Published in final edited form as:
Nat Chem Biol. 2020 April ; 16(4): 423–429. doi:10.1038/s41589-019-0435-y.

Chain Alignment of Collagen I Deciphered using Computationally Designed Heterotrimers

Abhishek A. Jalan[1,*], **Douglas Sammon**[2], **Jeffrey D. Hartgerink**[3], **Paul Brear**[1], **Katherine Stott**[1], **Samir W. Hamaia**[1], **Emma J. Hunter**[1], **Douglas R. Walker**[3], **Birgit Leitinger**[2], **Richard W. Farndale**[1]

[1]Department of Biochemistry, University of Cambridge, Cambridge, UK

[2]National Heart and Lung Institute, Imperial College London, London, UK

[3]Department of Chemistry and Bioengineering, Rice University, Houston, USA

Abstract

The most abundant member of the collagen protein family, collagen I (COL1), is composed of two similar (chain A) and one unique (chain B) 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 permutated 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.

[*]**Correspondence and request for materials** should be addressed to A.A.J. jalan@cantab.net.
[5]Present address: Department of Biochemistry, University of Bayreuth, Bayreuth, Germany

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 and 3 can be provided by A.A.J. upon reasonable request.

Code availability The code for computational design of heterotrimers can be requested from A.A.J.

Author contributions
A.A.J. and R.W.F. conceived the project. A.A.J. synthesized and characterized the peptides, obtained heterotrimer crystals and solved their crystal structures, developed the methodology for covalent capture of heterotrimers and their subsequent purification, performed solid phase binding assays and analysed CD, NMR, mass spectrometric 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 analysed cellular activation experiments performed by D.S. S.W.H. expressed 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 NMR experiments and co-analysed 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.

Europe PMC Funders Author Manuscripts

Introduction

Collagens are a large superfamily of 28 known mammalian proteins (COL1–28) that bind transmembrane receptors[1], secreted extracellular matrix (ECM)[2] proteins and blood serum proteins[3], collectively called collagen-binding proteins (CBP), to regulate cell signaling, matrix homeostasis and thrombosis. For example, fibrillar collagens COL1 and COL3, exposed in the subendothelium during vascular injury, bind to the blood plasma protein Von Willebrand Factor (VWF).[3] Subsequent collagen-mediated recruitment of platelets via the receptors α2β1 integrin and glycoprotein VI (GPVI) is responsible for the deposition of life-saving thrombi[4] as well as life-threatening ischemic myocardial damage.[5] Collagens also bind and activate discoidin domain receptors, DDR1 and DDR2, a subfamily of receptor tyrosine kinases (RTKs), which results in intracellular signaling events critical for cell survival and tissue remodeling.[6] DDR1-deficient mice show reduction in renal function[7], anomalous mammary gland development[8] and defective arterial wound repair[9], while DDR2-deficient mice exhibit dwarfism[10]. DDR1 and DDR2 also remodel extracellular matrix during tissue maturation and thus like most other RTKs, play a key role in cancer progression.[11]

Collagens are highly complex multidomain proteins that contain more than 1000 amino acids and often form hydrogels or non-specific aggregates when isolated from native tissues. Thus, the structural basis for the collagen–CBP interaction is 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 helix.[12] 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 CBP. This Toolkit approach[13] has revealed the structural basis for the interaction of COL2 and COL3 to integrins $\alpha_1\beta_1$[14], $\alpha_2\beta_1$[15] and $\alpha_{10}\beta_1$[16], thrombospondin-1 (TSP-1)[2], VWF A3 domain[17], DDR1[18], DDR2[19], matrix metalloproteinase 1 (MMP-1)[20], secreted protein acidic and rich in cysteine (SPARC)[21], osteoclast-associated receptor (OSCAR)[22], glycoprotein VI (GPVI)[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, i.e. 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. In contrast, designing peptide mimics of heterotrimeric collagens containing either two (AAB-type) or three (ABC-type) unique chains is challenging due 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 collagen[25], and given rise to a vexing 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 permutated 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 fifty years ago.[25] To add to

the mystery, all three combinations have variously been proposed based on computational analysis of COL1 sequence[26], interchain interactions[27], molecular packing[28], and fibrillar architecture[29]. 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 stretch of sequence from chain A and B of COL1 predicted to bind DDR1, DDR2 and VWF[17], 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 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[18], DDR2[19] and VWF[30] recognize a homotrimeric RGQOGVMGFO sequence (O = 4(*R*)-hydroxyproline, Hyp) conserved in human COL2 and COL3. A comparison to human COL1 revealed a similar site, ARGQAGVMGFO, at sequence positions 573–583 in chain A and ARGEOGNIGFO at the corresponding positions 485–495 of chain B, that could potentially bind these three CBPs if the two sequence motifs could be reconstituted in the native-like chain alignment.[17] Thus, our primary design challenge was to incorporate two copies of COL1 chain A and one copy of 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 collagen[34]. Briefly, multiple Lys and Asp/Glu salt bridges are introduced rationally or computationally such that all of these form salt bridges only in the target heterotrimeric register. By design, the competing triple-helical states contain multiple unpaired Lys and Asp/Glu and are thus unstable with respect to the target state. Paired and unpaired Lys and Asp/Glu residues represent elements of positive and negative design, respectively, and ensure high degree of specificity during heterotrimer self-assembly.

In our case, the design of the three registers is based on a host-guest scheme where the putative VWF and DDR recognition epitope in COL1 mentioned above is incorporated as guest in the midst of two flanking host domains (Fig. 1). The host domains contain multiple Lys–Asp salt bridges that lock the triple helix into AAB, ABA or BAA registers (Fig. 1a). Flanking domains are computationally designed using a modified genetic algorithm developed to find ABC heterotrimers.[35] 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 1 point and then any Lys-Asp pairs that are found in the appropriate geometry to form a salt-bridge are awarded 2 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, with ties being broken by the best stability, for the desired register, is saved and reproduced with mutation in a subsequent generation. This process iterates through multiple generations until a target specificity and stability is 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 via 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 permutated in either the leading (BAA), middle (ABA) or the 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)[36] with a distance cutoff of 3.2 Å between the Lys Nξ and Asp Oδ atoms revealed that all 12 Lys and Asp residues incorporated by the computational design formed salt bridges in BAA and ABA, while AAB contained 10 salt bridges.

The homotrimeric RGQOGVMGFO 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, VWF A3 domain[17] and DDR2[37] bind the COL2 sequence selectively through the leading and middle chain Phe, respectively. Since both chain A (ARGQAGVMGFO) and B (ARGEOGNIGFO) of the COL1 recognition epitope contain a Phe residue, we designed three variants of AAB containing Phe to Ala mutation in the two A chains (AAB-alaA), B chain only (AAB-alaB) or both A and B chains (AAB-alaAB), in order to understand their chain-specific role in CBP recognition. These mutant heterotrimers were not crystallized and were only analyzed via 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 functional CBP assays.

Salt bridges direct covalent capture of registers

The equilibrium population of salt bridge-stabilized collagen heterotrimers is extremely susceptible to total peptide concentration, ionic strength of the buffer, pH, temperature and

amino acid sequence.[38] In order to remove any bias in the binding affinities due 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). The covalently captured constructs were purified and characterized using size-exclusion chromatography (SEC) and mass spectrometry (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 circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy, respectively. CD spectrographs of all native and covalently captured heterotrimers show a strong minimum at ~195 nm and a weakly positive maximum at ~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, suggesting the presence of well-folded and stable triple helices in solution (Supplementary Fig. 4b). AAB is thermally less stable than BAA, ABA and the alanine mutant constructs of AAB. The lower thermal stability of AAB in comparison to 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 $[(GPO)_8]_3$ triple helix less than Phe[39]. The denaturation curves of the covalently captured heterotrimers show large dispersion in the CD signal during unfolding. This is expected as the solution of covalently captured heterotrimers contains an ensemble of triple helices with identical CBP recognition sequence but different fraction of Lys-Asp isopeptide bonds in the flanking sequence. In addition, the triple helices in this ensemble contain variable numbers of amides due to the contribution of isopeptide bonds and thus would have variable stability and rigidity along the triple-helical axis. Thus, the observed melting curves are 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 within 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 ^{1}H, ^{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 A3 homotrimers with

their many unpaired Lys and Asp residues are not amenable to efficient covalent capture. Alternatively, it may mean that they are removed during size exclusion chromatography purification owing to difference 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 characterization confirm that the triple helical fold and heterotrimer register remains 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 comprising the entire extracellular regions of DDR1 and DDR2 fused C-terminally to the Fc sequence of human IgG2 (hereafter called DDR1-Fc and DDR2-Fc), recombinant A3 domain of VWF (VWF A3) and 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, recombinant VWF A3 domain, DDR1-Fc and DDR2-Fc with significantly higher affinity than 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, recombinant VWF A3 domain, which is recognized as the primary locus of COL1–3 interaction in full-length VWF,[3,40] discriminated the least between the three registers with only a 2-fold 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 RGQOGVMGFO homotrimer[17]. The trend in binding affinities continued with DDR1-Fc showing more than 25-fold higher affinity for AAB than BAA or ABA (Fig. 2c). DDR2-Fc affinity for all registers was higher than 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 as it has been shown to recognize multiple loci in COL2 and COL3, compared with DDR1-Fc which recognizes a unique site with high affinity in both collagens.[18,19]

AAB-alaB bound all four proteins with affinity similar to AAB, except DDR1 which showed 2-fold reduced affinity, but 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 highest levels of DDR phosphorylation

The DDRs respond to collagen binding with intracellular kinase activation, which is primarily manifested as receptor autophosphorylation.[41] 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 with phospho-specific antibodies (see Online Methods). As anticipated from the binding data presented in Figure 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 only induce marginal DDR1 phosphorylation and no detectable DDR2 phosphorylation. AAB-alaB stimulated DDR1 phosphorylation with similar potency as AAB, while AAB-alaA did not elicit a 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 Phe residue in chain A of COL1 plays a crucial role in 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[17] and DDR2[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 Figure 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 while Hyp residues in the trailing peptide chain B are denoted as T:Hyp24 and T:Hyp30.

Since both chains A and B of the COL1 recognition sequence contain a Phe, each CBP can potentially recognize AAB, BAA and ABA via 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 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 case of DDR2, binding through the M:Phe of ABA replaces hydrophobic M:Val with bulkier and hydrophilic Asn. Similarly, in 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 to CBPs by moderate change 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 distinctly 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 resolve another vexing issue in collagen research. Both DDR1 and DDR2 recognize fibrillar collagens COL1-3. However, DDR1 selectively recognizes network forming COL4 present in the basal lamina.[41] The structural basis for this selectivity is not known.[18] 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 fibre diffraction data[29] 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 disease[42]. Structural perturbations in COL1 due to mutations in either or both chains can be correctly modelled and correlated to the resulting phenotypic severity only if the alignment of chain B is precisely known. Thus, our work could potentially advance understanding of both genetic and physiological aspects of collagen function.

Online Methods

Synthesis and purification of peptides

The peptides were synthesized on a CEM Liberty Blue microwave-assisted peptide synthesizer using standard Fmoc chemistry on a solid phase support, cleaved from the resin and purified via high performance liquid chromatography, as described previously.[13] Arg was coupled twice at 20°C to prevent gamma-lactam formation, which leads to Arg deletion. A ^{15}N-isotopically enriched Gly residue was coupled at Gly25 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 alpha and beta-peptides. We observed aspartimide formation in peptides containing Lys on the C-terminal side of Asp, i.e. in Asp-Lys-Gly triplets. In order to prevent the formation of aspartimide, we chose

5% piperazine containing 0.1 M 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 as compared to the t-butyl ester prevents a nucleophilic attack from the NH attached to the alpha-carboxy group. Synthesis using these methods yielded peptides free of aspartimide. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) spectra of all peptides are shown in Supplementary Fig. 1.

Protein expression and purification

Soluble recombinant His_6-tagged VWF A3 domain was expressed as described previously[30]. Soluble recombinant extracellular proteins comprising the entire extracellular regions of discoidin domain receptor 1 and 2 (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[18,43]. Native human full-length von Willebrand factor was purchased from Abcam (cat# ab88533).

Preparation of samples for NMR, CD and covalent capture

Purified peptides were dissolved in MilliQ H_2O at 25-30 mg/ml, pH adjusted to 7.0 using 1M NaOH solution and their concentrations determined spectrophotometrically using Nanodrop 1000. 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 total peptide concentration of 3 mM for the heterotrimers and 1 mM for 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 days before taking measurements. The 3 mM stock solutions of heterotrimer or 1 mM stock solutions of homotrimer were used for all subsequent measurements using NMR and CD 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 2:1 molar ratio at 20 mg/ml total peptide concentration, pH adjusted to 7.8, 8.6 and 9.2 for AAB, ABA and BAA, respectively, using 1M NaOH and annealed at 85°C for 15 minutes. The peptides were initially stocked in aqueous solutions without any buffer because the two buffer conditions, potassium phosphate and Tris-HCl, initially used to stock the peptides, were found to drastically lower the number of positive hits obtained during crystallization screens. The annealed solutions were incubated at room temperature for at least 3 days, diluted to 10, 6 and 10 mg/ml 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 PEGs I and 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 uL. Crystals were grown using sitting drop vapor diffusion method in 96-well 2-drop MRC crystallization plates. 200 nL of 6-10 mg/ml peptide solutions and 200 nL of reservoir solutions was aliquoted using mosquito® Crystal from TTP Labtech, immediately sealed with UV-transparent Scotchtape, stored in Formulatrix Rock Imager 1000 and scanned every 4 hours for the first day, and

then every day. Crystals appeared within a few hours to 2 days 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 N_2. 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. Crystallographic data collection and refinement statistic and crystallization conditions are provided in Supplementary Table 1 and 2, respectively.

Strategy for solving the crystal structures

The data for AAB and ABA were indexed and integrated using AutoProc[44]. However, due to unsatisfactory indexing, the data for BAA was indexed and integrated using DIALS[45]. Subsequently, all data were scaled and merged in AIMLESS[46] and truncated using Ctruncate[47]. The crystal structures of heterotrimers were solved using molecular replacement in Phaser[48]. 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 The-BuScr[49]. A second copy of each triple helix with N- and C-terminal end residues removed to create blunt ended termini was also generated. These 16 models along with crystal structures deposited under 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 Phaser[48]. In general, molecular replacement with the blunt-ended triple helices gave fewer solutions in each search with higher translation function Z-scores (TFZ) and log-likehood gain (LLG) values compared to the triple helices with end offsets. The TFZ score 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 single solution and highest TFZ scores was selected for subsequent model building and refinement. For example, in case of the AAB a 16 amino acid blunt ended model gave a single solution with TFZ score of 13.1. After a round of rigid body and restrained isotropic refinement in Refmac[50], the R-free dropped to 48% and density for the backbone atoms on the N- and C-termini of the placed model could be clearly observed. Subsequent rounds of refinement revealed side chain densities which were modelled in an incremental fashion. Density for all residues except the acetyl and amide group on one of the three chains could be unambiguously modelled after multiple rounds of isotropic refinement in Refmac and model building in Coot[51]. A few rounds of anisotropic refinement and model building with automatic weight calculation was performed in Phenix [52] which allows calculation of polder maps[53] 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 validated using Molprobity[54] and the Protein Anisotropic Refinement Validation and Analysis Tool (PARVATI) webserver[55]. Hydrogens were added to the structures and refined in the so-called 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 TFZ

score 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 was anisotropically refined, validated, and chain registration assigned as described in case of AAB. For ABA, none of the 17 models used gave a single solution. PDB 1v7H gave seven solutions with highest TFZ score of 10.9 and 9.8 for the two top scoring solutions. The solution with highest TFZ of 10.9 was used for further model building and refinement. R-free dropped to 53% after 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 modelled. 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 done 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 μM with 10 mM sodium phosphate buffer at pH 7.0 on ice. 200 μl 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, 2 s averaging at each data point in 3 scans. The spectrographs were averaged and molar residue ellipticity (MRE) calculated as previously [31]. 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 converted to MRE.

NMR spectroscopy

50 μl D_2O containing sodium d_6-trimethylsilyl propionate (TSP) as internal proton standard was added to 450 μl of 3 mM heterotrimer stock solution and transferred into a 5 mm NMR tube (Wilmad 507-PP-7). ^{15}N-Heteronuclear Single Quantum Coherence (HSQC) spectra of homotrimer, heterotrimer and covalently-captured peptide solutions were recorded at 25°C on a Bruker cryogenic probe operating at ^{1}H larmor frequency of 600 MHz. Typically, 128 increments in the ^{15}N dimension were acquired in 4 scans with a sweep width of 10 and 20 ppm in the direct and indirect dimension, respectively. The data were processed in NMRPipe[56] and analyzed in analysis2.4 package of CcpNmr[57].

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 due to highly non-specific nature of the crosslinking reaction and formation of unproductive N-acylurea intermediate.

The specificity of the reaction has been shown to improve substantially in presence of excess N-hydroxybenzotriazole (HOBt)[58]. 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 aliquot of freshly prepared 100 µl of 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 solution of 1 M hydroxylamine at room temperature for 2 h and desalted using a 3000 MWCO Vivaspin 2 polyetheresulfone (PES) centrifugal protein concentrator and washed thrice with 400 µl MQ water. The desalted reaction mixture was then purified in two steps on an S75 10 300 GL analytical column running Dulbecco's PBS buffer at 0.8 mg/ml 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 ~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 on MassLynx4.1 software accompanying 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 ~12 ml were collected and pooled and a final ESI-MS analysis performed to ascertain that only trimer peaks are observed. A two-step process was necessary to obtain covalently-captured heterotrimers of more than 95% purity. Deconvoluted ESI-MS spectra shown in Supplementary Fig. 3b show peaks corresponding to the covalently-captured trimer while monomer and higher order oligomers 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 whether they were present in both N- and C-terminal flanking sequences. CD and NMR analysis of both native and covalent-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 His_6-tag, DDR1-Fc and DDR2-Fc was determined using enzyme-linked immunosorbent assay (ELISA). In a typical assay, 96-well Nunc Maxisorp polystyrene plates (Thermofisher Scientific, Paisley, UK) were coated with 100 µl/ well of 10 µg/ml covalently-captured heterotrimers, fibrous collagen type I from Ethicon™ or inert triple helical peptide $(GPP)_{10}$ dissolved in 1X Dulbecco's phosphate buffered saline (DPBS) for 1 h at room temperature under static conditions. Unbound plate surface was blocked with 200 µl/well of 5% bovine serum albumin (BSA) in DPBS for 1 h and washed thrice with 200 µl DPBS containing 1mg/ml 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/well full-length VWF (1µg/ml), VWF-A3 (5 µg/ml), DDR1-Fc (100 ng/ml) or DDR2-Fc (50 ng/ml) for 1 h and washed thrice with 200 µl washing buffer followed by incubation with 100 µl/well of antibody for 1 h for DDR1-Fc and DDR2-Fc and 1 h each of primary and secondary antibody 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 and colour was developed using 100 μl of a 1:1 volumetric mixture of H_2O_2 and 3, 3', 5, 5'-tetramethylbenzidine (TMB) liquid substrate system from ThermoFisher Scientific (cat# 34021). The reaction was quenched with 50 μl 2 M H_2SO_4 and the intensity of the colour detected at 450 nm. A similar protocol was used to obtain dose-response curves. Decreasing concentrations of DDR1-Fc (~ 40, 13, 4, 1.5, 0.5, 0.2, 0.1, and 0.02 nM), DDR2-Fc (~ 20, 6, 2, 0.7, 0.3, 0.1, 0.03 and 0.01 nM), full-length VWF (~ 9.00, 3.00, 1.00, 0.33, 0.11, 0.04, 0.012, 0.004 nM) and VWF A3 domain (~ 1, 0.33, 0.11, 0.04, 0.012, 0.004 and 0.001 μM) prepared by serial dilution starting 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, cat# 9378) followed by goat anti-rabbit HRP (Dako, P0448) both at 1:2,000 dilution. His_6-tagged VWF-A3 was detected using monoclonal anti-polyhistidine (Sigma-Aldrich, cat# H1029) at 1:3,000 dilution followed by polyclonal goat anti-mouse Immunoglobulins-HRP-conjugate (Dako, cat#P0447) at 1:6,000 dilution. DDR1-Fc and DDR2-Fc were detected using HRP-conjugated goat anti-human Fc (Jackson ImmunoResearch, cat # 109-036-008) at 1:10,000 dilution.

Inhibition of VWF A3 binding

VWF-A3 adhesion was determined colorimetrically. Toolkit peptide III-23[12], which contains a high affinity VWF binding motif RGQOGVMGFO, was coated at 10 μg/ml for 1 h at 22°C on Immulon-2 HB 96- well plates (Thermo Life Sciences, Basingstoke, UK), and blocked for 1 h with 200 μl of Tris-buffered saline (TBS) containing 50 mg/ml bovine serum albumin. 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 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, cat#RPN1236V) at 1:10,000 dilution for 1 h at room temperature. After washing, colour 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

HEK 293 cells were transfected with expression constructs for DDR1 and DDR2 as described previously[43]. Two days later, cells were stimulated with heterotrimeric peptides for 90 minutes 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 aprotinin, 5 mM NaF and 1 mM $NaVO_3$. Cell lysates were analysed by reducing SDS-PAGE, followed by blotting onto nitrocellulose membranes. Blots were first probed with a 1:1,000 dilution of phospho-specific antibodies (rabbit anti-phospho-DDR1 (Tyr 513) from Cell Signalling, cat# E1N8F, or rabbit anti-phospho-DDR2 (Tyr 740) from R&D Systems, clone 1119D, cat# MAB25382),[59,60] then with antibodies against DDR1 (rabbit anti-DDR1, cat# SC-532, from Santa Cruz; 1:500) or DDR2 (goat anti-DDR2, cat# AF2538, from R&D Systems; 1:2,000). Blots with DDR1 samples were stripped in Antibody Stripping solution (Alpha Diagnostic International), 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- horseradish peroxidase-conjugated (cat# P0448 from DAKO) and rabbit anti-goat Ig-horseradish peroxidase-conjugated (Zymed Laboratories, cat# 31402 from Life Technologies). Signal detection was performed on a Typhoon FLA9500 Imager (GE Healthcare Bioscience) using ECL 2 Western blotting substrate (Pierce). Densitometry analysis of protein band intensities was performed using ImageStudio™ Lite (LI-COR Biosciences, UK). Statistical analysis was carried out using GraphPad Prism 8.00 for Windows (GraphPad software, LA Jolla, CA). Statistical significance was set at a p-value <0.05. The cropped and uncropped Western blots are shown in Supplementary Figs. 6 and 7, respectively, and precise p-values for each comparison are shown in Supplementary Tables 5 and 6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

A.A.J. was supported by Newton International Fellowship (NF140721) granted jointly by the Royal Society, British Academy and the Academy of Medical Sciences, UK. 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 British Heart Foundation programme, RG/15/4/31268. We 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 Science at Imperial College London for helpful discussion on solid phase binding assays, J-D. Malcor and A. Bonna in the Department of Biochemistry at the University of Cambridge for support in peptide synthesis. We 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.

References

1. Leitinger B. Transmembrane collagen receptors. Annu Rev Cell Dev Biol. 2011; 27:265–290. [PubMed: 21568710]
2. Rosini S, et al. Thrombospondin-1 promotes matrix homeostasis by interacting with collagen and lysyl oxidase precursors and collagen cross-linking sites. Sci Signal. 2018; 11
3. Lankhof H, et al. A3 domain is essential for interaction of von Willebrand factor with collagen type III. Thrombosis and haemostasis. 1996; 75:950–958. [PubMed: 8822592]
4. Santoro SA. Preferential binding of high molecular weight forms of von Willebrand factor to fibrillar collagen. Biochim Biophys Acta. 1983; 756:123–126. [PubMed: 6402037]
5. De Meyer SF, Stoll G, Wagner DD, Kleinschnitz C. von Willebrand factor: An emerging target in stroke therapy. Stroke. 2012; 43:599–606. [PubMed: 22180250]
6. Fu HL, et al. Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling. J Biol Chem. 2013; 288:7430–7437. [PubMed: 23335507]
7. Gross O, et al. DDR1-deficient mice show localized subepithelial GBM thickening with focal loss of slit diaphragms and proteinuria. Kidney Int. 2004; 66:102–111. [PubMed: 15200417]
8. Vogel WF, et al. Discoidin domain receptor 1 tyrosine kinase has an essential role in mammary gland development. Society. 2001; 21:2906–2917.
9. Hou G, Vogel W, Bendeck MP. The discoidin domain receptor tyrosine kinase DDR1 in arterial wound repair. J Clin Invest. 2001; 107:727–35. [PubMed: 11254672]
10. Labrador JP, et al. The collagen receptor DDR2 regulates proliferation and its elimination leads to dwarfism. EMBO Rep. 2001; 2:446–452. [PubMed: 11375938]
11. Valiathan RR, Marco M, et al. Discoidin domain receptor tyrosine kinases : new players in cancer progression. Cancer Metastasis Rev. 2013; 31:295–321.

12. Raynal N, et al. Use of synthetic peptides to locate novel integrin α2β1-binding motifs in human collagen III. J Biol Chem. 2006; 281:3821–3831. [PubMed: 16326707]
13. Farndale RW, et al. Cell–collagen interactions: the use of peptide Toolkits to investigate collagen–receptor interactions. Biochem Soc Trans. 2008; 36:241–250. [PubMed: 18363567]
14. Kim JK, et al. A novel binding site in collagen type III for integrins α1β1 and α2β1. J Biol Chem. 2005; 280:32512–20. [PubMed: 16043429]
15. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. Structural basis of collagen recognition by integrin α2β1. Cell. 2000; 101:47–56. [PubMed: 10778855]
16. Hamaia SW, et al. Unique charge-dependent constraint on collagen recognition by integrin α10β1. Matrix Biol. 2017; 59:80–94. [PubMed: 27569273]
17. Brondijk THC, Bihan D, Farndale RW, Huizinga EG. Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. Proc Natl Acad Sci. 2012; 109:5253–5258. [PubMed: 22440751]
18. 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. 2011; 30:16–26. [PubMed: 21044884]
19. Konitsiotis AD, et al. Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. J Biol Chem. 2008; 283:6861–6868. [PubMed: 18201965]
20. Manka SW, et al. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. Proc Natl Acad Sci. 2012; 109:12461–12466. [PubMed: 22761315]
21. Hohenester E, Sasaki T, Giudici C, Farndale RW, Bächinger HP. Structural basis of sequence-specific collagen recognition by SPARC. Proc Natl Acad Sci U S A. 2008; 105:18273–18277. [PubMed: 19011090]
22. Zhou L, et al. Structural basis for collagen recognition by the immune receptor OSCAR. Blood. 2016; 127:529–537. [PubMed: 26552697]
23. Munnix ICA, et al. Collagen-mimetic peptides mediate flow-dependent thrombus formation by high- or low-affinity binding of integrin α2β1 and glycoprotein VI. J Thromb Haemost. 2008; 6:2132–2142. [PubMed: 18826391]
24. Lebbink RJ, et al. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. Matrix Biol. 2009; 28:202–210. [PubMed: 19345263]
25. Piez KA, Eigner EA, Lewis MS. The chromatographic separation and amino acid composition of the subunits of several collagens. Biochemistry. 1963; 2:58–66.
26. Piez KA, Trus BL. Sequence regularities and packing of collagen molecules. J Mol Biol. 1978; 122:419–432. [PubMed: 691048]
27. Traub W, Fietzek PP. Contribution of the α2 chain to the molecular stability of collagen. FEBS Lett. 1976; 68:245–249. [PubMed: 976476]
28. Bender E, Silver H, Hayashi K, Trelstad RL. Collagen segment long spacing banding patterns. 1982; 257:9653–9657.
29. Orgel JPRO, Irving TC, Miller A, Wess TJ. Microfibrillar structure of type I collagen in situ. Proc Natl Acad Sci. 2006; 103:9001–9005. [PubMed: 16751282]
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. 2006; 108:3753–3756. [PubMed: 16912226]
31. Jalan AA, Hartgerink JD. Simultaneous control of composition and register of an AAB-type collagen heterotrimer. Biomacromolecules. 2013; 14:179–185. [PubMed: 23210738]
32. Xu F, Zhang L, Koder RL, Nanda V. De novo self-assembling collagen heterotrimers using explicit positive and negative design. Biochemistry. 2010; 49:2307–16. [PubMed: 20170197]
33. Zheng H, et al. How electrostatic networks modulate specificity and stability of collagen. Proc Natl Acad Sci. 2018; 115:6207–6212. [PubMed: 29844169]
34. Gauba V, Hartgerink JD. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. J Am Chem Soc. 2007; 129:2683–90. [PubMed: 17295489]
35. Fallas JA, Hartgerink JD. Computational design of self-assembling register-specific collagen heterotrimers. Nat Commun. 2012; 3:1087–1088. [PubMed: 23011141]

36. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996; 14:33–38. [PubMed: 8744570]
37. Carafoli F, et al. Crystallographic insight into collagen recognition by discoidin domain receptor 2. Structure. 2009; 17:1573–1581. [PubMed: 20004161]
38. Jalan AA, Demeler B, Hartgerink JD. Hydroxyproline-free single composition ABC collagen heterotrimer. J Am Chem Soc. 2013; 135:6014–6017. [PubMed: 23574286]
39. Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry. 2000; 39:14960–14967. [PubMed: 11101312]
40. Houdijk WPM, Sakariassen KS, Nievelstein PFEM, Sixma JJ. 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. 1985; 75:531–540. [PubMed: 3919060]
41. Vogel W, Gish GD, Alves F, Pawson T. The discoidin domain receptor tyrosine kinases are activated by collagen. Mol Cell. 1997; 1:13–23. [PubMed: 9659899]
42. Bodian DL, Madhan B, Brodsky B, Klein TE. Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. Biochemistry. 2008; 47:5424–32. [PubMed: 18412368]
43. Leitinger B. Molecular analysis of collagen binding by the human discoidin domain receptors, DDR1 and DDR2. J Biol Chem. 2003; 278:16761–16769. [PubMed: 12611880]
44. Vonrhein C, et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr Sect D Biol Crystallogr. 2011; 67:293–302. [PubMed: 21460447]
45. Winter G, et al. DIALS: Implementation and evaluation of a new integration package. Acta Crystallogr Sect D Struct Biol. 2018; 74:85–97. [PubMed: 29533234]
46. Evans P. Scaling and assessment of data quality. Acta Crystallogr Sect D: Biological Crystallography. 2006; 62:72–82. [PubMed: 16369096]
47. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr Sect D Biol Crystallogr. 2011; 67:282–292. [PubMed: 21460446]
48. McCoy AJ, et al. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
49. Rainey JK, Goh MC. An interactive triple-helical collagen builder. Bioinformatics. 2004; 20:2458–2459. [PubMed: 15073022]
50. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D Biol Crystallogr. 1997; 53:240–255.
51. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
52. Adams PD, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
53. Liebschner D, et al. Polder maps: Improving OMIT maps by excluding bulk solvent. Acta Crystallogr Sect D Struct Biol. 2017; 73:148–157. [PubMed: 28177311]
54. Chen VB, et al. MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
55. Merritt EA. Comparing anisotropic displacement parameters in protein structures. Acta Crystallogr Sect D Biol Crystallogr. 1999; 55:1997–2004. [PubMed: 10666575]
56. Delaglio F, et al. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
57. Fogh R, et al. The ccpn project: An interim report on a data model for the nmr community. Nat Struct Biol. 2002; 9:416–418. [PubMed: 12032555]
58. Schlick TL, Ding Z, Kovacs EW, Francis MB. Dual-surface modification of the tobacco mosaic virus. J Am Chem Soc. 2005; 127:3718–3723. [PubMed: 15771505]
59. 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. 2014; 289:13565–13574. [PubMed: 24671415]

60. Juskaite V, Corcoran DS, Leitinger B. Collagen induces activation of DDR1 through lateral dimer association and phosphorylation between dimers. Elife. 2017; 6:e25716. [PubMed: 28590245]

Europe PMC Funders Author Manuscripts

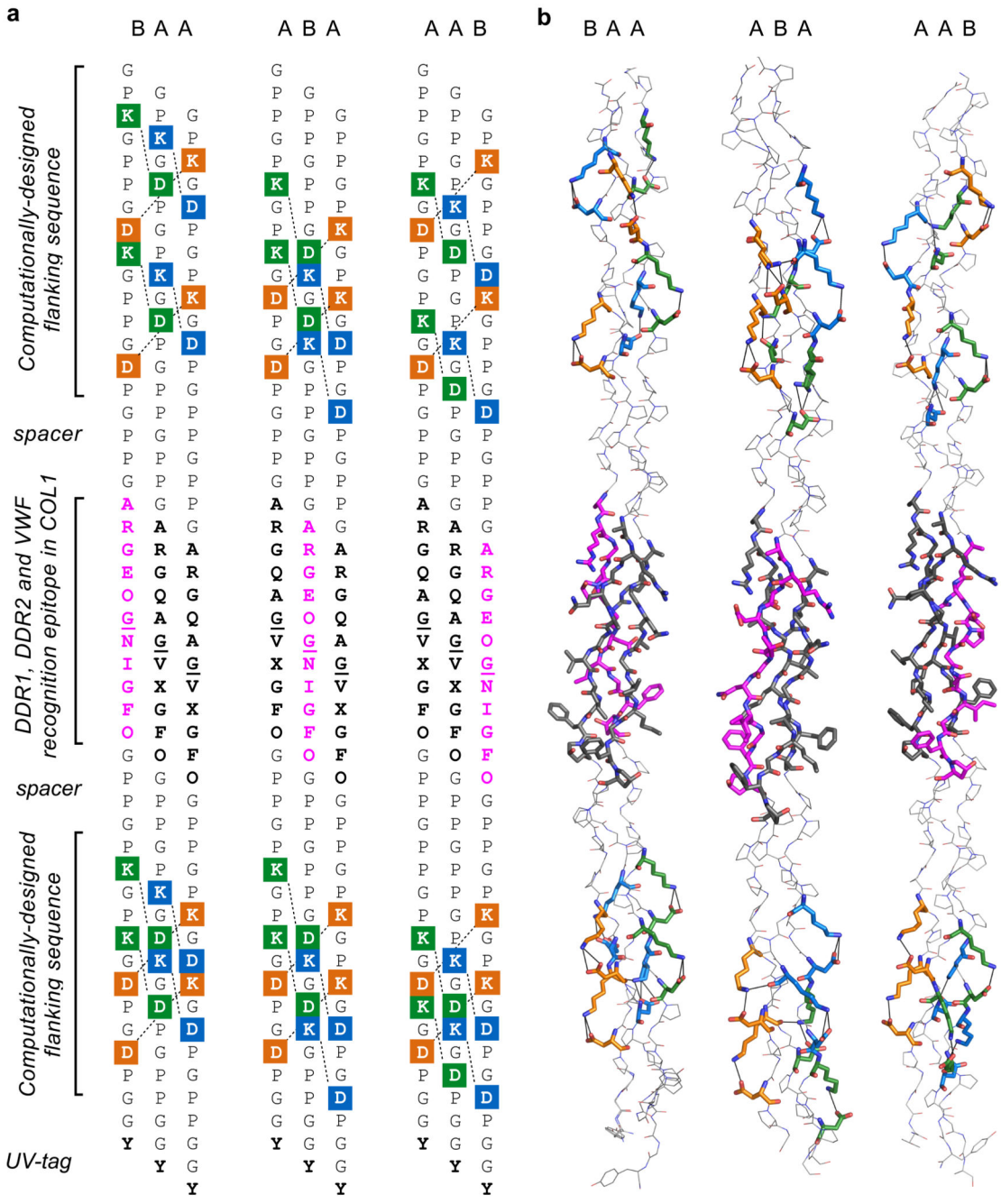

Figure 1. Design and structure of register-specific heterotrimers.

a,b, Design of heterotrimeric registers (**a**) and corresponding crystal structures (**b**) indicating 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. Chain A and B of the putative COL1 recognition epitope are coloured black and magenta, respectively. Hydrogen atoms and solvent molecules are not shown for clarity. Lys-Asp salt bridges between the leading-middle, middle-trailing and trailing-leading chains are coloured green, blue and orange, respectively. ^{15}N-isotopically

enriched Gly residues are underlined; Tyr used for determining peptide concentration is shown in bold; X denotes the methionine bioisostere, norleucine. Electron density maps of the recognition epitope and flanking sequences are presented in Supplementary Figure 2.

Europe PMC Funders Author Manuscripts

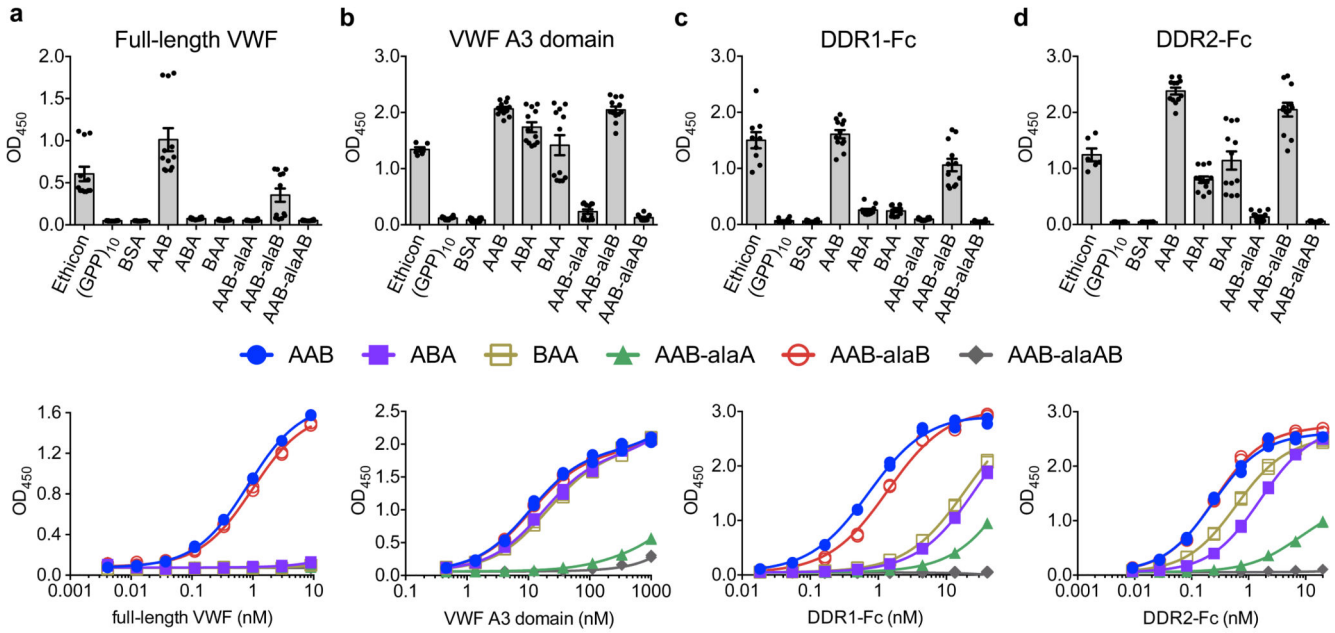

Figure 2. Binding affinity of registers for collagen-binding proteins.

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 alanine mutants AAB-alaA, AAB-alaB and AAB-alaAB. Ethicon™ collagen I fibers were used as 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 OD_{450} measurement from a single well. Error bars indicate the mean ± SEM. The binding isotherms were fit to a model of specific and non-specific binding in GraphPad Prism 6 and non-specific binding curves were removed for clarity. All replicates are shown in dose-response curves. Dissociation constants obtained from dose-response are shown in Supplementary Table 4.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

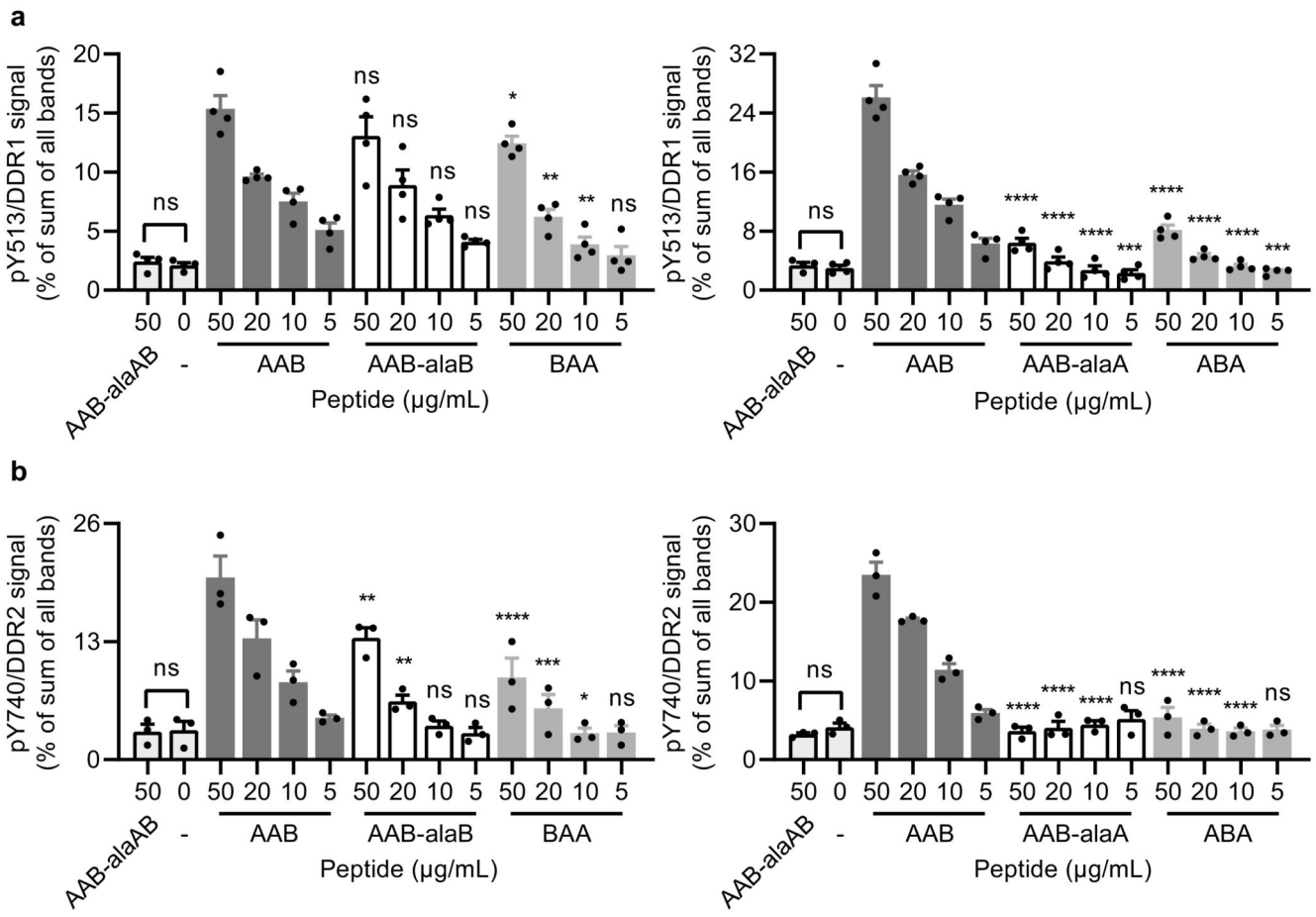

Figure 3. Peptide-induced DDR1 and DDR2 autophosphorylation.

a,b, HEK293 cells transiently expressing DDR1 (**a**) or DDR2 (**b**) were stimulated with the indicated peptides (concentration in μg/mL) or left unstimulated (-) for 90 min at 37°C. Cell lysates were analysed for phospho-tyrosine (anti-pY513 for DDR1; anti-pY740 for DDR2) and total DDR1 or DDR2. AAB and AAB-alaAB stimulated samples were included as positive and negative controls, respectively, on all blots. Representative blots are shown in Supplementary Figs. 6 and 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 SEM 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-alaAB 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 Bonferonni post hoc test). Details of statistical analysis, including precise p values, are presented in Supplementary Tables 5 and 6.

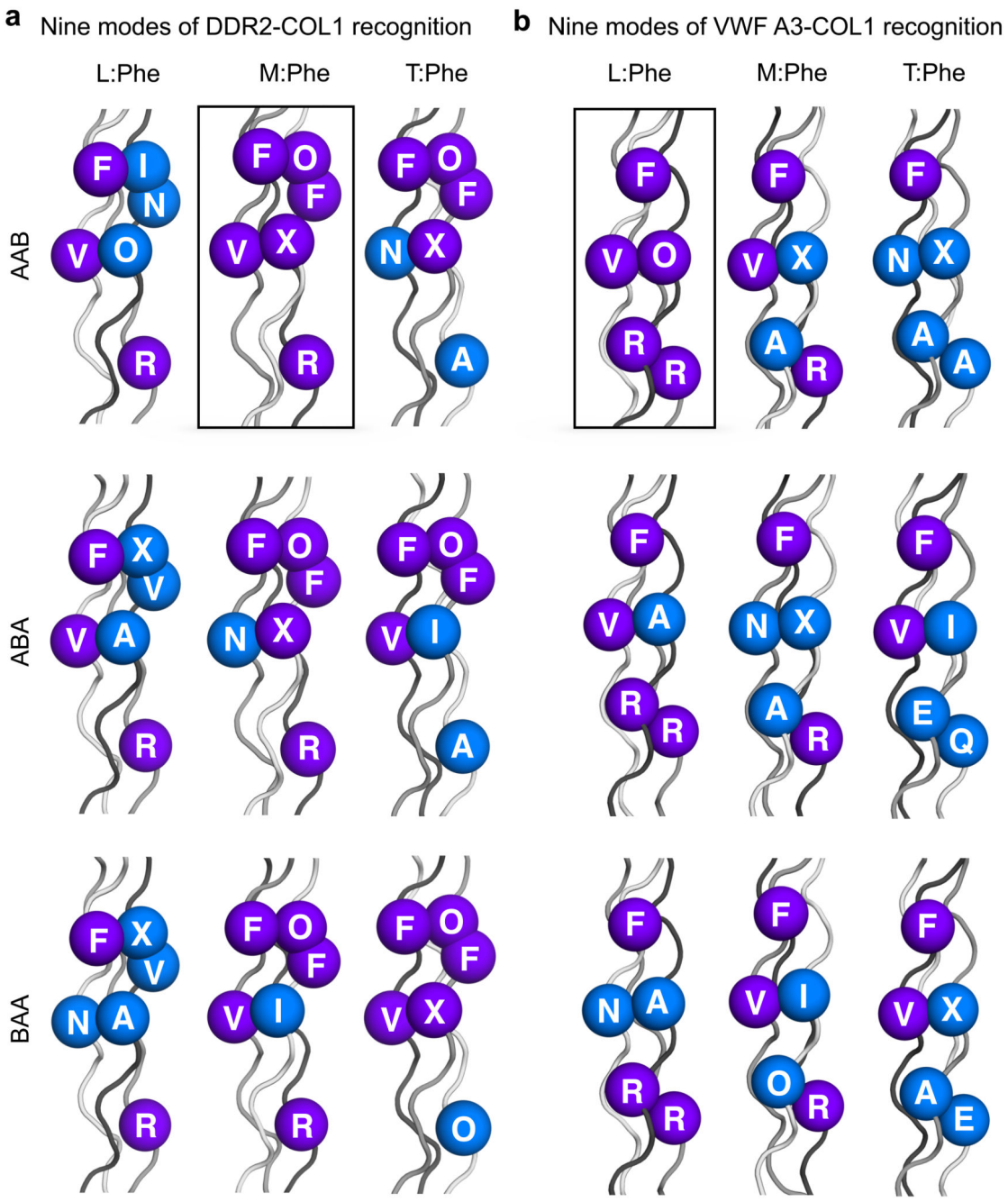

Figure 4. Nine possible modes for binding of COL1 to DDR2 and VWF A3.

a,b, Schematic drawing of AAB, ABA and BAA core residues that would contact DDR2 (**a**) or 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 120º 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 VWF A3 domain (depicted in Supplementary Fig. 8) are shown in purple and those replaced

with suboptimal interactions are shown in blue. Single letter amino acid code O denotes 4-(*R*)-hydroxyproline and X denotes the methionine bioisostere, norleucine. The binding mode that results in conservation of all interfacial contacts is shown in a box.

Europe PMC Funders Author Manuscripts