction of the collagen III interactome

We constructed the interactome of a collagen type III D-period, examined its structural features, charge distribution, ligand-binding sites, and missense sequence variant distribution, and compared these features with those for collagen I.

The reference sequences for the human pro-collagen  $\alpha 1$ (III) (UniProt P02461) and collagen  $\alpha 1$ (I) and  $\alpha 2$ (I) chains (UniProt P02452 and P08123 respectively) were used.

The collagen  $\alpha 1$ (III) chain is 1466 amino acids, with a signal peptide (23 aa), N-terminal propeptide (130 aa), collagenous sequence (1068 aa), and C-terminal propeptide (245 aa). Here we renumbered the amino acid from one D-period as 1 to 1068, with N- and C-telopeptides (14 and 25 aa respectively) and a triple helix (1029 aa) [21] corresponding to amino acids 154 to 1221 in the reference sequence. Numbers from the interactome [Reference sequence number -153], were used to describe variants so that numbering began at the start of the mature protein.

### Alignment of charge residues in collagen III and collagen I interactomes

The cross-link K residues in the collagen  $\alpha 1$ (III) chain were aligned with those in the collagen  $\alpha 1$ (I) and  $\alpha 2$ (I) chains, and the D-periods examined for positively- (R, K) and negatively-charged (E, D) residues.

### Analysis of fibril stability and structure

The distribution of atypical amino acid triplets, that confer flexibility such as GAA or GGY, and of GPP-rich regions that promote rigidity, was examined for randomness. The relationship of atypical and GPP triplets was then examined in the overlapping D-period regions. The numbers of atypical and GPP triplets were counted in ten 'bins' of equal size, starting from the N-terminus, summed over the D-period and compared.

The location of regions of high and low-stability within the collagen III D-period were examined using the Collagen Stability Calculator <http://compbio.cs.princeton.edu/csc/>[94].

### Sites related to structure, assembly, turnover, modification, cleavage and ligand-binding

Binding sites and structural motifs described for collagen III were identified using the search terms 'collagen III', 'ligand', 'binding partner' etc, from the scientific literature, open access web sites (UniProt, UCSC, Biogrid, Reactome, STRING, MINT, IntAct, and Ex-PASy Peptide Cutter, see later for websites), and the collagen I interactomes (.) [21,22].

In general, previously-reported ligand-binding sites were derived from experiments using the native collagen III molecule, triple helical mimetic peptides or fragments thereof, or from measurements based on rotary shadowing electron microscopy of type III collagen-ligand complexes assuming that type III collagen residues were spaced an average of 0.286 nm apart [95]. Binding sites on fragments and peptides derived from type III collagen were included because, *in vivo*, some ligands bind only to the denatured collagen.

The collagen III reference sequence was also examined for short motifs with biological functions described in other proteins (ELM, MnM databases).

### Functional and disease-associated sites

The map was examined for sites that reflected the normal biological functions of collagen III (hemostasis, angiogenesis, cell-binding, etc) and disease associations (infection, glycation, inherited disease).

Non-synonymous DNA missense variants were identified from a search of the open access COL3A1 web databases ([http://eds.gene.le.ac.uk/home.php?select\\_db=COL3A1](http://eds.gene.le.ac.uk/home.php?select_db=COL3A1)) on 6 July 2016). The locations of missense variants was examined for randomness. A Gaussian kernel smoother was used to calculate a smoothed variant frequency at each residue in the coding

sequence corresponding to the D-period [96]. Ten thousand random variant maps were generated using the number of unique variant locations in the protein sequence. The peaks found using the data from the *COL3A1* variant database distribution were compared with proportion of maps with a randomly-generated variant peak of this magnitude [97]. A similar randomization analysis was performed to examine the lack of variants in any region. These analyses were performed using the statistical software R [98].

## Supporting information

**S1 Table. Numbering systems for reference sequence and mature collagen III.** This compares the systematic numbering and numbering from the start of the mature collagen as shown in the interactome.

(DOCX)

**S2 Table. C-propeptide features and sequence variants.** This indicates features in the C-propeptide which is cleaved from the collagen III chain before it forms the interactome and the effect of pathogenic sequence variants.

(DOCX)

## Acknowledgments

We are grateful to Dr Richard Farndale provided novel unpublished ligands and their binding motifs on collagen III.

## Author Contributions

**Conceptualization:** JDP JDS JAS.

**Data curation:** JDP JDS AP HD RD STJ XJ JS.

**Formal analysis:** JDP JDS AP STJ JS.

**Investigation:** JDP JDS AP STJ JS.

**Methodology:** JDP JDS AP STJ JS.

**Project administration:** JDP JDS AP HD RD STJ XJ JS.

**Resources:** JDP JDS AP RD STJ XJ JS.

**Software:** STJ RD.

**Supervision:** JDS JS.

**Validation:** JDS JS JDP.

**Visualization:** JDS JS JDP.

**Writing – original draft:** JDS JS JDP.

**Writing – review & editing:** JDP JDS AP HD RD STJ XJ JS.

## References

1. Ricard-Blum S (2011) The collagen family. *Cold Spring Harb Perspect Biol* 3: a004978. <https://doi.org/10.1101/cshperspect.a004978> PMID: 21421911
2. Rich A, Crick FH (1961) The molecular structure of collagen. *J Mol Biol* 3: 483–506. PMID: 14491907



3. Meek KM, Chapman JA, Hardcastle RA (1979) The staining pattern of collagen fibrils. Improved correlation with sequence data. *J Biol Chem* 254: 10710–10714. PMID: [91606](#)
4. Chapman JA (1974) The staining pattern of collagen fibrils. I. An analysis of electron micrographs. *Connect Tissue Res* 2: 137–150. PMID: [4138005](#)
5. Fleischmajer R, MacDonald ED, Perlish JS, Burgeson RE, Fisher LW (1990) Dermal collagen fibrils are hybrids of type I and type III collagen molecules. *J Struct Biol* 105: 162–169. PMID: [2100147](#)
6. Takahara K, Hoffman GG, Greenspan DS (1995) Complete structural organization of the human alpha 1 (V) collagen gene (COL5A1): divergence from the conserved organization of other characterized fibrillar collagen genes. *Genomics* 29: 588–597. <https://doi.org/10.1006/geno.1995.9961> PMID: [8575750](#)
7. Wu JJ, Weis MA, Kim LS, Eyre DR (2010) Type III collagen, a fibril network modifier in articular cartilage. *J Biol Chem* 285: 18537–18544. <https://doi.org/10.1074/jbc.M110.112904> PMID: [20404341](#)
8. Wolman M, Kasten FH (1986) Polarized light microscopy in the study of the molecular structure of collagen and reticulin. *Histochemistry* 85: 41–49. PMID: [3733471](#)
9. Keene DR, Sakai LY, Bachinger HP, Burgeson RE (1987) Type III collagen can be present on banded collagen fibrils regardless of fibril diameter. *J Cell Biol* 105: 2393–2402. PMID: [2445760](#)
10. Miller EJ, Furuto DK, Narkates AJ (1991) Quantitation of type I, III, and V collagens in human tissue samples by high-performance liquid chromatography of selected cyanogen bromide peptides. *Anal Biochem* 196: 54–60. PMID: [1888036](#)
11. Avery NC, Bailey AJ (2006) The effects of the Maillard reaction on the physical properties and cell interactions of collagen. *Pathol Biol (Paris)* 54: 387–395.
12. Kanzawa S, Endo H, Shioya N (1993) Improved in vitro angiogenesis model by collagen density reduction and the use of type III collagen. *Ann Plast Surg* 30: 244–251. PMID: [7684210](#)
13. Goldberg SR, Quirk GL, Sykes VW, Kordula T, Lanning DA (2007) Altered procollagen gene expression in mid-gestational mouse excisional wounds. *J Surg Res* 143: 27–34. <https://doi.org/10.1016/j.jss.2007.05.013> PMID: [17950069](#)
14. Puchtler H (1964) On the Original Definition of the Term "Reticulin". *J Histochem Cytochem* 12: 552. <https://doi.org/10.1177/12.7.552> PMID: [14209992](#)
15. Volk SW, Wang Y, Mauldin EA, Liechty KW, Adams SL (2011) Diminished type III collagen promotes myofibroblast differentiation and increases scar deposition in cutaneous wound healing. *Cells Tissues Organs* 194: 25–37. <https://doi.org/10.1159/000322399> PMID: [21252470](#)
16. Beighton P, De Paepe A, Steinmann B, Tsipouras P, Wenstrup RJ (1998) Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). *Am J Med Genet* 77: 31–37. PMID: [9557891](#)
17. Pope FM, Narcisi P, Nicholls AC, Germaine D, Pals G, Richards AJ (1996) COL3A1 mutations cause variable clinical phenotypes including acrogeria and vascular rupture. *Br J Dermatol* 135: 163–181. PMID: [8881656](#)
18. Chen J, Zhu Y, Jiang Y, Yu H, Sun K, Song W, et al. (2012) A functional variant of the collagen type III alpha1 gene modify risk of sporadic intracranial aneurysms. *Hum Genet* 131: 1137–1143. <https://doi.org/10.1007/s00439-012-1138-6> PMID: [22241462](#)
19. Asling B, Jirholt J, Hammond P, Knutsson M, Walentinsson A, Davidson G, et al. (2009) Collagen type III alpha 1 is a gastro-oesophageal reflux disease susceptibility gene and a male risk factor for hiatus hernia. *Gut* 58: 1063–1069. <https://doi.org/10.1136/gut.2008.167353> PMID: [19398442](#)
20. Kluivers KB, Dijkstra JR, Hendriks JC, Lince SL, Vierhout ME, van Kempen LC (2009) COL3A1 2209G>A is a predictor of pelvic organ prolapse. *Int Urogynecol J Pelvic Floor Dysfunct* 20: 1113–1118. <https://doi.org/10.1007/s00192-009-0913-y> PMID: [19444361](#)
21. Di Lullo GA, Sweeney SM, Korkko J, Ala-Kokko L, San Antonio JD (2002) Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J Biol Chem* 277: 4223–4231. <https://doi.org/10.1074/jbc.M110709200> PMID: [11704682](#)
22. Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, et al. (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J Biol Chem* 283: 21187–21197. <https://doi.org/10.1074/jbc.M709319200> PMID: [18487200](#)
23. Parkin JD, San Antonio JD, Pedchenko V, Hudson B, Jensen ST, Savige J (2011) Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum Mutat* 32: 127–143. <https://doi.org/10.1002/humu.21401> PMID: [21280145](#)
24. Bodian DL, Klein TE (2009) COLdb, a database linking genetic data to molecular function in fibrillar collagens. *Hum Mutat* 30: 946–951. <https://doi.org/10.1002/humu.20978> PMID: [19370761](#)

25. Weis MA, Hudson DM, Kim L, Scott M, Wu JJ, Eyre DR (2010) Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J Biol Chem* 285: 2580–2590. <https://doi.org/10.1074/jbc.M109.068726> PMID: 19940144
26. EJ M, editor (1984) *Chemistry of the collagens and their distribution*. New York: Elsevier pp41–82
27. Brodsky B, Persikov AV (2005) Molecular structure of the collagen triple helix. *Adv Protein Chem* 70: 301–339. [https://doi.org/10.1016/S0065-3233\(05\)70009-7](https://doi.org/10.1016/S0065-3233(05)70009-7) PMID: 15837519
28. Persikov AV, Ramshaw JA, Brodsky B (2000) Collagen model peptides: Sequence dependence of triple-helix stability. *Biopolymers* 55: 436–450. [https://doi.org/10.1002/1097-0282\(2000\)55:6<436::AID-BIP1019>3.0.CO;2-D](https://doi.org/10.1002/1097-0282(2000)55:6<436::AID-BIP1019>3.0.CO;2-D) PMID: 11304671
29. Persikov AV, Brodsky B (2002) Unstable molecules form stable tissues. *Proc Natl Acad Sci U S A* 99: 1101–1103. <https://doi.org/10.1073/pnas.042707899> PMID: 11830649
30. Bourhis JM, Mariano N, Zhao Y, Harlos K, Exposito JY, Jones EY, et al. (2012) Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat Struct Mol Biol* 19: 1031–1036. <https://doi.org/10.1038/nsmb.2389> PMID: 23001006
31. Boot-Handford RP, Tuckwell DS (2003) Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *Bioessays* 25: 142–151. <https://doi.org/10.1002/bies.10230> PMID: 12539240
32. Murwantoko, Yano M, Ueta Y, Murasaki A, Kanda H, Oka C, et al. (2004) Binding of proteins to the PDZ domain regulates proteolytic activity of HtrA1 serine protease. *Biochem J* 381: 895–904. <https://doi.org/10.1042/BJ20040435> PMID: 15101818
33. Dong Z, Lewis RV, Middaugh CR (1991) Molecular mechanism of spider silk elasticity. *Arch Biochem Biophys* 284: 53–57. PMID: 1989503
34. Kar K, Ibrar S, Nanda V, Getz TM, Kunapuli SP, Brodsky B (2009) Aromatic interactions promote self-association of collagen triple-helical peptides to higher-order structures. *Biochemistry* 48: 7959–7968. <https://doi.org/10.1021/bi900496m> PMID: 19610672
35. Brondijk TH, Bihan D, Farndale RW, Huizinga EG (2012) Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc Natl Acad Sci U S A* 109: 5253–5258. <https://doi.org/10.1073/pnas.1112388109> PMID: 22440751
36. Hohenester E, Sasaki T, Giudici C, Farndale RW, Bachinger HP (2008) Structural basis of sequence-specific collagen recognition by SPARC. *Proc Natl Acad Sci U S A* 105: 18273–18277. <https://doi.org/10.1073/pnas.0808452105> PMID: 19011090
37. Carafoli F, Bihan D, Stathopoulos S, Konitsiotis AD, Kvsankul M, Farndale RW, et al. (2009) Crystallographic insight into collagen recognition by discoidin domain receptor 2. *Structure* 17: 1573–1581. <https://doi.org/10.1016/j.str.2009.10.012> PMID: 20004161
38. Boudko SP, Engel J, Okuyama K, Mizuno K, Bachinger HP, Schumacher MA (2008) Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J Biol Chem* 283: 32580–32589. <https://doi.org/10.1074/jbc.M805394200> PMID: 18805790
39. Seyer JM, Kang AH (1977) Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides from the amino-terminal segment of type III collagen of human liver. *Biochemistry* 16: 1158–1164. PMID: 557335
40. Kronenberg D, Bruns BC, Moali C, Vadon-Le Goff S, Sterchi EE, Traupe H, et al. (2010) Processing of procollagen III by mepriins: new players in extracellular matrix assembly? *J Invest Dermatol* 130: 2727–2735. <https://doi.org/10.1038/jid.2010.202> PMID: 20631730
41. Nagase H (2001) Metalloproteases. *Curr Protoc Protein Sci Chapter 21: Unit 21 24*.
42. Bigg HF, Rowan AD, Barker MD, Cawston TE (2007) Activity of matrix metalloproteinase-9 against native collagen types I and III. *FEBS J* 274: 1246–1255. <https://doi.org/10.1111/j.1742-4658.2007.05669.x> PMID: 17298441
43. Birkedal-Hansen H, Taylor RE, Bhowan AS, Katz J, Lin HY, Wells BR (1985) Cleavage of bovine skin type III collagen by proteolytic enzymes. Relative resistance of the fibrillar form. *J Biol Chem* 260: 16411–16417. PMID: 3905816
44. Goldfarb RH, Liotta LA (1986) Thrombin cleavage of extracellular matrix proteins. *Ann N Y Acad Sci* 485: 288–292. PMID: 2952037
45. Widmer C, Gebauer JM, Brunstein E, Rosenbaum S, Zaucke F, Drogemuller C, et al. (2012) Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc Natl Acad Sci U S A* 109: 13243–13247. <https://doi.org/10.1073/pnas.1208072109> PMID: 22847422
46. Giudici C, Raynal N, Wiedemann H, Cabral WA, Marini JC, Timpl R, et al. (2008) Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. *J Biol Chem* 283: 19551–19560. <https://doi.org/10.1074/jbc.M710001200> PMID: 18487610

47. Ruiz PA, Jarai G (2011) Collagen I induces discoidin domain receptor (DDR) 1 expression through DDR2 and a JAK2-ERK1/2-mediated mechanism in primary human lung fibroblasts. *J Biol Chem* 286: 12912–12923. <https://doi.org/10.1074/jbc.M110.143693> PMID: 21335558
48. Scott JE, Glanville RW (1993) Homologous sequences in fibrillar collagens may be proteoglycan binding sites. *Biochem Soc Trans* 21: 123S. PMID: 8359379
49. Orgel JP, Eid A, Antipova O, Bella J, Scott JE (2009) Decorin core protein (decoron) shape complements collagen fibril surface structure and mediates its binding. *PLoS One* 4: e7028. <https://doi.org/10.1371/journal.pone.0007028> PMID: 19753304
50. Pellegrini L, Burke DF, von Delft F, Mulloy B, Blundell TL (2000) Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* 407: 1029–1034. <https://doi.org/10.1038/35039551> PMID: 11069186
51. Guan M, Pang CP, Yam HF, Cheung KF, Liu WW, Lu Y (2004) Inhibition of glioma invasion by overexpression of pigment epithelium-derived factor. *Cancer Gene Ther* 11: 325–332. <https://doi.org/10.1038/sj.cgt.7700675> PMID: 15044958
52. Sekiya A, Okano-Kosugi H, Yamazaki CM, Koide T (2011) Pigment epithelium-derived factor (PEDF) shares binding sites in collagen with heparin/heparan sulfate proteoglycans. *J Biol Chem* 286: 26364–26374. <https://doi.org/10.1074/jbc.M111.252684> PMID: 21652703
53. Schonherr E, Witsch-Prehm P, Harrach B, Robenek H, Rauterberg J, Kresse H (1995) Interaction of biglycan with type I collagen. *J Biol Chem* 270: 2776–2783. PMID: 7852349
54. Fiedler LR, Eble JA (2009) Decorin regulates endothelial cell-matrix interactions during angiogenesis. *Cell Adh Migr* 3: 3–6. PMID: 19372733
55. Romanic AM, Adachi E, Kadler KE, Hojima Y, Prockop DJ (1991) Copolymerization of pNcollagen III and collagen I. pNcollagen III decreases the rate of incorporation of collagen I into fibrils, the amount of collagen I incorporated, and the diameter of the fibrils formed. *J Biol Chem* 266: 12703–12709. PMID: 2061335
56. Young RD, Lawrence PA, Duance VC, Aigner T, Monaghan P (2000) Immunolocalization of collagen types II and III in single fibrils of human articular cartilage. *J Histochem Cytochem* 48: 423–432. <https://doi.org/10.1177/002215540004800312> PMID: 10681396
57. Liden A, van Wieringen T, Lannergard J, Kassner A, Heinegard D, Reed RK, et al. (2008) A secreted collagen- and fibronectin-binding streptococcal protein modulates cell-mediated collagen gel contraction and interstitial fluid pressure. *J Biol Chem* 283: 1234–1242. <https://doi.org/10.1074/jbc.M704827200> PMID: 18003607
58. Galvin NJ, Vance PM, Dixit VM, Fink B, Frazier WA (1987) Interaction of human thrombospondin with types I-V collagen: direct binding and electron microscopy. *J Cell Biol* 104: 1413–1422. PMID: 3571333
59. Narouz-Ott L, Maurer P, Nitsche DP, Smyth N, Paulsson M (2000) Thrombospondin-4 binds specifically to both collagenous and non-collagenous extracellular matrix proteins via its C-terminal domains. *J Biol Chem* 275: 37110–37117. <https://doi.org/10.1074/jbc.M007223200> PMID: 10956668
60. Kim JK, Xu Y, Xu X, Keene DR, Gurusiddappa S, Liang X, et al. (2005) A novel binding site in collagen type III for integrins alpha1beta1 and alpha2beta1. *J Biol Chem* 280: 32512–32520. <https://doi.org/10.1074/jbc.M502431200> PMID: 16043429
61. Siljander PR, Hamaia S, Peachey AR, Slatter DA, Smethurst PA, Ouwehand WH, et al. (2004) Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J Biol Chem* 279: 47763–47772. <https://doi.org/10.1074/jbc.M404685200> PMID: 15345717
62. Raynal N, Hamaia SW, Siljander PR, Maddox B, Peachey AR, Fernandez R, et al. (2006) Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *J Biol Chem* 281: 3821–3831. <https://doi.org/10.1074/jbc.M509818200> PMID: 16326707
63. Hamaia SW, Pugh N, Raynal N, Nemoz B, Stone R, Gullberg D, et al. (2012) Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin alpha1beta1 using collagen toolkits II and III. *J Biol Chem* 287: 26019–26028. <https://doi.org/10.1074/jbc.M112.353144> PMID: 22654115
64. Barczyk M, Carracedo S, Gullberg D (2010) Integrins. *Cell Tissue Res* 339: 269–280. <https://doi.org/10.1007/s00441-009-0834-6> PMID: 19693543
65. Lebbink RJ, de Ruiter T, Adelmeijer J, Brenkman AB, van Helvoort JM, Koch M, et al. (2006) Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *J Exp Med* 203: 1419–1425. <https://doi.org/10.1084/jem.20052554> PMID: 16754721
66. Lebbink RJ, van den Berg MC, de Ruiter T, Raynal N, van Roon JA, Lenting PJ, et al. (2008) The soluble leukocyte-associated Ig-like receptor (LAIR)-2 antagonizes the collagen/LAIR-1 inhibitory immune interaction. *J Immunol* 180: 1662–1669. PMID: 18209062

67. Barrow AD, Raynal N, Andersen TL, Slatter DA, Bihan D, Pugh N, et al. (2011) OSCAR is a collagen receptor that costimulates osteoclastogenesis in DAP12-deficient humans and mice. *J Clin Invest* 121: 3505–3516. <https://doi.org/10.1172/JCI145913> PMID: 21841309
68. Lawton DM, Andrew JG, Marsh DR, Hoyland JA, Freemont AJ (1997) Mature osteoblasts in human non-union fractures express collagen type III. *Mol Pathol* 50: 194–197. PMID: 9350302
69. Mainardi CL, Hasty DL, Seyer JM, Kang AH (1980) Specific cleavage of human type III collagen by human polymorphonuclear leukocyte elastase. *J Biol Chem* 255: 12006–12010. PMID: 6254989
70. Luo GP, Ni B, Yang X, Wu YZ (2012) Von Willebrand factor: more than a regulator of hemostasis and thrombosis. *Acta Haematol* 128: 158–169. <https://doi.org/10.1159/000339426> PMID: 22890291
71. Lisman T, Raynal N, Groeneveld D, Maddox B, Peachey AR, Huizinga EG, et al. (2006) A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* 108: 3753–3756. <https://doi.org/10.1182/blood-2006-03-011965> PMID: 16912226
72. Calvo E, Tokumasu F, Marinotti O, Villeval JL, Ribeiro JM, Francischetti IM (2007) Aegyptin, a novel mosquito salivary gland protein, specifically binds to collagen and prevents its interaction with platelet glycoprotein VI, integrin alpha2beta1, and von Willebrand factor. *J Biol Chem* 282: 26928–26938. <https://doi.org/10.1074/jbc.M705669200> PMID: 17650501
73. Calvo E, Tokumasu F, Mizurini DM, McPhie P, Narum DL, Ribeiro JM, et al. (2010) Aegyptin displays high-affinity for the von Willebrand factor binding site (RGQOGVMGF) in collagen and inhibits carotid thrombus formation in vivo. *FEBS J* 277: 413–427. <https://doi.org/10.1111/j.1742-4658.2009.07494.x> PMID: 20015075
74. Harsfalvi J, Stassen JM, Hoylaerts MF, Van Houtte E, Sawyer RT, Vermynen J, et al. (1995) Calin from *Hirudo medicinalis*, an inhibitor of von Willebrand factor binding to collagen under static and flow conditions. *Blood* 85: 705–711. PMID: 7833474
75. Depraetere H, Kerekes A, Deckmyn H (1999) The collagen-binding leech products rLAPP and calin prevent both von Willebrand factor and alpha2beta1 (GPIIb/IIIa)-I-domain binding to collagen in a different manner. *Thromb Haemost* 82: 1160–1163. PMID: 10494781
76. Hayashi Y, Katada J, Sato Y, Igarashi K, Takiguchi Y, Harada T, et al. (1998) Discovery and structure—activity relationship studies of a novel and specific peptide motif, Pro-X-X-X-Asp-X, as a platelet fibrinogen receptor antagonist. *Bioorg Med Chem* 6: 355–364. PMID: 9568289
77. Smethurst PA, Joutsu-Korhonen L, O'Connor MN, Wilson E, Jennings NS, Garner SF, et al. (2004) Identification of the primary collagen-binding surface on human glycoprotein VI by site-directed mutagenesis and by a blocking phage antibody. *Blood* 103: 903–911. <https://doi.org/10.1182/blood-2003-01-0308> PMID: 14504096
78. Jarvis GE, Raynal N, Langford JP, Onley DJ, Andrews A, Smethurst PA, et al. (2008) Identification of a major GpVI-binding locus in human type III collagen. *Blood* 111: 4986–4996. <https://doi.org/10.1182/blood-2007-08-108472> PMID: 18305222
79. Maurice P, Legrand C, Fauvel-Lafeve F (2004) Platelet adhesion and signaling induced by the octapeptide primary binding sequence (KOGEOGPK) from type III collagen. *FASEB J* 18: 1339–1347. <https://doi.org/10.1096/fj.03-1151com> PMID: 15333577
80. San Antonio JD, Zoeller JJ, Habursky K, Turner K, Pimpong W, Burrows M, et al. (2009) A key role for the integrin alpha2beta1 in experimental and developmental angiogenesis. *Am J Pathol* 175: 1338–1347. <https://doi.org/10.2353/ajpath.2009.090234> PMID: 19700757
81. Iozzo RV, San Antonio JD (2001) Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena. *J Clin Invest* 108: 349–355. <https://doi.org/10.1172/JCI13738> PMID: 11489925
82. Tada Y, Riedl E, Lowenthal MS, Liotta LA, Briner DM, Crouch EC, et al. (2006) Identification and characterization of endogenous Langerin ligands in murine extracellular matrix. *J Invest Dermatol* 126: 1549–1558. <https://doi.org/10.1038/sj.jid.5700283> PMID: 16557233
83. Hu Y, Webb E, Singh J, Morgan BA, Gainor JA, Gordon TD, et al. (2002) Rapid determination of substrate specificity of *Clostridium histolyticum* beta-collagenase using an immobilized peptide library. *J Biol Chem* 277: 8366–8371. <https://doi.org/10.1074/jbc.M111042200> PMID: 11724807
84. Eckhard U, Schonauer E, Ducka P, Briza P, Nuss D, Brandstetter H (2009) Biochemical characterization of the catalytic domains of three different Clostridial collagenases. *Biol Chem* 390: 11–18. <https://doi.org/10.1515/BC.2009.004> PMID: 18937627
85. Leo JC, Elovaara H, Bihan D, Pugh N, Kilpinen SK, Raynal N, et al. (2010) 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* 78: 3226–3236. <https://doi.org/10.1128/IAI.01057-09> PMID: 20439473
86. Wagner C, Khan AS, Kamphausen T, Schmausser B, Unal C, Lorenz U, et al. (2007) Collagen binding protein Mip enables *Legionella pneumophila* to transmigrate through a barrier of NCI-H292 lung



- epithelial cells and extracellular matrix. *Cell Microbiol* 9: 450–462. <https://doi.org/10.1111/j.1462-5822.2006.00802.x> PMID: 16953800
87. Unal C, Schwedhelm KF, Thiele A, Weiwad M, Schweimer K, Frese F, et al. (2011) Collagen IV-derived peptide binds hydrophobic cavity of *Legionella pneumophila* Mip and interferes with bacterial epithelial transmigration. *Cell Microbiol* 13: 1558–1572. <https://doi.org/10.1111/j.1462-5822.2011.01641.x> PMID: 21794054
  88. Reiser KM, Amigable MA, Last JA (1992) Nonenzymatic glycation of type I collagen. The effects of aging on preferential glycation sites. *J Biol Chem* 267: 24207–24216. PMID: 1447170
  89. Stembridge NS, Vandersteen AM, Ghali N, Sawle P, Nesbitt M, Pollitt RC, et al. (2015) Clinical, structural, biochemical and X-ray crystallographic correlates of pathogenicity for variants in the C-propeptide region of the COL3A1 gene. *Am J Med Genet A* 167A: 1763–1772. <https://doi.org/10.1002/ajmg.a.37081> PMID: 25846194
  90. Frank M, Albuissou J, Ranque B, Golmard L, Mazzella JM, Bal-Theoleyre L, et al. (2015) The type of variants at the COL3A1 gene associates with the phenotype and severity of vascular Ehlers-Danlos syndrome. *Eur J Hum Genet* 23: 1657–1664. <https://doi.org/10.1038/ejhg.2015.32> PMID: 25758994
  91. Pepin M, Schwarze U, Superti-Furga A, Byers PH (2000) Clinical and genetic features of Ehlers-Danlos syndrome type IV, the vascular type. *N Engl J Med* 342: 673–680. <https://doi.org/10.1056/NEJM200003093421001> PMID: 10706896
  92. Zafarullah K, Kleinert C, Tromp G, Kuivaniemi H, Kontusaari S, Wu YL, et al. (1990) G to A polymorphism in exon 31 of the COL3A1 gene. *Nucleic Acids Res* 18: 6180. PMID: 2235526
  93. Wu J, Hatamochi A (2004) Acrogeria with decreased gene expression of alpha1 (I) and alpha1 (III) collagen in cultured dermal fibroblasts. *J Dermatol* 31: 535–539. PMID: 15492417
  94. Persikov AV, Ramshaw JA, Brodsky B (2005) Prediction of collagen stability from amino acid sequence. *J Biol Chem* 280: 19343–19349. <https://doi.org/10.1074/jbc.M501657200> PMID: 15753081
  95. Piez KA, Reddi AH (1984) *Extracellular matrix biochemistry*. New York: Elsevier. xvi, 473 p. p.
  96. Wand MPaJ, M.C., editor (1995) *Kernel Smoothing*. London: Chapman & Hall/CRC.
  97. Edgington ES, editor (1995) *Randomization tests*. 3rd edition ed. New York: Marcel-Dekker.
  98. Team RDC, editor (2009) *R: A Language and Environment for Statistical Computing*: R Foundation for Statistical Computing.
  99. Henkel W, Glanville RW (1982) Covalent crosslinking between molecules of type I and type III collagen. The involvement of the N-terminal, nonhelical regions of the alpha 1 (I) and alpha 1 (III) chains in the formation of intermolecular crosslinks. *Eur J Biochem* 122: 205–213. PMID: 6120835
  100. Piez KA, Reddi AH (1984) *Extracellular matrix biochemistry*. New York; Oxford: Elsevier.
  101. Terajima M, Perdivara I, Sricholpech M, Deguchi Y, Pleshko N, Tomer KB, et al. (2014) Glycosylation and cross-linking in bone type I collagen. *J Biol Chem* 289: 22636–22647. <https://doi.org/10.1074/jbc.M113.528513> PMID: 24958722
  102. Koide T, Asada S, Takahara Y, Nishikawa Y, Nagata K, Kitagawa K (2006) Specific recognition of the collagen triple helix by chaperone HSP47: minimal structural requirement and spatial molecular orientation. *J Biol Chem* 281: 3432–3438. <https://doi.org/10.1074/jbc.M509707200> PMID: 16326708
  103. Halila R, Peltonen L (1986) Purification of human procollagen type III N-proteinase from placenta and preparation of antiserum. *Biochem J* 239: 47–52. PMID: 3099780
  104. Wang WM, Lee S, Steiglitz BM, Scott IC, Lebares CC, Allen ML, et al. (2003) Transforming growth factor-beta induces secretion of activated ADAMTS-2. A procollagen III N-proteinase. *J Biol Chem* 278: 19549–19557. <https://doi.org/10.1074/jbc.M300767200> PMID: 12646579
  105. Miller EJ, Harris ED Jr., Chung E, Finch JE Jr., McCroskery PA, Butler WT (1976) Cleavage of Type II and III collagens with mammalian collagenase: site of cleavage and primary structure at the NH<sub>2</sub>-terminal portion of the smaller fragment released from both collagens. *Biochemistry* 15: 787–792. PMID: 174719
  106. Shi L, Ermis R, Garcia A, Telgenhoff D, Aust D (2010) Degradation of human collagen isoforms by *Clostridium* collagenase and the effects of degradation products on cell migration. *Int Wound J* 7: 87–95. <https://doi.org/10.1111/j.1742-481X.2010.00659.x> PMID: 20529148
  107. Cowan KN, Leung WC, Mar C, Bhattacharjee R, Zhu Y, Rabinovitch M (2005) Caspases from apoptotic myocytes degrade extracellular matrix: a novel remodeling paradigm. *FASEB J* 19: 1848–1850. <https://doi.org/10.1096/fj.05-3706fje> PMID: 16123172
  108. Scott PG, Pearson H (1981) Cathepsin D: specificity of peptide-bond cleavage in type-I collagen and effects on type-III collagen and procollagen. *Eur J Biochem* 114: 59–62. PMID: 6783404



109. Miller EJ, Finch JE Jr., Chung E, Butler WT, Robertson PB (1976) Specific cleavage of the native type III collagen molecule with trypsin. Similarity of the cleavage products to collagenase-produced fragments and primary structure at the cleavage site. *Arch Biochem Biophys* 173: 631–637. PMID: [179470](#)
110. Wang HM, Chan J, Pettigrew DW, Sodek J (1978) Cleavage of native type III collagen in the collagenase susceptible region by thermolysin. *Biochim Biophys Acta* 533: 270–277. PMID: [205266](#)
111. San Antonio JD, Karnovsky MJ, Gay S, Sanderson RD, Lander AD (1994) Interactions of syndecan-1 and heparin with human collagens. *Glycobiology* 4: 327–332. PMID: [7949658](#)
112. Sweeney SM, Guy CA, Fields GB, San Antonio JD (1998) Defining the domains of type I collagen involved in heparin-binding and endothelial tube formation. *Proc Natl Acad Sci U S A* 95: 7275–7280. PMID: [9636139](#)
113. Meyer C, Notari L, Becerra SP (2002) Mapping the type I collagen-binding site on pigment epithelium-derived factor. Implications for its antiangiogenic activity. *J Biol Chem* 277: 45400–45407. <https://doi.org/10.1074/jbc.M208339200> PMID: [12237317](#)
114. Eleftheriou F, Exposito JY, Garrone R, Lethias C (2001) Binding of tenascin-X to decorin. *FEBS Lett* 495: 44–47. PMID: [11322944](#)
115. Lethias C, Carisey A, Comte J, Cluzel C, Exposito JY (2006) A model of tenascin-X integration within the collagenous network. *FEBS Lett* 580: 6281–6285. <https://doi.org/10.1016/j.febslet.2006.10.037> PMID: [17078949](#)
116. Konitsiotis AD, Raynal N, Bihan D, Hohenester E, Farndale RW, Leitinger B (2008) Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J Biol Chem* 283: 6861–6868. <https://doi.org/10.1074/jbc.M709290200> PMID: [18201965](#)
117. Probstmeier R, Fahrig T, Spiess E, Schachner M (1992) Interactions of the neural cell adhesion molecule and the myelin-associated glycoprotein with collagen type I: involvement in fibrillogenesis. *J Cell Biol* 116: 1063–1070. PMID: [1370837](#)
118. Fahrig T, Landa C, Pesheva P, Kuhn K, Schachner M (1987) Characterization of binding properties of the myelin-associated glycoprotein to extracellular matrix constituents. *EMBO J* 6: 2875–2883. PMID: [2446864](#)
119. Somasundaram R, Schuppan D (1996) Type I, II, III, IV, V, and VI collagens serve as extracellular ligands for the isoforms of platelet-derived growth factor (AA, BB, and AB). *J Biol Chem* 271: 26884–26891. PMID: [8900172](#)
120. Luo R, Jeong SJ, Jin Z, Strokes N, Li S, Piao X (2011) G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc Natl Acad Sci U S A* 108: 12925–12930. <https://doi.org/10.1073/pnas.1104821108> PMID: [21768377](#)
121. Chiang TM, Cole F, Woo-Rasberry V (2002) Cloning, characterization, and functional studies of a 47-kDa platelet receptor for type III collagen. *J Biol Chem* 277: 34896–34901. <https://doi.org/10.1074/jbc.M205311200> PMID: [12110681](#)
122. Zhu Y, Wu JJ, Weis MA, Mirza SK, Eyre DR (2011) Type IX collagen neo-deposition in degenerative discs of surgical patients whether genotyped plus or minus for COL9 risk alleles. *Spine (Phila Pa 1976)* 36: 2031–2038.
123. Zhu J, Cole F, Woo-Rasberry V, Fang XR, Chiang TM (2007) Type I and type III collagen-platelet interaction: inhibition by type specific receptor peptides. *Thromb Res* 119: 111–119. <https://doi.org/10.1016/j.thromres.2005.11.012> PMID: [16472843](#)



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Des Parkin, J; San Antonio, JD; Persikov, AV; Dagher, H; Dalgleish, R; Jensen, ST;  
Jeunemaitre, X; Savige, J

**Title:**

The collagen III fibril has a "flexi-rod" structure of flexible sequences interspersed with rigid bioactive domains including two with hemostatic roles

**Date:**

2017-07-13

**Citation:**

Des Parkin, J., San Antonio, J. D., Persikov, A. V., Dagher, H., Dalgleish, R., Jensen, S. T., Jeunemaitre, X. & Savige, J. (2017). The collagen III fibril has a "flexi-rod" structure of flexible sequences interspersed with rigid bioactive domains including two with hemostatic roles. PLOS ONE, 12 (7), <https://doi.org/10.1371/journal.pone.0175582>.

**Persistent Link:**

<http://hdl.handle.net/11343/256555>

**File Description:**

published version

**License:**

CC BY