Collagen-binding proteins: Insights from the Collagen Toolkits

Richard W. Farndale

Keywords: Synthetic peptide; triple helix; binding sites; extracellular matrix; receptor

Abbreviations: COL domain, triple-helical domain of a collagen; DDR, discoidin domain receptor; FMOD, fibromodulin; FN, fibronectin; GPR56, G protein-coupled receptor 56; MMP, matrix metalloproteinase; OSCAR, osteoclast-associated receptor; SPARC, secreted protein, acidic, and rich in cysteine; TKP, Toolkit peptide; TSP, thrombospondin; VWF, von Willebrand factor.

Abstract (188 words)

The Collagen Toolkits are libraries of 56 and 57 triple-helical synthetic peptides spanning the length of the collagen II and collagen III helices. These have been used in solid-phase binding assays to locate sites where collagen receptors and extracellular matrix components bind to collagens. Truncation and substitution allowed exact binding sites to be identified, and corresponding minimal peptides to be synthesised for use in structural and functional studies. 170 sites where over 30 proteins bind to collagen II have been mapped, providing firm conclusions about the amino acid distribution within such binding sites. Protein binding to collagen II is not random, but displays a periodicity of about 28nm, with several prominent nodes where multiple proteins bind. Notably, the vicinity of the collagenase-cleavage site in Toolkit peptide II-44 is highly promiscuous, binding over 20 different proteins. This may reflect either the diverse chemistry of that locus or its diverse function, together with the interplay between regulatory binding partners. Peptides derived from Toolkit studies have been used to determine atomic level resolution of interactions between collagen and several of its binding partners and are finding practical application in tissue engineering.

Summary Points

- The Collagen Toolkits (systematic synthetic peptide libraries spanning the COL domains of collagens II and III) simplify the mapping of receptor and other protein binding sites on the collagen triple helix.
- Compilation of data from over 30 collagen-binding proteins reveals binding activity across the length of the mature collagen II triple-helical domain.
- Binding activity is periodic, being concentrated in nodes with about 28nm spacing (corresponding to about 5 Toolkit peptides) across all the D-periods of collagen II.
- Amino acids are not equally distributed between nodes and inert peptides; nodes are enriched in F, L, R and O, whilst non-binding peptides contain A, D, K, P and S in excess.
- Synthetic peptides derived from the Toolkits underpin structural studies and provide ligands to manipulate cell and protein function, and will in future find application in tissue engineering and regenerative medicine.

Introduction 4430 words (Summary to Acknowledgments)

The human collagens are a family of 28 triple-helical proteins that form, collectively, the most abundant protein group in vertebrates. Collagens contain a defining Gly-x-x' repeating sequence essential for the assembly of a right-handed triple-helix (COL domain); lacking a sidechain, glycine (G) alone can occupy the axial position of the triple helix. The x and x' positions of the constituent α-chains often contain proline (P) and hydroxyproline (O), respectively, that support the left-handed polyproline helix II which is adopted by each strand of the collagen superhelix. Ricard-Blum has reviewed the collagen family[1], illustrating for each member the location of its COL domains and of other structures occurring in the non-helical regions of the molecule, and the diversity of their supramolecular organisation.

The structural properties of the collagens are fundamental to the integrity of the extracellular matrix. Most obviously, the major fibrillar collagens I, II and III provide tensile strength to skin, bone, tendon, cartilage and blood vessel wall. To fulfil this seemingly inert mechanical role, collagen must offer specific interaction sites for other matrix components, which implies evolutionary diversification from the primordial collagen GPO repeat[2]. Collagens also contribute to the cellular niche, a nidus which offers the cell both anchorage and survival, and within which connective tissue cells fulfil their normal functions. Such regulatory roles of collagen require engagement of diverse cell surface receptors[3], and the evolution of complex signatures within the primary Gxx' sequence of the collagen α-chains in concert with the maintenance of triple helix stability. A recent review[4] expands on much of the foregoing.

Early methods of mapping binding sites for cells or proteins within the collagens were not straightforward. Typically, purified collagens extracted from tissue would be fragmented using cyanogen bromide (CB), cleaving at Met (M) residues. The resulting defined linear fragments (CB peptides) would be separated chromatographically, then, if sufficiently long to be thermally stable, reassembled as triple helices, and their capacity to bind target proteins established. This reductive approach allowed protein binding to be mapped, but those sites containing M (e.g. some integrin sites and the von Willebrand factor (VWF) locus – see below) were disrupted and so not identifiable. Barnes employed this method extensively 30 years ago to map platelet receptor binding to collagen[5,6]. Further progress towards exact binding sites might use blocking antibodies and epitope mapping[7,8]. Alternatives included rotary shadowing, using transmission electron microscopy to locate target proteins bound to collagen monomers. Unambiguous identification of the tropocollagen molecule orientation required a marker (e.g. the N-propeptide of collagen III, or a bound antibody)[9,10]. Rotary shadowing yielded relatively low-resolution data, but, unlike CB peptide mapping, could be applied to intact heterotrimeric species such as collagen I.

Development of the Collagen Toolkits

The synthesis of collagen-like peptides in several laboratories opened the way to rapid and independent verification of putative protein-binding sites[11,12]. The Barnes group pioneered this approach by addressing the platelet-reactivity of blood vessel wall collagens; having located an integrin α2β1-binding site in CB3 from the collagen I α1 chain, they went on to synthesise seven overlapping triple-helical peptides spanning 150 residues of primary sequence, and so located the sequence GFOGER as the minimal integrin-recognition motif. This strategy neglected the α2 chain, and its success depended upon binding activity being expressed by the α1 chain alone. This peptide set was the prototype Toolkit, and identification of GFOGER led to the first integrin-ligand co-crystal (PDB: 1DZI), and by sequence homology, the location of several further integrin-binding GxOGER and related motifs. This breakthrough guided others, notably the Höök group, who used asymmetric binding to three sites in chicken collagen I and recombinant human collagen III, visualised by rotary shadowing, to locate integrin binding motifs, and proved authenticity of these sites by synthesis of triple-helical peptides[13,14].

With Wellcome Trust funding, the synthesis of the systematic collagen III Toolkit commenced in 2003 and was completed after 18 months by Nicolas Raynal and Graham Knight. Collagen III was selected first for its role in blood vessel wall and because it is homotrimeric, promoting authentic peptide self-assembly, not then feasible for the heterotrimeric collagens such as the other key vessel wall species, collagen I. Toolkit peptides (TKPs) contained 27 residues from the sequence of the collagen III COL domain, starting at its N-terminus and advancing successively by 18 residues, allowing a 9-residue overlap between adjacent peptides. Primary (guest) sequence was flanked on each side by 5 GPP (host) triplets with sufficient propensity to form triple helices to drive this overall conformation regardless of unfavourable guest sequence. At each end, the peptide terminated with a GPC extension to allow crosslinking of peptides as desired (an elaboration required for platelet activation by the GPO polymer, collagen-related peptide, CRP-XL). Thus, each TKP is 63 residues long.

Collagen III is anomalous; alignment of the fibrillar collagens reveals that its COL domain protrudes by 9 and 6 extra residues beyond the N- and C-termini of the corresponding 1014 residues of collagens I and II, so that it contains 1029 residues. Hence, Toolkit III comprises 57 peptides, and Toolkit II, completed a year later, 56 peptides. With hindsight, we would have aligned the peptides in Toolkits II with those of Toolkit III by sequence homology rather than starting at the N-terminus of each of their COL domains. The α1(II) chain enjoys 78.9% identity with α1(I) but rather less (65.5%) with α2(I), so that, unless the α2(I) chain is critically involved in binding, results obtained with Toolkit II are likely to apply to collagen I.

The Toolkit project was envisaged as essentially collaborative; whilst our focus remained on platelet surface collagen receptors beginning with integrin $\alpha 2\beta 1$, peptides were distributed to colleagues in other laboratories to explore their own targets. To date, we have mapped 38 different proteins binding to collagen, several as yet unpublished. Our approach has been to locate the TKP exhibiting highest binding of target protein in ELISA-like assays, check whether adjacent peptides showed activity which might reside in their 9-residue overlap, then to synthesise truncated peptides until activity was lost. Next, an Alanine-scan of x and x' residues within the $[\text{Gxx'}]_n$ minimal binding motif would identify the residues involved in binding, allowing short GPP- or GPO-flanked triple-helical ligands to be made which could be used in structural studies[15-21], or to manipulate the target protein in biological settings[22-26].

This strategy has been applied to the major collagen receptors: the four collagen-binding integrins[17,27-31], the three known immune receptors[32-34], both discoidin domain receptors (DDRs)[35,36] and GPR56[§]. Matrix proteins have also been investigated: VWF, SPARC, several matrix metalloproteinases (MMPs), fibronectin (FN), thrombospondin 1[26], small leucine-rich repeat proteins (SLRPs) including fibromodulin (FMOD)[37] and chondroadherin (CHAD)[38], multimerin 1[§] and dermatopontin[§]. In addition, three bacterial adhesins have been studied[39,40].

Recently, the Baumann group reported a binding site for the thrombospondin-4 and -5, using a recombinant collagen II Toolkit, with its triple-helical structure stabilised by a viral foldon domain[41]. This elegant approach suffers some drawbacks, notably the absence of proline hydroxylation in the recombinant Toolkit. Thus, it will necessarily fail to detect binding sites for which O is crucial, such as for VWF and the collagen-binding immune receptors, whilst others, such as the integrin $\alpha 1\beta 1$, display reduced affinity for motifs lacking O, i.e., GFPGER[42]. This approach, however, readily allows the insertion of sequences of any length. A similar use of foldon-templated collagen fragments allowed Zwolanek et al to identify novel Gxx'GER integrin-binding motifs in collagen XXII[43]. Similarly, a recombinant collagen II Toolkit was used to probe human autoantibodies that may be causal in rheumatoid arthritis[44]. This library suffers from the same drawbacks as the foldon approach, lacking O, and as a consequence, the melting temperature of some members was quite low, limiting their application.

Limitations of the Toolkit approach

Whilst any binding site up to 9-residues will in principle be located using the Toolkits, longer binding sites that happen to span the overlap region will be disrupted and may be missed. Further, hypothetical composite binding sites comprising adjacent helices in a collagen fibre will not be identified using a random TKP coating in a solid phase ELISA-like assay. We have restricted our focus to the 1014- or 1029-residue COL domains, excluding the contiguous non-helical telopeptides and the more distal propeptides. Our findings should be applicable to homologous motifs in other collagens, at present excluding the heterotrimeric collagens, although recent progress in this area has been made. It should be noted that even some homotrimeric collagens, e.g. the fibrillar XXIV and XXVII, contain short interruptions to the Gxx' repeat sequence, and it is not yet clear whether synthetic Toolkit peptides would properly reflect these anomalies without the support of the underpinning fibrous assembly. The inclusion of GPC terminal triplets in our Toolkits allows us to crosslink the peptides specifically as required, but may result in spontaneous oxidation and limited disulphide bond-mediated polymerisation of triple helices. This may be a drawback in some settings, e.g. for use in kinetic binding studies, but has the unexpected advantage that immobilisation of peptides on plastic ELISA wells is increased by the presence of the GPC extensions, possibly by increasing their avidity for the hydrophobic surface. This makes for more consistent coatings and binding activity[45].

Periodic distribution of sites across the collagens.

Many target proteins bind several peptides across the Toolkits, and a compilation of 168 binding sites mapped for 30 proteins on Toolkit II is shown in Figure 1. Several conclusions can be drawn. Most prominent, TKP II-44 supports the binding of two-thirds of proteins examined to date, most with relatively high affinity. Competition between these species would be expected in vivo. Some of these binding partners are known to be involved in either the assembly of collagen fibres or their proteolysis, whereas others may simply exploit the diverse chemistry of II-44, which displays a hydrophobic N-terminal tract and a charged or polar C-terminus, whilst a GPO triplet close to the cleavage site located in TKP II-43 may introduce flexibility[46] and further enhance binding opportunities for the native collagen.

Binding activity appears periodic rather than uniformly-distributed across the tropocollagen molecule. Fourier transform of the binding distribution suggests a periodicity of around 10 TKPs, corresponding to a half-wave of about 26nm, or 0.42 D-periods, close to the reported length of the overlap region in Collagen II[47]. This might indicate that higher-order fibre structure dictates the evolution of binding activity in collagen.

Orgel has proposed that the intrinsic twist of the tropocollagen molecules within each microfibril leads to the burial of much of D-periods 1, 2 and 3 within the assembled fibre[48,49]. However, one might expect collagen-binding proteins crucial for mature tissue function to have evolved to recognise motifs that are available on the fibre surface. We have argued previously that concealment of crucial binding sites, for VWF and for integrin $\alpha 2\beta 1$ located in D1 to D3, for example, would render them unable to perform their essential roles in haemostasis[50]. The presence of binding nodes extending from D1 into D4 is at odds with crucial sites therein being inaccessible within the collagen fibre. Just 4 of the 12 nodes discovered using the Toolkits lie within the D-period overlap that Orgel dubbed the “Master Control Region” where important binding sites were proposed to reside[48]. The number of proteins that bind each D-period does not differ greatly, although the promiscuity of TKP II-44 skews the distribution of sites within D4. Our findings may inform debate on the exposure of the D-periods on the surface of the collagen fibre.

In contrast, sites utilised only in free tropocollagen molecules in situ, either prior to fibre assembly or released from fibres during collagen turnover, suffer from no such evolutionary constraints. Such sites might include those at or close to both the GxKGHR crosslinking motifs (located at residues 87 and 930 of the COL domain, in D1 and D4), where regulatory proteins may operate prior to or during fibre assembly, and the collagenase cleavage site, where proteins may bind collagen fragments after hydrolysis.

Which amino acids predominate in binding sites?

Comparison of the primary sequence of the 12 binding nodes and the 13 inert TKPs reveals significantly different compositions (Figure 2). Of hydrophobic residues, F and L are more abundant in nodes, reflecting their crucial contribution to hydrophobic interactions that have been defined in the structural studies outlined below. I, M, V and Y relatively sparse but equally-distributed. The polar residues E, N, Q and T are also equally represented, whilst S is more abundant in inert peptides. The imino acids, P and O, are more abundant in inert and node peptides, respectively. Since O is almost confined to the collagens, a role in specific binding would be expected and is confirmed below. Other charged residues show large differences, with D and K markedly under-represented in binding nodes, and R over-represented. It is not surprising that A occurs less frequently in binding peptides, its short sidechain, like those of D and S, not protruding far from the collagen helix, minimising binding opportunities. These global conclusions are derived from the full 27-residue guest sequence, extending beyond the binding motifs themselves. Local echoing of the binding residue properties may serve as a means of recruiting targets to the vicinity of the more specific motif, increasing the probability of productive interaction.

Some interactions can be examined more precisely: crystallography of proteins in complex with peptides (short motifs identified using Toolkit assays outlined above) yields positive atomic-level data on key interactions, subject to the usual artefacts of crystallisation. In contrast, alanine-scanning of such motifs generally yields negative data: loss of interaction through ala-substitution. Both datasets are available in some cases allowing consensus to be reached (see Table 1). However, amino acid substitution may alter peptide conformation, introducing a structural artefact. For example, an O residue destabilises the endo-exo ring pucker of an adjacent P, introducing flexibility into the helix [46] which may be essential for candidate protein binding. It follows that Ala-scanning at O may reduce binding, without O contributing directly. In contrast, we have observed increased binding of both MMP-1 and MMP-13 to a truncated version of TKP II-44, close to the collagenase cleavage site, when E within its RGER motif was replaced with A[51,52]. It may be that structural effects, e.g. elimination of the E–R sidechain interaction[53] altering the effective diameter of the helix, improve contact with the MMP at the apposed hemopexin domain.

Long side chains extending away from the axis of the helix dominate protein–TKP interactions, illustrated in Figure 3, which shows the structure (PDB: 1Q7D) of the free integrin-binding peptide[53], with F, E and R sidechains clearly resolved. Thus, F occurs commonly within binding nodes, although under-represented in collagen with just 13 and 8 occurrences in collagens II and III. Random codon usage in collagen genes would result in F occurring 21 times. R also occurs frequently in binding sites, whilst the similarly-abundant K has been located just once (see below). E is required in integrin-binding GxOGER motifs, but we have observed only one other contribution of E to binding, in the FN-binding site (TKP II-44) that overlaps with the MMP cleavage site. The over-occurrence of E in collagens II and III (53 and 48 times) relative to F indicates an important role for E in collagen which appears unrelated to its binding activity. This may reflect charge-stabilisation between adjacent α-chains and triple-helices in the collagen fibre.

Hydrophobic interactions of collagen residues other than F frequently stabilise the binding of proteins. Thus, L, V and M are key contributors to several interactions, such as with integrins, DDRs, VWF, SPARC, OSCAR (PDB: 5EIV) and with MMPs and fibronectin. Two aromatic residues, W and Y, form part of the apposed binding pocket in several proteins, with which collagen F in particular can interact. The aliphatic stems of R may also contribute to such hydrophobic pockets[15].

Integrins represent a special case, where the canonical E carboxylate anion of the GxOGER motif directly coordinates Mg^{2+} in the metal ion-dependent adhesion site of the collagen-binding integrin I-domains, in α1-, α2-, α10- and $\alpha 11\beta 1$. This interaction is supplemented by the hydrophobic x residues in both the leading and middle strands of collagen which contact the surface of the I domain. These hydrophobic residues determine overall affinity, with L and M able to complete relatively high-affinity motifs[30]. The hydrophobic stem of R in GROGER may assume the same function, although no crystal structure exists[29]. Ala-substitution at x results in near-basal affinity for the resting integrin; GAOGER, naturally occurring in collagens II and III, becomes important only for the activated integrin[30].

The collagen-binding immune receptors, Glycoprotein (GP) VI, LAIR1 and OSCAR, rely on the unique abundance of O residues within collagen for binding specificity. Such tracts are found at the C-terminus, TKPs II-56 and III-57. GPO polymers are sufficient for recognition by GPVI[54], and its subsequent activation provided higher-order structure is introduced into the collagen peptide, whilst LAIR1[55] and OSCAR[20,32] require a greater diversity, with OSCAR in particular requiring an F residue (PDB: 5EIV). Although both GPVI and LAIR1 bind TKP III-30 with high affinity, there is little overlap between the repertoire of Toolkit peptides recognised by the three receptors despite conservation of 3D structure and primary sequence.

Specific Binding Nodes

1. Integrin sites: GxOGER

The integrin-binding GxOGER and related motifs occur at conserved loci across collagens I, II and III[30]. In Toolkit II, these occur at peptides 7 and 8, 28, with a low affinity site in 44. These motifs bind few other proteins, indicating the importance of the cell-integrin interaction. Note that an atypical positive charge [27] on its surface prevents $\alpha 10\beta 1$ from binding GROGER, present in collagens I and III, but not II. These sites display selectivity for the different integrins, with α1- and $\alpha 10\beta 1$ preferring GLOGEN, in collagen III, whilst α2- and $\alpha 11\beta 1$ express higher affinity for GFOGER, in collagens I, II, IV and others[56]. There is only limited evidence for absolute specificity, but analysis in a cellular setting is confounded by the activation state of the integrin, which may override intrinsic selectivity.

2. Crosslinking sites: GxKGHR

The only observed lysine-containing binding motif within the Toolkits is GxKGHR, the key inter-helix crosslinking site. In nature, covalent crosslinks can form through the condensation of a lysine aldehyde located within the telopeptide of an α-chain in one helix with a target lysine residue in a nearby helix COL domain[57] . Thus, in the staggered assembly of the native collagen fibre, the *N-telo*-lysine aldehyde condenses with K-930, and the *C-telo*-lysine aldehyde with K-87. The target GxKGHR motif (K-87) is found in the unique sequence of TKP II-5 and II-52 and, as a result of the 9-residue offset between Toolkits II and III, in the overlaps of TKP III-5/6 and III-52/53 (K-930). The lysine complement of native collagens is at least partially hydroxylated[58], whereas we have included no sidechain modification other than O in the Toolkits. This may account for the under-representation of K in the observed binding sites, since it is plausible that hydroxylysine contributes to binding in nature although absent from our experiments. Acetylation of K compromised binding of FMOD to GxKGHR motifs[37], suggesting that its natural covalent modification through hydroxylation or further derivatisation, e.g. glycosylation and glycation, may be disruptive. On the other hand, Ala-scanning of the unmodified GxKGHR motif abolished target molecule binding, showing both K and R to be key contributors to the interaction with FMOD. Similarly, Gebauer et al found that substitution of K by P reduced binding of cartilage oligomeric protein (COMP or TSP-5) to a T4 foldon-guided peptide set containing GxKGHR[41]. No crystal complex between a target protein and a full KGHR-containing motif is yet available, which might elucidate the binding mechanism, although R-containing peptides have been shown both to bind and yield structures with HSP47[59,60]. Several molecules known to be involved in the regulation of collagen crosslinking have been shown to bind at, or close to, this motif, including FMOD and TSP-1, both of which may help recruit the lysyloxidase which effects the hydroxylation of the telopeptide lysine residues needed for crosslink formation.

3. Collagenase site G~LAGQRGIVGLOGQRGER

This site, comprising much of TKP II-44, contains the unique collagenase cleavage site. Notably, II-44 differs in only two conserved residues from the corresponding tract of collagen α1(I), so Toolkit II results will likely apply to collagen I. II-44 binds both collagenases investigated in detail, MMP-1 and MMP-13[51,52] along with many other molecules, including low-affinity interactions with the DDRs[61], FMOD[37] and TSP-1[26], whilst the GQRGER motif that interacts with the hemopexin domain of full-length MMP-13 is also a weak integrin-binding site. Interestingly, the free hemopexin domain binds to both the N-terminal GLAGQR and the central GLOGQR tracts, established by ala-scanning[51], offering a second nearby locus for recruitment of MMP-13. MMPs and FN each interact with a long tract of collagen at this locus, at least four triplets, so that all three α-chains are likely to be involved[52,62]. Data is summarised in Table 1. A linear collagen peptide adds a strand to a β-sheet of FN, which might imply unwinding of helical collagen as FN binds[62]. Although a similar mechanism has been mooted for MMP-1 binding to this locus, crystallography shows only minor relaxation of the helix in an MMP-1–peptide complex[52]. Comparison with an FN-helical peptide complex is needed. This marked overlap of binding motifs suggests that competition between FN and MMP-1 or MMP-13 for this site in collagens I and II would be inevitable.

4 VWF A3 site: GxRGQOGVMGFO

This site, recently reviewed by Chen and Lin[63], was identified as binding the VWF A3 domain[64], along with SPARC[9,19] and DDRs 1 and 2[16,35,61]. It is found at TKP II-22 and III-23. The DDRs and SPARC utilise only the last 2 triplets, as indicated by co-crystal structures with DDR2 DS domain (PDB: 2WUH) and with SPARC (PDB: 2V53), whilst VWF A3 also interacts with the preceding 2 triplets (PDB: 4DMU). This preceding sequence may be required for DDR signalling. Like the extended collagenase site, all three α-chains contribute to the interaction with A3. Importantly, this site provided insight into the chain register of collagen I, since neither α1(I) nor α2(I) contain a complete A3-binding motif, and the likely composite site is best assembled with α2(I) in the trailing position[15].

Bacterial adhesins

Three collagen-binding adhesins have been studied. CNE, an adhesin of *S. equi equi*, binds with highest affinity to TKP II-44, and also to TKP II-1. CNE may compete with the host proteins that recognise II-44, such as FN, perhaps so contributing to the virulence of the pathogen by disrupting endothelial cell $\alpha V\beta 3$ interaction with FN[40]. In marked contrast, two other adhesins, *Yersinia* adhesin A[39] and *S. pyogenes* M3[§] are promiscuous, binding many sites across both Toolkits. The large number of peptides binding YadA allows statistical analysis of its binding propensity, which revealed a marked preference for hydrophobic residues, and for Pro and Hyp. This lack of specificity across the Toolkits may indicate that non-selective collagen binding has a role in virulence, where escape from the bloodstream and sequestration within the host tissue may be more important, by allowing the pathogen to evade host immune defences, than the specific disruption of host tissue homeostasis proposed for CNE.

Inter-D-period interactions.

Several proteins have been observed to bind sites in different D-periods that prove to be aligned within the fibrillar collagen assembly. For example, FMOD binds TKPs III-5 and III-44[37], and TSP-1 binds II-5, II-20 and II-45[26]. Both FMOD and TSP-1 have a role in regulating fibre crosslinking (which involves K in the GxKGHR motif found in II-5 and III-5) and in fibre assembly. It seems likely that one collagen helix provides a platform within a fibre from which these molecular chaperones can deliver the required activity to the site in the adjacent D-period. As well as this functional role, such alignments may also permit co-operative binding of the trimeric TSP-1. Chondroadherin may perform a similar function, binding a single site, II-26, which aligns perfectly with the GxKGHR crosslinking site in II-52[38], shown in Figure 4. This would imply simultaneous accessibility of the relevant sections of D2 and D4.

Forward Developments

The identification of specific binding motifs within collagen has provided tools to manipulate cell function in research and other applications. The Leitinger group used integrin and DDR-specific peptides to probe integrin-mediated cell adhesion[23]. We used peptide-coated surfaces to characterise the determinants of thrombus deposition[24,25,65,66], a method being developed elsewhere for diagnostic applications[67,68]. The Garcia group exploited the integrin motif GFOGER as a surface coating for orthopaedic devices[69], whilst Koide discusses the use of collagen peptide-based biomaterials[70]. Bacterial collagens have enjoyed some popularity as alternatives to synthetic peptides, and Toolkit-derived motifs have been introduced into bacterial collagens with a view to producing materials for tissue engineering[71-73]. In this laboratory, we have used photochemical coupling of specific motifs to enhance the cell-reactivity of collagen scaffolds, a generic technique which could be applied to derivatise many organic substrates[74,75].

The next big step in the field will be to map binding sites in heterotrimeric collagens. Hartgerink and colleagues have designed heterotrimeric peptides using charge-complementarity to align the different collagen strands correctly, to produce collagen I-like self-assembling peptides [76,77], an approach recently brought to fruition by Jalan[78]. The synthetic challenge this strategy presents is substantial, requiring at least three discrete peptides for each heterotrimer along with knowledge of the register prevailing in the corresponding collagen.

We believe the Toolkit peptide-derived collagenous materials will find application in drug discovery, diagnostics and regenerative medicine as well as being valuable and specific research reagents.

Acknowledgments

The Toolkit project owes much to the pioneering work of Michael Barnes, and to the dedication and perseverance of the peptide chemists, Graham Knight, Nicolas Raynal, Loraine Foley, Dominique Bihan, Jean-Daniel Malcor and Arkadiusz Bonna. Samir Hamaia made a huge contribution through protein expression and the development of solid-phase binding measurement throughout. Funding was provided by Wellcome Trust, Medical Research Council and British Heart Foundation.

Footnote

§ represents an unpublished protein-Toolkit interaction.

References

[1] Ricard-Blum S. (2011) The Collagen Family. Cold Spring Harb. Perspect. Biol., **3**: a004978.

[2] Slatter DA, Farndale RW. (2015) Structural constraints on the evolution of the collagen fibril: convergence on a 1014-residue COL domain. Open Biol, **5**: 140220.

[3] Leitinger B. (2011) Transmembrane collagen receptors. Annu Rev Cell Dev Biol, **27**: 265-90.

[4] An B, Lin YS, Brodsky B. (2016) Collagen interactions: Drug design and delivery. Adv Drug Deliv Rev, **97**: 69-84.

[5] Morton LF, Peachey AR, Barnes MJ. (1989) Platelet-reactive sites in collagens type I and type III. Evidence for separate adhesion and aggregatory sites. Biochem J, **258**: 157-63.

[6] Zijenah LS, Barnes MJ. (1990) Platelet-reactive sites in human collagens I and III: evidence for cell-recognition sites in collagen unrelated to RGD and like sequences Thromb. Res., **59**: 553-566.

[7] Glattauer V, Werkmeister JA, Kirkpatrick A, Ramshaw JA. (1997) Identification of the epitope for a monoclonal antibody that blocks platelet aggregation induced by type III collagen. Biochem J, **323**: 45-9.

[8] Werkmeister JA, Ramshaw JAM. (1991) Multiple antigenic determinants on type III collagen. Biochem J, **274**: 895-898.

[9] 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**: 19551-60.

[10] Tenni R, Viola M, Welser F, Sini P, Giudici C, Rossi A, et al. (2002) Interaction of decorin with CNBr peptides from collagens I and II. Evidence for multiple binding sites and essential lysyl residues in collagen. Eur J Biochem, **269**: 1428-37.

[11] Fields GB. (2010) Synthesis and biological applications of collagen-model triple-helical peptides. Org Biomol Chem, **8**: 1237-58.

[12] Ramshaw JAM, Shah NK, Brodsky B. (1998) Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. J. Structural Biol., **122**: 86-91.

[13] Xu Y, Gurusiddappa S, Rich RL, Owens RT, Keene DR, Mayne R, et al. (2000) Multiple binding sites in collagen type I for the integrins alpha1beta1 and alpha2beta1. J Biol Chem, **275**: 38981-9.

[14] Kim JK, Xu Y, Xu X, Keene DR, Gurusiddappa S, Liang X, et al. (2005) A novel binding site in collagen type III for the integrins, alpha 1beta 1 and alpha 2beta 1. J Biol Chem, **280**: 32512-32520.

[15] 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**: 5253-8.

[16] Carafoli F, Bihan D, Stathopoulos S, Konitsiotis AD, Kvansakul M, Farndale RW, et al. (2009) Crystallographic insight into collagen recognition by discoidin domain receptor 2. Structure, **17**: 1573-81.

[17] Carafoli F, Hamaia SW, Bihan D, Hohenester E, Farndale RW. (2013) An activating mutation reveals a second binding mode of the integrin alpha2 I domain to the GFOGER motif in collagens. PLoS One, **8**: e69833.

[18] Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. (2000) Structural basis of collagen recognition by integrin alpha2beta1. Cell, **101**: 47-56.

[19] 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**: 18273-7.

[20] Zhou L, Hinerman JM, Blaszczyk M, Miller JL, Conrady DG, Barrow AD, et al. (2016) Structural basis for collagen recognition by the immune receptor OSCAR. Blood, **127**: 529-37.

[21] Chin YK, Headey SJ, Mohanty B, Patil R, McEwan PA, Swarbrick JD, et al. (2013) The Structure of Integrin alpha1I Domain in Complex with a Collagen-mimetic Peptide. J Biol Chem, **288**: 36796-809.

[22] Gigout A, Jolicoeur M, Nelea M, Raynal N, Farndale R, Buschmann MD. (2008) Chondrocyte aggregation in suspension culture is GFOGER-GPP- and beta1 integrin-dependent. J Biol Chem, **283**: 31522-30.

[23] Xu H, Bihan D, Chang F, Huang PH, Farndale RW, Leitinger B. (2012) Discoidin domain receptors promote alpha1beta1- and alpha2beta1-integrin mediated cell adhesion to collagen by enhancing integrin activation. PLoS One, **7**: e52209.

[24] Pugh N, Simpson AM, Smethurst PA, de Groot PG, Raynal N, Farndale RW. (2010) Synergism between platelet collagen receptors defined using receptor-specific collagen-mimetic peptide substrata in flowing blood. Blood, **115**: 5069-79.

[25] Siljander PR, Munnix IC, Smethurst PA, Deckmyn H, Lindhout T, Ouwehand WH, et al. (2004) Platelet receptor interplay regulates collagen-induced thrombus formation in flowing human blood. Blood, **103**: 1333-41.

[26] Rosini S, Pugh N, Bonna AM, Hulmes DJS, Farndale RW, Adams JC. (2018) Thrombospondin-1 promotes matrix homeostasis by interacting with collagen and lysyl oxidase precursors and collagen cross-linking sites. Sci Signal, **11**: aar2566.

[27] Hamaia SW, Luff D, Hunter EJ, Malcor JD, Bihan D, Gullberg D, et al. (2016) Unique charge-dependent constraint on collagen recognition by integrin alpha10beta1. Matrix Biol, **59**: 80-94.

[28] 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**: 26019-28.

[29] 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**: 3821-31.

[30] 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**: 47763-72.

[31] Zhang WM, Kapyla J, Puranen JS, Knight CG, Tiger CF, Pentikainen OT, et al. (2003) alpha 11beta 1 integrin recognizes the GFOGER sequence in interstitial collagens. J Biol Chem, **278**: 7270-7.

[32] 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**: 3505-16.

[33] 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**: 4986-96.

[34] 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.

[35] 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**: 6861-8.

[36] Xu Y, Gurusiddappa S, Rich RL, Owens RT, Keene DR, Mayne R, et al. (2000) Multiple binding sites in collagen type I for the integrins a1b1 and a2b1. J Biol Chem, **275**: 38981-38989.

[37] Kalamajski S, Bihan D, Bonna A, Rubin K, Farndale RW. (2016) Fibromodulin Interacts with Collagen Cross-linking Sites and Activates Lysyl Oxidase. J Biol Chem, **291**: 7951-60.

[38] Paracuellos P, Kalamajski S, Bonna A, Bihan D, Farndale RW, Hohenester E. (2017) Structural and functional analysis of two small leucine-rich repeat proteoglycans, fibromodulin and chondroadherin. Matrix Biol, **63**: 106-116.

[39] 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**: 3226-36.

[40] van Wieringen T, Kalamajski S, Liden A, Bihan D, Guss B, Heinegard D, et al. (2010) The streptococcal collagen-binding protein CNE specifically interferes with alphaVbeta3-mediated cellular interactions with triple helical collagen. J Biol Chem, **285**: 35803-13.

[41] Gebauer JM, Kohler A, Dietmar H, Gompert M, Neundorf I, Zaucke F, et al. (2018) COMP and TSP-4 interact specifically with the novel GXKGHR motif only found in fibrillar collagens. Sci Rep, **8**: 17187.

[42] Perret S, Eble JA, Siljander PR, Merle C, Farndale RW, Theisen M, et al. (2003) Prolyl hydroxylation of collagen type I is required for efficient binding to integrin alpha 1 beta 1 and platelet glycoprotein VI but not to alpha 2 beta 1. J Biol Chem, **278**: 29873-9.

[43] Zwolanek D, Veit G, Eble JA, Gullberg D, Ruggiero F, Heino J, et al. (2014) Collagen XXII binds to collagen-binding integrins via the novel motifs GLQGER and GFKGER. Biochem J, **459**: 217-27.

[44] Lindh I, Snir O, Lonnblom E, Uysal H, Andersson I, Nandakumar KS, et al. (2014) Type II collagen antibody response is enriched in the synovial fluid of rheumatoid joints and directed to the same major epitopes as in collagen induced arthritis in primates and mice. Arthritis Res Ther, **16**: R143.

[45] Slatter DA, Bihan DG, Jarvis GE, Stone R, Pugh N, Giddu S, et al. (2012) The properties conferred upon triple-helical collagen-mimetic peptides by the presence of cysteine residues. Peptides, **36**: 86-93.

[46] Chow WY, Forman CJ, Bihan D, Puszkarska AM, Rajan R, Reid DG, et al. (2018) Proline provides site-specific flexibility for in vivo collagen. Sci Rep, **8**: 13809.

[47] Antipova O, Orgel JP. (2010) In situ D-periodic molecular structure of type II collagen. J Biol Chem, **285**: 7087-96.

[48] Orgel JP, Antipova O, Sagi I, Bitler A, Qiu D, Wang R, et al. (2011) Collagen fibril surface displays a constellation of sites capable of promoting fibril assembly, stability, and hemostasis. Connect Tissue Res, **52**: 18-24.

[49] Perumal S, Antipova O, Orgel JP. (2008) Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. Proc Natl Acad Sci USA, **105**: 2824-9.

[50] Herr AB, Farndale RW. (2009) Structural insights into the interactions between platelet receptors and fibrillar collagen. J Biol Chem, **284**: 19781-5.

[51] Howes JM, Bihan D, Slatter DA, Hamaia SW, Packman LC, Knauper V, et al. (2014) The recognition of collagen and triple-helical Toolkit peptides by MMP-13: Sequence specificity for binding and cleavage. J Biol Chem.

[52] Manka SW, Carafoli F, Visse R, Bihan D, Raynal N, Farndale RW, et al. (2012) Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. Proc Natl Acad Sci U S A, **109**: 12461-6.

[53] Emsley J, Knight CG, Farndale RW, Barnes MJ. (2004) Structure of the integrin alpha2beta1-binding collagen peptide. J Mol Biol, **335**: 1019-28.

[54] Morton LF, Hargreaves PG, Farndale RW, Young RD, Barnes MJ. (1995) Integrin α2β1-independent activation of platelets by simple collagen-like peptides: collagen tertiary (triple-helical) and quaternary (polymeric) structures are sufficient alone for α2β1-independent platelet reactivity. Biochem J, **306 (Pt 2)**: 337-44.

[55] Lebbink RJ, Raynal N, de Ruiter T, Bihan DG, Farndale RW, Meyaard L. (2009) Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. Matrix Biol, **28**: 202-10.

[56] Hamaia S, Farndale RW. (2014) Integrin recognition motifs in the human collagens. Adv Exp Med Biol, **819**: 127-42.

[57] Knott L, Bailey AJ. (1998) Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance. Bone, **22**: 181-7.

[58] Yamauchi M, Sricholpech M. (2012) Lysine post-translational modifications of collagen. Essays Biochem, **52**: 113-33.

[59] 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**: 13243-7.

[60] Tasab M, Jenkinson L, Bulleid NJ. (2002) Sequence-specific recognition of collagen triple helices by the collagen-specific molecular chaperone HSP47. J Biol Chem, **277**: 35007-12.

[61] Xu H, Raynal N, Stathopoulos S, Myllyharju J, Farndale RW, Leitinger B. (2011) 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, **30**: 16-26.

[62] Erat MC, Slatter DA, Lowe ED, Millard CJ, Farndale RW, Campbell ID, et al. (2009) Identification and structural analysis of type I collagen sites in complex with fibronectin fragments. Proc Natl Acad Sci U S A, **106**: 4195-200.

[63] Chen EA, Lin YS. (2019) Using synthetic peptides and recombinant collagen to understand DDR-collagen interactions. Biochim Biophys Acta Mol Cell Res.

[64] 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**: 3753-6.

[65] Munnix IC, Gilio K, Siljander PR, Raynal N, Feijge MA, Hackeng TM, et al. (2008) Collagen-mimetic peptides mediate flow-dependent thrombus formation by high- or low-

affinity binding of integrin alpha2beta1 and glycoprotein VI. J Thromb Haemost, **6**: 2132-42.

[66] Pugh N, Bihan D, Perry DJ, Farndale RW. (2014) Dynamic analysis of platelet deposition to resolve platelet adhesion receptor activity in whole blood at arterial shear rate. Platelets, **26**: 216-9.

[67] de Witt SM, Swieringa F, Cavill R, Lamers MM, van Kruchten R, Mastenbroek T, et al. (2014) Identification of platelet function defects by multi-parameter assessment of thrombus formation. Nat Commun, **5**: 4257.

[68] Geffen JPV, Brouns SLN, Batista J, McKinney H, Kempster C, Nagy M, et al. (2018) High-throughput elucidation of thrombus formation reveals sources of platelet function variability. Haematologica.

[69] Reyes CD, Garcia AJ. (2003) Engineering integrin-specific surfaces with a triple-helical collagen-mimetic peptide. J Biomed Mater Res A, **65**: 511-23.

[70] Koide T. (2007) Designed Triple-Helical Peptides as Tools for Collagen Biochemistry and Matrix Engineering. Philosophical Transactions: Biological Sciences, **362**: 1281-1291.

[71] An B, Abbonante V, Xu H, Gavriilidou D, Yoshizumi A, Bihan D, et al. (2015) Recombinant Collagen Engineered to Bind to Discoidin Domain Receptors Functions as a Receptor Inhibitor. J Biol Chem.

[72] An B, Kaplan DL, Brodsky B. (2014) Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. Front Chem, **2**: 40.

[73] Seo N, Russell BH, Rivera JJ, Liang X, Xu X, Afshar-Kharghan V, et al. (2010) An engineered alpha1 integrin-binding collagenous sequence. J Biol Chem, **285**: 31046-54.

[74] Malcor JD, Bax D, Hamaia SW, Davidenko N, Best SM, Cameron RE, et al. (2016) The synthesis and coupling of photoreactive collagen-based peptides to restore integrin reactivity to an inert substrate, chemically-crosslinked collagen. Biomaterials, **85**: 65-77.

[75] Malcor JD, Juskaite V, Gavriilidou D, Hunter EJ, Davidenko N, Hamaia S, et al. (2018) Coupling of a specific photoreactive triple-helical peptide to crosslinked collagen films restores binding and activation of DDR2 and VWF. Biomaterials, **182**: 21-34.

[76] Jalan AA, Demeler B, Hartgerink JD. (2013) Hydroxyproline-free single composition ABC collagen heterotrimer. J Am Chem Soc, **135**: 6014-7.

[77] Russell LE, Fallas JA, Hartgerink JD. (2010) Selective assembly of a high stability AAB collagen heterotrimer. J Am Chem Soc, **132**: 3242-3.

[78] Jalan AA, Sammon D, Hartgerink JD, Brear P, Stott K, Hamaia SW, et al. Deciphering the Chain Alignment of the Collagen Heterotrimer. Nature Cell Biol (under review).

[79] Reynolds CR, Islam SA, Sternberg MJE. (2018) EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures. J Mol Biol, **430**: 2244-2248.

Figure Legends

Figure 1

Figure 1a shows the number of proteins found to bind to each peptide from Toolkit II. 30 proteins recorded 170 "hits" across the 56 peptides of Toolkit II. Those with 5 or more "hits" are shaded in red and provide the sequences to be analysed as described in the legend to Figure 2. The location of D-periods is indicated below the x-axis, with the number of hits per D-period in brackets.

Figure 1b shows the output of a Fourier transform (Sinc function) of the data in Figure 1a. The x-axis shows periodicity, where 1 unit = the length of one peptide. The series converged as shown, indicating that the binding pattern was periodic, with a "wavelength" of 10.6 peptides; r^2 = 0.988. This value corresponds to [10.6 x 18] residues (the incremental length of each Toolkit peptide), i.e. 191 residues, and in length to [300 x 191/1014] nanometres, since the COL domain is 300nm in length and contains 1014 residues. Thus, nodes (half-waves) are separated by 28nm. Analysis was performed using GraphPad Prism 8.1.1 for Macintosh.

Figure 2

Figure 2 shows the numbers of each amino acid found in the x and x' positions of the 12 peptides that reach the node threshold of 5 binding proteins (red bars), and the corresponding information for the 13 peptides with no partners (blue bars). Residues in excess in inert peptides are at the left of the figure, and in binding nodes at the right. These data were used to set up contingency tables, allowing the distribution of amino acids to be compared between peptide sets; the distribution differed significantly (χ^2 test, p = .003).

Figure 3

Figure 3 shows the structure of the free integrin-binding peptide, $[GPO]_2GFOGER[GPO]_3$, using PDB file 1Q7D[53]. The key amino acids, F, E and R, protrude into the environment away from the axis of the helix, and are readily available for interaction with the surface of the integrin I domain. Long residues such as these are over-represented in the Toolkit peptides that interact with their binding partners. Rendering was performed using EzMol[79].

Figure 4

Figure 4 shows a simple model of the alignment of tropocollagen molecules and D-periods in a collagen fibre (above) and an expanded section (below) showing the perfect alignment of the sole CHAD-binding sequence in TKP II-26 (at the D2-D3 boundary) with the KGHR crosslinking motif in D4. This proximity implies involvement of CHAD in regulating the crosslinking process. For reference, the second KGHR motif shown in D1 lies about 40nm towards the N-terminus of the molecule. The Figure was published in Matrix Biology[38] and is reproduced in accordance with Creative Commons Attribution License (CC BY).

Table 1. Peptide II-44 residues critical for binding to FN, MMP1 and MMP-13

II-44	G~LAGQRGIVGLOGQRGERGFOGLOGPS
FN (crystal contacts[62])	GQRGIVGLOGQRGERGFOGLOG
FN (Ala-scan§)	GPQG~LAGQRGIVGLOGQRGER
MMP-13 (Ala-scan[51])	GPQG~LAGQRGIVGLOGQRGER
MMP-1 (Ala-scan[52])	GPQG~LAGQRGIVGLOGQRGER
MMP-1 (crystal contacts[52])	GPOGPQG~LAGQRGIVGLOGQRGER

Table 1 shows the residues found to be crucial for binding of FN and MMPs to collagen-derived peptides. Erat[62] used a linear peptide (sequence as indicated) to form a complex (PDB: 3EJH) with the 8-9FnI module pair, and found a series of interactions (highlighted turquoise) that stabilised the β-strand adopted by the peptide on the FN surface. We used ala-scanning of a Toolkit-derived triple-helical peptide, including the collagenase cleavage site, and found the first Lx'GxR motif to be critical for the binding of full-length FN. The same peptide set was used to test binding of MMP-1 and MMP-13, and similar residues were involved, highlighted in magenta. Notably, some ala-substitutions of charged or polar residues enhanced binding, highlighted in yellow, perhaps for structural reasons discussed in the text. This applied even to MMP-1 although the hemopexin domain that aligns with the RGER motif makes little contact with the peptide (PDB: 4AUO). The MMP-1 co-crystal showed more extensive contacts, highlighted in turquoise, on both sides of the scissile G~L bond than were revealed by ala-scanning.

Figure 1a

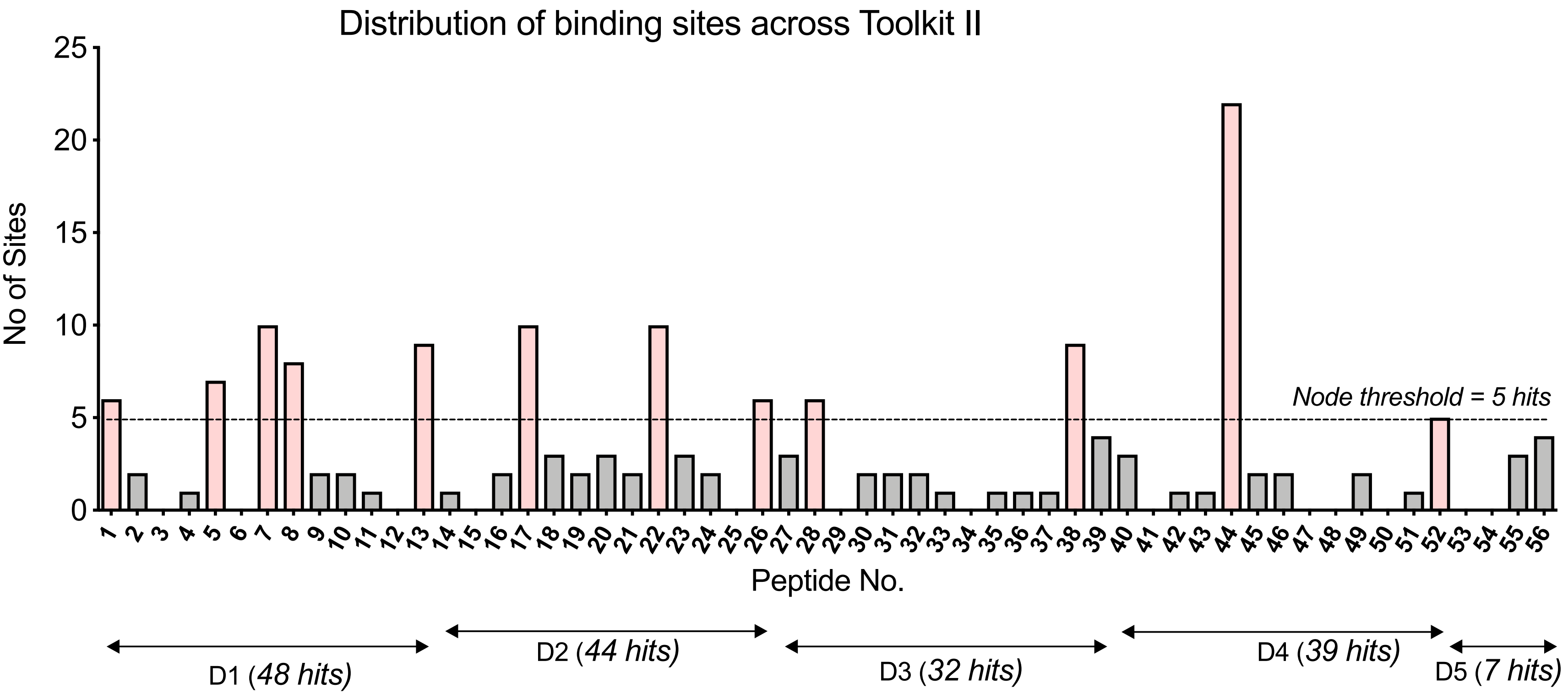


Figure 1b

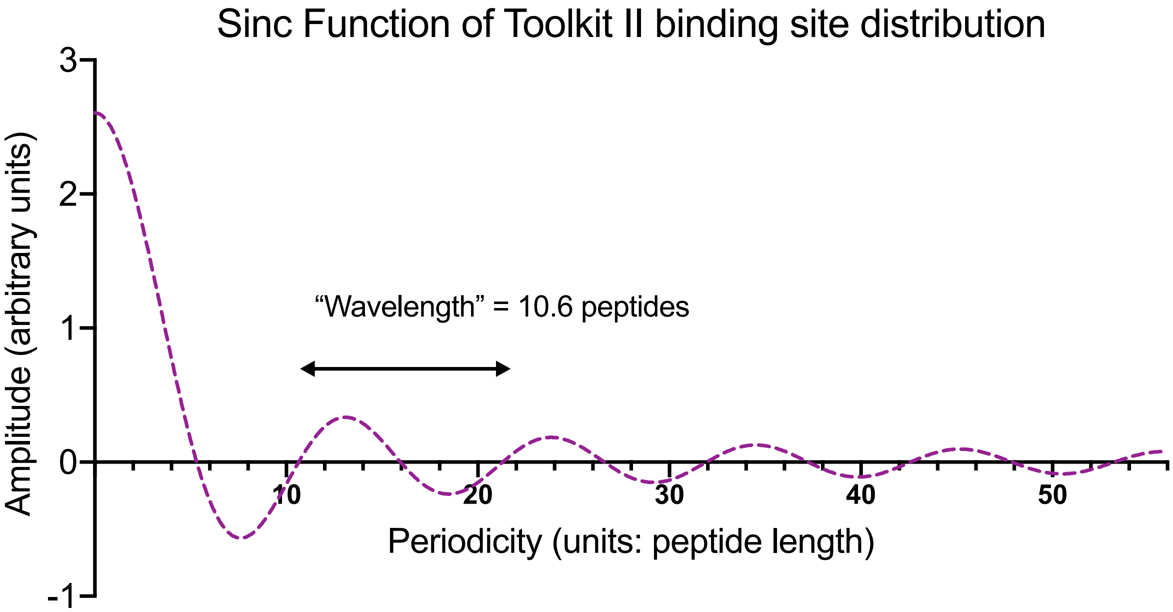


Figure 2

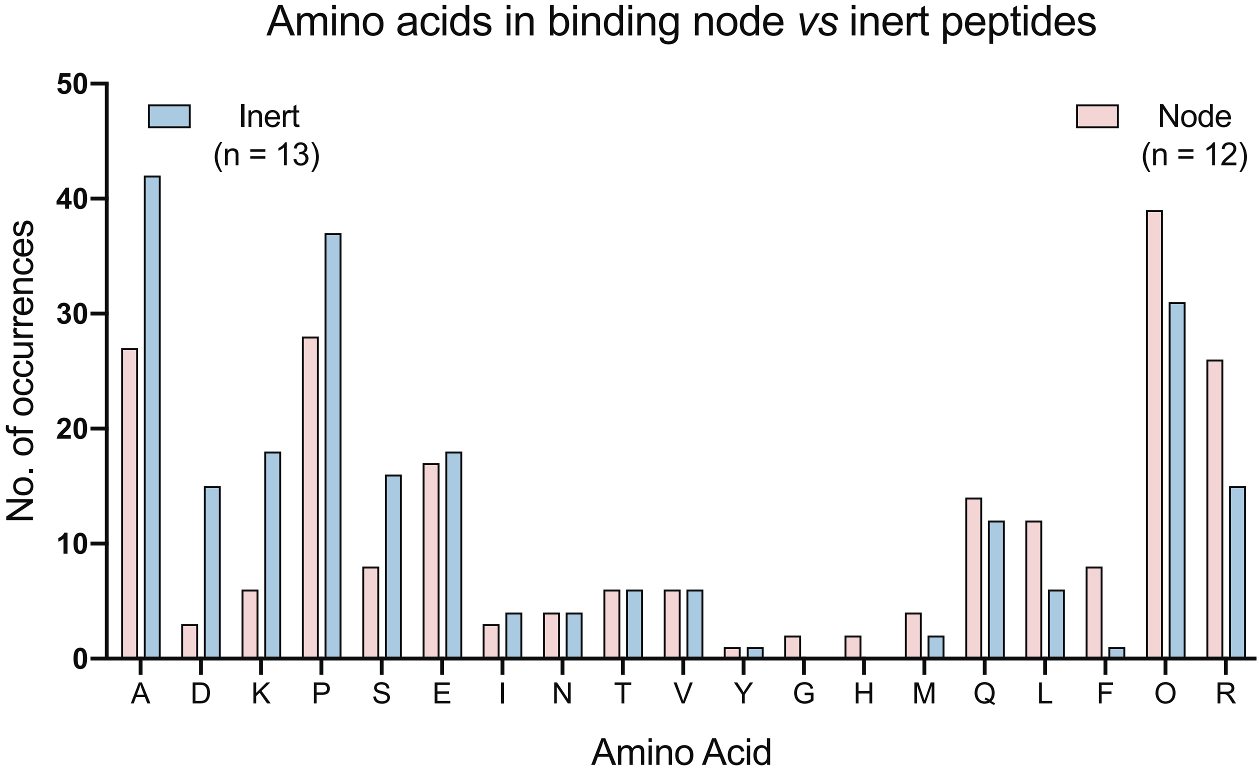

Amino acids in binding node *vs* inert peptides
Inert (n = 13)
Node (n = 12)
No. of occurrences
0
10
20
30
40
50
A
D
K
P
S
E
I
N
T
V
Y
G
H
M
Q
L
F
O
R
Amino Acid

Figure 3

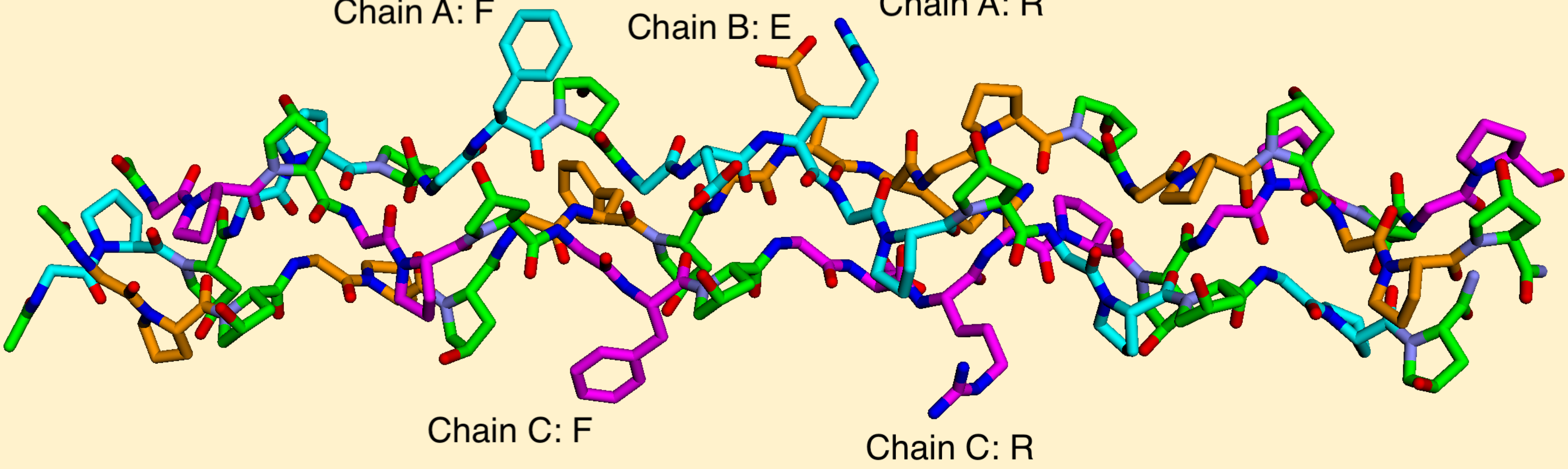
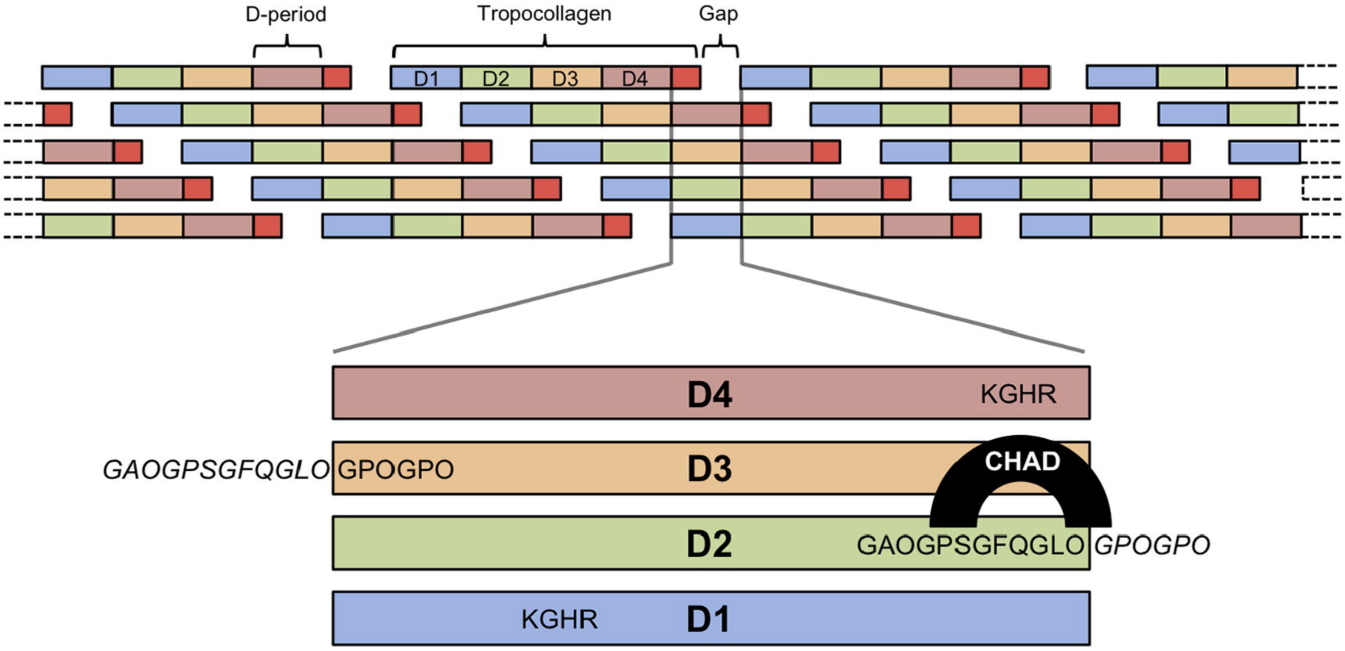


Figure 4


D-period
Tropocollagen
Gap
D1
D2
D3
D4
KGHR
GAOGPSGFQGLO
GPOGPO
CHAD
GAOGPSGFQGLO
GPOGPO
KGHR