

SCIENTIFIC REPORTS



OPEN

COMP and TSP-4 interact specifically with the novel GXKGHR motif only found in fibrillar collagens

Jan M. Gebauer¹, Anna Köhler², Helen Dietmar², Monika Gompert¹, Ines Neundorff¹, Frank Zaucke^{2,3}, Manuel Koch^{2,4} & Ulrich Baumann¹

COMP (cartilage oligomeric matrix protein) is a member of the thrombospondin family and forms homopentamers as well as mixed heterooligomers with its closely related family member TSP-4. COMP is long known to bind to collagens and to influence collagen fibril formation. Recent work indicates that already intracellular interaction with collagen is important for collagen secretion. However, the exact binding site of COMP on the collagen triple helix has not been described up to now. In this study we have identified a GXKGHR motif on the collagen II helix to bind to COMP, using a recombinantly expressed collagen II peptide library. This binding sequence is conserved throughout evolution and we demonstrate that TSP-4 binds to the same sequence. The identified binding motif overlaps with the recognition sites of many other collagen-binding partners (e.g. PEDF, Heparin) and also spans the lysine residues, which form collagen cross-links. COMP might thereby protect collagen helices from premature modification and cross-linking. Interestingly, this motif is only found in classical fibrillar collagens, although COMP is known to also bind other types. This might indicate that COMP has a unique interface for fibrillar collagens, thus making it an interesting target for the development of antifibrotic drugs.

Collagens are an essential component of the extracellular matrix (ECM) of all higher organisms. 28 family members are currently recognized, numbered with roman numbers from I to XXVIII¹. Collagens always form trimers and the individual polypeptides are called α chains. Some collagens form homotrimers consisting of three identical polypeptides, while others form heterotrimers with two or three different α chains. All of them share as defining feature the so-called collagen triple helix; a structural element that is formed by three individual polypeptide chains (reviewed in¹⁻³). The three-dimensional structure of this extended, non-globular domain requires a glycine in every third position, leading to a repetition of the triplet motif Gly-Xaa-Yaa. Additionally, the Xaa and Yaa position are often prolines and hydroxyprolines, respectively (approx. 27% and 38%). Biosynthesis of collagen is a complicated process involving extensive post-translational modifications⁴. After secretion into the endoplasmic reticulum (ER), a large network of modifying enzymes like proline-hydroxylase or glycotransferases as well as molecular chaperones, such as protein disulphide isomerase, binding immunoglobulin protein (BIP), or heat shock protein 47 (HSP47)^{4,5} are involved in the formation of a properly modified and folded trimeric procollagen molecule that is secreted via the Golgi compartment to the extracellular space.

Collagens form a great diversity of superstructures. Besides the well-known collagen fibrils, they also form hexagonal networks (VIII & X), beaded filaments (VI), basement membrane networks (IV), and anchoring fibrils (VII). Some collagens are even membrane bound (XIII, XVII, XXIII, XXV). Collagens have a vast variety of functions, as demonstrated by their causative role in various inherited diseases, with *Epidermolysis bullosa* (e.g. collagen VII) and *Osteogenesis imperfecta* (e.g. collagen I) being amongst the most well-known collagen linked diseases. The interactions of the collagen helix with other extracellular proteins is very important for the understanding of the extracellular matrix and its related diseases; however, to date the exact nature of many interactions is only poorly understood.

¹Institute of Biochemistry, University of Cologne, Cologne, Germany. ²Center for Biochemistry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ³Dr. Rolf Schwiete Research Unit for Osteoarthritis, Orthopedic University Hospital Friedrichsheim gGmbH, Frankfurt, Germany. ⁴Institute for Dental Research and Oral Musculoskeletal Biology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. Correspondence and requests for materials should be addressed to J.M.G. (email: jan.gebauer@uni-koeln.de)

Received: 29 September 2016

Accepted: 5 November 2018

Published online: 21 November 2018

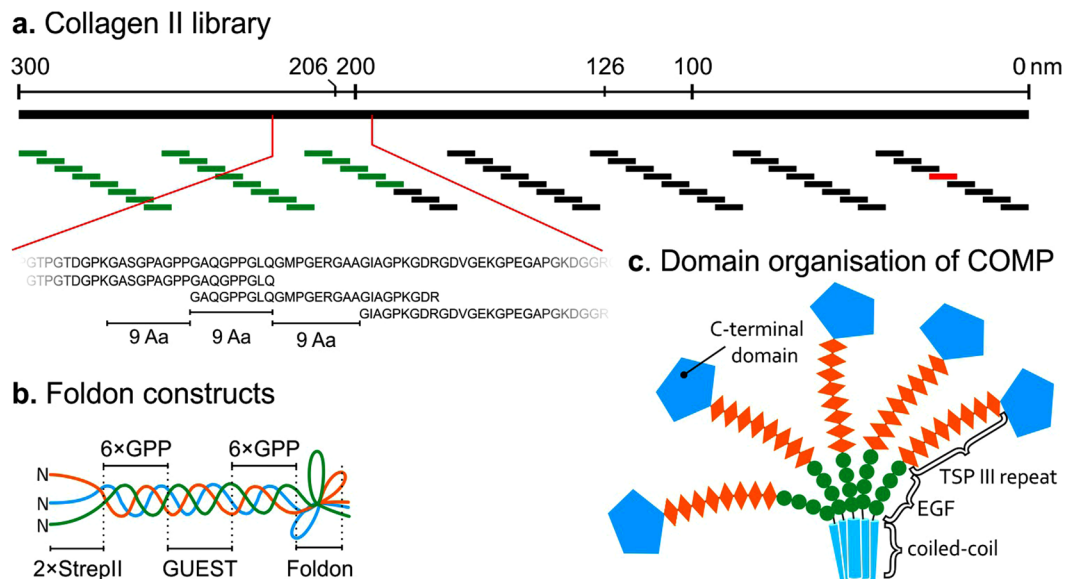


Figure 1. Organisation of the collagen library and protein constructs. **(a)** The 300 nm long collagen II triple helical domain was divided into 56 overlapping peptide sequences. The amino acid (Aa) sequences were backtranslated and optimised for *E. coli* expression⁴³. The current library covers the N-terminal region of collagen II covering the sequence stretch between ₂₁₀GPMGPM to GEDGRP₅₈₇ of the collagen helix (peptides in green/ Table 1). The red marked peptide (pep52) was also expressed later in the study. **(b)** The sequences were inserted into the GUEST position of a vector coding for a GPP collagen host sequence with a C-terminal foldon domain. **(c)** COMP is a pentameric protein with a multi-domain structure.

No	Sequence
1	<u>GPMGPMGPRGPPGAGAPGPFQGNP</u>
2	GPQGFQGNP <u>GEPGEPGVS</u> GPMGPRGPP
3	GPMGPRGPPGPPGKPGDDGEAGKPGKA
4	GEAGKPGKAG <u>ERPPG</u> QGARGFPGTP
5	GARGFPPTPGLPGVKGHRGYPLDGAK
6	GYPGLDGAKGEAGAPGVKGESGSPGEN
7	GESGSPGENSGPGMGRPLGERGRT
8	GLPGERGRT <u>GAGAAGAR</u> NDGQPGPA
9	GNDGQPGPAGPPGPGVGPAGGPGFPGAP
10	GGPGFPGAP <u>GAKGEAGPT</u> GARGPEGAQ
11	GARGPEGAQ <u>GPRGEP</u> TPGSPGPAGAS
12	GSPGPAGASGN <u>P</u> GTGIPGAKGSAGAP
13	GAKGSAGAP <u>GIAGAP</u> GFPGRGPPGPQ
14	GPRGPPG <u>PQATG</u> PLGPKGQTGEPGIA
15	GQTGEPGIAG <u>FKGEQ</u> GPKGEPGPAGPQ
16	GEPGPAGPQ <u>GAPGAGE</u> GKRGARGEP
17	GKRGARGEP <u>GGVGP</u> IGPPGERGAPGNR
18	GERGAPGNR <u>GFP</u> QDGLAGPKGAPGER
19	GPKGAPGER <u>GPS</u> LAGPKGANGDPGRP
20	GANGDPGRP <u>GEP</u> GLPGARGLTGRPGDA
21	GLTGRPGDAG <u>PQ</u> GKVGPSGAPGEDGRP
52	GEAGEPGERLKGHRGFTGLQGLPGPP

Table 1. Sequences of the collagen II N-terminal foldon library. All peptide sequences are presented in an NStrep-(GPP)₆-GUEST-(GPP)₆-foldon context. Underlined sequences are unique and are only present in one peptide.

This is partly because collagen domains are difficult to express recombinantly and even more difficult to express as fragments. For other protein interactions, truncations and site directed mutagenesis normally lead to a general idea of a binding interface. Ideally, this knowledge is further used to crystallize the minimal binding complex, leading to detailed understanding of the interaction. While this method is very suitable for the

non-collagenous side of the interaction, collagenous domains are difficult to manipulate. Breaking collagenous domains into pieces is not trivial, as the triple helical structure needs to be retained. Current techniques involve the chemical synthesis of collagen model peptides with a host-guest-approach. Typically, the sequence of interest is flanked by approximately six glycine-proline-hydroxyproline (GPO) repeats on both sides. The GPO repeats serve as a host and facilitate the formation of a triple helical collagen structure. However, only peptides of a certain length can be efficiently produced and the synthesis is time consuming and expensive. With collagenous domains of more than 300 triplets, screening the whole domain with overlapping fragments is only feasible for the most abundant collagens. Indeed peptide libraries, called “collagen toolkits”, are only available for collagen II and III⁶.

Besides chemical synthesis, collagen model peptides can also be expressed recombinantly. Often trimerisation domains, like the T4 foldon domain, are used in these systems in order to stabilise the truncated triple helix⁷. The collagenous part is normally designed in the same host-guest approach mentioned above. Typically expressed in *E. coli*, these fragments can be much longer and are produced at lower costs. However, incorporation of hydroxyproline in *E. coli* is difficult, and the use of eukaryotic cell lines to overcome this problem increases the cost and time significantly. Recent improvements in the post-translational generation of hydroxyproline in prokaryotes⁸ might alleviate this problem. However, to our knowledge, recombinantly expressed collagen peptide libraries were not systematically used before this study.

One of the important interaction partners of collagens in cartilage is the cartilage oligomeric matrix protein (COMP) a member of the thrombospondin family of proteins, also referred to as TSP-5. COMP forms homopentamers induced by an N-terminal coiled-coil domain, which is followed by four epidermal growth factor-like domains (EGF), eight thrombospondin type III repeats, and a C-terminal COMP (TC) domain (Fig. 1c)⁹. In healthy situations COMP has been described to fulfil numerous functions by cross-linking various components in the extracellular matrix and might therefore play an important role in the organisation of the different matrix protein networks⁹. In addition to other extracellular matrix proteins (for review see⁹), COMP is described to bind directly to collagen type I and II¹⁰, collagen IX¹¹, as well as collagen XII and XIV¹². Binding of pentameric COMP to the fibrillar collagens leads to an increased rate of fibril formation *in vitro*¹³, while the interaction of COMP with collagens XII and XIV may be important for the dermal-epidermal junction¹². Recently, the importance of intracellular COMP/collagen interaction for collagen secretion was described¹⁴, a function also attributed to HSP47¹⁵.

Mutations in COMP can lead to skeletal disorders in humans, ranging from mild to severe forms of multiple epiphyseal dysplasia (MED) or pseudoachondroplasia (PSACH)¹⁶. Most disease-causing mutations are not secreted and the pathology is believed to be caused by a co-retention of fibrillar collagen. For some mutations, however, the secretion is normal and for one particular mutation (pH587R) an aberrant extracellular fibril formation is observed¹⁷. Interestingly, this mutation (and others), lie in the C-terminal domain of COMP¹⁸, which is responsible for the collagen binding.

Despite this well documented functional importance of the collagen–COMP interaction, the exact binding interface on the collagen triple helix is unknown up to now. In this study, we identified for the first time the direct binding epitope for COMP and TSP-4 on the collagen helix.

Results

Generation of the N-terminal collagen library. From electron micrographs it is known that COMP binds to four distinct sites, located at approximately 0, 126, 206, and 300 nm distance from the C-terminus of the triple helix¹⁰. To elucidate the N-terminal binding site of COMP, we generated a foldon stabilised, *E. coli* expressed, collagen model peptide library spanning the first 100 nm of collagen II (Fig. 1 and Table 1). Although the design of the library was inspired by the collagen II toolkit⁶, by using a foldon tag and a prokaryotic expression system, we hoped to reduce costs and increase speed for identification and characterisation of binding sites. Furthermore, the existence of a foldon tag and an affinity tag enables certain biochemical assays (e.g. biolayer interferometry), difficult to perform with synthesised peptides. The library was generated using the golden gate assembly method^{19,20}, which uses Type IIS restriction endonuclease. These enzymes cut outside of their recognition sequence. With a proper design of the vectors and PCR primers (respectively inserts) the recognition sites can be excluded from the final product. Cycling between cutting and ligation steps will thus accumulate the only stable product. In our study we modified an existing pET vector²¹ already containing several GPP repeats, an N-terminal 2xStrep tag and a C-terminal foldon domain. Via a central XhoI site, we introduced the necessary acceptor sites (pCMP-3b, Fig. S2). This vector together with synthetically synthesised DNA oligos allowed rapid cloning, with nearly 100% positive clones. The purified peptides run as a partially SDS-stable trimer on SDS-PAGE.

COMP binds to discrete peptides on the collagen library. To elucidate the binding site of COMP on the collagen II helix we incubated pentameric COMP at 200 nM concentration with our generated collagen II foldon library (Fig. 2). Only one peptide (pep5) showed a similar signal for the COMP binding in comparison with the collagen I interaction, which served as a positive control. To confirm our finding and get proper binding affinities we coated the peptide 5 and incubated it with indicated concentration of COMP (Fig. 3a). The dissociation constant for COMP and pep5 (19.3 ± 12 nM) was slightly higher than reported for collagen II (1.72 ± 0.16 nM)¹⁰. It is worthwhile mentioning, that COMP (as a pentamer) will exhibit significant avidity effect in solid-phase binding assays, which are dependent of the density of the coated binding partner.

COMP is described to interact with collagen I and II in a zinc dependent fashion. In our experiments, adding zinc was not strictly necessary for the interaction, but increased the affinity towards the peptides (Fig. 3b). However, we also noted an increase in background binding with 0.1 mM zinc. If zinc is important for the interaction, recombinantly expressed COMP needs to already carry zinc as we observe binding even

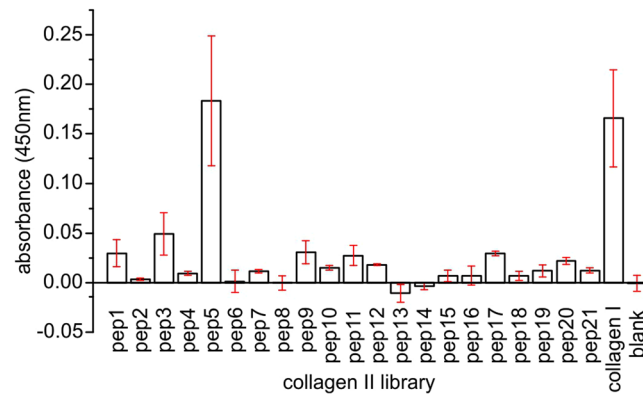


Figure 2. Binding of COMP to the N-terminal collagen II foldon library. Foldon stabilised collagen mimetic peptides and rat-tail collagen I as positive control were coated to an ELISA plate and incubated with 200 nM pentameric COMP. Binding was detected with a specific anti-COMP antibody. Bars indicate standard deviation of triplicates. The peptide sequences are depicted in Table 1.

in pure TBS. Following this line of argument, additional zinc would only fill empty binding pockets in COMP, thus increasing the active population of COMP in the solution. To test this hypothesis, we incubated COMP with 1,10-phenanthroline a high-affinity zinc chelator and EDTA a more general chelator of divalent cations. Interestingly, phenanthroline does not reduce the binding compared to pure TBS, indicating that there is no “pre-charged zinc” present or at least not necessary for the interaction. EDTA, however, reduced the binding significantly (Fig. 3b), but this can easily be explained by the necessity of calcium ions for the overall structural integrity of COMP’s TSP repeats. If we incubate COMP with various concentrations of zinc and observe the hydrodynamic radius via DLS, we could not detect any difference between 0.0 mM and 0.1 mM ZnCl₂, but COMP starts to aggregate dramatically upon addition of 1 mM ZnCl₂ (Fig. 3c).

The minimal binding site of COMP on the N-terminus on collagen I is GPKGHR. The peptide 5 of the library has a total length of 27 amino acids. However, as the adjacent peptides 4 and 6 are not interacting, the amino acids responsible for COMP interactions has to be in or at least part of the unique region of pep5. We first tried, whether the unique region of pep5 “GLPGVKGHR” is the binding epitope for COMP and indeed this motif was sufficient for binding, as shown by one-point measurements (Fig. 4a). To further narrow down the exact prerequisites for COMP binding, we systematically exchanged all non-glycine amino acids in the GLPGVKGHR motif to proline residues (Table 2). Interestingly, neither the exchange of the leucine nor the valine did reduce the affinity significantly (Fig. 4b). The exchange of the arginine or the lysine residue, however, reduced the binding affinities below measurable values. The exchange of the histidine showed an intermediate effect.

Knowing that the leucine is not important for binding, we tried to further narrow down the interaction motif to GVKGHR and observed a similar binding as to original peptides (Fig. 4).

Collagen II has a second binding site close to the C-terminus. From earlier studies it was known that COMP binds to the collagen helix at least twice¹⁰. We used our consensus sequence GXXKGHR to search for additional binding sites on collagen II. Close to the C-terminal propeptide, the sequence ₁₁₂₈GLKGHR₁₁₃₃ matches this consensus. We cloned, expressed, and purified the library peptide 52, which includes the GXXKGHR motif in its unique sequence (GEAGEPGERGLKGHRGFTGLQLPGPP). In ELISA-style binding assays this peptide bound with a good, although slightly reduced affinity (K_D 4.14 ± 0.77 nM) in comparison to peptide 5 (Fig. 5).

Collagen XI also binds COMP. By searching other collagens for this collagen:COMP interaction sequence, we found the GXXKGHR motif in collagen I, II, III, and additionally in collagen V, and XI (Table 3). The latter are to our knowledge currently not known to bind COMP. Interestingly, all motifs are located at the same positions in the collagenous domain. The N-terminal motifs always lies 28 GXX triplets from the start while the C-terminal motifs is placed around 27-28 triplets from the end of the collagen. The large numerical difference in numbering of collagen $\alpha 1(V)$ is due to the big N-terminal domain of this particular chain.

We tested binding of COMP to collagen XI in an ELISA-style binding assay (Fig. 6). COMP binds to collagen XI, although the affinity is significantly lower compared to collagen I and II (20 nM ± 3 nM). The motifs found in collagen XI are GEKGHR and GDKGHR in the $\alpha 1$ chain and $\alpha 2$ chain, respectively. In contrast to the motifs in collagen I, II and III the X position is occupied by a negatively charged amino acids. It is tempting to speculate that this charge is responsible for the observed lower affinity.

TSP-4 also interact with the GXXKGHR motif. COMP is a member of the Thrombospondin family. For TSP-4 it is known that it also binds collagens²². To test if our motif might be a common motif for collagen binding, we investigated whether TSP-4 binds to the GVKGHR peptide as well. Indeed, the affinity of TSP-4 (0.56 ± 0.05 nM) to our collagen model peptide was very similar to the affinity of COMP (Fig. 7).

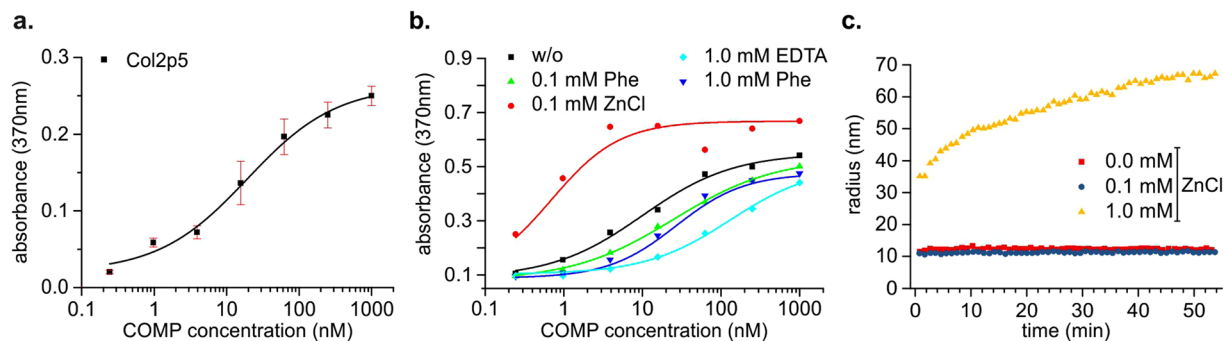


Figure 3. Binding of COMP to peptide 5 and the dependency of Zn^{2+} . Peptide 5 of the collagen II foldon library was coated to an ELISA plate and incubated with COMP protein (a) or with COMP in interaction buffer containing indicated chemical compounds (b). Binding was detected with a specific anti-COMP antibody. Bars indicate standard deviation of triplicates. The peptide sequence is depicted in Table 1. The calculated K_D for the pep5 interaction in (a) is 19.3 ± 12.4 nM. (c) The radius of particles (obtained by cumulant analysis) measured of a COMP solution with addition of the indicated $ZnCl$ ions were monitored over time using dynamic light scattering. *Phe*: 1,10-phenanthroline.

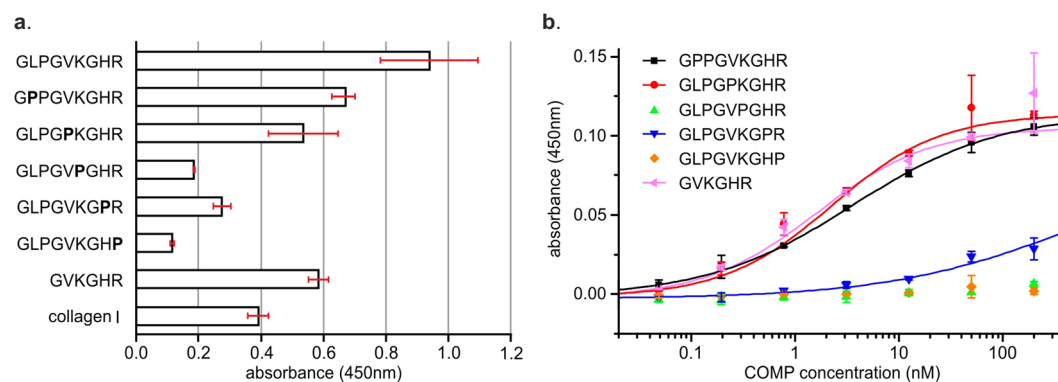


Figure 4. Characterisation of important residues for COMP binding. Amino acids, with the exception of proline or glycine, were exchanged to proline (bold indicated). COMP was incubated either at 400 nM (a) or at indicated concentrations (b). Binding was detected with a specific anti-COMP antibody. Bars indicate standard deviation of triplicates.

Sequence	Apparent K_D (nM)	Melting point ($^{\circ}C$)
GARGFP <u>GTPLPGVKGHR</u> GYPGLDGA	19.3 ± 12.4	36.1 ± 0.5
GLPGVKGHR	n.d.	48.2 ± 0.6
GPPGVKGHR	3.57 ± 0.37	54.2 ± 0.7
GLPGPKGHR	2.26 ± 0.39	52.5 ± 0.6
GLPGVPGHR	n.a.	51.1 ± 0.3
GLPGVKGPR	496 ± 3203	n.d.
GLPGVKGHP	n.a.	47.9 ± 0.4
GVKGHR	1.61 ± 0.29	56.5 ± 0.6

Table 2. Peptides used for further analysis of the minimal binding site. The sequence unique to peptide 5 is underlined and mutations in comparison to the wt sequence are marked in bold. All peptide sequences are presented in an N-Strep-(GPP)₆-GUEST-(GPP)₆-foldon context. Melting points were determined by CD spectroscopy at 210 nm. Exemplary spectra can be found in the Fig. S1. (n.d. - not determined; n.a. - not applicable/no measurable signals observed to fit K_D values).

Docking of COMP to a collagen model peptide. To elucidate possible binding interfaces we used an unpublished crystal structure of a collagen peptide with the sequence Ac-(POG)₄LKGHRGFTGLQG-(POG)₄-NH₂, which resembles the C-terminal COMP binding motif of collagen II. The statistics for data collection and structural solution of this peptide can be found in Table 4. Docking was performed with ClusPro and the best 10 models analysed by manual inspection. The models clustered into two groups, whereas one group with 7 of 10 structures,

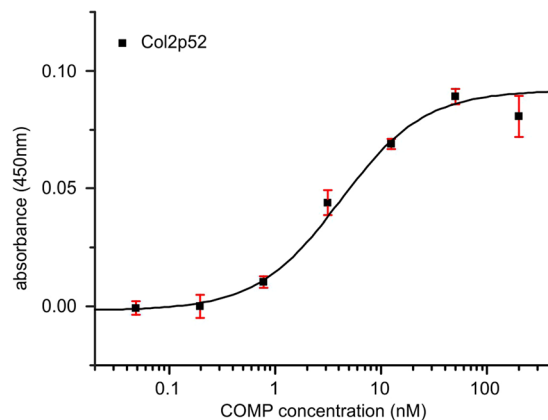


Figure 5. Binding of COMP to the C-terminal motif of collagen II contained in peptide 52. Peptide 52 of the collagen II foldon library was coated to an ELISA plate and incubated with COMP protein. Binding was detected with a specific anti-COMP antibody. Bars indicate standard deviation of triplicates. The peptide sequence is depicted in Table 1. The fitted K_D is 4.1 ± 0.8 nM.

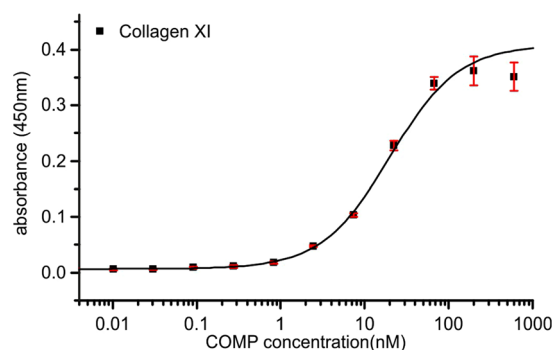


Figure 6. Binding of COMP to collagen type XI. Collagen XI was coated to an ELISA plate and incubated with COMP protein. Binding was detected with a specific anti-COMP antibody. Bars indicate standard deviation of triplicates. The peptide sequence is depicted in Table 1. The fitted K_D is 20.0 ± 3.0 nM.

Collagen $\alpha 1$ (I)	²⁶³ GMKGHR ₂₆₈ ¹¹⁰⁶ GIKGHR ₁₁₁₁
Collagen $\alpha 1$ (II)	²⁸⁵ GVKGHR ₂₉₀ ¹¹²⁸ GLKGHR ₁₁₃₃
Collagen $\alpha 1$ (III)	²⁶¹ GMKGHR ₂₆₆ ¹¹⁰⁴ GIKGHR ₁₁₀₉
Collagen $\alpha 1$ (V)	⁶⁴⁰ GEKGHR ₆₄₅
Collagen $\alpha 2$ (V)	²⁹⁷ GEKGHR ₃₀₂ ¹¹⁴⁰ GQKGHR ₁₁₄₅
Collagen $\alpha 1$ (XI)	⁶¹⁰ GDKGHR ₆₁₅
Collagen $\alpha 2$ (XI)	⁵⁶⁸ GEKGHR ₅₇₃

Table 3. Binding sites in other collagens.

located the collagen helix around the aspartate triplet ⁵⁹³DDD₅₉₅, which was described earlier as a potential MIDAS motif²³. Interestingly, the two calcium ions in the crystal structure would be replaced by the collagen side chains. Furthermore, no metal ion would be necessary for this interaction (Fig. 8). In the docked model, Asp₅₉₃ of COMP form hydrogen bonds to the arginine and lysine of the middle strand (Fig. 8 dark blue) and the histidine of the middle chain (Fig. 8 light blue). Asp₅₉₄ and Asp₅₉₅ interact with the lysine of the middle strand. To test this model, we mutated all three aspartates to alanine residues and expressed and purified the protein. To our surprise the mutant behaved similar to the wildtype in binding assays (Fig. 8b), falsifying the proposed model.

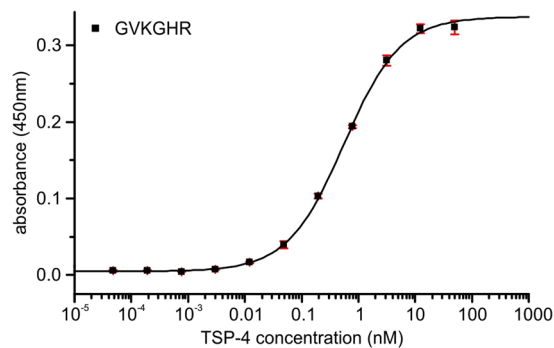


Figure 7. Binding of TSP4 to the GVKGHR motif. The minimal binding site GVKGHR expressed as a foldon construct was coated to an ELISA plate and incubated with TSP4. Binding was detected with a specific anti-TSP4 antibody. Bars indicate standard deviation of triplicates. The fitted K_D is 0.56 ± 0.05 nM.

Data collection	CMP
Beamline	X06DA (PXIII)/SLS
Wavelength (Å)	0.800
Space group	I2 (No. 5)
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	31.97, 23.64, 107.94
α , β , γ (°)	90.00, 92.70, 90.00
Resolution (Å)	31.02–1.00 (1.06–1.00)
<i>R</i> _{sym}	0.046 (0.559)
<i>R</i> _{meas}	0.024 (0.602)
<i>CC</i> _{1/2}	1.00 (0.963)
<i>I</i> / σ <i>I</i>	18.92 (4.17)
Completeness (%)	99.7 (98.5)
Multiplicity	6.38 (6.25)
Refinement	
Resolution (Å)	31.02–11.00
No. reflections (test set)	43886 (598)
<i>R</i> _{work} / <i>R</i> _{free} (%)	21.8/23.4
No. atoms	
Protein	686
Ligand/ion	17
Water	184
B-factors	
Protein	22.48
Ligand/ion	31.62
Water	34.89
Ramachandran	
Favoured (%)	100
Outliers (%)	0
R.M.S deviations	
Bond lengths (Å)	0.004
Bond angles (°)	1.25

Table 4. Data collection and refinement statistics. Highest resolution shell is shown in parenthesis.

Discussion

It is well known that cartilage oligomeric matrix protein (COMP) interacts with collagen type I and II¹⁰, the exact sequence however has been unknown up to now. We expressed overlapping peptides covering the first third of the collagen II helix, which is known to harbour at least one COMP binding site. The peptides were stabilised by flanking GPP repeats and a C-terminal T4 foldon domain. The overlapping peptides were designed similarly to the well-established collagen II toolkit⁶. Only the peptide 5 interacted specifically with pentameric COMP. We checked the trimeric nature of the collagen helix for peptide 5 and its derivatives via CD spectroscopy and could confirm a proper collagen helix at room temperature (Table 2 and Fig. S1). There may be further binding sites as

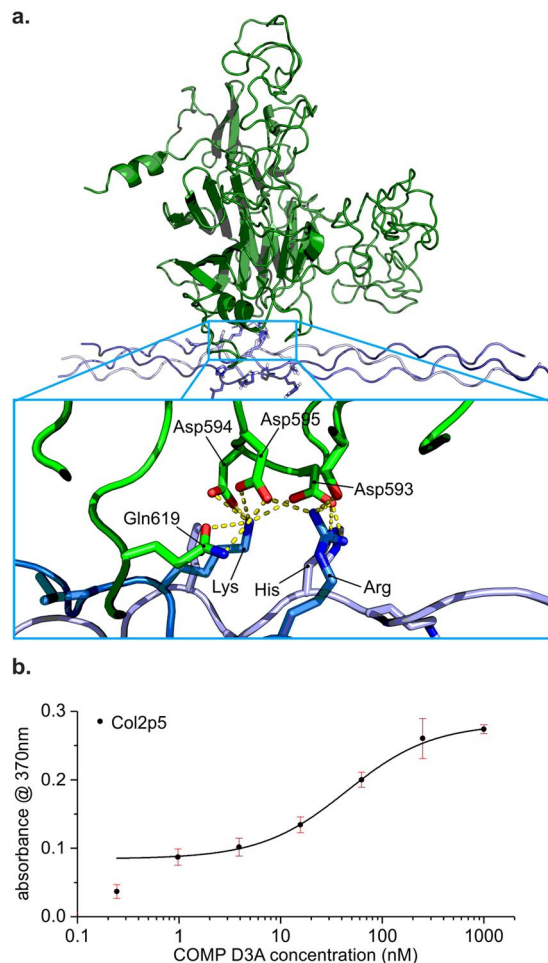


Figure 8. Docking of COMP to a collagen model peptide. **(a)** The COMP C-terminal domain (3FBY) was docked to our collagen model peptide carrying a GLKGHR motif using ClusPro2. The lysine and arginine residues of the leading and middle chain were defined as attractive. Best-rated models included the above shown complex, where the COMP interacts with the presumed MIDAS motif. Interestingly, this would displace the bound calcium ions and does not need a metal ion for binding. **(b)** ELISA-style-binding assay with Col2p5 and COMP D3A shows similar binding of COMP D3A to wt. Bars indicate standard deviation of triplicates.

we have only examined a library spanning the N-terminal part of collagen II and only checked the interacting peptides for proper triple-helix formation.

We investigated the unique sequence of peptide 5 (GLPGVKGHR) for binding by mutating single amino acids to proline (Fig. 4). The exchange of either the lysine or the arginine residue lead to a complete loss of binding at the measured concentrations, while the exchange of the histidine residue significantly reduced the affinity by several orders of magnitude. The exchange of the leucine and the valine residue did not alter affinity significantly and the shortened version GVKGHR bound collagen with similar affinity. Therefore, we propose that GXKGHR is a sufficient motif for collagen binding by COMP and crucial for high affinity interactions. The slight decrease in affinities observed with the shortened peptides in comparison to the peptide 5 might indicate some additional beneficial interactions further away from the core sequence.

This proposed core sequence occurs twice in collagen II. Firstly, as the originally found GVKGHR motif starting at position 285, and secondly as a GLKGHR motif starting at position 1128. This sequence is contained in the collagen II peptide 52, for which we also showed binding to COMP (Fig. 5).

To study how well these two sites are conserved during evolution, we reviewed all currently annotated orthologues of collagen II in the Ensemble database. Due to the repetitive sequence of the collagen helix, the overall conservation rates are high, with more than 80 percent identity amongst distant species (e.g. human and zebrafish). However, although the N- and C-terminal GXKGHR binding motif is completely conserved in all species, there are variation at the undefined position X_1 (X in the motif above). N-terminally the GVKGHR motif in mammals (*Mammalia*) is converted to a GIKGHR motif in fish (*Sarcopterygi*), while the GLKGHR at the C-terminus, is exchanged to GVKGHR in guinea pig and kangaroo rat, and to a GQKGHR motif in fish. This strongly supports the finding that the X_1 position is not of strict importance for COMP binding, and that at least glutamine and isoleucine are tolerated at this position.

We then searched for other collagen types bearing a GXKGHR motif. We found this motif to be present in all classical fibrillar collagens (collagen I, II, III, V, and XI; compare Tab. 3), which might indicate that all fibrillar collagens are clients for COMP mediated fibril formation assistance. However we were not able to find our binding motif in the FACIT collagens, IX, XII, and XIV, which are also known to interact with COMP^{12,24}. For collagen IX it is known that COMP's C-terminal domain only interacts with the non-collagenous domains (NC 1–4)^{11,24}. It is not unlikely that this interactions involves a different interface on COMP's C-terminal domain. For collagen XII and XIV the exact binding sites are not yet determined, and initial studies indicated an involvement of the collagenous domain or its interspersed short NC domain. Whether COMP binds to a different collagen sequence or parts of the NC domains in these two collagens needs to be further investigated.

Interestingly, in collagen I, the GXKGHR motifs are present only in the $\alpha 1$ chain and not in the $\alpha 2$ chain. As collagen I is a heterotrimer build from two $\alpha 1$ chains and one $\alpha 2$ chain, there will be a significant difference in the COMP binding site of collagen I compared to the homotrimeric collagen II, i.e. collagen I only present two binding motifs instead of three. Although it is tempting to speculate that this implies a lower affinity of COMP to collagen I, two important aspects of collagen recognition have to be considered. Firstly, the exact stoichiometry of the collagen-COMP interaction is currently unknown. Other known collagen interfaces suggest a typical stoichiometry of 1 or 2 globular proteins per collagen triple helix. This is easy understandable, due to the tight packing of the collagen helix and the resulting steric hindrance for more than 2 binding partners (compare crystal structures of HSP47-collagen complex (PDB: 4AU2)²⁵ or integrin $\alpha 2$ -collagen complex (PDB: 4BJ3)²⁶). Considering the relatively big and bulky size of the C-terminal domain and the adjacent T3 repeats (PDB: 3FBY²³) it is more likely that only one COMP binds per collagen triple helix. This would also better explain the observed 1:5 ratio for COMP's activity in fibril formation¹⁰. Secondly, collagen motifs are presented as three-dimensional epitopes. As the collagen superhelix turns 9° per amino acid, a typical interaction surface of collagen is generated by more than one chain. The prototypical collagen-binder integrin $\alpha 2$ (recognition sequence: GFOGER) for example recognises a glutamic acid, a phenylalanine, and an arginine on the trailing strand, and a phenylalanine, hydroxyproline, and arginine on the middle strand^{26,27}. Similarly, the von Willebrand Factor recognises its collagen motif (GxRGQOGVMGFx) via the valine and phenylalanine from the leading strand, and the arginines from middle and trailing strand²⁸. As neither the three-dimensional binding interface of collagen to COMP nor the stagger of the collagen I helix is known, we can currently not evaluate whether the absence of the GXKGHR motif from the $\alpha 2$ has any effect on the collagen binding. However, by conversion we could also conclude that our identified binding interface formed by a homotrimeric GXKGHR motif might also be formed by a heterotrimer with a different linear sequence, i.e. the lysine and arginine might come from different chains. Therefore, our current search motif potentially does not identify all possible interaction sites in heterotrimeric collagens – a problem present with all currently described binding motifs in collagens.

Although all thrombospondins share a common domain structure, only TSP-1, TSP-4 and COMP (TSP-5) are known to directly interact with collagenous domains^{10,22,29}. The latter two proteins are also reported to form heterooligomers³⁰. In the thrombospondin family, TSP3, TSP-4, and COMP are closely related (52%, for TSP-3 vs TSP-4, 46% TSP-4 vs COMP, 42% COMP vs TSP-3 sequence identity) and form a subgroup in the protein family, characterised by their ability to form (homo-)pentamers. It is tempting to speculate that this group shares a common binding motif to collagen, especially as the identity in the C-terminal region is even higher (83%, 86%, 80%). Indeed, we could show that TSP-4 binds to the same motif as COMP, indicating that this motif might be a general motif for collagen recognition by thrombospondins. Whether TSP-3 also binds collagen is to our knowledge not yet tested. TSP-1 is very different in its C-terminal domain (only approx. 50 percent sequence identity to TSP-3, -4 or COMP) and is more similar to TSP-2. Structurally, however, the C-terminal domain of TSP-1 (PDB: 1UX6³¹) and COMP (PDB: 3FBY²³) are very similar and only differ by an rmsd of 0.89 Å, which might indicate that it binds collagen via a similar interface, despite their low sequence identity. Indeed very recently, others showed that TSP-1 interacts with the same binding site on collagens as COMP³².

Very interestingly, the GXKGHR motif is located in a region on collagen I and II, which seems to be a hotspot for collagen interactions. Recently, it was described that PEDF and heparin are binding in this region³³. Furthermore, the lysine residue is known to be important for cross-linking collagen fibrils^{34,35}. It is known that PEDF, as most collagen binding molecules, interfere with fibril formation *in vitro*³⁶. The interaction of COMP with collagen might therefore not only regulate fibril formation, but also protect this region from prematurely binding to other molecules. Similarly, the interaction with COMP might prevent incorrect crosslinking of not finally assembled collagen fibrils.

Recently it was shown that COMP also interacts with collagens intracellularly¹⁴. Binding of COMP to collagen is important for collagen secretion and lack of COMP leads to intracellular retention of procollagen molecule, which is similar to the lack of HSP47³⁷. Consequently, both proteins are described to be involved in fibrotic diseases. Recent studies showed that COMP ablation reduces negative effects of fibrotic condition in skin and liver^{14,38} and propose COMP as a good target for future anti-fibrotic drugs. Our current study reveals for the first time that COMP recognises fibrillar collagens specifically by a sequence motif not existent in other collagens. By inhibiting this interaction, a potential drug might specifically target the secretion of fibrillar collagens only, which are abundant in fibrotic tissues. This might be beneficial over the inhibition of HSP47, which also interacts with non-fibrillar collagens, at least collagen IV¹⁵.

Interestingly, the exact binding interface on the COMP site is still unknown. The susceptibility of the interaction to EDTA lead to the idea that divalent cations – presumably Zn ions – are important for the collagen interaction. However, only calcium ions were detected in the crystal structure of COMP²³, or the structurally related TSP-1³¹ or TSP-2³⁹, although a potential MIDAS motif was postulated in the former. Studies with a disease-relevant mutation in the C-terminal domain, showed an effect of mutations at position 587 on fibrillogenesis, which might be a result of defective collagen binding¹⁷. However, this mutation is more than 16 Å from the proposed MIDAS motif and a defective binding was not yet shown.

To get an idea of possible interaction sites, we docked a crystallised model peptide containing the C-terminal COMP binding motif of collagen II to the previously described crystal structure of COMP's C-terminal domain. Interestingly, in most predicted complexes the collagen is recognised by three consecutive aspartate residues (Asp593–Asp595, Fig. 8) which were previously predicted to form a MIDAS motif²³. The docking process does not support the addition of divalent cations and indeed none is necessary for the predicted interaction. Zn²⁺ is normally recognised by histidine, cysteine, or acidic residues (like aspartate or glutamate)⁴⁰. Although there is a histidine present in the recognition motif on the collagen side, the more important lysine and arginine residues are unknown to play a role in metal coordination. By testing a generated triple mutant replacing all three aspartate residues by alanines, we could show that this potential MIDAS motif is not involved in the binding of COMP to collagen. This agrees with our earlier observation that zinc is not necessary for the interaction of COMP with collagen. Earlier studies identified the DDD motif to be important for the integration of thrombospondins into the extracellular matrix via direct TSP-TSP interactions^{41,42}.

A final answer to this problem will best be provided by an experimental structural model of the complex, which is also of great importance for proper drug development. Our study characterised the collagen site and identified the minimal binding sequence necessary for COMP interaction. This should enable structural characterisation of the complex by high-resolution methods, which were previously hampered by the absence of defined collagen mimetic peptides.

In summary, our study showed for the first time, the minimal sequence requirements in collagen for COMP and TSP-4 binding. Interestingly, this motif (GXKGR) overlaps with the recognition sequence of various other collagen binders (heparin, PEDF). Due to its close proximity to other important interactions sites, this sheds new light on the role of thrombospondin interaction for collagen fibril formation and matrix organization and will facilitate the further investigation of COMP as a fibrotic drug target.

Experimental Procedures

Generation of a foldon stabilized collagen II library. Peptide sequences were backtranslated into a DNA sequence and the sequence optimised for *E. coli* expression using the OPTIMIZER webpage⁴³. For facilitating cloning, an adapted golden gate assembly system was used^{19,20}. Recognition sites for BsaI were attached to the 5' and 3' end of the sequence in that way that the recognition site of BsaI lies outside of the coding sequence (5'prime: TACATGGTCTCA|CGG, 3'prime: AGATGAGGTCTCA|GTCC; recognition site underlines, cut position indicated by a |). The sequences were ordered as a forward and reverse oligo with a central overlapping region of approximately 20 bases, the oligos were hybridised and filled up with a Pfu polymerase. A pET based vector containing an N-terminal double Strep-tag followed by 12 GPP repeats and a C-terminal T4 foldon domain²¹ was altered to include an insert with two internal BsaI restriction sites (pCMP-3b, Fig. S2).

100 ng of vector and equimolar amounts of the double stranded DNA insert was incubated with 30 units BsaI-HF (NEB) and 1000 units of T4 Ligase (NEB) in a total volume of 15 µl and incubated in a thermocycler with 15 alternating cycles of 37 °C for 3 min and 16 °C for 4 min. After a final digestion at 50 °C for 5 min all enzymes were inactivated at 80 °C for 5 min. 5 µl of the reaction was transformed into DH5α and the plasmid DNA from single clones sequenced (GATC Biotech).

Expression and purification of recombinant proteins. Collagen model peptides were expressed in BL21 (DE3) cells as described earlier⁵. Shortly, cultures inoculated from an overnight culture were initially allowed to grow at 37 °C and subsequently cooled down to 20 °C. At an OD₆₀₀ of approx. 1.2, cells were induced by 1 mM IPTG and protein was expressed overnight at 20 °C. Cells were harvested, washed, and stored frozen at –20 °C. For protein purification, cells were ruptured in presence of 10 µg/ml DNase in a cell disruptor (Constant Systems, UK), the lysis solution cleared by ultracentrifugation, and the supernatant loaded on a Strep-Tactin column (IBA). After washing with at least 10 column volumes, the bound protein was eluted with 2 mM des-thiobiotin in Tris buffered saline (TBS). The proteins were dialysed against TBS overnight and stored in aliquots at –80 °C.

Recombinant COMP and TSP-4 was expressed as described before^{12,44}. Essentially, COMP wt, D3A (Asp593Ala, Asp594Ala, Asp595Ala) and TSP-4 coding pCEP vectors were transfected into HEK293 (EBNA) cells and expressed under serum free conditions for 3 days. The supernatant was cleared by centrifugation and the protein purified via a Strep-Tactin column at room temperature. The protein was dialysed overnight against pure TBS and stored in aliquots at –80 °C.

CD Spectroscopy and Dynamic Light Scattering (DLS). CD spectra of the col2p5 peptide and its derivatives were measured using a Jasco J-715 equipped with a peltier-element cuvette cooler. The samples were heated at 1 °C/min in a 1 mm path length quartz cell at approx. 0.1 mg/ml. Thermal unfolding was measured in a range between 20 °C and 70–90 °C and specific ellipticity recorded at 210 nm every 1 °C. Additionally, every 5 °C far-UV spectra (200–250 nm) were recorded in triplicates. Data was fitted with a 4 parameter logistic fit using Origin 2017⁴⁵.

Dynamic light scattering (DLS) was measured with a Wyatt DynaPro Nanostar using the disposable cuvettes (UVettes, Eppendorf). Data was acquired as a mean of 5 acquisitions, measured each for 5 s at a fixed temperature of 20 °C. The laser was set to auto-attenuation and the data analysed with the Dynamics software from Wyatt. Graph show the cumulant radius versus time for COMP with indicated ZnCl₂ concentrations.

ELISA-Style Binding Assays. For ELISA-style binding assays, rat tail collagen I (BD Bioscience) or the indicated collagen model peptides were diluted in TBS and coated at 10 µg/ml (500 ng per well) overnight at 4 °C. Further steps were performed at room temperature. After washing with TBS-T (TBS containing 0.05% (vol/vol) Tween-20), plates were blocked for 1 h with 1% (wt/vol) BSA in TBS. Strep-tagged COMP was added at indicated

concentrations in TBS-T containing 0.25% (wt/vol) BSA) and incubated for 1 h. For assays with addition of cations pure TBS-T was used as interaction buffer. After washing away unbound protein with TBS-T, bound COMP was detected with a 1:3000 dilution of a specific polyclonal antibody against bovine COMP in interaction buffer. The antibody was raised in rabbit against bovine COMP purified from articular cartilage and characterised before in various studies^{46–48}. The TSP-4 antibody was raised in guinea pigs and was previously described and characterised^{30,49}. The primary antibodies were detected with an HRP-conjugated swine anti-rabbit (DAKO, P0399) or rabbit anti-guinea pig antibody (Sigma, A5545). Bound secondary antibodies were visualised with tetramethylbenzidine as substrate, the reaction was either measured directly at 370 nm or stopped with 10% H₂SO₄ and measured at 450 nm. Data were analysed using Origin 2017 and fitted using a four-parameter logistic model⁴⁵.

Crystallization and Structure Solution of the collagen model peptide and Docking. A collagen model peptide coding for the N-terminal binding site of COMP, flanked by 4 GPOs (Ac-POG POG POG LKG HRG FTG LQG POG POG POG POG-NH₂) were ordered from Peptide 2.0 (USA). The peptide was dissolved at approximately 250 μM in Water and mixed 1:2, 1:1 and 2:1 in various crystallisation screens using a Mosquito Nanodrop. Crystals were finally grown in 1.5 M Ammonium sulphate and 0.1 M BIS-TRIS propane pH 7.0 at 20 °C by vapour diffusion, cryoprotected with the addition of 20% glycerol to the mother liquor and flash frozen in liquid nitrogen. Data up to 1.4 Å were collected at beamline X06DA, Swiss Light Source (Paul Scherer Institute, Villigen, Switzerland). The data were processed in I2 using the program iMosflm⁵⁰ and the Pointless-Aimless-cTruncate^{51,52} pipeline from within CCP4i⁵³ and the structure solved by molecular replacement using PHASER⁵⁴, using a collagen model peptide solved earlier (PDB: 4AXY⁵⁴) as search model. The initial model was completely rebuilt manually using COOT⁵⁵ and refined using iterative cycles of model building and refinement with phenix.refine⁵⁶. The data collection and model refinement statistics can be found in Table 4. The model and diffraction data, was deposited at the PDB under the accession number 6HG7. Docking was performed using the ClusPro Server 2.0^{57,58} using parts of the model of the COMP's C-terminal domain (PDB Code 3FBY²³, residues Asp₅₃₀–Ala₇₅₇). The lysine, arginine, and histidine residue of the trailing chain and the lysine residue of the leading chain of the collagen model peptide were defined to be attractive in the algorithm. These four amino acids are accessible from one side of the helix and might, thereby constitute one binding interface. In the generated models, the isolated C-terminal domain was replaced by the full-length model (3FBY) to detect possible clashes of the collagen molecule with the TSP repeats.

References

- Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* **3** (2011).
- Kadler, K. E., Baldock, C., Bella, J. & Boot-Handford, R. P. Collagens at a glance. *J. Cell Sci.* **120**, 1955–1958 (2007).
- Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **78**, 929–958 (2009).
- Koide, T. & Nagata, K. Collagen Biosynthesis. In *Collagen* (eds Brinckmann, J., Notbohm, H. & Müller, P. K.) 85–114, doi:10.1007/b103820 (Springer Berlin Heidelberg, 2005).
- Oecal, S. *et al.* The pH-dependent Client Release from the Collagen-specific Chaperone HSP47 Is Triggered by a Tandem Histidine Pair. *J. Biol. Chem.* **291**, 12612–12626 (2016).
- 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).
- Du, C. *et al.* Improvement of thermostability of recombinant collagen-like protein by incorporating a foldon sequence. *Appl. Microbiol. Biotechnol.* **79**, 195–202 (2008).
- Pinkas, D. M., Ding, S., Raines, R. T. & Barron, A. E. Tunable, Post-translational Hydroxylation of Collagen Domains in *Escherichia coli*. *ACS Chem Biol* **6**, 320–324 (2011).
- Acharya, C. *et al.* Cartilage oligomeric matrix protein and its binding partners in the cartilage extracellular matrix: Interaction, regulation and role in chondrogenesis. *Matrix Biol.* **37**, 102–111 (2014).
- Rosenberg, K., Olsson, H., Mörgelin, M. & Heinegård, D. Cartilage oligomeric matrix protein shows high affinity zinc-dependent interaction with triple helical collagen. *J. Biol. Chem.* **273**, 20397–20403 (1998).
- Thur, J. *et al.* Mutations in Cartilage Oligomeric Matrix Protein Causing Pseudoachondroplasia and Multiple Epiphyseal Dysplasia Affect Binding of Calcium and Collagen I, II, and IX. *J. Biol. Chem.* **276**, 6083–6092 (2001).
- Agarwal, P. *et al.* Collagen XII and XIV, New Partners of Cartilage Oligomeric Matrix Protein in the Skin Extracellular Matrix Suprastructure. *J. Biol. Chem.* **287**, 22549–22559 (2012).
- Halász, K., Kassner, A., Mörgelin, M. & Heinegård, D. COMP Acts as a Catalyst in Collagen Fibrillogenesis. *J. Biol. Chem.* **282**, 31166–31173 (2007).
- Schulz, J.-N. *et al.* COMP-assisted collagen secretion – a novel intracellular function required for fibrosis. *J Cell Sci* **129**, 706–716 (2016).
- Natsume, T., Koide, T., Yokota, S., Hirayoshi, K. & Nagata, K. Interactions between collagen-binding stress protein HSP47 and collagen. Analysis of kinetic parameters by surface plasmon resonance biosensor. *J. Biol. Chem.* **269**, 31224–31228 (1994).
- Briggs, M. D., Brock, J., Ramsden, S. C. & Bell, P. A. Genotype to phenotype correlations in cartilage oligomeric matrix protein associated chondrodysplasias. *Eur. J. Hum. Genet.* **22**, 1278–1282 (2014).
- Hansen, U. *et al.* A secreted variant of cartilage oligomeric matrix protein carrying a chondrodysplasia-causing mutation (p.H587R) disrupts collagen fibrillogenesis. *Arthritis Rheum.* **63**, 159–167 (2011).
- Posey, K. L., Yang, Y., Veerisetty, A. C., Sharan, S. K. & Hecht, J. T. Thrombospondins: from structure to therapeutics. *Cell. Mol. Life Sci.* **65**, 687 (2008).
- Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *Plos one* **3**, e3647 (2008).
- Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *Plos one* **4**, e5553 (2009).
- Zwolaneck, D. *et al.* Collagen XXII binds to collagen-binding integrins via the novel motifs GLQGER and GFKGER. *Biochem. J.* **459**, 217–227 (2014).
- Narouz-Ott, L., Maurer, P., Nitsche, D. P., Smyth, N. & Paulsson, M. Thrombospondin-4 Binds Specifically to Both Collagenous and Non-collagenous Extracellular Matrix Proteins via Its C-terminal Domains. *J. Biol. Chem.* **275**, 37110–37117 (2000).
- Tan, K., Duquette, M., Joachimiak, A. & Lawler, J. The crystal structure of the signature domain of cartilage oligomeric matrix protein: implications for collagen, glycosaminoglycan and integrin binding. *FASEB J.* **23**, 2490–2501 (2009).

24. Holden, P. *et al.* Cartilage Oligomeric Matrix Protein Interacts with Type IX Collagen, and Disruptions to These Interactions Identify a Pathogenetic Mechanism in a Bone Dysplasia Family. *J. Biol. Chem.* **276**, 6046–6055 (2001).
25. Widmer, C. *et al.* Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci.*, <https://doi.org/10.1073/pnas.1208072109> (2012).
26. Carafoli, F., Hamaia, S. W., Bihan, D., Hohenester, E. & Farndale, R. W. An Activating Mutation Reveals a Second Binding Mode of the Integrin $\alpha 2$ I Domain to the GFOGER Motif in Collagens. *PLoS ONE* **8**, e69833 (2013).
27. Emsley, J., King, S. L., Bergelson, J. M. & Liddington, R. C. Crystal Structure of the I Domain from Integrin $\alpha 2\beta 1$. *J. Biol. Chem.* **272**, 28512–28517 (1997).
28. 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.* **109**, 5253–5258 (2012).
29. Cockburn, C. G. & Barnes, M. J. Characterization of Thrombospondin Binding to Collagen (Type I) Fibres: Role of Collagen Telopeptides. *Matrix* **11**, 168–176 (1991).
30. Södersten, E., Ekman, S., Schmitz, M., Paulsson, M. & Zaucke, F. Thrombospondin-4 and Cartilage Oligomeric Matrix Protein form Heterooligomers in Equine Tendon. *Connect. Tissue Res.* **47**, 85–91 (2006).
31. Kvensakul, M., Adams, J. C. & Hohenester, E. Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J.* **23**, 1223–1233 (2004).
32. 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**, ear2566 (2018).
33. Sekiya, A., Okano-Kosugi, H., Yamazaki, C. M. & Koide, T. Pigment Epithelium-derived Factor (PEDF) Shares Binding Sites in Collagen with Heparin/Heparan Sulfate Proteoglycans. *J. Biol. Chem.* **286**, 26364–26374 (2011).
34. Hanson, D. A. & Eyre, D. R. Molecular Site Specificity of Pyridinoline and Pyrrole Cross-links in Type I Collagen of Human Bone. *J. Biol. Chem.* **271**, 26508–26516 (1996).
35. Yamauchi, M. & Sricholpech, M. Lysine post-translational modifications of collagen. *Essays Biochem.* **52**, 113–133 (2012).
36. Okano-Kosugi, H. *et al.* Development of a high-throughput screening system for the compounds that inhibit collagen–protein interactions. *Anal. Biochem.* **394**, 125–131 (2009).
37. Ishida, Y. *et al.* Type I Collagen in Hsp47-null Cells Is Aggregated in Endoplasmic Reticulum and Deficient in N-Propeptide Processing and Fibrillogenesis. *Mol. Biol. Cell* **17**, 2346–2355 (2006).
38. Magdaleno, F. *et al.* Cartilage oligomeric matrix protein participates in the pathogenesis of liver fibrosis. *J. Hepatol.* <https://doi.org/10.1016/j.jhep.2016.06.003>.
39. Carlson, C. B. *et al.* Structure of the calcium-rich signature domain of human thrombospondin-2. *Nat. Struct. Mol. Biol.* **12**, 910–914 (2005).
40. Dokmanić, I., Šikić, M. & Tomić, S. Metals in proteins: correlation between the metal-ion type, coordination number and the amino acid residues involved in the coordination. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 257–263 (2008).
41. Adams, J. C., Bentley, A. A., Kvensakul, M., Hatherley, D. & Hohenester, E. Extracellular matrix retention of thrombospondin 1 is controlled by its conserved C-terminal region. *J. Cell Sci* **121**, 784–795 (2008).
42. Kim, D. J., Christofidou, E. D., Keene, D. R., Milde, M. H. & Adams, J. C. Intermolecular interactions of thrombospondins drive their accumulation in extracellular matrix. *Mol. Biol. Cell* **26**, 2640–2654 (2015).
43. Puigbò, P., Guzmán, E., Romeu, A. & Garcia-Vallvé, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* **35**, W126–W131 (2007).
44. Crosby, N. D. *et al.* Thrombospondin-4 and excitatory synaptogenesis promote spinal sensitization after painful mechanical joint injury. *Exp. Neurol.* **264**, 111–120 (2015).
45. Findlay, J. & Dillard, R. Appropriate calibration curve fitting in ligand binding assays. *AAPS J.* **9**, E260–E267 (2007).
46. Hedbom, E. *et al.* Cartilage matrix proteins. An acidic oligomeric protein (COMP) detected only in cartilage. *J. Biol. Chem.* **267**, 6132–6136 (1992).
47. DiCesare, P., Hauser, N., Lehman, D., Pasumarti, S. & Paulsson, M. Cartilage oligomeric matrix protein (COMP) is an abundant component of tendon. *FEBS Lett.* **354**, 237–240 (1994).
48. Blumbach, K., Niehoff, A., Paulsson, M. & Zaucke, F. Ablation of collagen IX and COMP disrupts epiphyseal cartilage architecture. *Matrix Biol.* **27**, 306–318 (2008).
49. Dunkle, E. T., Zaucke, F. & Clegg, D. O. Thrombospondin-4 and matrix three-dimensionality in axon outgrowth and adhesion in the developing retina. *Exp. Eye Res.* **84**, 707–717 (2007).
50. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
51. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
52. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011).
53. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* **67**, 235–242 (2011).
54. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
57. Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**, 45–50 (2004).
58. Kozakov, D. *et al.* How good is automated protein docking? *Proteins Struct. Funct. Bioinforma.* **81**, 2159–2166 (2013).

Acknowledgements

The work was partially funded by the Deutsche Forschungsgemeinschaft SFB829/B11 and the Deutsche Forschungsgemeinschaft - Forschungsgruppe 2722. We thank the staff of the beamline X06DA at the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland for their support during data collection and Uwe Hansen for fruitful discussions regarding collagen fibril formation. Crystals were grown in the Cologne Crystallization facility (<http://C2f.uni-koeln.de/>).

Author Contributions

H.D. expressed and purified the TSP-4. M.G. and J.M.G. did all the other cloning and protein purification. A.K. performed the binding studies with TSP-4 and collagen XI. J.M.G. performed all the other binding studies, the structural work, and designed the research. M.K. and F.Z. supplied valuable reagents. U.B., M.K., F.Z., I.N. and J.M.G. analysed the data and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-35447-8>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018