

Chain alignment of collagen I deciphered using computationally designed heterotrimers

Abhishek A. Jalan^{1,4*}, Douglas Sammon², Jeffrey D. Hartgerink³, Paul Brear¹, Katherine Stott¹, Samir W. Hamaia¹, Emma J. Hunter¹, Douglas R. Walker³, Birgit Leitinger² and Richard W. Farndale¹

The most abundant member of the collagen protein family, collagen I (also known as type I collagen; COL1), is composed of one unique (chain B) and two similar (chain A) polypeptides that self-assemble with one amino acid offset into a heterotrimeric triple helix. Given the offset, chain B can occupy either the leading (BAA), middle (ABA) or trailing (AAB) position of the triple helix, yielding three isomeric biomacromolecules with different protein recognition properties. Despite five decades of intensive research, there is no consensus on the position of chain B in COL1. Here, three triple-helical heterotrimers that each contain a putative von Willebrand factor (VWF) and discoidin domain receptor (DDR) recognition sequence from COL1 were designed with chain B permuted in all three positions. AAB demonstrated a strong preference for both VWF and DDR, and also induced higher levels of cellular DDR phosphorylation. Thus, we resolve this long-standing mystery and show that COL1 adopts an AAB register.

Collagens are a large superfamily of 28 known mammalian proteins (COL1–COL28) that bind transmembrane receptors¹, secreted extracellular matrix proteins² and blood serum proteins³, collectively called collagen-binding proteins (CBPs), to regulate cell signaling, matrix homeostasis and thrombosis. For example, fibrillar collagens COL1 and COL3, exposed in the sub-endothelium during vascular injury, bind to the blood plasma protein von Willebrand factor (VWF)³. Subsequent collagen-mediated recruitment of platelets via the receptors integrin $\alpha_v\beta_3$ and glycoprotein VI is responsible for the deposition of life-saving thrombi⁴ as well as life-threatening ischemic myocardial damage⁵. Collagens also bind and activate the discoidin domain receptors DDR1 and DDR2, a subfamily of receptor tyrosine kinases, which results in intracellular signaling events that are critical for cell survival and tissue remodeling⁶. DDR1-deficient mice show loss of renal function⁷, anomalous mammary gland development⁸ and defective arterial wound repair⁹, and DDR2-deficient mice exhibit dwarfism¹⁰. DDR1 and DDR2 also remodel the extracellular matrix during tissue maturation and thus, like most other receptor tyrosine kinases, have a key role in cancer progression¹¹.

Collagens are highly complex multidomain proteins that contain more than 1,000 amino acids and often form hydrogels or non-specific aggregates when isolated from native tissues. The structural basis for the collagen–CBP interaction is therefore studied using a library of synthetic peptides. Each peptide in the library contains a short stretch of native collagen sequence flanked on both termini by an inert sequence that induces the peptide to fold into a collagen-like triple helix¹². Thus, putative CBP recognition sequences are displayed in a native-like triple-helical fold without the accompanying complexity of native collagen and can be readily tested for activity against potential CBPs. This Toolkit approach¹³ has revealed the structural basis for the interaction of COL2 and COL3 with integrins $\alpha_1\beta_1$ (ref. ¹⁴), $\alpha_2\beta_1$ (ref. ¹⁵) and $\alpha_{10}\beta_1$ (ref. ¹⁶), as well as thrombospondin 1 (TSP-1) (ref. ³), the VWF A3 domain¹⁷, DDR1 (ref. ¹⁸), DDR2 (ref. ¹⁹), matrix metalloproteinase 1 (MMP-1)

(ref. ²⁰), secreted protein acidic and rich in cysteine (SPARC)²¹, osteoclast-associated receptor (OSCAR)²², glycoprotein VI (ref. ²³) and leukocyte-associated Ig receptor 1 (LAIR1)²⁴.

The broad success of the Toolkit approach is possible in part because both COL2 and COL3 are homotrimers; that is, all three polypeptide chains of the triple helix are identical. Thus, a synthetic peptide containing a collagen-like sequence of sufficient length rapidly folds into a triple helix without the need for additional design intervention. By contrast, designing peptide mimics of heterotrimeric collagens containing either two (AAB-type) or three (ABC-type) unique chains is challenging owing to the combinatorial explosion of possible triple helices in a mixture of two or three peptides. This has restricted the *in vitro* study of heterotrimeric collagens such as COL1 (an AAB-type heterotrimer), the most abundant mammalian collagen²⁵, and has given rise to a prolonged debate in collagen research. Peptides in a collagen triple helix self-assemble with one amino acid offset to optimize molecular packing. Thus, chain B in COL1 can be permuted in either the leading (BAA), middle (ABA) or trailing (AAB) position, resulting in isomeric triple helices that would bind CBPs with varying affinities. The precise position of chain B in COL1 is unknown and has been intensely debated since its heterotrimeric nature became known 50 years ago²⁵. To add to the mystery, all three combinations have variously been proposed based on computational analysis of the COL1 sequence²⁶, interchain interactions²⁷, molecular packing²⁸ and fibrillar architecture²⁹. Here, we provide an empirical demonstration that chain B in COL1 resides in the trailing position.

Three defined-register collagen heterotrimers containing permutations of a portion of sequence from chain A and chain B of COL1 that is predicted to bind DDR1, DDR2 and VWF¹⁷ were computationally designed. Of the three permutations, AAB demonstrated a clear and strong preference for DDR1 and VWF in solid-phase binding assays and also induced distinctly higher levels of cellular DDR1 and DDR2 kinase activation. AAB also selectively inhibited binding of VWF to a high-affinity surface-coated homotrimeric

¹Department of Biochemistry, University of Cambridge, Cambridge, UK. ²National Heart and Lung Institute, Imperial College London, London, UK.

³Department of Chemistry and Bioengineering, Rice University, Houston, TX, USA. ⁴Present address: Department of Biochemistry, University of Bayreuth, Bayreuth, Germany. *e-mail: jalana@cantab.net

peptide. These results provide the first direct proof that chain alignment in COL1 is AAB and resolve a five-decade-old conundrum in collagen research.

Results

Salt bridges direct heterotrimeric register. CBPs recognize highly specific amino acid motifs within collagen. For example, DDR1 (ref. ¹⁸), DDR2 (ref. ¹⁹) and VWF³⁰ recognize a homotrimeric RGQGVGMGFO sequence (O=4(R)-hydroxyproline, Hyp) conserved in human COL2 and COL3. A comparison with human COL1 revealed a similar site, ARGQAGVMGFO at sequence positions 573–583 in chain A and ARGEONGNIGFO at the corresponding positions 485–495 in chain B, that could potentially bind these three CBPs if the two sequence motifs could be reconstituted in the native-like chain alignment¹⁷. Thus, our primary design challenge was to incorporate two copies of COL1 chain A and one copy of COL1 chain B sequence within a triple helix such that chain B is aligned in the three possible AAB, ABA and BAA registers.

Previously, proof-of-principle defined-register AAB-³¹ and ABC-type^{32,33} heterotrimers were designed by exploiting the geometric and sequence specificity of Lys–Asp and Lys–Glu salt bridges demonstrated within the context of collagen³⁴. In brief, multiple Lys–Asp and Lys–Glu salt bridges are introduced rationally or computationally such that they all form salt bridges only in the target heterotrimeric register. By design, the competing triple-helical states contain multiple unpaired Lys and Asp or Glu, and are thus unstable with respect to the target state. Paired and unpaired Lys and Asp or Glu residues represent elements of positive and negative design, respectively, and ensure a high degree of specificity during heterotrimer self-assembly.

In our case, the design of the three registers is based on a host-guest scheme in which the putative VWF and DDR recognition epitope in COL1 mentioned above is incorporated as a guest between two flanking host domains (Fig. 1). The host domains contain multiple Lys–Asp salt bridges that lock the triple helix into the AAB, ABA or BAA register (Fig. 1a). Flanking domains are computationally designed using a modified genetic algorithm developed to find ABC heterotrimers³⁵. Our genetic algorithm generates a population of 100 pairs of peptides A and B whose sequence is restricted to the use of Pro–Pro–Gly, Asp–Pro–Gly, Pro–Lys–Gly and Asp–Lys–Gly amino acid triplets but is otherwise random. These pairs of peptides are then assigned all eight possible triple-helical compositions and registers (AAA, BBB, AAB, ABA, BAA, ABB, BAB and BBA). Each composition and register is scored in the following fashion: any instance of unpaired Asp or Lys is penalized by one point, and any Lys–Asp pairs that are found in the appropriate geometry to form a salt bridge are awarded two points. Each pair of peptides is then assigned a stability score based on its best register and a specificity score based on the difference between the best and second-best register. From the initial population of 100 pairs of peptides, the triple helix with the highest specificity score for the desired register (with ties broken by the best stability) is saved and reproduced with mutation in a subsequent generation. This process iterates through multiple generations until a target specificity and stability are found. Here, the algorithm was run to design AAB, ABA or BAA registers, as needed.

Computational design resulted in two peptides for self-assembling ABA. However, only three peptides were needed for self-assembling the other two registers, as the sequence of peptide B was found to be common to both AAB and BAA registers. These five peptides were synthesized by automated solid-phase peptide synthesis, purified, self-assembled and crystallized from aqueous solutions using commercial screens (Online Methods, Supplementary Fig. 1 and Supplementary Tables 1–3). Crystal structures of the heterotrimers solved to near-atomic resolution confirmed the chain registration intended by the computational design (Fig. 1b). 2Fo–Fc maps of

the flanking sequences and recognition epitope in Supplementary Fig. 2 show clear electron densities for each residue of the recognition sequence. The peptide chains in each register were offset by one residue with respect to each other, and chain B was permuted in either the leading (BAA), middle (ABA) or trailing (AAB) position. Thus, the putative VWF, DDR1 and DDR2 recognition epitope from COL1 was obtained in all three possible registers. Salt bridge analysis using Visual Molecular Dynamics (VMD) software³⁶ with a distance cutoff of 3.2 Å between the Lys N ϵ and Asp O δ atoms revealed that all 12 Lys and Asp residues incorporated by the computational design formed salt bridges in BAA and ABA, while ten salt bridges formed in AAB.

The homotrimeric RGQGVGMGFO sequence has been shown to bind a partially conserved amphipathic pocket in DDR2 and VWF A3 via the Phe residues, and its mutation to Ala completely abrogates binding^{19,30}. Furthermore, the VWF A3 domain¹⁷ and DDR2 (ref. ³⁷) bind the COL2 sequence selectively through the leading and middle chain Phe, respectively. As both chain A (ARGQAGVMGFO) and chain B (ARGEONGNIGFO) of the COL1 recognition epitope contain a Phe residue, we designed three variants of AAB containing a Phe to Ala mutation in the two A chains (AAB–alaA), the B chain only (AAB–alaB) or both the A and B chains (AAB–alaAB) to understand their chain-specific role in CBP recognition. These mutant heterotrimers were not crystallized and were only analyzed by NMR spectroscopy (Online Methods). Thus, we obtained a total of six heterotrimers, AAB, AAB–alaA, AAB–alaB, AAB–alaAB, ABA and BAA, which were covalently captured and then used in functional CBP assays.

Salt bridges direct covalent capture of registers. The equilibrium population of salt-bridge-stabilized collagen heterotrimers is extremely sensitive to total peptide concentration, ionic strength of the buffer, pH, temperature and amino acid sequence³⁸. To remove any bias in the binding affinities owing to the broad concentration range and variable buffer conditions needed for solid-phase and cellular activation assays, we covalently captured all registers and alanine mutants as single triple helices. Crystal structures show that each heterotrimer contains between 10 and 12 salt bridges in the sequence flanking the protein recognition epitope. We converted these salt bridges into isopeptide bonds in which the ammonium group of the Lys side chain forms an amide bond with the carboxylate group of Asp using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) amide coupling chemistry (Online Methods)³⁹. The covalently captured constructs were purified and characterized using size exclusion chromatography and MS, respectively (Supplementary Fig. 3).

Covalent capture does not alter heterotrimer register. The triple-helical fold and chain registration of the heterotrimers before and after covalent capture were assessed for consistency using CD and NMR spectroscopy, respectively. CD spectrographs of all native and covalently captured heterotrimers show a strong minimum at approximately 195 nm and a weakly positive maximum at approximately 225 nm, characteristic of collagen triple helices (Supplementary Fig. 4a). This suggests that the triple-helical fold is preserved after covalent capture. Thermal denaturation curves of native heterotrimers monitored at 225 nm show a cooperative unfolding transition, indicating the presence of well-folded and stable triple helices in solution (Supplementary Fig. 4b). AAB is less thermally stable than BAA, ABA and the alanine mutant constructs of AAB. The lower thermal stability of AAB compared with the alanine mutants is consistent with the experimental amino acid propensities determined previously, in which substitution of Pro or Hyp with Ala destabilizes a canonical [(GPO) α_1]₃ triple helix less than substitution with Phe⁴⁰. The denaturation curves of the covalently captured heterotrimers show large dispersion in the CD

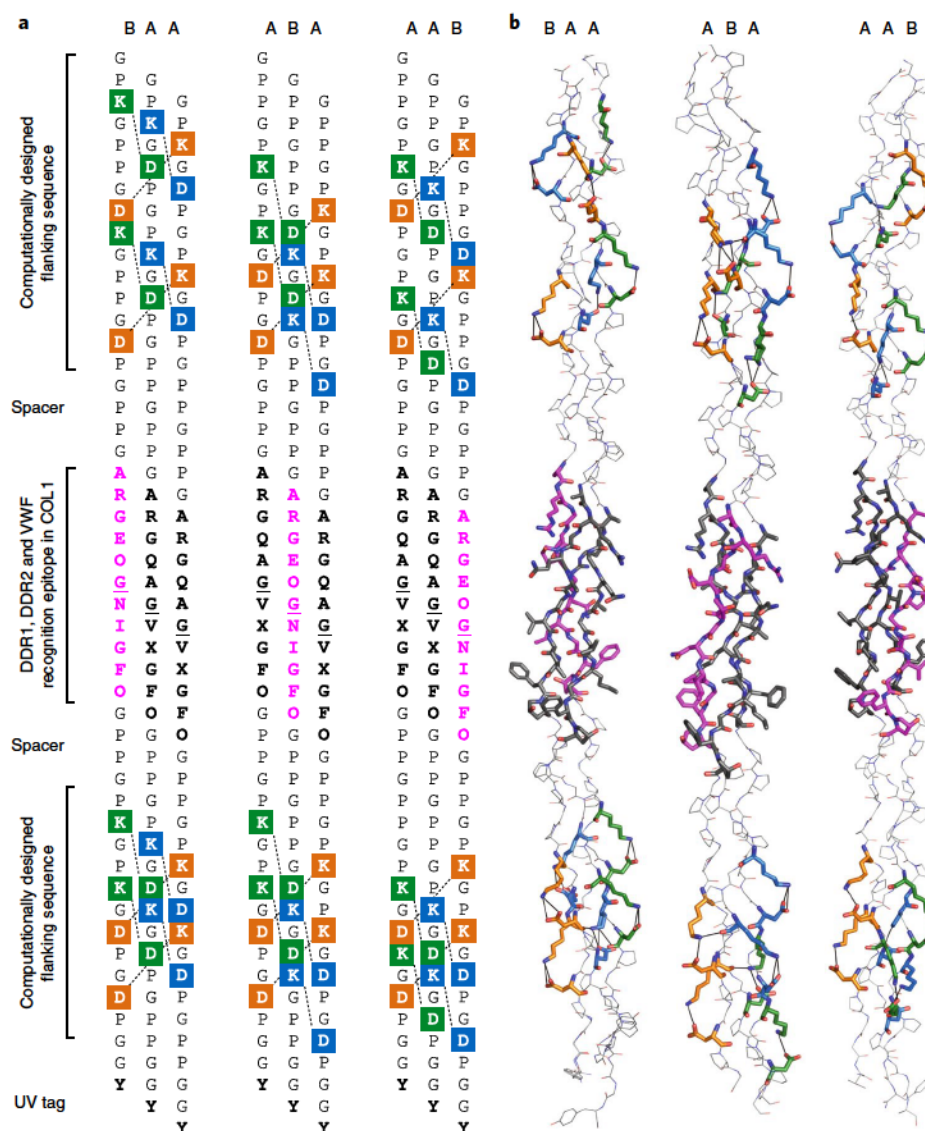


Fig. 1 | Design and structure of register-specific heterotrimers. a, b, Design of heterotrimeric registers (**a**) and corresponding crystal structures (**b**), showing the putative DDR1, DDR2 and VWF recognition epitope in COL1 (in bold) and the computationally designed flanking sequences that drive self-assembly of peptides into BAA, ABA or AAB registers. Chains A and B of the putative COL1 recognition epitope are colored black and magenta, respectively. For clarity, hydrogen atoms and solvent molecules are not shown. Lys–Asp salt bridges between the leading-middle, middle-trailing and trailing-leading chains are colored green, blue and orange, respectively. ^{15}N -isotopically enriched Gly residues are underlined, the Tyr used to determine peptide concentration is shown in bold, and X denotes the methionine bioisostere, norleucine. Electron density maps of the recognition epitope and flanking sequences are presented in Supplementary Fig. 2.

signal during unfolding. This is expected, as the solution of covalently captured heterotrimers contains an ensemble of triple helices with identical CBP recognition sequences but different fractions of Lys–Asp isopeptide bonds in the flanking sequences. In addition, the triple helices in this ensemble contain variable numbers of amides owing to the contribution of isopeptide bonds and thus would have variable stability and rigidity along the triple-helical axis. The observed melting curves are therefore consistent with simultaneous denaturation of an ensemble of covalently captured triple helices with varying molar residue ellipticity, stability and effective length of the spectroscopic unit.

The ^{15}N isotopically enriched Gly residue present at sequence position 25 in the recognition epitope was used to study native and covalently captured heterotrimers in the solution state by NMR

and detect any change in their register after covalent capture. Two-dimensional ^1H , ^{15}N -HSQC spectra of the native heterotrimers AAB, ABA, BAA and AAB-alaB show three trimer cross peaks and two monomer cross peaks (Supplementary Fig. 4c), consistent with a single heterotrimeric species in solution. In contrast, the HSQC spectra of alanine mutants AAB-alaA and AAB-alaAB show six trimer cross peaks, three for the heterotrimer and an additional three for the A_3 homotrimer. Reassuringly, the HSQC spectra of all covalently captured heterotrimers and the alanine mutants show only three heterotrimer cross peaks. Absence of the homotrimer cross peaks in the HSQC spectra of the covalently captured AAB-alaA and AAB-alaAB may mean that the A_3 homotrimers with their many unpaired Lys and Asp residues are not amenable to efficient covalent capture. Alternatively, it may mean that they are removed

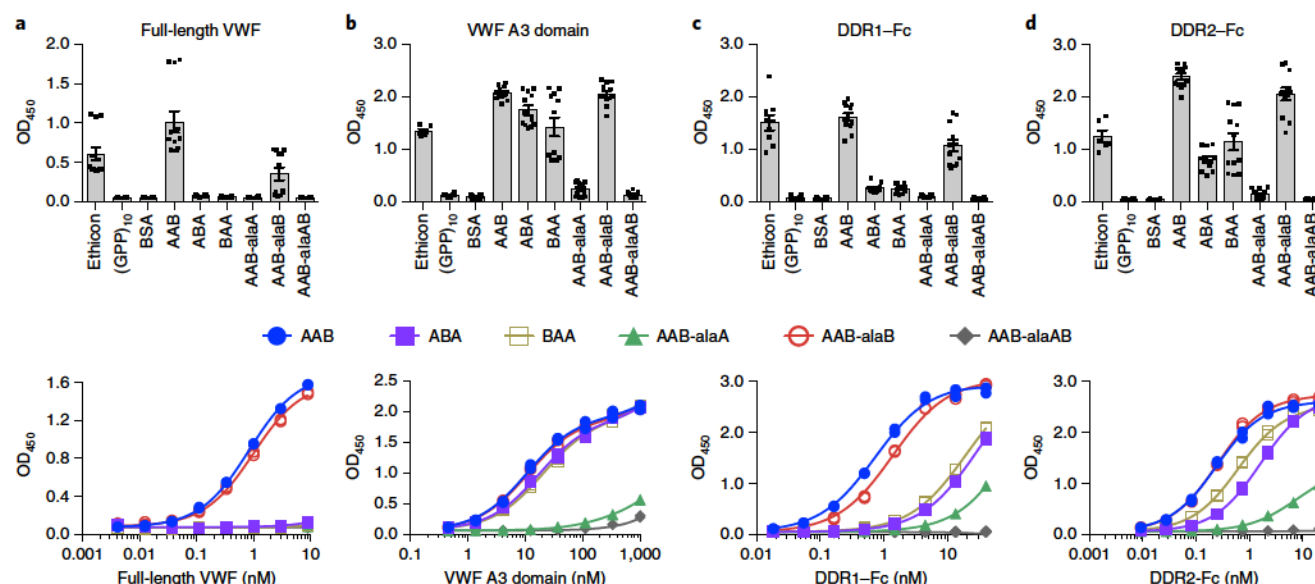


Fig. 2 | Binding affinity of registers for collagen-binding proteins. **a–d**, Solid-phase assay (top) and dose-response curves (bottom) for the binding of full-length VWF (**a**), recombinant VWF A3 domain (**b**), DDR1-Fc (**c**) and DDR2-Fc (**d**) to the covalently captured heterotrimers AAB, ABA and BAA and the alanine mutants AAB-alaA, AAB-alaB and AAB-alaAB. Ethicon collagen I fibers were used as a positive control, and the inert triple-helical peptide (GPP)₁₀ and bovine serum albumin (BSA) were used as negative controls. Data represent the mean of $n=4$ independent experiments performed in triplicate with two independently prepared batches of covalently captured heterotrimers. Each point represents an optical density at 450 nm (OD_{450}) measurement from a single well. Error bars indicate the mean $\pm$ s.e.m. The binding isotherms were fit to a model of specific and non-specific binding in GraphPad Prism 6.0, and non-specific binding curves were removed for clarity. All replicates are shown in dose-response curves. Dissociation constants obtained from the dose-response curves are shown in Supplementary Table 4.

during size exclusion chromatography purification owing to differences in the surface properties of the electrostatically neutral heterotrimer and charged homotrimers.

Two of the three cross peaks of covalently captured heterotrimers have chemical shifts equivalent to those in native heterotrimers, but a small deviation is observed for the third cross peak. Amide chemical shifts are sensitive to chemical exchange and protection from the solvent. Given that these would be altered substantially after covalent capture, the small deviation observed is not surprising. Taken together, the CD and NMR characterizations confirm that the triple-helical fold and heterotrimer register remain unchanged upon covalent capture.

AAB shows higher affinity for all three CBPs. The covalently captured constructs were used in solid-phase and dose-response assays with recombinant proteins that comprised the entire extracellular regions of DDR1 and DDR2 fused carboxy-terminally to the Fc sequence of human IgG2 (hereafter called DDR1-Fc and DDR2-Fc), the recombinant A3 domain of VWF (VWF A3) or the full-length VWF. Cellular kinase activation assays were performed with DDRs expressed transiently in HEK293 cells (see Online Methods).

Solid-phase binding assays show that AAB binds full-length VWF, the recombinant VWF A3 domain, DDR1-Fc and DDR2-Fc with significantly higher affinity than do BAA and ABA (Fig. 2 and Supplementary Table 4). For example, full-length VWF bound AAB with sub-nanomolar affinity but did not elicit any detectable response in ELISA with either BAA or ABA (Fig. 2a). AAB also selectively inhibited VWF binding to a surface coated with the high-affinity RGQOGVMGFO sequence (Supplementary Fig. 5). Surprisingly, the recombinant VWF A3 domain, which is recognized as the primary locus of COL1–COL3 interaction in full-length VWF^{34,41}, discriminated the least between the three registers, with only a twofold difference in affinity between AAB and BAA or

ABA (Fig. 2b). This suggests a more complex mechanism of VWF recognition by fibrillar collagens, perhaps involving neighbouring VWF A domains, than is currently understood based on the structure of the VWF A3 domain in complex with the RGQOGVMGFO homotrimer¹⁷. A similar trend in binding affinities was observed with DDR1-Fc, which showed more than 25-fold higher affinity for AAB than for BAA or ABA (Fig. 2c). The DDR2-Fc affinity for all registers was higher than that of DDR1-Fc, but in this case too, DDR2-Fc bound AAB with 2.5-fold and 6-fold higher affinity than BAA and ABA, respectively (Fig. 2d). The lower specificity of DDR2-Fc is not surprising; it has been shown to recognize multiple loci in COL2 and COL3, in contrast to DDR1-Fc, which recognizes a unique site with high affinity in both collagens^{18,19}.

AAB-alaB bound all four proteins with an affinity similar to that of AAB, with the exception of DDR1, for which it showed a twofold lower affinity. However, mutant constructs AAB-alaA and AAB-alaAB did not bind any of the CBPs. Although structural determinants of DDR1–collagen interaction are not known, these results suggest that Phe residues on chain A of COL1 are critical for recognition of all three CBPs.

AAB induces the highest levels of DDR phosphorylation. The DDRs respond to collagen binding with intracellular kinase activation, which is primarily manifested as receptor autophosphorylation⁴². To test the ability of the heterotrimeric peptides to stimulate DDR autophosphorylation, cells expressing full-length DDR1 or DDR2 were stimulated with the peptides, and DDR phosphorylation was detected on western blots of cell lysates using phospho-specific antibodies (see Online Methods). As anticipated from the binding data presented in Fig. 2, AAB induced the highest levels of DDR1 and DDR2 autophosphorylation (Fig. 3 and Supplementary Figs. 6 and 7). BAA was able to stimulate some DDR phosphorylation, albeit to significantly lower levels than AAB, while ABA could

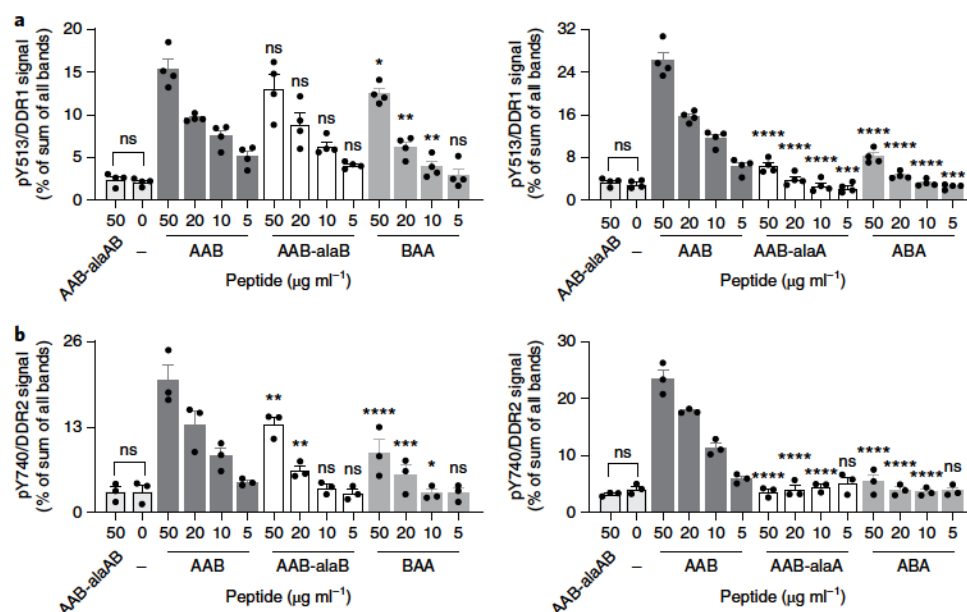


Fig. 3 | Peptide-induced DDR1 and DDR2 autophosphorylation. **a, b**, HEK293 cells transiently expressing DDR1 (**a**) or DDR2 (**b**) were stimulated with the indicated peptides (concentration in $\mu\text{g ml}^{-1}$) or left unstimulated (–) for 90 min at 37 °C. Cell lysates were analyzed for phospho-tyrosine (anti-pY513 for DDR1; anti-pY740 for DDR2) and total DDR1 or DDR2. AAB- and AAB-alaB-stimulated samples were included on all blots as positive and negative controls, respectively. Representative blots are shown in Supplementary Figs. 6, 7. Each point on the graph shows quantitation of phospho-DDR signals for one measurement relative to total DDR levels, expressed as a percentage of the sum of all the bands on a blot, with the mean and s.e.m. shown ($n=4$ independent experiments for DDR1; $n=3$ independent experiments for DDR2). Statistical significance for each peptide concentration compared with the corresponding concentration of AAB is presented. Significance between AAB-alaB and the unstimulated control is also shown. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns (not significant, $P > 0.05$), two-way ANOVA followed by Bonferroni post hoc test. Details of the statistical analysis, including precise P values, are presented in Supplementary Tables 5, 6.

induce only marginal DDR1 phosphorylation and no detectable DDR2 phosphorylation. AAB-alaB stimulated DDR1 phosphorylation with similar potency to AAB, while AAB-alaA did not elicit detectable DDR1 phosphorylation. Peptide stimulation of DDR2 with the alanine mutant AAB peptides led to similar results, but AAB-alaB was significantly less potent than AAB. These results corroborate the findings of the solid-phase binding assays that the Phe residue in chain A of COL1 plays a crucial part in the binding and activation of DDR.

Chain alignment determines COL1–CBP binding specificity. Crystal structures show that the homotrimeric sequence RGQOGVMGFO, conserved in COL2 and COL3, recognizes VWF A3 (ref. 17) and DDR2 (ref. 37) via the leading and middle chain Phe, respectively. Although it could potentially bind both CBPs through Phe on either of the other two chains, these alternative binding modes are not observed. As shown in Supplementary Fig. 8, Phe binds a partly conserved amphipathic pocket in both VWF A3 and DDR2, and peripheral collagen residues make polar contacts that supplement the binding interface. Binding through Phe on the other two chains leads to substitution of residues that form polar contacts with those that result in suboptimal interfacial interaction. Similar analysis reveals the reason for the discrimination observed in binding of CBPs to the three registers.

Henceforth, residues within the recognition epitope are denoted with a prefix L (leading), M (middle) or T (trailing) followed by their three-letter residue code and residue position. Unique residues within each chain are denoted using only their three-letter residue codes. For example, Phe residues in the three chains are denoted as L:Phe, M:Phe and T:Phe, and Hyp residues in the trailing peptide chain B are denoted as T:Hyp24 and T:Hyp30.

As both chains A and B of the COL1 recognition sequence contain a Phe, each CBP can potentially recognize AAB, BAA and ABA through any of the nine possible binding modes (Fig. 4). Assuming that the mechanism of VWF and DDR2 recognition in COL1 and COL2 is conserved, the interfacial interactions observed in crystal structures of RGQOGVMGFO in complex with DDR2 and VWF A3 shown in Supplementary Fig. 8 are replicated only when AAB binds DDR2 through the M:Phe (Fig. 4a) and the VWF A3 domain through the L:Phe (Fig. 4b). The two remaining binding modes of AAB and all six binding modes of BAA and ABA result in loss of one or more interfacial interactions. For example, in the case of DDR2, binding through the M:Phe of ABA replaces hydrophobic M:Val with bulkier and hydrophilic Asn. Similarly, in the case of the VWF A3 domain, binding through the L:Phe of ABA replaces polar T:Hyp with non-polar Ala, resulting in the loss of a hydrogen bond. Thus, heterotrimeric collagens have evolved a high degree of specificity of interaction with CBPs by moderate changes in amino acid composition and alignment of peptide chains within the triple helix.

Discussion

Our results provide direct proof that chain B in COL1 resides in the trailing position of the triple helix, based on the markedly high affinity of AAB for two classes of CBPs with distinct protein folds and different tissue localization and function. In addition, we also begin to appreciate how collagens have evolved to exploit triple-helical peptide composition and register to achieve high levels of CBP recognition specificity.

Besides advancing our understanding of COL1, our design strategy can be adapted to determine the register of the other major AAB-type heterotrimer, COL4. This would help to resolve another

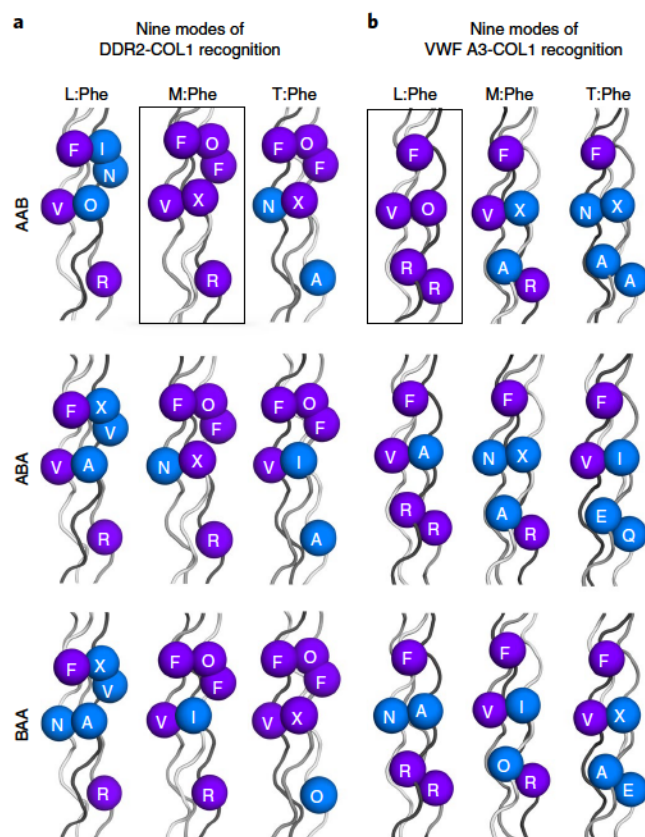


Fig. 4 | Nine possible modes for binding of COL1 to DDR2 and VWF A3. **a,b.** Schematic drawing of AAB, ABA and BAA core residues that would contact DDR2 (**a**) or the VWF A3 domain (**b**) when binding through L:Phe, M:Phe or T:Phe at a distance of 4.5 Å. CBPs can access the Phe residues on the leading, middle or trailing chain (shown in increasing shades of grey) by an approximately 120° rotation and 4 Å C-terminal translation through the triple-helical axis (not shown). Residues that correspond to the binding site observed in the crystal structures of COL2 homotrimer peptides in complex with DDR2 or the VWF A3 domain (depicted in Supplementary Fig. 8) are shown in purple, and those replaced with suboptimal interactions are shown in blue. The single-letter amino acid code O denotes Hyp, and X denotes the methionine bioisostere, norleucine. The binding mode that results in conservation of all interfacial contacts is shown in a box.

longstanding issue in collagen research. Both DDR1 and DDR2 recognize fibrillar collagens COL1–COL3. However, DDR1 selectively recognizes network-forming COL4 present in the basal lamina⁴². The structural basis for this selectivity is not known¹⁸. The DDR recognition epitope in COL1 identified here can be used as a guide to determine putative binding sites in COL4. Subsequent design of register-specific heterotrimers containing these epitopes and mapping their interaction with DDR1 would reveal the structural basis of COL4 selectivity for DDR1. In addition, a low-resolution structural model of COL1 derived from fiber diffraction data²⁹ can now be improved with the knowledge of the correct chain alignment.

Therapeutic agents that disrupt DDR–collagen and VWF–collagen interactions are potential drug targets in preventing cancer progression and hemostasis, respectively. Moreover, numerous point mutations in COL1 cause genetic diseases such as osteogenesis imperfecta, or brittle bone disease⁴³. Structural perturbations in COL1 due to mutations in either or both chains can be correctly modeled and correlated with the resulting phenotypic severity only if the alignment of chain B is precisely known. Thus, our work can

potentially advance understanding of both genetic and physiological aspects of collagen function.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-019-0435-y>.

Received: 13 March 2019; Accepted: 20 November 2019;

Published online: 06 January 2020

References

- Leitinger, B. Transmembrane collagen receptors. *Annu. Rev. Cell Dev. Biol.* **27**, 265–290 (2011).
- Rosini, S. et al. Thrombospondin-1 promotes matrix homeostasis by interacting with collagen and lysyl oxidase precursors and collagen cross-linking sites. *Sci. Signal.* **11**, eaar2566 (2018).
- Lankhof, H. et al. A3 domain is essential for interaction of von Willebrand factor with collagen type III. *Thromb. Haemost.* **75**, 950–958 (1996).
- Santoro, S. A. Preferential binding of high molecular weight forms of von Willebrand factor to fibrillar collagen. *Biochim. Biophys. Acta* **756**, 123–126 (1983).
- De Meyer, S. F., Stoll, G., Wagner, D. D. & Kleinschnitz, C. von Willebrand factor: an emerging target in stroke therapy. *Stroke* **43**, 599–606 (2012).
- Fu, H. L. et al. Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling. *J. Biol. Chem.* **288**, 7430–7437 (2013).
- Gross, O. et al. DDR1-deficient mice show localized subepithelial GBM thickening with focal loss of slit diaphragms and proteinuria. *Kidney Int.* **66**, 102–111 (2004).
- Vogel, W. F. et al. Discoidin domain receptor 1 tyrosine kinase has an essential role in mammary gland development. *Mol. Cell. Biol.* **21**, 2906–2917 (2001).
- Hou, G., Vogel, W. & Bendeck, M. P. The discoidin domain receptor tyrosine kinase DDR1 in arterial wound repair. *J. Clin. Invest.* **107**, 727–735 (2001).
- Labrador, J. P. et al. The collagen receptor DDR2 regulates proliferation and its elimination leads to dwarfism. *EMBO Rep.* **2**, 446–452 (2001).
- Valiathan, R. R. et al. Discoidin domain receptor tyrosine kinases: new players in cancer progression. *Cancer Metastasis Rev.* **31**, 295–321 (2012).
- Raynal, N. et al. Use of synthetic peptides to locate novel integrin $\alpha\beta_1$ -binding motifs in human collagen III. *J. Biol. Chem.* **281**, 3821–3831 (2006).
- Farndale, R. W. et al. Cell–collagen interactions: the use of peptide Toolkits to investigate collagen–receptor interactions. *Biochem. Soc. Trans.* **36**, 241–250 (2008).
- Kim, J. K. et al. A novel binding site in collagen type III for integrins $\alpha\beta_1$ and $\alpha\beta_2$. *J. Biol. Chem.* **280**, 32512–32520 (2005).
- Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. Structural basis of collagen recognition by integrin $\alpha\beta_1$. *Cell* **101**, 47–56 (2000).
- Hamaia, S. W. et al. Unique charge-dependent constraint on collagen recognition by integrin $\alpha 10\beta 1$. *Matrix Biol.* **59**, 80–94 (2017).
- Brondijk, T. H. C., Bihan, D., Farndale, R. W. & Huizinga, E. G. Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc. Natl Acad. Sci. USA* **109**, 5253–5258 (2012).
- Xu, H. et al. 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 (2011).
- Konitsiotis, A. D. et al. Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **283**, 6861–6868 (2008).
- Manka, S. W. et al. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Natl Acad. Sci. USA* **109**, 12461–12466 (2012).
- Hohenester, E., Sasaki, T., Giudici, C., Farndale, R. W. & Bächinger, H. P. Structural basis of sequence-specific collagen recognition by SPARC. *Proc. Natl Acad. Sci. USA* **105**, 18273–18277 (2008).
- Zhou, L. et al. Structural basis for collagen recognition by the immune receptor OSCAR. *Blood* **127**, 529–537 (2016).
- Munnix, I. C. A. et al. Collagen-mimetic peptides mediate flow-dependent thrombus formation by high- or low-affinity binding of integrin $\alpha 2\beta 1$ and glycoprotein VI. *J. Thromb. Haemost.* **6**, 2132–2142 (2008).
- Lebbink, R. J. et al. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol.* **28**, 202–210 (2009).
- Piez, K. A., Eigner, E. A. & Lewis, M. S. The chromatographic separation and amino acid composition of the subunits of several collagens. *Biochemistry* **2**, 58–66 (1963).

26. Piez, K. A. & Trus, B. L. Sequence regularities and packing of collagen molecules. *J. Mol. Biol.* **122**, 419–432 (1978).
27. Traub, W. & Fietzek, P. P. Contribution of the $\alpha 2$ chain to the molecular stability of collagen. *FEBS Lett.* **68**, 245–249 (1976).
28. Bender, E., Silver, H., Hayashi, K. & Trelstad, R. L. Type I collagen segment long spacing banding patterns. *J. Biol. Chem.* **257**, 9653–9657 (1982).
29. Orgel, J. P. R. O., Irving, T. C., Miller, A. & Wess, T. J. Microfibrillar structure of type I collagen in situ. *Proc. Natl Acad. Sci. USA* **103**, 9001–9005 (2006).
30. Lisman, T. et al. A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **108**, 3753–3756 (2006).
31. Jalan, A. A. & Hartgerink, J. D. Simultaneous control of composition and register of an AAB-type collagen heterotrimer. *Biomacromolecules* **14**, 179–185 (2013).
32. Xu, F., Zhang, L., Koder, R. L. & Nanda, V. De novo self-assembling collagen heterotrimers using explicit positive and negative design. *Biochemistry* **49**, 2307–2316 (2010).
33. Zheng, H. et al. How electrostatic networks modulate specificity and stability of collagen. *Proc. Natl Acad. Sci. USA* **115**, 6207–6212 (2018).
34. Gauba, V. & Hartgerink, J. D. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.* **129**, 2683–2690 (2007).
35. Fallas, J. A. & Hartgerink, J. D. Computational design of self-assembling register-specific collagen heterotrimers. *Nat. Commun.* **3**, 1087–1088 (2012).
36. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
37. Carafoli, F. et al. Crystallographic insight into collagen recognition by discoidin domain receptor 2. *Structure* **17**, 1573–1581 (2009).
38. Jalan, A. A., Demeler, B. & Hartgerink, J. D. Hydroxyproline-free single composition ABC collagen heterotrimer. *J. Am. Chem. Soc.* **135**, 6014–6017 (2013).
39. Li, L.-C. et al. *Org. Lett.* **21**(14), 5480–5484 (2019). <https://doi.org/10.1021/acs.orglett.9b01771>.
40. Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A. & Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960–14967 (2000).
41. Houdijk, W. P. M., Sakariassen, K. S., Nievelstein, P. F. E. M. & Sixma, J. J. Role of factor VIII-von Willebrand factor and fibronectin in the interaction of platelets in flowing blood with monomeric and fibrillar human collagen types I and III. *J. Clin. Invest.* **75**, 531–540 (1985).
42. Vogel, W., Gish, G. D., Alves, F. & Pawson, T. The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol. Cell* **1**, 13–23 (1997).
43. Bodian, D. L., Madhan, B., Brodsky, B. & Klein, T. E. Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry* **47**, 5424–5432 (2008).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Synthesis and purification of peptides. The peptides were synthesized on a CEM Liberty Blue microwave-assisted peptide synthesizer using standard Fmoc (9-fluorenylmethoxycarbonyl) chemistry on a solid-phase support, cleaved from the resin and purified via HPLC, as described previously¹³. Arg was coupled twice at room temperature (20 °C) to prevent γ -lactam formation, which leads to Arg deletion. A ¹⁵N isotopically enriched Gly residue was coupled at the Gly 25 position in all peptides to assist in NMR spectroscopic analysis. Asp frequently undergoes intramolecular cyclization whereby its side chain ester undergoes a nucleophilic attack from an amide group to form aspartimide when it is succeeded on the C-terminal side by Gly, Asn, Ser or Ala. The racemized aspartimide can undergo hydrolysis to give a mixture of α - and β -peptides. We observed aspartimide formation in peptides containing Lys on the C-terminal side of Asp, that is in Asp-Lys-Gly triplets. To prevent the formation of aspartimide, we chose 5% piperazine containing 0.1 M N-hydroxybenzotriazole (HOBT) for deprotection and also used 3-methylpent-3-yl ester (OMpe) of the Asp side chain instead of *t*-butyl ester wherever Asp was followed by an amino acid. The steric bulk of the OMpe ester compared with that of the *t*-butyl ester prevents a nucleophilic attack from the NH attached to the α -carboxy group. Synthesis using these methods yielded peptides free of aspartimide. MALDI-TOF spectra of all peptides are shown in Supplementary Fig. 1.

Protein expression and purification. Soluble recombinant His₆-tagged VWF A3 domain was expressed as described previously³⁰. Soluble recombinant extracellular proteins comprising the entire extracellular regions of DDR1 and DDR2 fused C-terminally to the Fc sequence of human IgG2 were produced in episomally transfected HEK293-EBNA cells and purified by affinity chromatography as described previously^{18,44}. Native human full-length VWF was purchased from Abcam (ab85533).

Preparation of samples for NMR, CD and covalent capture. Purified peptides were dissolved in MilliQ water at 25–30 mg ml⁻¹, adjusted to pH 7.0 using 1 M NaOH solution, and their concentrations determined spectrophotometrically using a NanoDrop 1000 (Thermo Fisher). Appropriate volumes of the different peptide stock solutions were mixed and diluted with 100 mM sodium phosphate buffer at pH 7.0 and deionized water to a final buffer concentration of 10 mM and a total peptide concentration of 3 mM for the heterotrimers and 1 mM for the homotrimers. The final heterotrimer solutions contained a 2:1 molar ratio of peptides A and B. The pH of the final peptide solutions was adjusted to 7.0 with 1 M NaOH, if necessary, and annealed at 85 °C for 15 min. Annealed solutions were incubated at room temperature for at least 3 d before taking measurements. The 3 mM stock solutions of heterotrimer or 1 mM stock solutions of homotrimer were used for all subsequent NMR and CD measurements as well as for the covalent capture of heterotrimers.

Preparation of samples for crystallization. Aqueous solutions of purified peptides A and B were mixed in a 2:1 molar ratio at 20 mg ml⁻¹ total peptide concentration, adjusted to pH 7.8, 8.6 and 9.2 with 1 M NaOH for AAB, ABA and BAA, respectively, and annealed at 85 °C for 15 min. The peptides were initially stocked in aqueous solutions without any buffer because the two buffer conditions first used to stock the peptides, potassium phosphate and Tris-HCl, were found to markedly lower the number of positive hits obtained during crystallization screens. The annealed solutions were incubated at room temperature for at least 3 d, diluted to 10, 6 and 10 mg ml⁻¹ concentrations for AAB, ABA and BAA, respectively, and subsequently used for crystallization trials. The crystallization conditions were obtained using commercial screens. For all three peptide constructs, we used PEG I and PEG II from Qiagen and Wizard Classic 1 & 2, Wizard Classic 3 & 4, JCSG Plus and MIDAS MD1–60 from Molecular Dimensions. The total reservoir volume in the wells was 200 μ L. Crystals were grown using the sitting drop vapor diffusion method in 96-well two-drop MRC crystallization plates. Two hundred nanoliters of 6–10 mg ml⁻¹ peptide solutions and 200 nl of reservoir solutions were aliquoted using the mosquito Crystal liquid handler from TTP Labtech, immediately sealed with UV-transparent Scotch tape, stored in Formulatrix Rock Imager 1000 and scanned every 4 h for the first day and daily thereafter. Crystals appeared within a few hours to 2 d at 20 °C. We obtained multiple hits for each register and chose one each for diffraction based on size and visual inspection of the crystal morphology. For cryoprotection, crystals were transferred to a 15% solution of glycerol in the mother liquor and immediately frozen in liquid nitrogen. The BAA crystals grown in 50% ethylene glycol were not cryo-protected in glycerol. Diffraction data were collected at 100 K on beamlines I03, I04 and I24 of the Diamond Light Source synchrotron facility. AAB, BAA and ABA datasets were collected at wavelengths of 0.7000, 0.8000 and 0.9795 Å, respectively. The crystallographic data collection and refinement statistics and the crystallization conditions are provided in Supplementary Tables 1 and 3, respectively.

Strategy for solving the crystal structures. The data for AAB and ABA were indexed and integrated using autoPROC⁴⁵. However, owing to unsatisfactory indexing, the data for BAA were indexed and integrated using DIALS⁴⁶. Subsequently, all data were scaled and merged in AIMLESS⁴⁷ and truncated

using CTRUNCATE⁴⁸. The crystal structures of heterotrimers were solved using molecular replacement in Phaser⁴⁹. Idealized collagen triple helices containing 7, 10, 13, 16, 19, 22, 25 and 40 amino acid peptide chains in a one-residue offset 7/2 helical conformation were generated in TheBuScr⁵⁰. A second copy of each triple helix was also generated with amino- and C-terminal end residues removed to create blunt-ended termini. These 16 models along with crystal structures deposited under Protein Data Bank (PDB) codes 1V4F, 1V6Q and 1V7H, which are 21-amino-acid triple helices at varying resolutions, were used to search for molecular replacement solutions in Phaser⁴⁹. In general, molecular replacement with the blunt-ended triple helices gave fewer solutions in each search with higher translation function Z-scores (TFZs) and log-likelihood gain values compared with the triple helices with end offsets. The TFZ of blunt-ended triple helices increased non-linearly with chain length, plateaued between 13 and 22 residues, and then decreased. In each case, the model that gave a single solution and the highest TFZ was selected for subsequent model building and refinement. For example, in the case of AAB, a 16-amino-acid blunt-ended model gave a single solution with a TFZ of 13.1. After a round of rigid body and restrained isotropic refinement in Refmac⁵¹, the R-free dropped to 48%, and the density for the backbone atoms on the N and C termini of the placed model could be observed clearly. Subsequent rounds of refinement revealed side chain densities, which were modeled in an incremental fashion. Density for all residues except the acetyl and amide group on one of the three chains could be unambiguously modeled after multiple rounds of isotropic refinement in Refmac and model building in Coot⁵². A few rounds of anisotropic refinement and model building with automatic weight calculation were performed in PHENIX⁵³, which allows calculation of polder maps⁵⁴ to reveal weak sidechain densities masked by bulk solvent. Special care was taken to use identical reflections for calculation of R-free, and the model was validated using MolProbity⁵⁵ and the Protein Anisotropic Refinement Validation and Analysis Tool (PARVATI) webserver⁵⁶. Hydrogens were added to the structures and refined in riding mode. In this mode, hydrogens are not refined individually and do not add any additional refinable parameters. Instead, adding hydrogens improves refinement of other atoms, resulting in better model parameters. Composite omit maps calculated with the anisotropically refined model in PHENIX were used to confirm single residue stagger and assign chain registration. Similarly, PDB coordinates 1v4f gave a single solution with a TFZ of 7.1 for BAA. After a single round of rigid body and isotropic refinement, R-free dropped to 49%. Clear density was observed for 10–12 amino acid backbone atoms at two different locations with density missing in between them. Isotropic refinement in Refmac and model building in Coot resulted in a complete model. The resulting model underwent anisotropic refinement, validation and assignment of chain registration as described for AAB. For ABA, none of the 17 models used gave a single solution. PDB 1v7H gave seven solutions with TFZs of 10.9 and 9.8 for the two top scoring solutions. The solution with the highest TFZ of 10.9 was used for further model building and refinement. R-free dropped to 53% after a single round of rigid body and isotropic refinement in Refmac, and clear density for a few residues on the N- and C-termini of the placed model could be seen. Iterative model building and refinement improved the maps, and nearly 95% of all residues could be unambiguously modeled. Density for the acetyl group of the leading and C-terminal Gly and Tyr of the middle and trailing chains could not be observed. Anisotropic refinement, validation and assignment of chain registration was performed as described for AAB. 2Fo-Fc maps of the flanking and recognition sequences are shown in Supplementary Fig. 2.

Circular dichroism. CD experiments were performed on an Aviv Model 400 spectropolarimeter equipped with a Peltier temperature-controlled stage. Appropriate volumes of 3 mM heterotrimer stock solutions stored at 5 °C were diluted to 25 μ M with 10 mM sodium phosphate buffer at pH 7.0 on ice. 200 μ L of the diluted solution was transferred into a quartz cuvette (path length = 0.1 cm) and equilibrated at 8 °C in the sample chamber of the spectropolarimeter for 30 min before recording spectrographs. Ellipticity was monitored as a function of wavelength between 185 and 250 nm with a 1 nm bandwidth, 1 nm step size and 2 s averaging at each data point in three scans. The spectrographs were averaged and molar residue ellipticity (MRE) calculated as described previously⁵¹. Thermal melts were recorded by monitoring ellipticity at 225 nm as a function of temperature between 8 and 70 °C with 1 nm bandwidth, 1 nm step size, 30 s equilibrations and 2 s averaging at each temperature, and observed ellipticity was converted to MRE.

Nuclear magnetic resonance spectroscopy. Fifty microliters of D₂O containing sodium *d*₆-trimethylsilyl propionate (TSP) as an internal proton standard was added to 450 μ L of 3 mM heterotrimer stock solution and transferred into a 5 mm NMR tube (Wilmad-LabGlass, 507-PP-7). The ¹⁵N-HSQC spectra of homotrimer, heterotrimer and covalently captured peptide solutions were recorded at 25 °C on a Bruker cryogenic probe operating at a ¹H Larmor frequency of 600 MHz. Typically, 128 increments in the ¹⁵N dimension were acquired in four scans with sweep widths of 10 and 20 ppm in the direct and indirect dimensions, respectively. The data were processed in NMRPipe⁵⁷ and analyzed in the analysis2.4 package of CcpNmr⁵⁸.

Covalent capture of heterotrimers. Typically, the amide coupling chemistry is accomplished using zero-length heterobifunctional linkers such as

1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC•HCl). However, EDC chemistry is fraught with challenges owing to the highly non-specific nature of the crosslinking reaction and the formation of unproductive *N*-acylurea intermediates. The specificity of the reaction has been shown to improve substantially in presence of excess HOBt²⁸. In a typical synthesis, 5 mg of HOBt and 100 µl of 30 mM EDC solution in 100 mM MES at pH 6.1 were added successively to a 1 ml 2:1 molar mixture of peptides A and B (3 mM total peptide concentration) in 10 mM phosphate buffer (pH 7.5) at 5 °C. The solution was vortexed and rotated on a mini rotator for 4 h. An additional 100 µl aliquot of freshly prepared 30 mM EDC solution was added at 6 and 8 h, and the reaction mixture was rotated overnight at 5 °C. The reaction mixture was quenched with 100 µl of 1 M hydroxylamine at room temperature for 2 h, desalted using a 3,000 MWCO Vivaspin 2 polyethersulfone (PES) centrifugal protein concentrator and washed three times with 400 µl MilliQ water. The desalted reaction mixture was then purified in two steps on an S75 10/300 GL analytical column running Dulbecco's PBS at a 0.8 mg ml⁻¹ flow rate with detection at 280 nm. In the first step, fractions corresponding to all major peaks were collected. In all cases, a major peak eluted at approximately 12 ml (Supplementary Fig. 3a) with shoulders observed at both higher and lower elution volumes. The fractions were individually analyzed by ESI-MS on a Waters Xevo G2-S in positive ion mode. Raw mass spectrographs were integrated and deconvoluted using the MassLynx 4.1 software that accompanied the instrument, according to the manufacturer's instructions. Fractions that showed covalently captured trimer peaks and small amounts of monomers were pooled in and repurified on the size exclusion column. Fractions under the major peak observed at approximately 12 ml were collected and pooled, and a final ESI-MS analysis was performed to ascertain that only trimer peaks were observed. A two-step process was necessary to obtain covalently captured heterotrimers of more than 95% purity. Deconvoluted ESI-MS spectra in Supplementary Fig. 3b show peaks corresponding to the covalently captured trimer, while monomer and higher order oligomer peaks are not observed. The heterogeneity of the expected dehydration products made it difficult to quantify the fraction of total Lys-Asp salt bridges that converted to isopeptides and to determine whether they were present in both N- and C-terminal flanking sequences. CD and NMR analysis of both native and covalently captured heterotrimers is provided in Supplementary Fig. 4.

Solid-phase binding and dose-response assays. Binding of covalently captured heterotrimers to full-length VWF, VWF A3 domain containing a C-terminal His₆-tag, DDR1-Fc and DDR2-Fc was determined using ELISA. In a typical assay, 96-well Nunc MaxiSorp polystyrene plates (Thermo Fisher Scientific) were coated with 100 µl per well of 10 µg ml⁻¹ covalently captured heterotrimers, fibrous collagen type I from Ethonic or the inert triple-helical peptide (GPP)₁₀ dissolved in 1× Dulbecco's PBS (DPBS) for 1 h at room temperature under static conditions. Unbound plate surface was blocked with 200 µl per well of 5% BSA in DPBS for 1 h and washed three times with 200 µl DPBS containing 1 mg ml⁻¹ BSA and 0.05% Tween 20 (washing buffer). After washing, the plate was inverted and blotted to remove residual buffer. Designated wells were incubated with 100 µl per well of full-length VWF (1 µg ml⁻¹), VWF-A3 (5 µg ml⁻¹), DDR1-Fc (100 ng ml⁻¹) or DDR2-Fc (50 ng ml⁻¹) for 1 h and washed three times with 200 µl washing buffer. This was followed by incubation with 100 µl per well of antibody for 1 h for DDR1-Fc and DDR2-Fc and of primary and secondary antibody for 1 h each for full-length VWF and recombinant VWF A3 domain. The wells were washed again four times with 200 µl washing buffer and blotted to remove residual buffer. Color was developed using 100 µl of a 1:1 volumetric mixture of H₂O₂ and 3,3',5,5'-tetramethylbenzidine (TMB) liquid substrate system (Thermo Fisher, 34021). The reaction was quenched with 50 µl of 2 M H₂SO₄, and the intensity of the color was detected at 450 nm. A similar protocol was used to obtain dose-response curves. Decreasing concentrations of DDR1-Fc (approximately 40, 13, 4, 1.5, 0.5, 0.2, 0.1 and 0.02 nM), DDR2-Fc (approximately 20, 6, 2, 0.7, 0.3, 0.1, 0.03 and 0.01 nM), full-length VWF (approximately 9, 3, 1, 0.33, 0.11, 0.04, 0.012 and 0.004 nM) and VWF A3 domain (approximately 1, 0.33, 0.11, 0.04, 0.012, 0.004 and 0.001 µM) prepared by serial dilution from the highest concentration in DPBS were incubated with coated peptides for 1 h before incubating with antibody and subsequent detection. Peptide-bound full-length VWF was detected with a combination of rabbit polyclonal anti-VWF primary antibody (Abcam, 9378) followed by goat anti-rabbit HRP (Dako, P0448), both at 1:2,000 dilution. His₆-tagged VWF-A3 was detected with monoclonal anti-polyhistidine (Sigma-Aldrich, H1029) at 1:3,000 dilution followed by polyclonal goat anti-mouse Ig-horseradish peroxidase (HRP)-conjugate (Dako, P0447) at 1:6,000 dilution. DDR1-Fc and DDR2-Fc were detected using HRP-conjugated goat anti-human Fc (Jackson ImmunoResearch, 109-036-008) at 1:10,000 dilution.

Inhibition of VWF A3 binding. VWF A3 adhesion was determined colorimetrically. Toolkit peptide III-23 (ref.²¹), which contains the high affinity VWF binding motif RGQOGVMGFO, was coated at 10 µg ml⁻¹ for 1 h at 22 °C on Immulon 2 HB 96-well plates (Thermo Fisher, 3455) and blocked for 1 h with 200 µl of Tris-buffered saline (TBS) that contained 50 mg ml⁻¹ BSA. Wells were washed four times with 200 µl of adhesion buffer (TBS plus 1 mg ml⁻¹ BSA) before adding 100 µl of adhesion buffer that contained 10 µg ml⁻¹ of recombinant glutathione-S-transferase (GST)-VWF-A3 domains that had been preincubated

for 20 min with increasing concentrations of heterotrimers. After 1 h at room temperature, wells were washed five times with 200 µl of adhesion buffer before adding 100 µl of adhesion buffer that contained the anti-GST-HRP conjugate (GE Healthcare, RPN1236V) at 1:10,000 dilution for 1 h at room temperature. After washing, color was developed using an ImmunoPure TMB Substrate kit (Pierce) according to the manufacturer's instructions. The inhibition plots are shown in Supplementary Fig. 5.

DDR1 and DDR2 activation assays. HEK293 cells were transfected with expression constructs for DDR1 and DDR2, as described previously⁴⁴. Two days later, cells were stimulated with heterotrimeric peptides for 90 min at 37 °C, followed by lysis in 150 mM NaCl, 50 mM Tris pH 7.4, 1 mM EDTA, 1 mM PMSF, 50 mg ml⁻¹ aprotinin, 5 mM NaF and 1 mM NaVO₃. Cell lysates were analyzed by reducing SDS-PAGE, followed by blotting onto nitrocellulose membranes. Blots were first probed with a 1:1,000 dilution of the phospho-specific antibodies rabbit anti-phospho-DDR1 (Tyr 513) (Cell Signaling, E1N8F) or rabbit anti-phospho-DDR2 (Tyr 740) (R&D Systems, clone 1119D, MAB25382)^{40,41}, then with antibodies against DDR1 (rabbit anti-DDR1, Santa Cruz, SC-532, 1:500) or DDR2 (goat anti-DDR2, R&D Systems, AF2538, 1:2,000). Blots with DDR1 samples were stripped in Antibody Strip Solution (Alpha Diagnostic, 90101) before reprobing with DDR1 antibodies. DDR2 samples were run on two separate gels. Secondary antibodies (used at 1:10,000) were goat anti-rabbit Ig-HRP-conjugated (Dako, P0448) and rabbit anti-goat Ig-HRP-conjugated (Thermo Fisher, 31402). Signal detection was performed on a Typhoon FLA 9500 Imager (GE Healthcare Bioscience) using ECL 2 western blotting substrate (Pierce). Densitometry analysis of protein band intensities was performed using Image Studio Lite (LI-COR Biosciences). Statistical analysis was carried out using GraphPad Prism 8.0 for Windows. Statistical significance was set at *P* < 0.05. The cropped and uncropped western blots are shown in Supplementary Figs. 6, 7 and precise *P* values for each comparison are shown in Supplementary Tables 5, 6.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates of AAB (6Q3P), ABA (6Q41) and BAA (6Q43) crystal structures have been deposited with the Protein Data Bank. Raw data associated with Figs. 1, 2, 3 can be provided by the corresponding author upon reasonable request.

Code availability

The code for computational design of heterotrimers may be requested from the corresponding author.

References

- Leitinger, B. Molecular analysis of collagen binding by the human discoidin domain receptors, DDR1 and DDR2. *J. Biol. Chem.* **278**, 16761–16769 (2003).
- Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67**, 293–302 (2011).
- Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D* **74**, 85–97 (2018).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Rainey, J. K. & Goh, M. C. An interactive triple-helical collagen builder. *Bioinformatics* **20**, 2458–2459 (2004).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Liebschner, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D* **73**, 148–157 (2017).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Merritt, E. A. Comparing anisotropic displacement parameters in protein structures. *Acta Crystallogr. D* **55**, 1997–2004 (1999).
- Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
- Fogh, R. et al. The ccpn project: an interim report on a data model for the nmr community. *Nat. Struct. Biol.* **9**, 416–418 (2002).
- Schlick, T. L., Ding, Z., Kovacs, E. W. & Francis, M. B. Dual-surface modification of the tobacco mosaic virus. *J. Am. Chem. Soc.* **127**, 3718–3723 (2005).

60. Xu, H. et al. Normal activation of discoidin domain receptor 1 mutants with disulfide cross-links, insertions, or deletions in the extracellular juxtamembrane region: mechanistic implications. *J. Biol. Chem.* **289**, 13565–13574 (2014).
61. Juskaite, V., Corcoran, D. S. & Leitinger, B. Collagen induces activation of DDR1 through lateral dimer association and phosphorylation between dimers. *elife* **6**, e25716 (2017).

Acknowledgements

A.A.J. was supported by Newton International Fellowship (NF140721) granted jointly by the Royal Society, the British Academy and the Academy of Medical Sciences. D.S. was supported by a PhD studentship from the Imperial College London–Royal Holloway BBSRC Doctoral Training Partnership. J.D.H. and D.R.W. were supported in part by the Welch Foundation (C1557) and the National Science Foundation (CHE1709631). R.W.F. was supported by a British Heart Foundation programme grant (RG/15/4/31268). The authors thank D. Chirgadze and M. Hyvonen in the Department of Biochemistry at the University of Cambridge for X-ray crystallography support and crystallographic data refinement, respectively; E. Hohenester in the Department of Life Sciences at Imperial College London for helpful discussion on solid-phase binding assays; and J.-D. Malcor and A. Bonna in the Department of Biochemistry at the University of Cambridge for support in peptide synthesis. The authors also thank Diamond Light Source for beamtime (proposal mx14043) and the staff of beamlines I03, I04 and I24 for assistance with crystal testing and data collection.

Author contributions

A.A.J. and R.W.F. conceived the project. A.A.J. synthesized and characterized the peptides, obtained the heterotrimer crystals and solved their crystal structures, developed the methodology for covalent capture of heterotrimers and their subsequent purification, performed the solid-phase binding assays and analyzed the CD, NMR, MS and solid-phase binding assay data. J.D.H. wrote the code for computational design of heterotrimers. B.L. expressed DDR–Fc fusion constructs and analyzed cellular activation experiments performed by D.S. S.W.H. expressed the recombinant VWF A3 domain, and E.J.H. assisted in optimization of solid-phase assays. P.B. co-solved and refined the crystal structure of AAB. K.S. planned the NMR experiments and co-analyzed the NMR and CD data. D.R.W. wrote the script for the analysis of the helical twist of heterotrimers. A.A.J., R.W.F. and B.L. co-wrote the manuscript with input from other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41589-019-0435-y>.

Correspondence and requests for materials should be addressed to A.A.J.

Reprints and permissions information is available at www.nature.com/reprints.