eLife

Reviewed Preprint
v1 • March 12, 2025
Not revised

Reviewed Preprint
v2 • November 28, 2025
Revised by authors

For correspondence:
Anna.Norrby-Teglund@ki.se
us6@st-andrews.ac.uk

*, #, †, €, ¶, &, ±, + **Author notes:** See page 32

Funding: See page 32

Reviewing editor: Wendy S Garrett, Harvard T.H. Chan School of Public Health, United States

Structural basis for collagen recognition by the *Streptococcus pyogenes* M3 protein and its involvement in biofilm

Marta Wojnowska[1,*,#], Takeaki Wajima[2,*,†], Tamas Yelland[3,€], Hannes Ludewig[1,¶], Robert M Hagan[1], Olivia F McCurry[1], Grant Watt[1], Samir W Hamaia[4], Dominique Bihan[4], Jean-Daniel Malcor[4,&], Arkadiusz Bonna[4,±], Helena Bergsten[2], Laura M Palma Medina[2], Mattias Svensson[2], Oddvar Oppegaard[5], Steinar Skrede[5], Per Arnell[6], Ole Hyldegaard[7], Richard W Farndale[4,±], Anna Norrby-Teglund[2,+] ✉, Ulrich Schwarz-Linek[1,+] ✉

[1]Biomedical Sciences Research Complex, University of St Andrews, St Andrews, United Kingdom • [2]Center for Infectious Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, Huddinge, Sweden • [3]CRUK Beatson Institute, Glasgow, United Kingdom • [4]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom • [5]Department of Medicine, Haukeland University Hospital, Department of Clinical Science, University of Bergen, Bergen, Norway • [6]Department of Anesthesiology and Intensive Care Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden • [7]Department of Anesthesiology, Hyperbaric Medicine Center, Head and Orthopedic Center, Copenhagen University Hospital, Rigshospitalet, Denmark and Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

eLife Assessment

M proteins are essential group A streptococci virulence factors that bind to numerous human proteins; a small subset of M proteins, such as M3, have been reported to bind collagen, which is thought to promote tissue adherence. In this **important** paper, the authors provide a **solid** characterization of M3 interactions with collagen. The work raises significant questions regarding the specificity of the structure and its interactions with different collagens, with implications for the variable actions of M protein collagen interactions on biofilm formation.

https://doi.org/10.7554/eLife.105539.2.sa3

Abstract

The M protein is an essential virulence factor of *Streptococcus pyogenes*, or group A streptococcus (GAS), one of the most common and dangerous human pathogens. Molecular and functional characterization of M protein variants and their interactions with host components is crucial for understanding streptococcal pathogenesis and vaccine development. The M3 protein is produced by the prevalent *emm*3 GAS serotype, which is frequently associated with severe invasive diseases. Here we characterize the interaction of M3 with human collagens through detailed structural and biochemical binding analysis. High-resolution structures of the N-terminal M3 domain in the free state as well as bound to a collagen peptide derived from the Collagen Ligands Collection reveal a novel T-shaped protein fold that presents binding sites complementing the characteristic topology of collagen triple helices. The structure of the M3/collagen peptide complex explains how *emm*3 GAS and related streptococci, such as the emerging human pathogen *Streptococcus dysgalactiae* subsp. *equisimilis*, can target collagens to enable colonization of various tissues. In line with this, we demonstrate that the M3/collagen interaction promotes enhanced biofilm formation of *emm*3 GAS in an *emm* type specific manner, which can be inhibited with the recombinant M3 N-terminal domain fragment. Further, *emm*3 GAS are shown to colocalize with collagen in tissue biopsies

from patients with necrotizing soft tissue infections, where GAS biofilms are common. This observation is reproduced in infected organotypic skin models. Together, these data provide detailed molecular insights into an important streptococcal virulence mechanism with implications for the understanding of invasive infections, strategies for treating biofilm and M-protein based vaccine design.

Introduction

Streptococcus pyogenes (group A streptococcus, GAS) is one of the most prevalent human pathogens. It causes acute diseases ranging from trivial to life-threatening, such as pharyngotonsillitis (strep throat), scarlet fever, impetigo, meningitis, necrotizing fasciitis and streptococcal toxic shock-like syndrome [1]. The highest disease burden is caused by the post-infection sequelae acute rheumatic fever, rheumatic heart disease and glomerulonephritis [2, 3] but there is also a worrying global rise of severe invasive infections [4]. A key virulence factor involved in many if not all these diseases is the GAS M protein. Encoded by the chromosomal *emm* gene, the M protein is covalently anchored at its C-terminus to the cell wall, covering the GAS surface at high density and acting as an adhesin and a potent immune evasion factor [5–7]. M proteins are known, or can confidently be predicted, to form linear fibrils extending approximately 50 nm from the bacterial surface [8, 9]. Their hair-like architecture is based on dimerization of α-helical chains into parallel coiled coils, the defining and dominant structural feature of this class of proteins. Whilst sharing similar overall structure, M proteins are sequentially highly diverse, with over 220 distinct known variants of the *emm* gene defining GAS serotypes [10]. The C-terminal regions of M protein variants are highly conserved and predicted to form well-defined coiled coils. Sequences diverge increasingly towards the N-terminal hypervariable region (HVR) that is projected away from the bacterial surface. Experimental high-resolution structural information for M proteins is limited [11–14].

In addition to their role in adhesion and subverting the host's immune response, M proteins have also been implicated in biofilm formation [15]. Bacterial biofilms are a major cause of difficult-to-treat infections largely contributed by their recalcitrance to antibiotics and immune responses [16]. They are classified as either surface-attached biofilms, typically associated with implants and medical device-associated infections, or non-surface-associated biofilms. The latter are commonly observed in respiratory infections of patients with impaired mucociliary function, and in persistent soft tissue infections seen in chronic wounds resulting from diabetes or impaired vascularization [17]. Biofilm is also a potentially complicating feature associated with severe invasive GAS diseases. Analyses of biopsies collected during the acute phase identified biofilm in over 30% of patients with necrotizing soft tissue infections caused by GAS [18]. Some GAS types, such as *emm*1 and *emm*3, are over-represented among isolates from severe invasive infections, i.e. necrotizing soft tissue infections and streptococcal toxic shock syndrome [1, 19, 20], but no clear association between biofilm formation and *emm* type is known [21]. Köller *et al.* highlighted that the propensity of GAS to form biofilm *in vitro* was directly dictated by the experimental setting where both the medium and coating with specific extracellular matrix proteins influenced the outcome [22].

M proteins have been reported to interact with a wide range of host proteins, such as fibrinogen [23], C4-binding protein [24], plasminogen [25], fibronectin [26], collagens [27] and immunoglobulins [28]. Importantly, binding activities vary between *emm* types. Phylogenetic clusters of M proteins broadly share binding propensities [10]. However, very few M protein interactions have been confirmed through biophysical and structural analyses. The M1 protein binds to the fibrinogen D domain in the variable "B-repeat" region [11]. Several M protein types have the ability to bind to the C4-binding protein through their HVRs [14]. While these two molecular complexes rely on the dimeric coiled-coil topology of the M proteins, the plasminogen kringle-2 domain binds to monomeric, non-coiled segments of certain M proteins [29, 30].

In a screening of GAS strains belonging to 43 *emm* types, only *emm3* and *emm18* strains were identified by Dinkla *et al.* to bind collagen IV to their surface [27]. For *emm3* isolates, this interaction was demonstrated to depend on direct binding of collagen to the M3 protein. For

emm18 GAS strains, binding was mediated by the hyaluronic acid capsule rather than the M18 protein. Direct collagen binding has also been reported for M1 protein [31]. Based on available evidence, collagen binding is not a universal property of GAS M proteins, but appears to be common among M proteins of group C and G streptococci (*Streptococcus dysgalactiae* subsp. *equisimilis*, SDSE) [32, 33], which are emerging as important human pathogens that share disease manifestations and virulence traits with GAS [34, 35].

Collagens are commonly targeted by a wide range of pathogenic bacteria for the purposes of tissue colonization and dissemination [36, 37]. They have been shown to interact directly with several bacterial adhesins, such as CNA from *Staphylococcus aureus* [38], YadA from *Yersinia enterocolitica* [39], and CNE from *Streptococcus equi* subsp. *equi* [40]. In the latter two cases, adhesin binding has been mapped using the peptide libraries applied in the present study. While the biological role of collagen binding by GAS is unknown, it has been linked to the induction of an autoimmune response associated with post-streptococcal sequelae [27, 41]. The collagen IV binding site in M3 was mapped to the N-terminal half of the protein [27]. Subsequently, a collagen-binding sequence motif that is conserved in the HVRs of some M proteins of group A, C and G streptococci was identified – the "peptide associated with rheumatic fever" (PARF) [33]. This motif is found in a region predicted to deviate from the canonical coiled coil structure in M3, several other M proteins and their homologs in non-group A streptococci (**Figure 1**). Binding of M3 protein to other collagen types has not been reported, but the M protein FOG (aka Stg11) of SDSE was found to bind to collagens I and IV [33, 42].

To gain insights into mechanism and biological role of streptococcal collagen binding, we characterised the interaction of M3 with the Collagen Ligand Collections (CLCs, triple-helical collagen peptide libraries formerly known as Toolkits), and determined crystal structures of the M3 N-terminal domain (M3-NTD), encompassing the HVR, alone and in complex with a CLC-derived collagen peptide. We find that M3-NTD folds into a novel T-shaped domain that binds promiscuously to collagen triple helices. Furthermore, we demonstrate that the M3-collagen interaction underpins biofilm formation by GAS *emm3* isolates from necrotizing soft tissue infections, which can be inhibited *in vitro* by recombinant M3 protein.

Results

Recombinant M3 protein binds to the triple-helical domain of collagens II and III

The first aim of this study was to confirm the collagen binding activity of streptococcal M3 protein, and to identify any specific binding sites in collagen. To reduce the complexity of the experimental system we chose collagens II and III, as these are homotrimers, in contrast to, for instance, heterotrimeric collagens I and IV [44]. Recombinant M3 protein, lacking the N-terminal secretion signal and the C-terminal membrane-spanning region was produced in *Escherichia coli* as a glutathione S-transferase (GST) fusion protein. Binding to collagens II and III was investigated using CLCs, libraries of 56 and 57 overlapping triple-helical peptides, respectively, covering the entire triple-helical tropocollagen domain of these two collagen types (1014 and 1029 amino acids) [45]. Recombinant M3 was found to bind to CLC peptides to various degrees, as quantified by an ELISA-like approach using an anti-GST antibody (**Figure 2**). While there were clear differences in binding between the CLC peptides, it was impossible to identify sequence features that were conserved in good binders (high apparent affinity) but absent from peptides that gave rise to signals comparable to negative controls. Neither sequence motifs, nor amino acid composition of the peptides (e.g., presence of hydroxyprolines, charged, hydrophobic or aromatic residues) were distinct in good binders. To investigate further the sequence requirements for M3 binding, we ranked the CLC peptides in decreasing apparent affinity, measured as A_{450} determined in the solid-phase binding assay, and then analyzed the distribution of particular amino acids in the entire set, pooling data from CLC-II and CLC-III. After background subtraction, we defined three binding groups: high affinity, having A_{450} between 0.5 and 0.75; medium affinity, A_{450} from 0.25 to 0.5; and low affinity, A_{450} from 0 to 0.25.

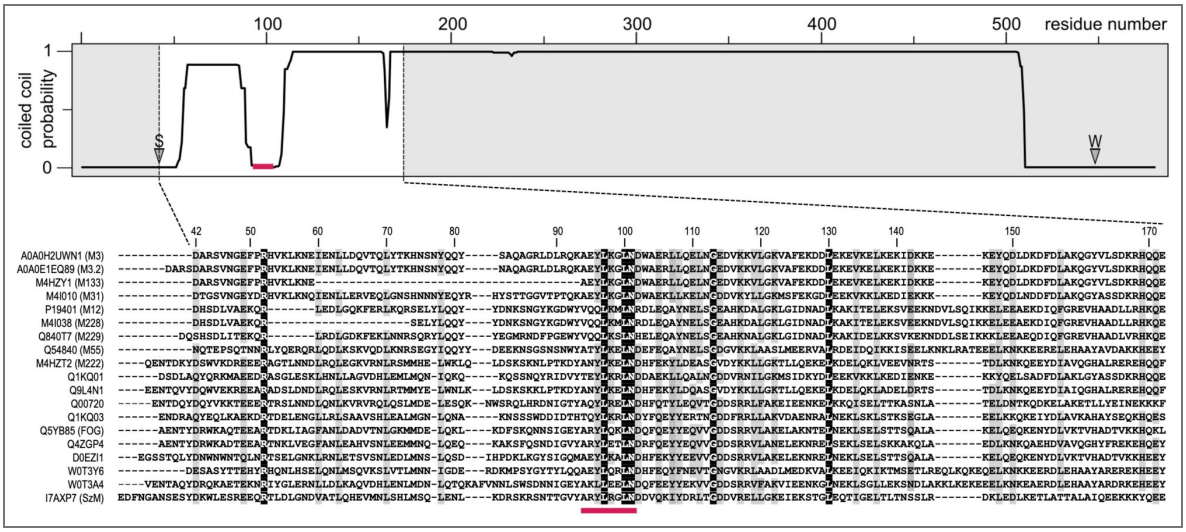


Figure 1. Multiple sequence alignment of the hypervariable regions of streptococcal M proteins.

The top panel shows a MARCOIL coiled coil prediction [43] for full-length M3 protein. Positions of signal and wall anchor cleavage sites are indicated by S and W, respectively. The PARF motif (position indicated by pink bar), previously suggested to be required for collagen binding, resides in a region of M3 that is predicted to not adopt coiled coil topology. In the alignment, sequence similarity is indicated by grey shading for residues conserved or similar in at least 75% of sequences. 100% conserved residues are highlighted by black shading. Uniprot identifiers are shown to the left of the alignment. Sequences are ordered by alignment score. The proteins are from the following species: *Streptococcus pyogenes*: A0A0H2UWN1 (M3), A0A0E1EQ89 (M3.2), M4HZY1 (M133), M4I010 (M31), P19401 (M12), M4I038 (M228), Q840T7 (M229), Q54840 (M55), M4HZT2 (M222); *Streptococcus dysgalactiae* subsp. *equisimilis*: Q1KQ01, Q9L4N1, Q00720, Q1KQ03, Q5YB85, Q4ZGP4, D0EZI1, W0T3Y6, W0T3A4; *Streptococcus equi* subsp. *zooepidemicus*: I7AXP7.

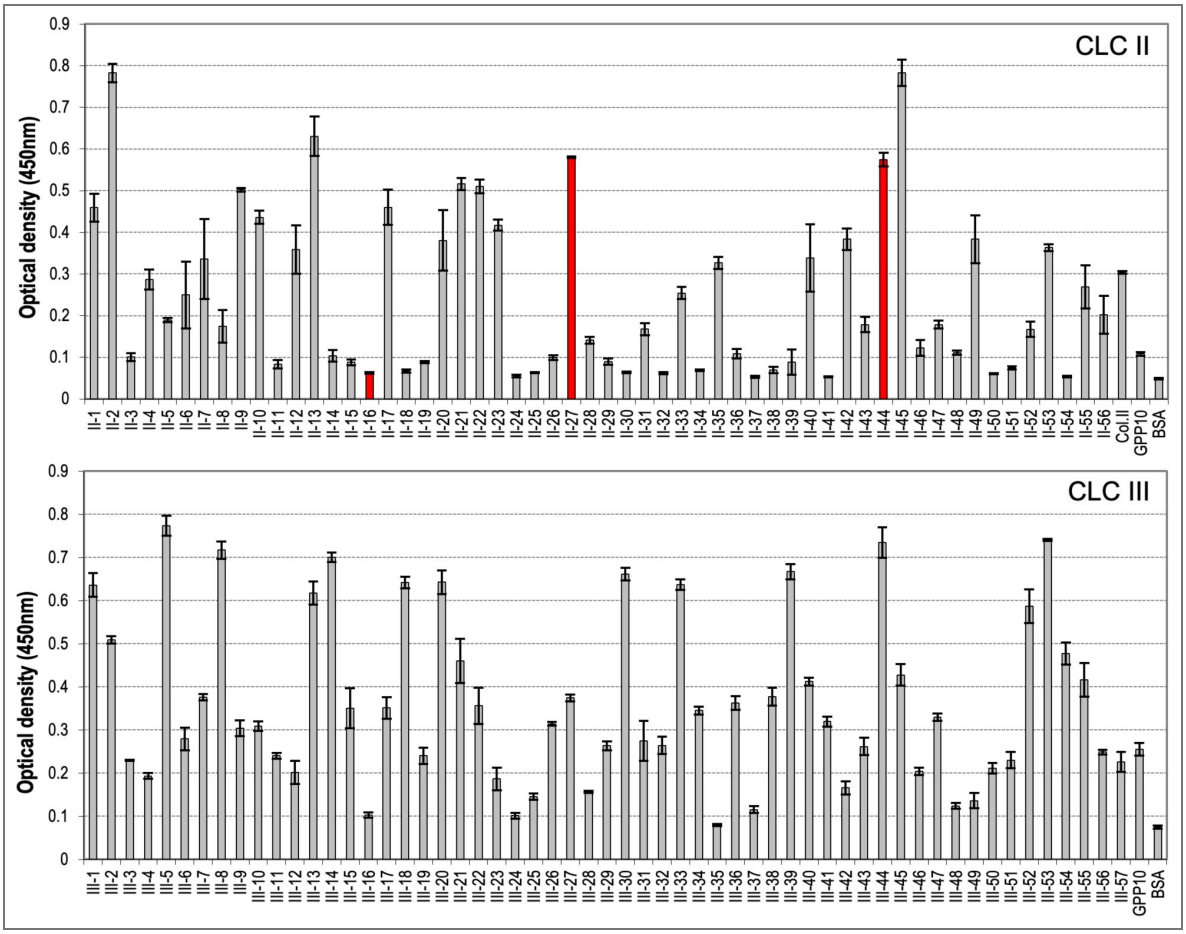


Figure 2. Binding of full-length M3 protein fused to GST to immobilised CLC II and III peptides.

ELISA signal based on GST antibody is plotted against peptides spanning the tropocollagen (triple-helical) regions of collagens II and III, as well as positive (collagen II) and negative (GPP10, BSA) controls. Signals for three peptides chosen for further binding experiments are highlighted in red.

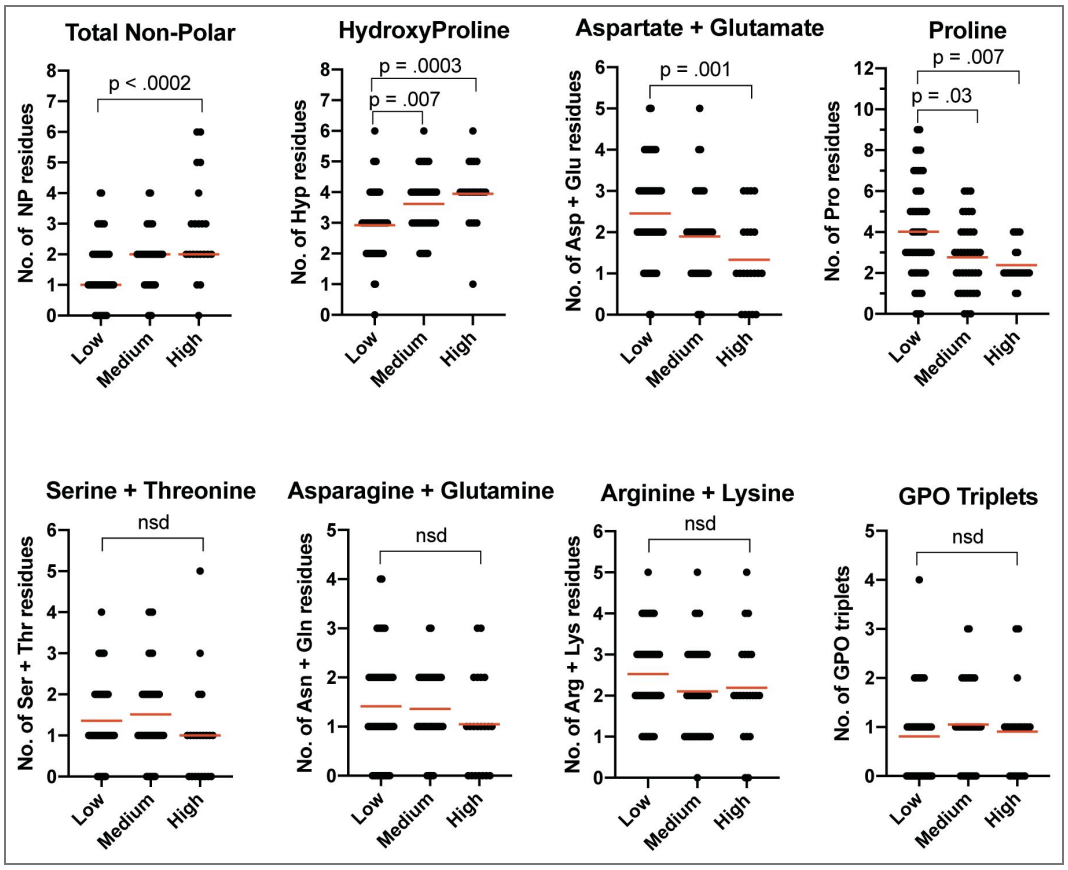


Figure 2–figure supplement 1. The Figure shows the number of specific residues or residue classes per CLC peptide (y-axis) plotted versus rank group (x-axis).

Groups are defined as low affinity (A450 = 0 to 0.25), medium affinity (A450 = 0.25 to 0.5) and high affinity (A450 = 0.5 to 0.5) in the solid phase binding assays. Statistical difference between numbers in each group was determined by non-parametric testing as defined in Methods.

We counted residues of each type in these three groups and compared their occurrence in peptides of the three binding groups using non-parametric tests (**Figure 2—figure supplement 1**). The outcomes are summarized in **Table 1**. Hydrophobic amino acids and hydroxyprolines appear to be more frequent in good binders, while prolines and acidic residues are underrepresented. The negative effect of proline on binding may explain why the $[GPP]_5$ flanking sequences of the CLC peptides do not dominate binding to M3, allowing marked sequence selectivity to be observed in the binding assays.

M3-NTD harbors the collagen binding site

Full-length M proteins are not amenable to high-resolution structural characterization due to their anisotropic shape and potential conformational dynamics. We therefore designed a construct representing the M3 N-terminal domain (M3-NTD), which comprises the HVR and a short region predicted to adopt a dimeric coiled-coil structure. This fragment would not only allow structural characterization but could also be used to confirm the previously suggested localization of the collagen binding site at the HVR [33]. M3-NTD includes the 110 N-terminal residues of mature M3 (residues 42-151 of the M3 protein precursor sequence). To stabilize a dimeric conformation, Leu151 was replaced with a cysteine for disulfide bond formation at the C-terminus. Leu151 is predicted to occupy a “d” position in the canonical coiled coil heptad pattern, forming part of the hydrophobic interface between the α-helices, a position ideally suited for disulfide bond formation [46]. A ^{15}N isotopically labelled version of M3-NTD was made for NMR spectroscopic analysis. From a comparison of ^{1}H, ^{15}N heteronuclear single-quantum coherence (HSQC) spectra it is evident that a significant structural change occurred upon oxidation (resulting in disulfide bond formation) of the protein. The higher dispersion of signals, most noticeably in the ^{1}H dimension, indicates disulfide-linked M3-NTD adopts a folded conformation (**Figure 3A**). In contrast, the reduced, monomeric form gives rise to a poorly resolved spectrum with signals falling within the random-coil chemical shift range, and big differences in crosspeak intensities that suggest dynamic behavior (**Figure 3B**). This demonstrates the C-terminal disulfide bond is required to stabilize a dimeric, folded structure that differs from extended unfolded or linear α-helical conformations.

To test if M3-NTD harbors the collagen binding site, binding to selected CLC triple-helical peptides was studied by isothermal titration calorimetry (ITC). We selected two peptides that showed medium to high apparent affinity in the CLC screening (II-27 and II-44) and a low affinity peptide (II-16). These three peptides all interacted with M3-NTD in solution (**Figure 4A-C**). II-27 and II-44 binding was characterized by dissociation constants (K_D) in the low micromolar range (7 and 5 μM, respectively). II-16 had an approximately ten times lower affinity (K_D = 70 μM) for M3-NTD. These K_Ds reflect the differences in binding of full-length M3 evident from the CLC solid-phase binding assay. Fitting of the sigmoidal binding curves for II-27 and II-44 suggested two indistinguishable collagen binding sites per M3 dimer. To assess if the interaction was dependent on the triple-helical conformation of the CLC peptides, we titrated M3-NTD into a monomeric version of II-44 with scrambled GPP repeat sequences at the termini [47]. No binding was observed (**Figure 4D**). To test if collagen peptide binding required the M3-NTD to adopt its dimeric fold, the reduced (monomeric) form of M3-NTD was titrated into trimeric II-44 (**Fig. 4E**). The titration curve reflected weak binding, with data not suitable for fitting. Weak binding is in line with the observation that peptides representing the PARF motif in collagen-binding M proteins have some ability to bind to collagen IV in a spot-membrane assay [33].

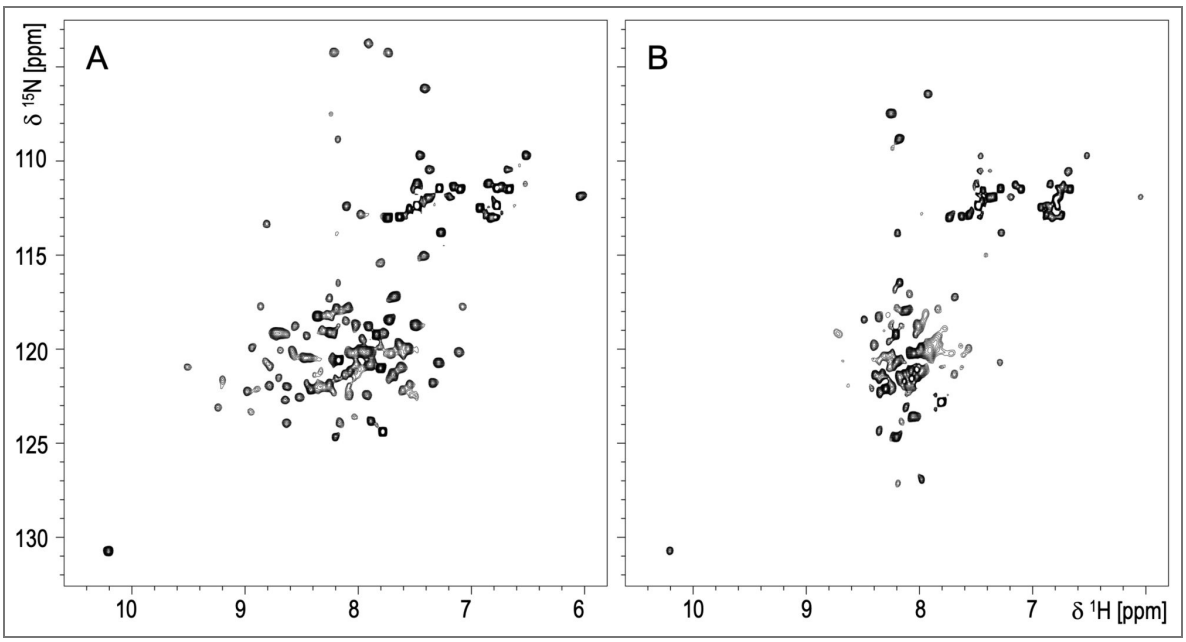
In conclusion, the collagen binding site of M3 was confirmed to reside in M3-NTD. CLC solid-phase binding results obtained for full-length M3 align with in-solution interaction analyses by ITC using M3-NTD. Binding of M3-NTD to collagen peptides depends on their triple-helical structure, and on the dimeric conformation of M3-NTD.

Table 1. **Effect of amino acid classes on binding of M3 to CLCs.**

Amino Acid class	Effect on binding rank	Significance
Hydrophobic (F, L, V, M, I)	Positive	p = .0002
Hydroxyproline (O)	Positive	p = .0003
Polar (S, T)	None	
Polar (Q, N)	None	
Basic (R, K)	None	
GPO triplets	None	
Acidic (D, E)	Negative	p = .001
Proline (P)	Negative	p = .0007

Figure 3.

$^{1}H,^{15}N$ HSQC spectra for (A) disulfide-bond stabilized dimeric M3-NTD and (B) the reduced, monomeric form.



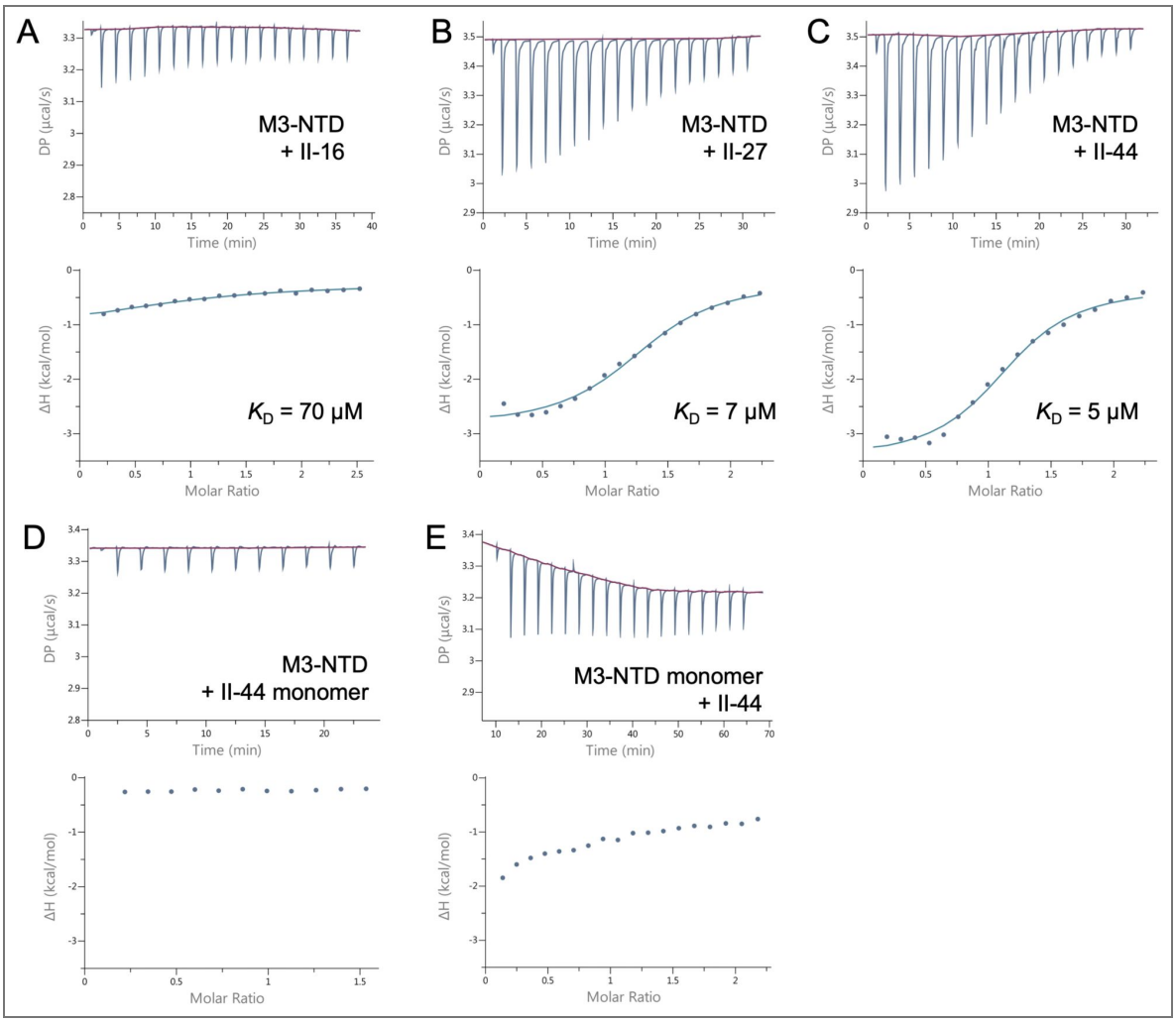


Figure 4. ITC binding curves for M3-NTD interactions with CLC peptides.

Top panels show heat responses to repeated injections of M3-NTD into collagen peptide solutions, with baselines shown in red. Bottom panels shown integrated signals (enthalpy changes) plotted against the molar ratio of binding partners. Where non-linear fitting gave meaningful results, the fits are shown as blue lines, and dissociation constants (K_D) are specified.

M3-NTD adopts a folded structure deviating from dimeric coiled coil

M3-NTD crystallized in several conditions with the best crystals diffracting X-rays to a resolution of 1.9 Å at the synchrotron radiation source. Phasing and structure determination was achieved using a selenium single-wavelength anomalous diffraction (Se-SAD) dataset at 2.6 Å resolution with the final model refined against the native data set at 1.9 Å resolution (**Table 2**).

M3-NTD is a symmetrical homodimer, linked at the C-terminus by a disulfide bond (**Figure 5A**). The first three residues of the construct (Gly-Ala-Met), a cloning artefact, are not resolved in the structure.

The fold represents a novel T-shaped architecture with no significantly similar structures identifiable by the protein structure comparison server DALI [48]. Each monomer is composed of three α-helices, H1-H3. The H3 helices comprise the C-terminal 38 residues of M3-NTD (M3 residues 114-151) and form a coiled coil stem. This, in the full-length protein, would be extended into a ~50 nm long coiled coil characteristic of M proteins (**Figure 5B**). Helices H1 and H2 form hairpins that pack against each other to form the slightly kinked bar of the T-shape, effectively a three-helix bundle. A key residue is Gly113, which resides at the T-junction, separating H2 and H3 (**Figure 5C**). It is completely conserved in M proteins identified in our sequence similarity search (**Figure 1**). The T-junction structure is defined by conserved leucine and isoleucine residues that form a small hydrophobic core (**Figure 5C**). The T-bar structure is stabilized by a network of polar interactions of several conserved residues, most notably inter-chain salt bridges of Arg52 with Asp102 and Glu105, and an inter-chain hydrogen bond between the side chains of Glu49 and Asn101 (**Figure 5D**). Asn101 is one of the completely conserved residues in the PARF motif, previously implicated in collagen binding. This motif is located on H2 at the bottom of the T-bar (**Figure 5A**). Some but not all conserved PARF residues contribute to the formation of the T-bar structure of M3-NTD. Leu97 and Asn101 are the only completely conserved residue of PARF with a structural role. Leu97 is buried at the interface between the H1-H2 hairpin of one monomer and H1 of the other monomer, packing against the Tyr81 sidechain (**Figure 5D**).

An N-terminal T-shaped fold is predicted to be a feature of other M proteins

AlphaFold3 [49] predictions were carried out for the proteins identified in our sequence similarity search (**Figure 1**) to support the structural role of the conserved residues. In validation of this approach, the AlphaFold3 predicted model of M3 is almost perfectly superimposable with the experimental M3-NTD structure (**Figure 6A**). All other proteins included in this study are predicted to have T-shaped N-terminal domains, with exception of proteins M133 and M228, where deletions break the topology of the T-bar fold (**Figure 6B**, **Figure 6—figure supplement 1**). Some of the predictions for GAS M protein variants (M12, M55, M229) carry low confidence and these proteins may well not adopt the T-fold. HVRs of M proteins that are phylogenetically more distant to M3, such as M1 and M28, are not predicted to fold into any distinct tertiary structure deviating from coiled coil (**Figure 6C**). On the other hand, M proteins of SDSE and the SzM protein of *Streptococcus equi* subsp. *zooepidemicus* (SESZ) are predicted with high confidence to adopt a structure similar to M3. This analysis structurally validates our sequence alignment of M3 homologs. It suggests the T-shaped structure of the NTD is a feature of a subclass of M proteins and is common in SDSE.

M3-NTD binds promiscuously to the collagen triple helix

M3-NTD is a structurally tractable fragment that retains the collagen binding activity of the full-length protein. We chose II-27, a CLC peptide with good affinity for M3-NTD, as the basis for structural characterization of an M3-collagen complex. II-27 contains an atypical α1β1 integrin-selective motif, GVOGEA [50]. To increase chances of crystallization, a shorter 24-residue peptide, JDM238, was synthesized which contains the GVOGEA motif flanked by three glycine-proline-4-

	M3-NTD	SeMet M3-NTD	M3-NTD+JDM238
Data collection	**8P6K**		**8P6J**
Spacegroup	C2	C2	P1
Cell dimensions			
a, b, c (Å)	150.5, 24.1, 78.0	150.4, 25.1, 77.0	31.8, 51.5, 80.2
a, b, c (°)	90, 105.3, 90	90, 104.4, 90	87.3, 84.8, 89.4
Resolution (Å)	37.62 - 1.92 (1.94 - 1.92)	37.3 – 2.67 (2.71 – 2.67)	79.77 - 2.32 (2.45 - 2.32)
R_{meas}	0.058	0.198	0.107
I/σ	14.8	10.2	2.1
$CC_{1/2}$	0.9 (0.5)	1 (0.9)	1 (0.3)
Completeness (%)	97.80 (94.07)	98.8	89.58 (89.66)
Multiplicity	6.2	18.6	1.7
Refinement statistics			
Resolution (Å)	37.6 - 1.92		79.77 - 2.32
No. reflections	21411 (1054)		19530 (947)
R_{work}/R_{free}	0.230/0.268		0.241/0.285
Ramachandran favoured (%)	97.7		97.7
Ramachandran outliers (%)	0		0.5
RMS deviations			
Bond lengths (Å)	0.039		0.013
Bond angles (°)	2.06		1.89

Table 2. **Data collection and refinement statistics.**

Values in brackets refer to the highest resolution shell.

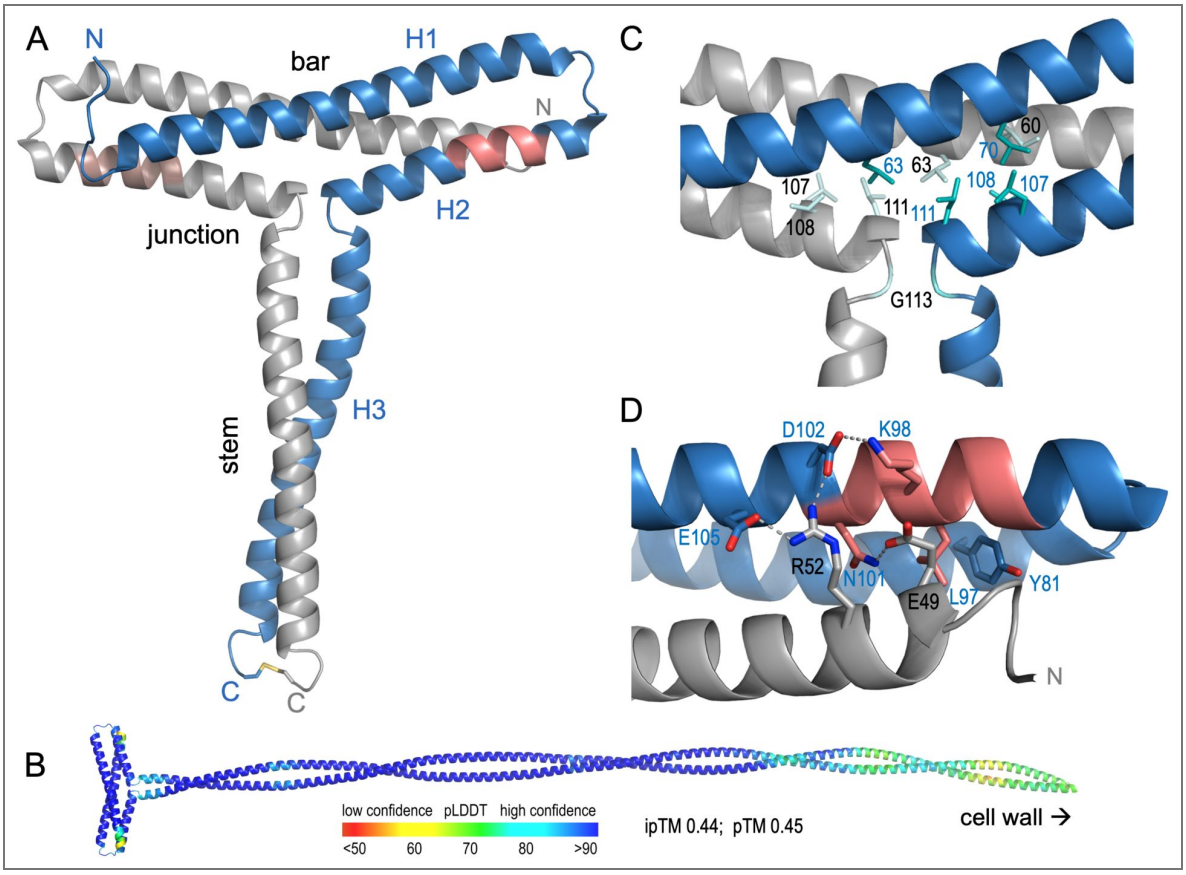


Figure 5. **Crystal structure of M3-NTD (PDB 8p6k).**

A) Ribbon diagram of the covalently stabilized M3-NTD dimer. Monomers are shown in blue and grey. The regions previously associated with collagen binding (PARF motif) are highlighted in red. N-termini (Asp42) and C-termini (Cys151) are labelled N and C, respectively. The three helices of the blue monomer are labelled H1-3. The C-terminal disulfide bond is shown as sticks. B) Ribbon diagram model for the full extracellular region of M3, predicted by AlphaFold3 [49] and colored by per-residue confidence (pLDDT). C) Conserved leucine, isoleucine (Ile60) and glycine (Gly103) residues define the T-junction structure of M3-NTD. D) Role of conserved residues around the PARF motif in stabilizing the T-bar region of M3-NTD. Polar contacts are shown as dashed lines.

hydroxyproline (GPO) repeats at each terminus. Crystals formed in conditions containing JDM238 and M3-NTD at a 3:2 molar ratio and diffracted synchrotron X-rays to 2.3 Å resolution. Molecular replacement with M3-NTD yielded a high-quality structural model for the complex (**Table 2**).

In agreement with our ITC data for CLC II-27 binding to M3-NTD, the M3-NTD dimer is bound to two copies of triple-helical JDM238. While both copies occupy equivalent binding sites on opposite faces of the M3-NTD T-bar, they contact the protein with two different regions (**Figure 7**). One copy contacts M3-NTD through residues 15-21 of the C-terminal GPO repeats, the other is bound at its more central GPOGVO region (residues 6-12) (**Figures 7** and **8A**). The presence of two different binding registers in the same crystal is likely an effect of crystal packing, as this binding mode results in a compact assembly required for crystallization (**Figure 7—figure supplement 1**), but it also reflects a lack of specificity of collagen binding by M3, in line with the CLC screening data. Superposition of the two binding sites highlights how the M3-NTD T-bar complements the characteristic and uniform surface topology of triple helices in both binding registers in the same fashion (**Figure 8B**).

The M3 collagen binding site includes the PARF region on H2 (residues 94-101), extending beyond it to include the C-terminal half of H1 and most of H2 (**Figure 8**). This confirms previous reports on the involvement of PARF in collagen binding [32, 33]. Interfaces between the binding partners largely comprise hydrophobic interactions between highly complementary surfaces. M3 residues Tyr96 and Trp103 play prominent roles: they pack against Hyp and Pro sidechains, filling grooves on the collagen triple helices and form polar contacts with the peptide backbone. The indole amino group of Trp103 interacts with the carbonyl of Ala15 or Hyp6 (depending on the register), while Tyr96 forms a water-mediated hydrogen bond with the carbonyl of Hyp9 or Hyp18. In one binding mode, M3 Gln42 hydrogen bonds with the same water molecule as Tyr96. Direct polar contacts also include M3 residues Lys35 and Arg49 which interact with the hydroxyl group of Hyp9 and the backbone carbonyl of Hyp12, respectively. In addition, in one binding mode M3 residue Gln46 forms a hydrogen bond with Gly22 of the peptide via a water molecule.

Two M3 residues are required for collagen peptide binding

Tyr96 is a conserved residue of the PARF motif. We tested its role in collagen binding using two site-specific variants, Tyr96Ala and Tyr96Phe. When titrating M3-NTD Tyr96Ala into CLC peptide II-44, which bound to wild type M3-NTD with low micromolar K_D, no binding was observed (**Figure 8— figure supplement 2**). The more conservative substitution of Tyr96 with Phe also greatly reduced the affinity for II-44 (K_D ~200 μM) (**Figure 4F**), suggesting that the water-mediated interaction between Tyr96 and the peptide backbone strongly contributes to binding.

Similar results were obtained for Trp103, the second prominent residue in the M3/collagen interface. Mutation to an Ala led to a loss of binding as detected by ITC. Replacement of Trp by the large hydrophobic Ile resulted in detectable but weak binding, with data not suitable for non-linear fitting to derive meaningful thermodynamic parameters.

To rule out any effect of mutations of Tyr96 and Trp103 on the structure of M3-NTD, ^{1}H NMR spectra were recorded (**Figure 8—figure supplement 3**). The spectra all showed very similar dispersion of resonances, reflecting a folded state with some hydrophobic core giving rise to resonances below 0.5 ppm. The only readily assignable resonance, due to its unique chemical shift, is the indole NH of Trp103, which, as expected, is lacking in the Trp mutants, but found at almost exactly the same chemical shift in M3-NTD and the Tyr96 mutants. These data support the loss of affinity of collagen peptides for the Tyr96 and Trp103 mutants being attributed to contribution of these residues to the binding interface rather than a loss of structure, as would be expected from the crystal structures. This supports the critical role of Tyr96 and Trp103 in collagen binding, both via forming van der Waals interactions with the collagen peptide triple helix as well as forming direct and water-mediated polar interactions with the peptide backbone.

In conclusion, we present the first structural evidence for an M-protein collagen complex. The structure indicates a general binding mechanism that might explain the promiscuity of M3 for diverse CLC peptides and explains how M3 is able to bind non-selectively to different collagen

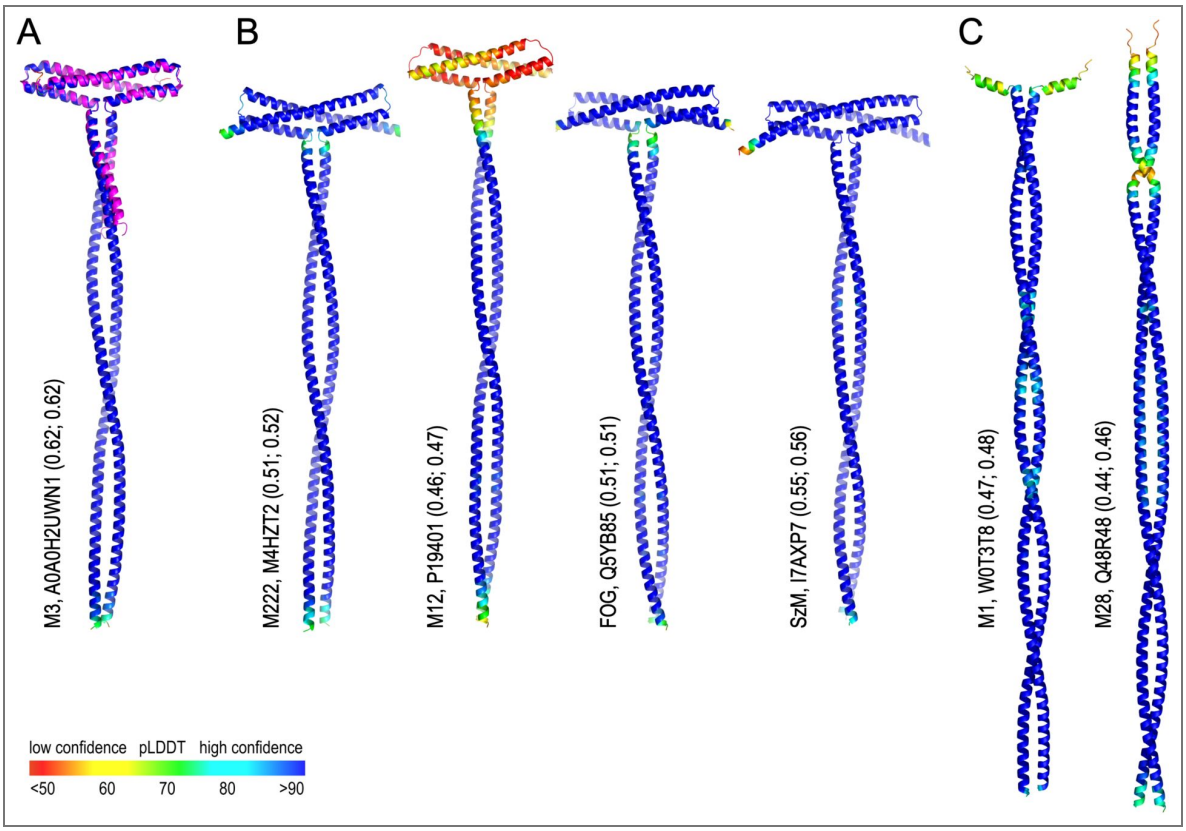


Figure 6. AlphaFold3 predictions for M proteins of GAS, SDSE and SESZ.

A) Overlay of the experimental M3-NTD structure (magenta) with a predicted structure for the N-terminal 230 residues of mature M3 (coloured by pLDDT). B) Predicted structures for the N-terminal 230 residues of other M proteins included in the sequence alignment (**Figure 1**). C) Predicted structures for the N-terminal 230 residues of M1 and M28 proteins, which are not known to interact with collagens. Values in parentheses are AlphaFold3 prediction quality parameters ipTM and pTM, respectively.

Figure 6–figure supplement 1. Additional AlphaFold3 predicted structures of proteins included in the sequence alignment (Figure 1).

The N-terminal 230 residues of mature proteins were included in predictions (colored by pLDDT). For M133 only the first 196 residues have been modelled to highlight of the effect of the deletion of residues that are integral to the T-shape structure of M3. Values in parentheses are AlphaFold3 prediction quality parameters ipTM and pTM, respectively.

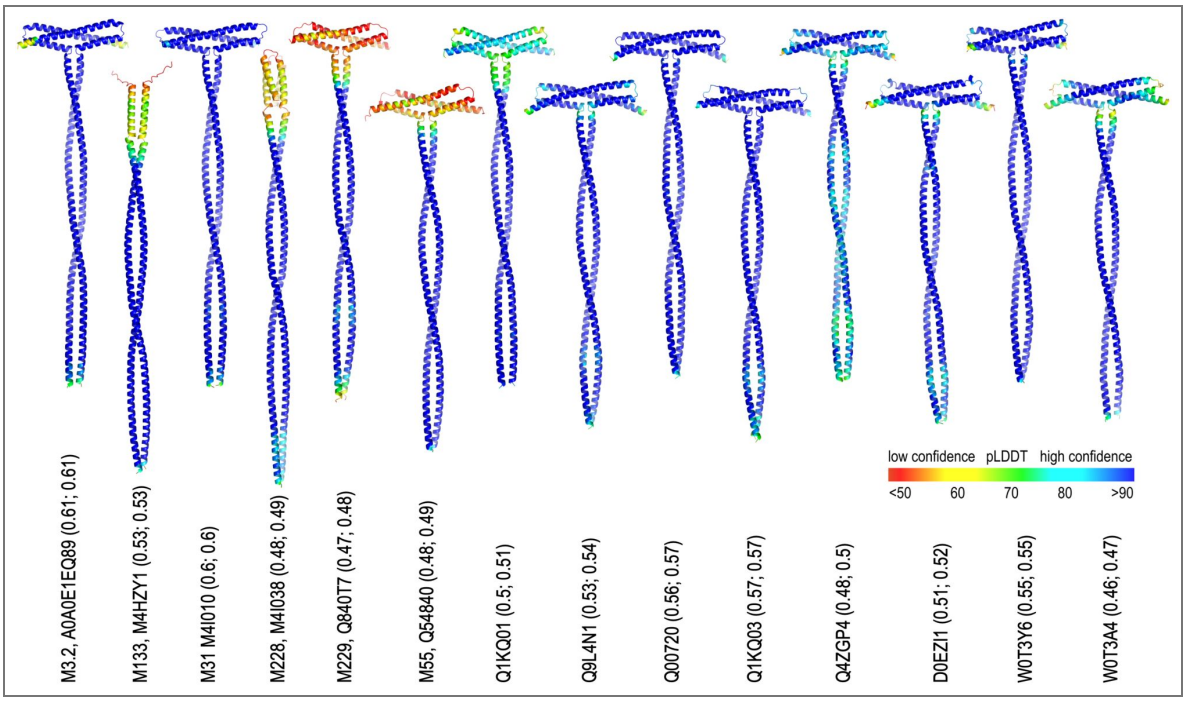
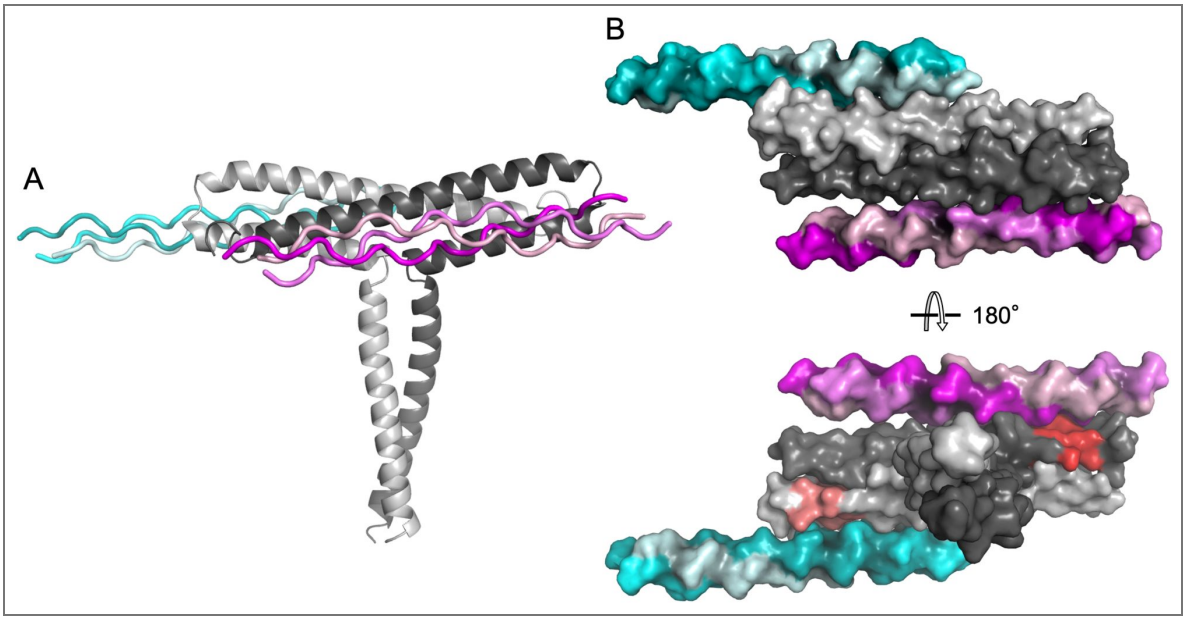


Figure 7. Structural basis of collagen binding by M3 (PDB 8p6j).

A) Crystal structure of the complex of M3-NTD (subunits shown in light and dark grey cartoon representation) with collagen-derived peptide JDM238. Two binding registers are observed bound to equivalent sites of the M3-NTD dimer, indicated by shades of magenta and cyan. B) Space-filling models of views from the top of the T-bar of M3-NTD (top) and looking down the stem towards the bottom of the T-bar (bottom). The PARF motif is shown in tones of red, otherwise coloring is as in A.



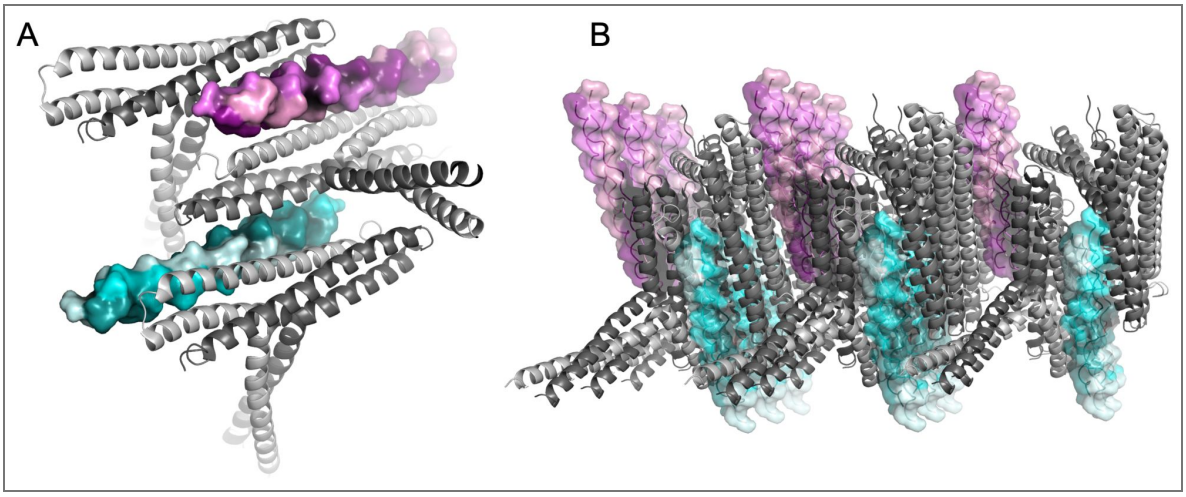


Figure 7–figure supplement 1. Crystal packing.

(A) Each triple-helical peptide (shades of magenta and cyan) is bound to two copies of M3-NTD (grey ribbons) at two distinct sites. (B) This results in the compact assembly observed in the crystal.

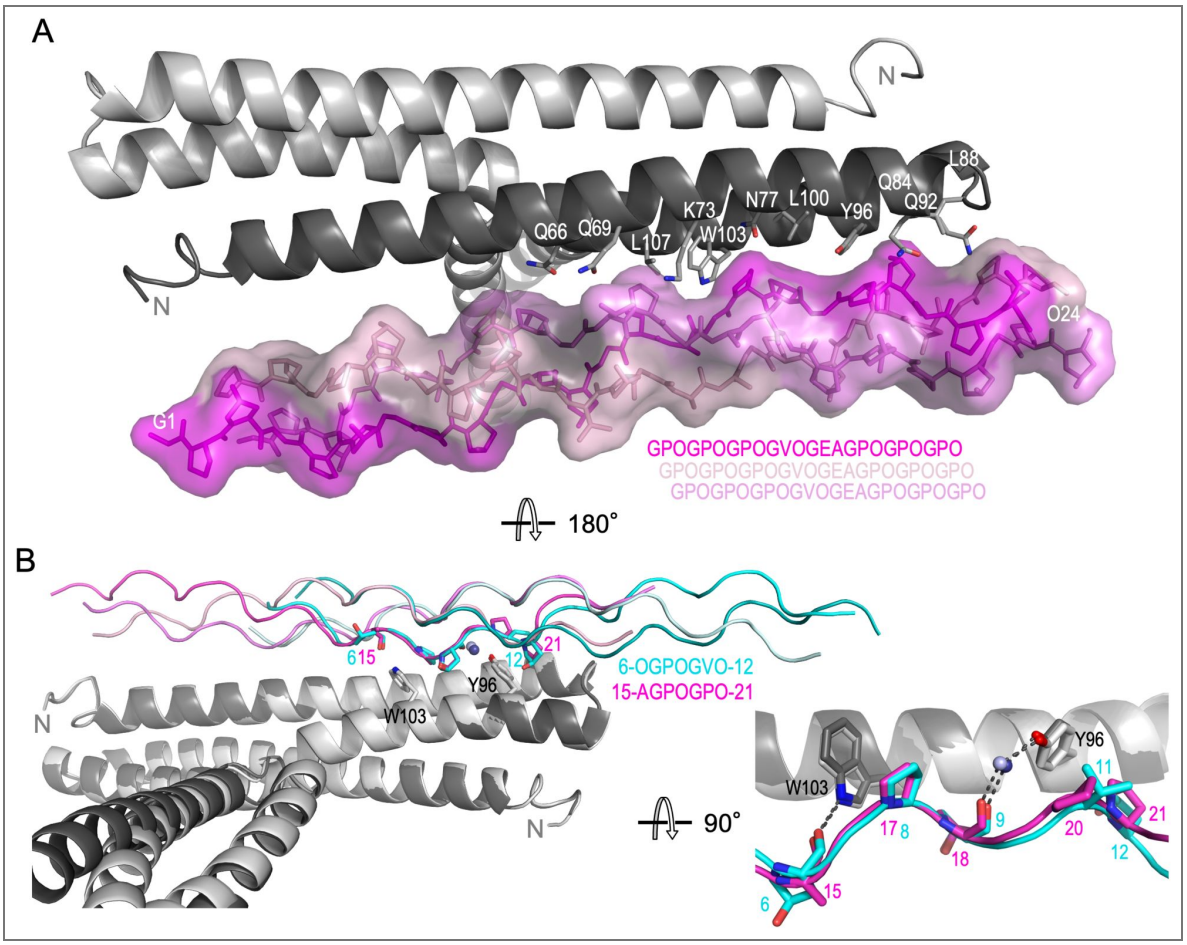


Figure 8. Collagen triple helix/M3-NTD interface.

A) Residues of M3-NTD (grey cartoon representation) that form the collagen binding site are shown as sticks and are labelled. The collagen peptide triple helix is shown in stick and surface representation in shades of magenta. The sequence is shown indicating the staggered arrangement of the chains in the triple helix. B) Superposition of the two peptide binding sites to highlight conservation of collagen binding mode despite sequence deviation between the two binding registers. Tyr96 and Trp103 are shown as sticks. The two monomers of M3-NTD are shown in light and dark shades of grey. Collagen peptides of the two binding modes are shown in magenta and cyan, with five equivalent sidechains shown as sticks. The equivalent sequences of the binding sites are shown. Water molecules in the binding interface are shown as blue spheres. In the zoomed-in image on the bottom right, only one of the collagen peptide chains is shown per complex for clarity. Hydrogen bonds are shown as grey dashed lines.

Figure 8–figure supplement 1. AlphaFold3 predicted structures of M proteins in complex with 24-residue collagen peptides.

(A) Predicted complexes with the model collagen peptide (GPO_8), representing the generic tropocollagen triple-helical structure. Two chains of the N-terminal 230 residues of mature proteins and, for clarity, only three collagen peptide chains were included in predictions. M proteins are shown in cartoon, GPO_8 peptide in surface representations, respectively (coloured by pLDDT). Values in parentheses are AlphaFold3 prediction quality parameters ipTM and pTM, respectively. (B) Predicted structure of an equivalent complex between M3 and a type I collagen peptide. (C) Closeup view of the predicted binding site of the type I collagen peptide (shades of orange) and M3 (shades of grey). Key residues, equivalent to the ones shown in **Figure 8B**, are shown in stick representation. The sequences of the hetero-triple-helical peptide (two α1 chains and one α2 chain), derived from PDB 5CTD, are also shown.

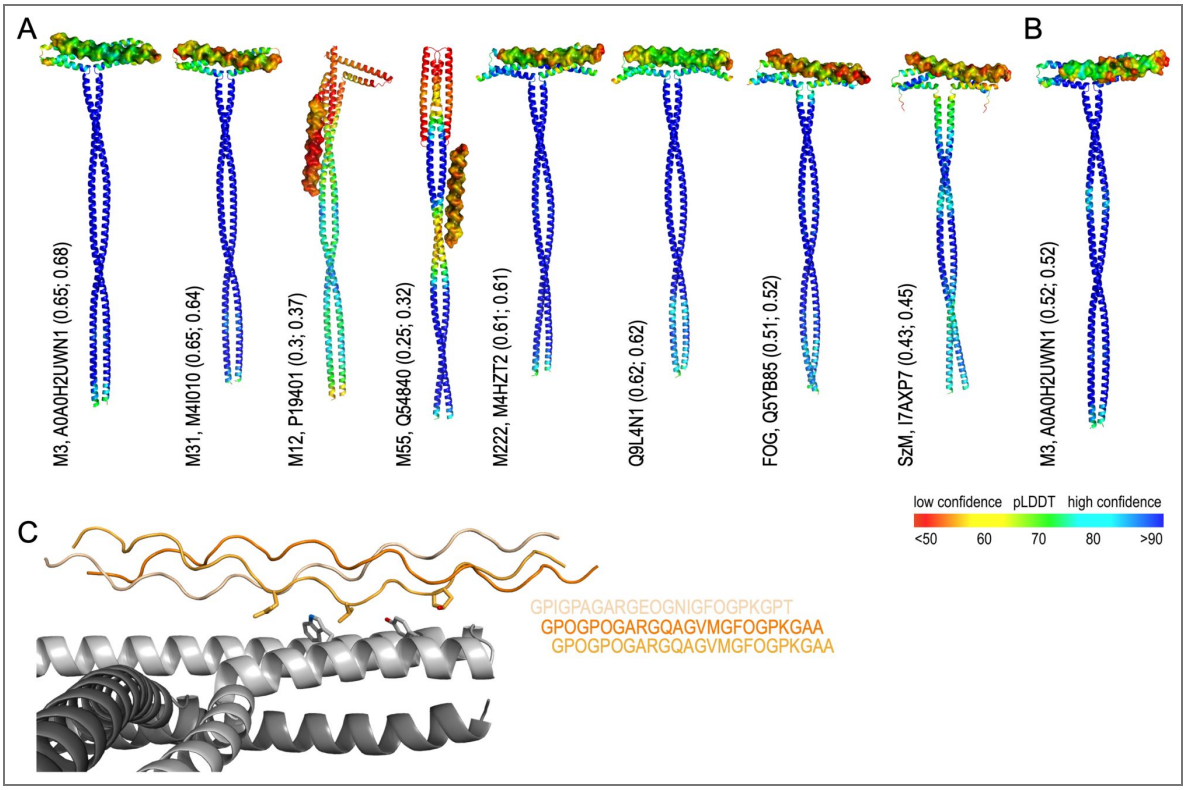
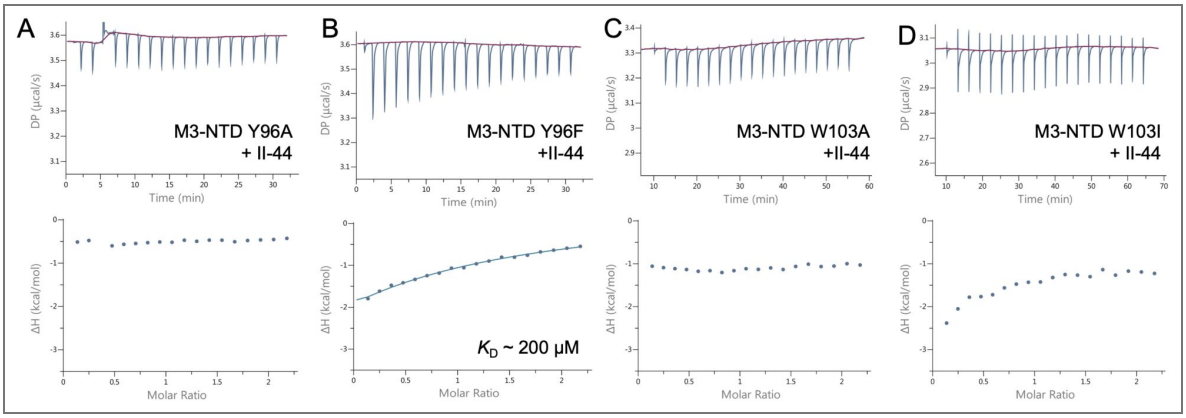


Figure 8–figure supplement 2. ITC binding curves for M3-NTD variants with peptide II-44.

Top panels show heat responses to repeated injections of M3-NTD into II-44 solutions, with baselines shown in red. Bottom panels shown integrated signals (enthalpy changes) plotted against the molar ratio of binding partners. For M3-NTD Y96F the fit is shown as a blue line, and dissociation constant (K_D) specified.



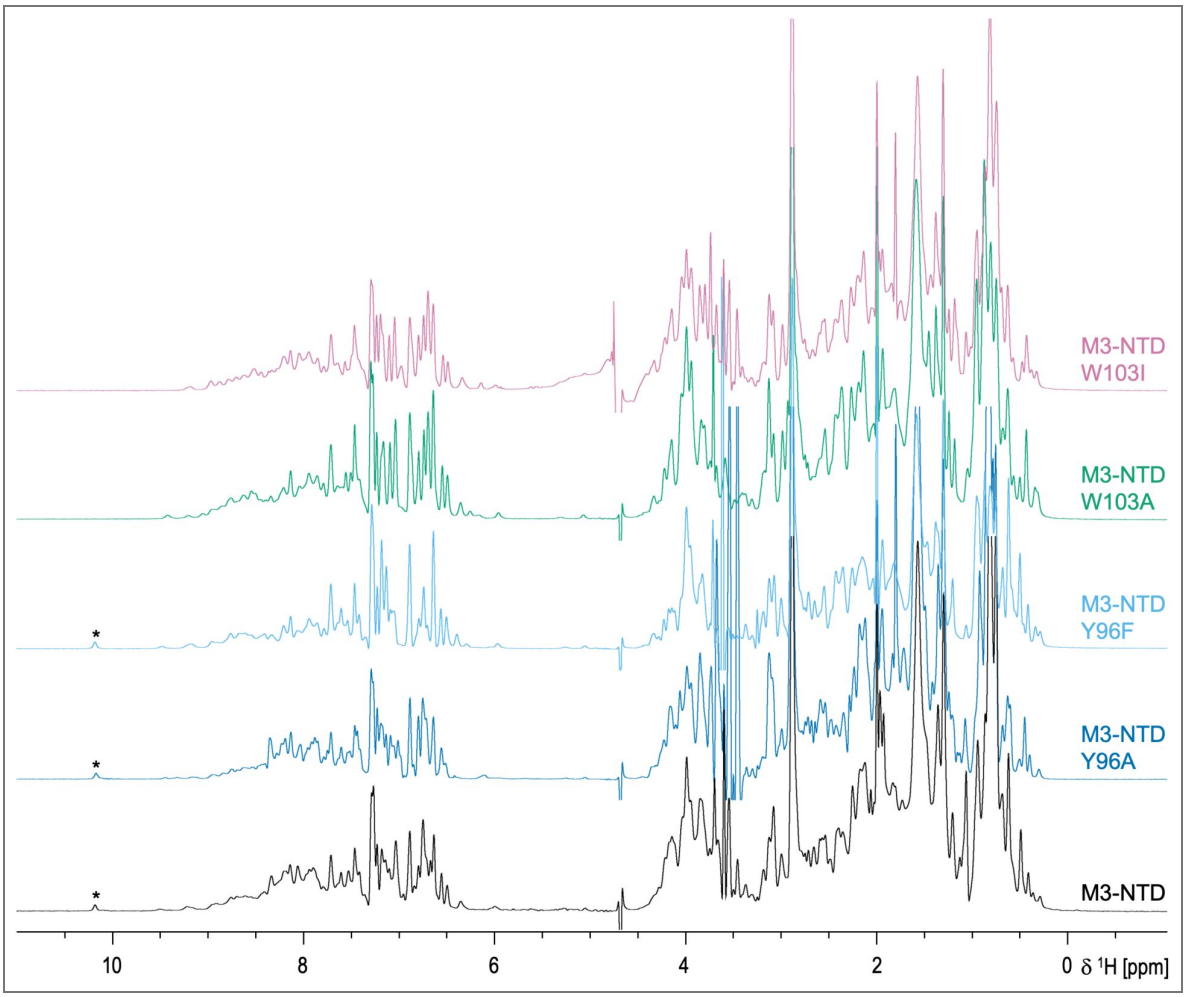


Figure 8–figure supplement 3. **^{1}H NMR spectra of M3-NTD and variants with key collagen-binding residues mutated.**

Asterisks, Trp103 indole NH signals.

types, which all share very similar triple-helical (COL) domains. This is supported by an AlphaFold3 prediction of a complex between M3 and a type I collagen peptide, which is predicted to bind to M3 in a way that is highly similar to the experimental complex structure (**Figure 8—figure supplement 1B,C**).

emm type-dependent effect of human collagen on biofilm formation by GAS strains

GAS *emm*3 strains are highly prevalent in necrotizing soft tissue infections [19, 20], where they have been found to form biofilm [18]. Based on our previous data demonstrating that the tissue milieu seems to promote GAS biofilm [18], and with collagen being a ubiquitous structural protein, it was of interest to assess whether the M3-collagen interaction could affect biofilm formation. For this purpose, we used crystal-violet based assay to determine biofilm formation by GAS strains from patients with necrotizing soft tissue infections on polystyrene plates, either uncoated or coated with human type I collagen. The three isolates, *emm*1, *emm*3 and *emm*28 strains, all formed biofilm on uncoated plates (**Figure 9A**). Notably, on collagen-coated plates, biofilm of isolates 2006 (*emm*1) and 5004 (*emm*28) was significantly reduced in a dose-dependent manner, while that of isolate 2028 (*emm*3) was significantly enhanced (**Figure 9B**, C). The collagen-enhancing effect of biofilm for the *emm*3 strain was seen even at inocula as low as 100 CFU/well (**Figure 9D**), where also a dose-response to collagen was evident (**Figure 9D**). Confocal microscopy was used to assess bacterial attachment and biofilm formation with bacteria grown on glass slides with or without collagen coating. In line with the crystal-violet assay, bacterial attachment of the *emm*1 (2006) and the *emm*28 (5004) strain was almost completely blocked by collagen whereas the *emm*3 (2028) strain formed strong robust biofilm in the presence of collagen (**Figure 9E**).

Taken together the data shows a collagen-enhanced biofilm formation for the *emm*3, but not the *emm*1 and *emm*28 strains. To test whether this effect was linked to the M3-protein/collagen interaction, we performed competition experiments where biofilm formation of the *emm*3 (2028) strain was assessed in the presence of increasing concentrations of M3-NTD in the medium. The enhancing effect of collagen was almost completely negated in presence of 20 μM M3-NTD (**Figure 10**).

These data indicate that the interaction between M3 protein and collagen enhances biofilm formation. In further support of this finding, two additional *emm*3 strains, GAS 5626 and 8003 showed a similar increase in biofilm formation on collagen-coated plates (**Figure 10—figure supplement 1A**). The expression of M3 protein by all three *emm*3 strains was confirmed using an M3-specific antibody (**Figure 10—figure supplement 1B**).

GAS M3-collagen interaction in patient biopsies and in a 3D skin tissue model

To demonstrate a potential interaction between GAS and collagen during infection, we stained tissue biopsies from patients with necrotizing soft tissue infections using specific anti-GAS and anti-collagen IV antibodies (**Figure 11**). In areas with presence of bacteria, collagen was observed to co-localize with the bacteria in biopsies from two patients infected with *emm*3 strains 2028 and 5020, and in one out of two patients infected with *emm*1 strains (2006 and 2068).

To further investigate this in a more controlled infection model, we used a human 3D organotypic skin model, which is based on a collagen type IV scaffold. We have previously employed this tissue model for studies of biofilm and GAS-elicited tissue pathology [51, 52]. Infection of the skin organotypic tissue with the *emm*1 (2006) and *emm*3 (2028) strains showed that both strains efficiently infected the skin model tissue and caused disruption of the epithelial layer (**Figure 12**). Bacterial load and tissue pathology increased over time, as did the colocalization of *emm*3 with collagen. However, no such increase was noted in the *emm*1 strain infected models. These data suggest that an interaction between *emm*3 GAS and collagen occurs in human tissue and might be an important factor for biofilm formation.

Figure 9. Effect of collagen type I on biofilm formation by necrotizing soft tissue infections strains.

A) Quantitative analysis of biofilm formation on polystyrene surface with or without human type I collagen. The inoculum of each bacterial strain was 10^5 CFU per well. Significance was determined by one-way ANOVA with Tukey's *post-hoc* test. ****, P<0.0001; ***, P< 0.001; **, P< 0.01; *, P< 0.05 B) Varying effect of collagen concentration. The inoculum of each bacterial strain was 10^5 CFU per well. C) Inoculum effects on biofilm of strain 2028. D) Effect of concentration of coating collagen on strain 2028 biofilm (100 CFU per well). A-D) Results are shown as Mean + SE. All assays were repeated at least three times in triplicate. E) Confocal microscopic analysis of biofilm formation 48 h after incubation. Fluorescence staining (WGA-Alexa 488, DAPI, and Nile red) of biofilm on uncoated and collagen-coated glass slides. Scale bars indicate 50 µm.

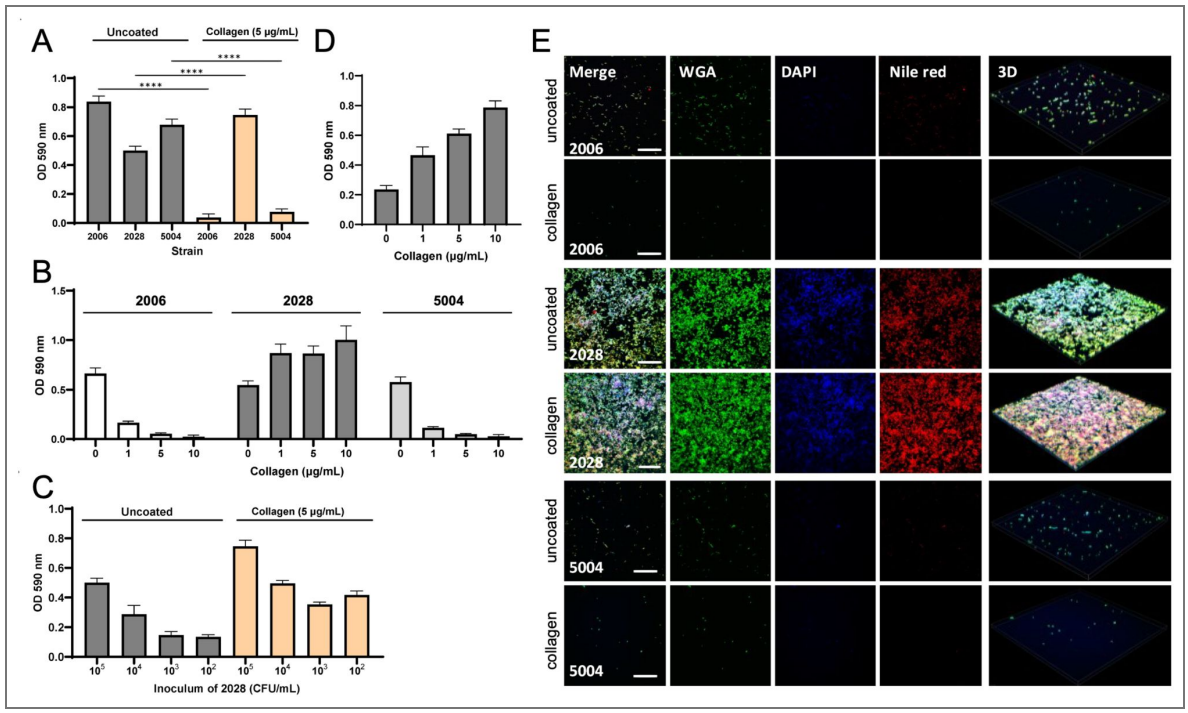
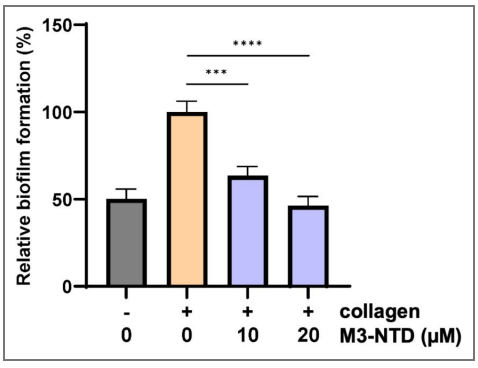


Figure 10. Competition analysis of biofilm formation of GAS isolate 2028 (100 CFU per well) in the presence of M3-NTD.

Results are shown as Mean + SE. All assays were repeated three times in triplicate. Significance was determined by one-way ANOVA, followed by Tukey's *post-hoc* test ****, P<0.0001; ***, P< 0.001.



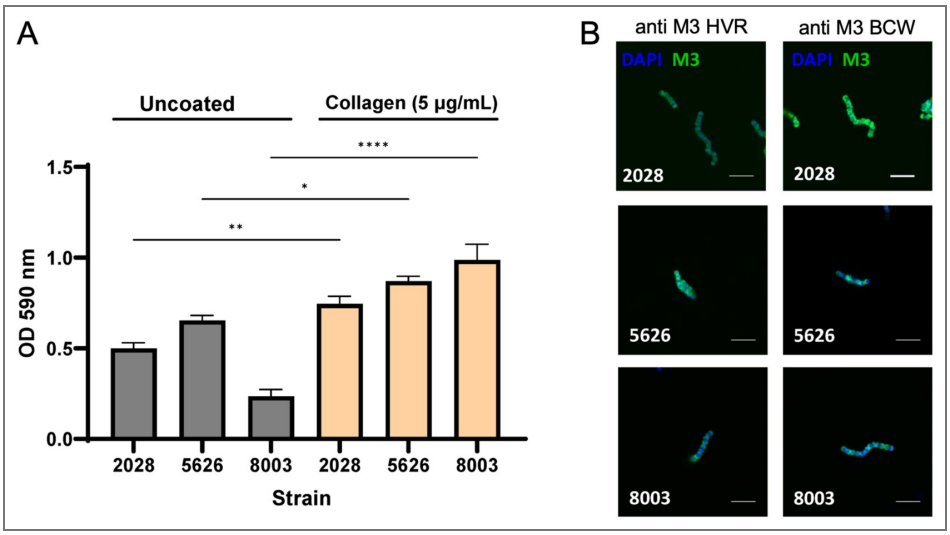


Figure 10–figure supplement a. Biofilm formation of *emm*3 GAS.

A) Quantitative analysis of biofilm formation on polystyrene surface with or without human type I collagen. The inoculum of each bacterial strain was 10^5 CFU per well. This assay was repeated at four times in triplicate and results are shown as Mean + SE. Significance was determined by one-way ANOVA with Tukey's *post-hoc* test. ****, P<0.0001; ***, P< 0.001; **, P< 0.01; *, P< 0.05. B) Confocal image of *emm*3 GAS stained with DAPI and M3-specific antibodies. BCW, B- and C-repeat and wall-spanning regions of M3. Scale bars indicate 5 µm.

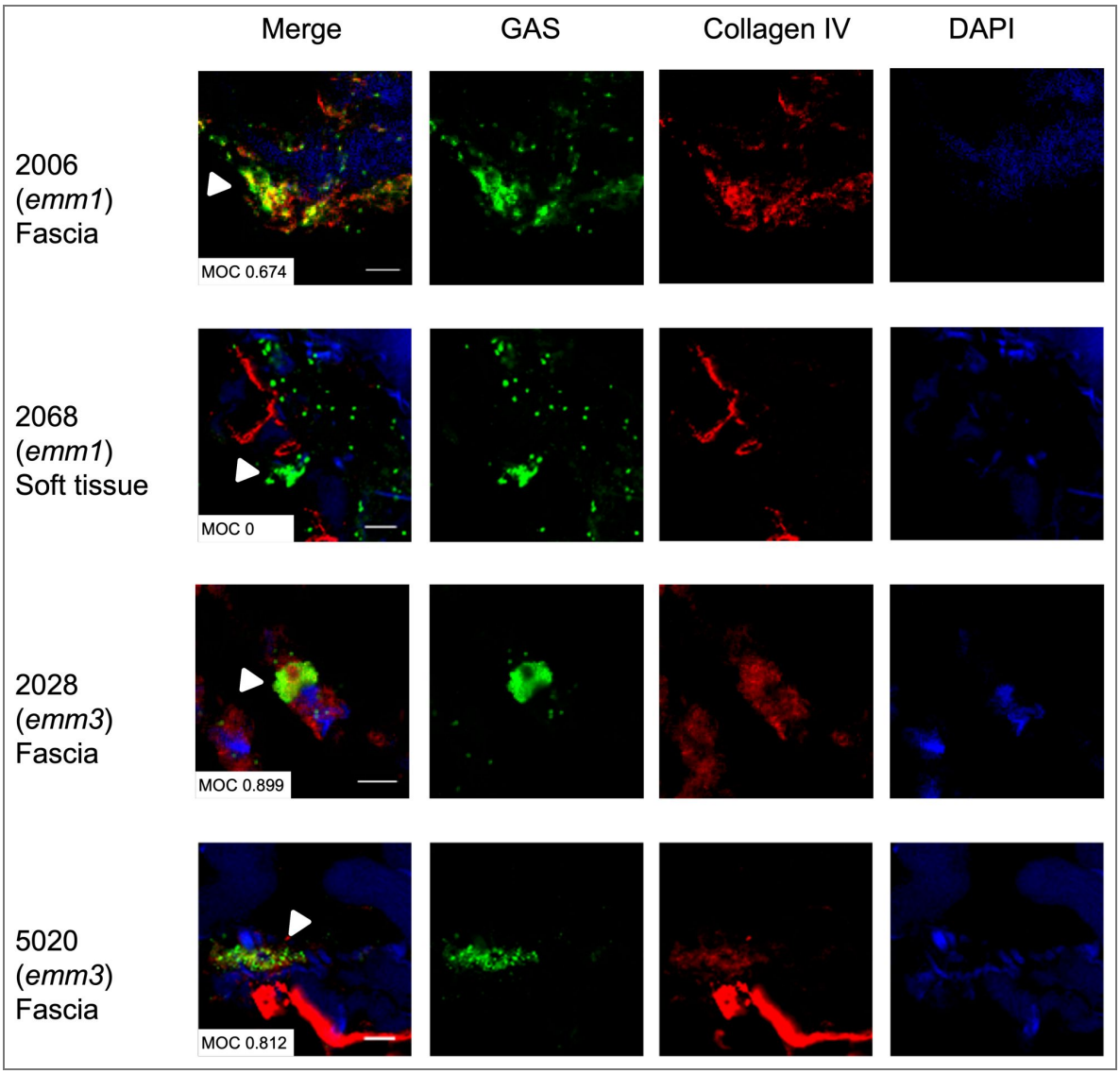


Figure 11. Colocalization of collagen with *emm3* GAS in patients' biopsies.

Frozen biopsy sections were stained with anti-GAS (green), anti-collagen IV (red) and DAPI (blue). Scale bars indicate 50 μm. The Mander's overlap coefficient (MOC) was used to quantify the colocalization of bacteria with collagen.

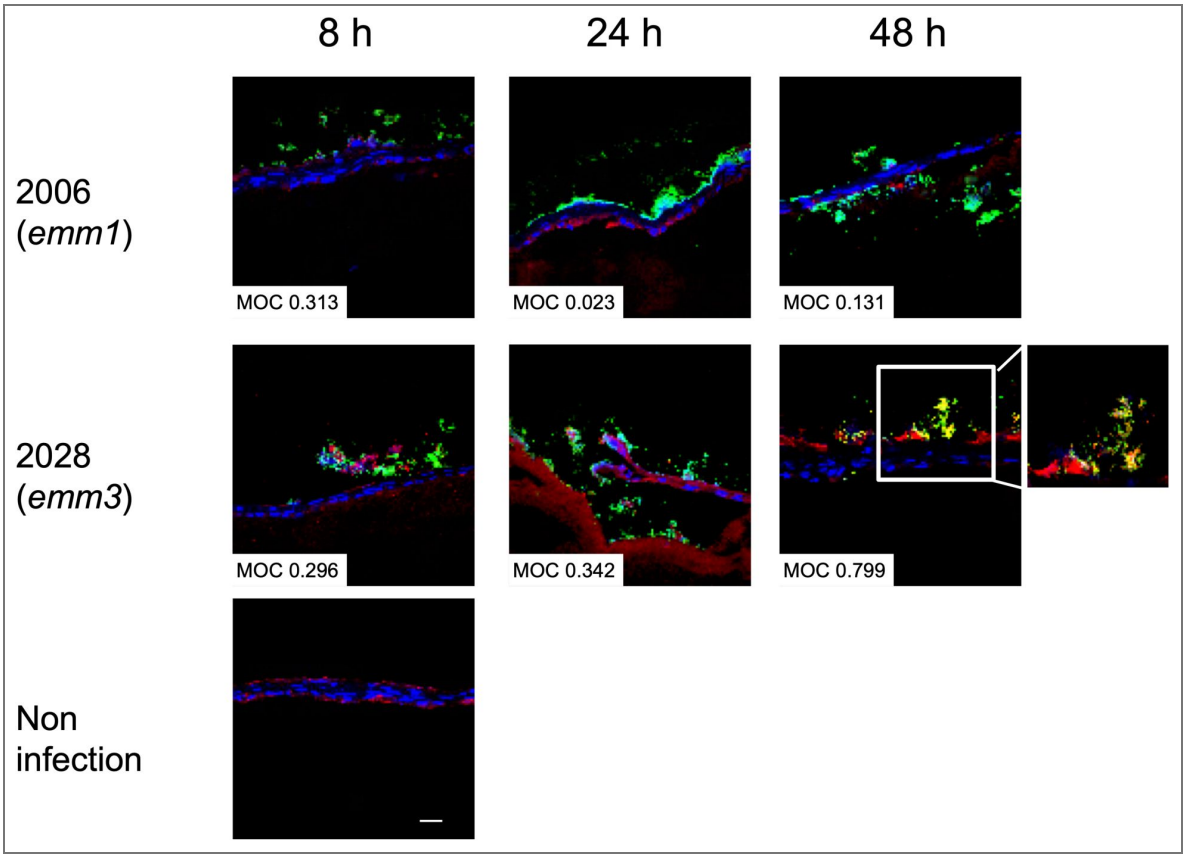


Figure 12. Colocalization of collagen with *emm3* GAS in 3D skin tissue model.

Skin tissue models were infected with GAS strains 2006 and 2028. At 8, 24 and 48 hours after infection, frozen sections were stained with anti-GAS (green), anti-collagen IV (red) and DAPI (blue). The Mander's overlap coefficient (MOC) was used to quantify the colocalization of bacteria with collagen. The merged images are shown.

Discussion

No obvious or simple connection can be made between M protein sequences and the differential abilities of the over 200 variants to interact with host factors. At least some activities of this major streptococcal virulence factor may be encoded in hidden sequence patterns, as identified for the C4-binding protein interaction with HVRs [14]. In this study, we show that the HVR of the M3 protein deviates from the canonical coiled-coil structure generally assumed to be adopted by M proteins. The novel T-shaped fold identified here is required for the collagen binding activity of M3. While this fold may be limited to a few phylogenetically close variants of M proteins in GAS, it is more abundant in SDSE. The discovery of a folded HVR may inform vaccine design. Although immunization with short and therefore monomeric M protein HVR peptides may be sufficient for most GAS serotypes, immunogenic epitopes of M3 may only be present in the folded, dimeric form of the protein.

The structures of M3-NTD alone and in presence of a collagen peptide give a rational basis for the role of the PARF motif, which was previously implicated in collagen binding [33]. The role of conserved residues within PARF is predominantly to stabilize the T-fold of M3-NTD. The motif alone, encompassing eight residues (Ala94 to Asn101) as defined by Barroso *et al.* [32], does not present all structural features involved in collagen binding based on our structure. This depends on a larger region spanning residues Gln92 to Leu107 on the H2 helix, with some minor contributions of residues in H1 (**Figure 8**). The complex structure identifies Tyr96 and Trp103 as key interface residues. They provide shape complementarity and stack against proline and hydroxyproline rings of the collagen triple helix. Our mutational analysis confirms a role in collagen binding for Tyr96 and Trp103. Conclusions from the crystal structure are consistent with our analysis of the binding propensities of the different amino acids in the CLC peptides. The observed prominent roles for hydrophobic residues and for hydroxyproline in the solid-phase binding assays would be predicted from the complex structure. The negative effects on binding rank of glutamate and of proline are less easy to explain. Proline is restricted to the X position of the GXY triplet, which is also the position usually occupied by glutamate. The effect of both residues may be to exclude more productive hydrophobic amino acids from the X position. The positive role of hydroxyproline in the Y position is amply demonstrated in the crystal structure. In addition to its ability to form hydrogen bonds, this may derive from its greater extension from the helix axis than proline. Our findings are reminiscent of those obtained for YadA [39], in that we also observed promiscuous binding to the CLC peptides, which was strongly dependent on hydrophobic residues. However, in the case of YadA, the number of GPO triplets and number of both Pro and Hyp also correlated with affinity.

The lack of specificity of CLC peptide binding and the alternative binding registers observed in the crystal structure of the M3-NTD complex provide a rationale for the ability of M3 and related proteins to target various types of collagens. The M3 fold recognizes the collagen triple helix through shape complementarity. While collagen types differ greatly in their structural organization and biological functions, their defining feature is the presence of triple-helical COL domains. The most abundant collagens are the fibrillar collagens I, II, III and V, which share high sequence homology and are rich in the hydrophobic and hydroxyproline residues needed to bind M3. The non-fibrillar collagens IV and VI have a similar amino acid content. All these abundant collagen types contain extensive triple-helical domains that we show can bind M3, and crucially, are found in the epithelial and endothelial surfaces where they may encounter GAS.

The molecular complex described here explains how GAS of the M3 type achieve very strong binding to collagens. While the affinity for the triple-helical peptides measured here by ITC is moderate, the presence of two binding sites on the HVR that is found at the distal end of each M3 protein suggests cooperative binding in the context of larger collagen assemblies. Most likely the bacteria do not encounter isolated triple helices, but higher-order collagen structures made of bundles of triple helices. It can be envisaged how M3 proteins probably intercalate between triple helices in such higher-order assemblies. Given the high density of M proteins on GAS surfaces, this would generate a polyvalent, high-avidity host-bacteria interface with the HVR acting as an anchor

binding M3 perpendicularly to collagen fibrils/fibers. It is also at this higher-order structural level that an explanation for the ability of M3 to induce an anti-collagen autoimmune response might be found. In the rare autoimmune disease Goodpasture syndrome, an anti-collagen response is thought to be caused by collagen IV neoepitopes, with infections being suggested as one possible cause for their exposure [53]. It is conceivable that such neoepitopes could be generated as a result of collagen structural changes caused by M3 protein binding. However, such structural rearrangements could not be observed in our simplified molecular system, and our data cannot provide evidence backing the anti-collagen hypothesis of post-streptococcal rheumatic sequelae [41] beyond providing the molecular basis for collagen recognition by M3 protein.

The structural data presented here allow us to conclude that collagen binding by M proteins relies on the presence of a T-shaped helical bundle domain combined with key residues to provide shape complementarity to tropocollagen triple helices. But even if all GAS M protein variants included in our sequence analysis bound collagens in a similar manner to M3, this would still indicate that M3-like collagen binding mechanism is rare among GAS M protein variants. Based on data published for M55, which was found to bind comparably weakly to collagen IV [54], it seems likely that M3 and closely related serotypes have a unique ability to bind collagens in the way described here. AlphaFold3 predictions support this, yielding converging structural models for the M3-collagen peptide complex that closely resembled our experimental structure. This was also true for M31 and M222. However, predictions with M12 and M55 resulted in low-confidence, non-converging structural models with poorly defined binding sites, and/or disruption of the T-fold (**Figure 6—figure supplement 1**). On the other hand, highly confident structural predictions together with conservation of key residues suggests the M3-like collagen binding mechanism is shared by SDSE and SESZ M proteins (**Figure 6—figure supplement 1**). This has previously been experimentally validated for the FOG (or Stg11) protein of SDSE, which was found to bind to collagens I and IV with high affinity [33, 42]. SDSE is increasingly recognized as a highly prevalent human pathogen closely resembling GAS in terms of virulence traits and pathologies, including invasive infections, post-infection autoimmune sequelae [35] and biofilm formation [55]. Collagen binding by M proteins may be therefore a far more common virulence mechanism in human streptococcal infections than the prevalence of *emm*3 GAS alone would suggest. It should be noted, too, that a new *emm*3 variant, *emm*3.93, is currently emerging in the UK and the Netherlands, with a high prevalence in invasive infections [56].

The interaction between M protein and collagen has implications for several biological functions, including bacterial attachment and biofilm formation. In this study, we investigated biofilm formation due to the high prevalence of *emm*3 strains in necrotizing soft tissue infections [20], in which biofilm is complicating feature [18, 57]. Our findings reveal that the M3 protein-collagen interaction promotes biofilm in *emm*3 strains. Notably, for other *emm* types the presence of collagen appears to have an opposing effect, reducing bacterial attachment and biofilm *in vitro* for other *emm* types. This type-specific difference is further supported by observations within infected tissue models where *emm*3 GAS showed a more pronounced colocalization with collagen fibers, as compared to *emm*1 GAS. These findings support the importance of collagen for *emm*3 biofilm development in the tissue setting, while other GAS *emm* types likely utilize other mechanisms. Further studies are warranted to explore the underlying mechanisms of biofilm formation in streptococcal tissue infections, including not only GAS but also SDSE strains.

Materials and Methods

Cloning and site-directed mutagenesis

M3-NTD DNA insert was amplified from a plasmid containing full M3 DNA sequence using primers 5'-GCTAGCCATGGATGCTAGGAGTGTTAATGG-3' and 5'-CTAGGGATCCCTAGCAGTCCTGATATTCCTTTTC-3', digested with NcoI and BamHI and ligated into the pEHISTEV vector [58], pre-digested with the same enzymes. The final M3-NTD construct, after digestion with TEV protease, comprised 113 amino acids starting with Gly-Ala-Met, an artefact of the purification tag. Site-directed mutagenesis was performed by PCR amplification of pEHISTEV-

M3NTD plasmid using mutagenic primers listed in **Table 3**. For selenomethionine labelling, two methionine residues were introduced by substituting Ile60 and Ile141 using two sequential rounds of site-directed mutagenesis.

Recombinant protein production

Recombinant M3 protein was produced as an GST fusion construct as previously described [59] from a pGEX6P-1 vector, a kind gift from Dr Susanne Talay, in *E. coli* BL21 (DE3) cells. The cells were grown in LB media containing 100 mg/mL ampicillin at 37 °C for 3 h after growth to an optical density at 600 nm of 0.6, and induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The fusion protein was purified from bacterial cell lysate using a GSTrap 4B column according to the protocol provided by the manufacturer (Cytiva). The recombinant protein comprised GST fused to a proteolytic cleavage site (not used in this study) and the N-terminus of M3 (UniProt entry A0A0H2UWN1), lacking N-terminal secretion signal, cell wall anchor and membrane spanning region (residues 42-546).

M3-NTD constructs were overexpressed from pEHISTEV plasmid in *E. coli* soluBL21 cells (Genlantis), in a standard LB medium supplemented with 50 mg/mL kanamycin. An overnight culture was used to inoculate flasks containing 1 L of the growth medium which were incubated at 37 °C with shaking. When the optical density at 600 nm was 0.5-0.8 expression was induced by adding IPTG to a final concentration of 1 mM. After 3-4 h of further incubation with shaking the cells were harvested and frozen. For selenomethionine labelling the pHT-M3-NTD [Ile60Met, Ile141Met] plasmid was introduced into the methionine auxotroph strain of *E. coli*, B834, which was obtained from Dr Clarissa Melo Czekster. The cells from the overnight culture were washed twice with M9 minimal media, resuspended in 0.2x the initial volume and used to inoculate (10 mL per L) of SelenoMet medium (Molecular Dimensions) supplemented with 0.04 mg/L L-selenomethionine (ThermoFisher). Following induction of expression, the cells were incubated with shaking for 5 h.

Cell pellets were resuspended in 50-100 mL 50 mM Tris-HCl pH 8, 500 mM NaCl lysis buffer supplemented with one cOmplete protease inhibitor tablet (Roche) and 1 mg DNAse I (Merck). The suspension was passed twice through a Cell Disrupter (Constant Systems) at 30 kpsi and the lysate was clarified by centrifugation at 20,000 x g for 25 min at 4 °C. The supernatant was applied to a HisTrap column (Cytiva) pre-equilibrated in the lysis buffer and the bound protein was washed with 10 column volumes of the wash buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 20 mM imidazole) and eluted in a linear gradient (20-250 mM imidazole). To remove imidazole and the N-terminal purification tag, the eluate was mixed with TEV protease (produced in-house) at approximately 30:1 stoichiometric ratio and dialyzed overnight at room temperature against phosphate-buffered saline (PBS) supplemented with 1 mM dithiothreitol (DTT). The digested protein solution was passed through the HisTrap column again, with PBS as the mobile phase, and M3-NTD collected in the flow-through. Following concentration in 10 kDa Amicon Ultra centrifugal filters (Millipore) the protein was further purified by size exclusion chromatography using Superdex75 10/300 column (Cytiva) pre-equilibrated in PBS. Purity and oxidation state were verified by SDS-PAGE. If some monomeric protein was still evident on the gel the fractions from size exclusion chromatography were left overnight at 4 °C. Fully oxidised protein was concentrated, aliquoted and flash frozen.

Collagen peptides

The Collagen Ligand Collection (CLC) peptides (formerly known as Collagen Toolkits) were obtained as C-terminal amides from Triple Helical Peptides Ltd, Cambridge, UK. They were synthesised on TentaGel R-Ram resin using Fmoc/tBu chemistry, either on an Applied Biosystems Pioneer peptide synthesiser as described previously [60], or a CEM Liberty or Liberty Blue microwave-assisted peptide synthesiser. Fractions containing homogeneous product were identified by analytical HPLC on an ACEphenyl300 (5 mm) column, characterised by MALDI-TOF mass spectrometry, pooled and freeze-dried. In the CLCs, the variable 27-residue primary collagen

MUTATION	PRIMER SEQUENCE (5'→3')
Tyr96Ala	GACAAAAGGCTGAAGCGCTAAAAGGCC
	GGCCTTTTAGCGCTTCAGCCTTTTGTC
Tyr96Phe	GACAAAAGGCTGAATTTCTAAAAGGCC
	GGCCTTTTAGAAATTCAGCCTTTTGTC
Trp103Ala	CCTTAATGATGCGGCTGAGAGGC
	CCTTTTAGATATTCAGCCTTTTG
Trp103Asp	CCTTAATGATGATGCTGAGAGGCTG
	CCTTTTAGATATTCAGCCTTTTG
Trp103Ile	CCTTAATGATATTGCTGAGAGGCTG
	CCTTTTAGATATTCAGCCTTTTG
Ile60Met	GTTAAATTAAAAAATGAAATGGAGAACTTGTTAGATC
	GATCTAACAAGTTCTCCATTTCATTTTTTAATTTAAC
Ile141Met	GAACTTAAGGAAAAAATGGACAAAAAGGAAAAGG
	CCTTTTCCTTTTTGTCCATTTTTTCCTTAAGTTC

Table 3. **List of mutagenic primers.**

structure (guest sequence) was flanked by GPC[GPP]$_5$- and-[GPP]$_5$GPC (host) peptides, to ensure stable triple-helical form. In JDM238, the guest sequence is GVOGEA and the host sequences [GPO]$_3$.

Solid phase CLC binding assay

CLC peptide solid-phase binding assays were performed following a previously published protocol [45, 47]. Briefly, CLC peptides, collagen II (positive control), GPP10 and bovine serum albumin (BSA) (negative controls) (10 µg/mL in 0.01 M acetic acid) were immobilized on Immulon2 HB 96-well plates (Nunc, Langenselbold, Germany) overnight at 4 °C. All subsequent incubation steps were for 1 h at 25 °C. The assay volume was 100 µL per well. Wells were washed three times with adhesion buffer (0.1% (v/v) Tween-20 and 1 mg/mL BSA in PBS) between incubation steps. The wells were blocked with 50 mg/mL BSA in PBS prior to the addition of recombinant GST-M3 at a concentration of 10 µg/mL in adhesion buffer. Bound protein was detected with biotin goat anti-GST antibody (Abcam) at a 1:500 dilution in adhesion buffer, followed by streptavidin-fused horseradish peroxidase (HRP) and 3,3',5,5'-tetramethylbenzidine liquid substrate system (Sigma), and plates read at 450 nm. Including Tween in the washing steps reduced background signal but did not change the overall outcome for CLC-II. It was omitted for the CLC-III assay.

NMR

M3-NTD was isotopically labeled by expression in M9 minimal media (0.25 M Na_2HPO_4, 0.13 M KH_2PO_4, 0.04 M NaCl, 18 mM $^{15}NH_4Cl$, 2.5 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1% (w/v) glucose, 2.125 g/L BDTM Difco^{TM} Yeast Nitrogen Base without Amino Acids and Ammonium Sulfate (Thermo Fisher) and 50 µg/mL kanamycin) and purified as described above for unlabeled protein. The NMR sample contained 0.4 mM protein in 10 mM phosphate, 50 mM NaCl, pH 6.5, 1.5% (v/v) D_2O, without or with addition of 10 mM DTT to generate monomeric M3-NTD. ^1H,^{15}N HSQC spectra were recorded on a Bruker Ascend 700 MHz spectrometer equipped with a Prodigy TCI probe and controlled by Bruker Topspin software at 30 °C. A standard Bruker pulse sequence with gradients and water flip back pulse (hsqcetfpf3gpsi) was used with 20 transients and spectral resolutions or 14.5 Hz and 41.2 Hz in the direct (^1H) and indirect (^{15}N) dimension, respectively. Spectra were processed with NMRPipe [61] and visualized with CCPN Analysis 2 [62]. For 1D 1H NMR spectra samples were prepared in PBS at 0.2 mM. A standard Bruker pulse sequence with gradients and excitation sculpting for water suppression (zgesgp) was used with 256 transients and a spectral resolution of 0.86 Hz. Spectra were processed and visualized with Topspin.

Isothermal titration calorimetry

Freeze-dried CLC peptides were reconstituted by dissolving in PBS (10 mg/mL), incubating the solution at 75 °C in a water bath that was left to cool slowly to room temperature overnight. M3-NTD was dialysed into PBS. Experiments were performed using a MicroCal PEAQ-ITC instrument (Malvern Panalytical) at 25 °C in PBS, with the reference power set to 3 µcal/s, stirring speed to 750 rpm and injection speed to 0.5 µL/s. The cell contained CLC peptide at 270 µM (monomer concentration) and the injector syringe contained M3-NTD variants at 1 mM (monomer concentration). The only exception was the M3-NTD Trp103Ile titration, where the cell and syringe solution concentrations were 165 µM and 0.61 mM, respectively. For control experiments determining the heats of dilution, the cell contained PBS buffer only. Titration involved a single injection of 0.4 µl followed by 18 injections of 2 µl syringe solution.

Protein crystallisation and structure determination

Crystallization trials were conducted using the vapor diffusion sitting drop method and several sparse matrix screens. M3-NTD dimer was freshly purified by size exclusion chromatography and concentrated to ~700 µM (monomer concentration). For co-crystallization with JDM238 peptide, the protein was concentrated to 1 mM and then diluted by adding reconstituted peptide so that the final concentration of M3-NTD was 700 µM with triple-helical peptide at 1.1 mM (~1:3 ratio of M3-NTD dimer to triple-helical peptide). The crystals of M3-NTD grew at 20 °C within 1-2 weeks in 40-

50% 2-methyl-2,4-pentanediol (MPD), and the best one diffracted to 1.92 Å resolution. They were used for streak-seeding the selenomethionine-labelled protein, which resulted in shard-like crystals diffracting to 2.67 Å with the data collected at the selenium K absorption edge. Phasing of the SAD dataset was conducted automatically by the Diamond Light Source (DLS) pipeline FastEP and initial model building was performed using the ARP/wARP software package. The crystals of M3-NTD in complex with the collagen peptide were obtained in 15% polythethylene glycol 10K, 0.1 M Tris-HCl pH 8.5, 0.29 M $MgSO_4$ and diffracted to 2.32 Å resolution. Indexing, scaling and merging of data was performed using the autoPROC pipeline at the DLS. The complex structure was solved by molecular replacement with PHASER [63] using M3-NTD and collagen peptide (PDB 3P46) structures as search models, with all non-proline residues in the collagen peptide model substituted with alanine. For both structures, iterative model building and refinement was performed using Coot [64] and Refmac5 [65]. MolProbity [66] was used for model quality assessment. Figures of protein structure models were produced using PyMOL (Schrödinger, LLC).

Accession codes

The M3-NTD and M3-NTD/JDM238 protein structures and the data used to derive these, have been deposited at the PDBe with accession codes 8p6k and 8p6j, respectively.

Bacterial strains

GAS strains 2006 (*emm1*), 2028 (*emm3*), and 5004 (*emm28*) are isolates from patients with necrotizing soft tissue infections from the INFECT project [18]. GAS 5262 and 8003 were provided by Donald E. Low (Mount Sinai Hospital, Toronto, Canada) and used as further *emm3* isolates from patients with necrotizing soft tissue infections [67, 68]. All isolates were cultured in either Todd-Hewitt broth supplemented with 1.5% yeast extract or brain heart infusion broth at 37 °C under a 5% CO_2 atmosphere.

Biofilm formation assay on a polystyrene surface

Biofilm formation was evaluated by crystal violet staining as described previously, with minor modifications [69]. Overnight cultures of the bacteria being tested were washed with PBS and diluted to approximately 10^6 CFU/mL. Bacterial suspensions (100 µL) were seeded into 96-well polystyrene plates (Thermo Fisher Scientific, Waltham, MA, USA). The plates were then incubated for 24 h at 37°C. After incubation, planktonic bacteria were removed by washing with PBS. Plates were then stained with a 0.1% crystal violet solution (Invitrogen, Waltham, MA, USA) for 30 min. Excess crystal violet was removed by washing with PBS, and then the crystal violet that was associated with bacterial cells was eluted with absolute ethanol. The amount of crystal violet (and by association, biofilm) was evaluated by measuring the absorbance at 590 nm using a spectrophotometer. Collagen-coated plates were prepared by incubating plates with 1-10 µg/mL of collagen I overnight at 4 °C.

Confocal microscopic analysis of biofilm formation

The biofilms were formed on an 8-well chamber slide (Lab-Tek, Thermo Fisher Scientific) in the same manner as for the biofilm assay. After removing planktonic bacteria by washing with PBS, the biofilms were fixed with 10% formalin and stained with wheat germ agglutinin (WGA)-Alexa Fluor 488 conjugate (Invitrogen) and Nile red (Invitrogen). The slides were mounted using ProLong™ Gold Antifade Mountant with DAPI (Invitrogen).

Biofilm competition assay

Recombinant M3-NTD was added at 10 and 20 µM concentrations in PBS to a collagen-coated 96-well plate (0.5 µg/well) and the plate was incubated for 1 h at room temperature. After removal of the protein solution, 100 µL of strain 2028 culture (ca. 10^3 CFU/mL) was seeded and incubated for 24 h. Biofilms were stained with crystal violet, as described above.

Immunofluorescent staining of GAS

A few colonies were suspended in PBS on a glass slide and then fixed in 3.7% formaldehyde in PBS for 15 min. The bacteria were stained with anti-M3 specific antibodies (kindly provided by Prof. Gunnar Lindahl, Lund University), followed by WGA-Alexa Fluor 488 conjugate antibodies (Invitrogen). The slides were mounted using ProLong™ Gold Antifade Mountant with DAPI (Invitrogen).

Three-dimensional organotypic skin model

The 3D skin models were constructed as described previously [18, 51, 52]. Briefly, 4.0 ξ 10^4 normal human dermal fibroblasts (NHDF) in a collagen matrix (Pure Col, Advanced Biomatrix, Carlsbad, CA, USA) were seeded onto a polymerized cell-free collagen layer in a 6-well filter insert (Corning, Corning, NY, USA). After culturing in Dulbecco's Modified Eagle Medium (DMEM) for one week, 1.0 ξ 10^6 human keratinocyte cells N/TERT-1 were seeded onto the NHDF layer. The models were cultured in EpiLife medium (Invitrogen) for three days and exposed to air for one week. For the infection assay, the models were infected with 1 × 10^6 CFU of bacteria for 8, 24, or 48 h.

Immunofluorescent staining of patient tissue biopsies and tissue models

Snapfrozen tissue biopsies from patients with necrotizing soft tissue infections caused by either *emm*1 GAS strains (patients 2006 and 2068) or *emm*3 GAS (patients 2028 and 5020) were available from the INFECT patient biobank [18]. Cryosectioning and staining were performed as previously described [52]. Briefly, the biopsies and models were embedded in an optimum cutting temperature compound (Sakura, Torrance, CA, USA) and frozen in liquid nitrogen. Cryosectioning (8 µm) was performed using a Leica CM3050 cryostat (Leica, Nußloch, Germany). Sections were then fixed in 3.7% formaldehyde in PBS for 15 min and stained with both anti-GAS (goat polyclonal antibody, Abcam, Cambridge, UK) and anti-collagen IV antibodies (Abcam, Cambridge, UK), followed by WGA-Alexa Fluor 488 and 546 conjugate antibodies (Invitrogen). Manders' correlation coefficient was used to quantify the degree of colocalization between fluorophores and assessed with ImageJ V.1.54p using the JACoP plugin V. 2.1.4. [70].

Statistical Analysis

To compare the propensity of specific CLC amino acids to contribute to M3 binding, the A_{450} value for M3 binding to BSA was subtracted from all values obtained from the solid-phase binding assays. This allowed the CLC peptides to be ranked in descending order of A_{450} and divided into three absorbance groups: high, from 0.75 to 0.5; medium, from 0.5 to 0.25; and low, from 0.25 to zero. Data from both CLCs were pooled, and the amino acid abundance in the three groups was compared. For each peptide from CLC-II and CLC-III, the number of occurrences of each amino acid of interest was noted. The non-parametric Kruskal-Wallis test was used to determine whether the amino acid abundance differed between the three groups. For the biofilm experiments, One-way ANOVA with the Tukey's *post-hoc* test was used to determine statistically significant differences.

Acknowledgements

M3-specific antibodies were kindly provided by Professor Gunnar Lindahl, Lund University. We are grateful to Dr Susanne Talay for the M3-encoding pGEX6P-1 vector. The authors thank Dr Conny Yu for preparing a batch of M3-NTD for biofilm experiments. We thank the i04 and i24 beamline staff at the Diamond Light Source synchrotron for their help with data collection.

Additional information

Funding

This research was funded by the Medical Research Council (MR/N009681/1), grants from European Union FP6 ASSIST (032390) and FP7 INFECT (305340), the Swedish Research Council (2022-01-202), and Region Stockholm, Center for Innovative Medicine (FoUI-975603).

Funding

Funder	Grant reference number
Medical Research Council	MR/N009681/1
European Union	FP6 ASSIST 032390
European Union	FP7 INFECT 305340
Swedish Research Council	2022-01-202
Region Stockholm, Center for Innovative Medicine	FoUI-975603

Author ORCID iDs

Ulrich Schwarz-Linek: http://orcid.org/0000-0003-0526-223X

Author notes

Marta Wojnowska, Takeaki Wajima: equal contribution first authors
Marta Wojnowska: Present address: Department of Biochemistry, University of Cambridge, United Kingdom
Takeaki Wajima: Present address: Department of Microbiology, Faculty of Pharmacy, Meijo University, Nagoya, Japan
Tamas Yelland: Present address: Evotec Ltd, Oxford, United Kingdom
Hannes Ludewig: Present address: Howard Hughes Medical Institute and Department of Biology, Brandeis University, United States
Jean-Daniel Malcor: Present address: Laboratory of Tissue Biology and Therapeutic Engineering, CNRS UMR 5305, University Claude Bernard-Lyon 1 and University of Lyon, Lyon, France
Arkadiusz Bonna, Richard W Farndale: Present address: Triple Helical Peptides Ltd, Cambridge, United Kingdom
Anna Norrby-Teglund, Ulrich Schwarz-Linek: equal contribution senior authors
Competing interests: No competing interests declared

References

1. **Brouwer S**, Rivera-Hernandez T, Curren BF, Harbison-Price N, De Oliveira DMP, Jespersen MG, Davies MR, Walker MJ (2023) Pathogenesis, epidemiology and control of Group A Streptococcus infection. *Nature Reviews Microbiology* **21**:431-447 https://doi.org/10.1038/s41579-023-00865-7
2. **Carapetis JR**, Steer AC, Mulholland EK, Weber M (2005) The global burden of group A streptococcal diseases. *The Lancet Infectious Diseases* **5**:685-94
3. **Jackson SJ**, Steer AC, Campbell H (2011) Systematic Review: Estimation of global burden of non-suppurative sequelae of upper respiratory tract infection: rheumatic fever and post-streptococcal glomerulonephritis. *Tropical Medicine & International Health* **16**:2-11 https://doi.org/10.1111/j.1365-3156.2010.02670.x
4. **Bagcchi S** (2023) Surge of invasive Group A Streptococcus disease. *Lancet Infectious Diseases* **23**:284-284 https://doi.org/10.1016/s1473-3099(23)00043-9

5. **Oehmcke S**, Shannon O, Morgelin M, Herwald H (2010) Streptococcal M proteins and their role as virulence determinants. *Clinica Chimica Acta; International Journal of Clinical Chemistry* **411**:1172-80 https://doi.org/10.1016/j.cca.2010.04.032
6. **Smeesters PR**, McMillan DJ, Sriprakash KS (2010) The streptococcal M protein: a highly versatile molecule. *Trends in Microbiology* **18**:275-282 https://doi.org/10.1016/j.tim.2010.02.007
7. **Walker MJ**, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V (2014) Disease manifestations and pathogenic mechanisms of Group A Streptococcus. *Clinical Microbiology Reviews* **27**:264-301 https://doi.org/10.1128/CMR.00101-13
8. **Phillips GN Jr**, Flicker PF, Cohen C, Manjula BN, Fischetti VA (1981) Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proceedings of the National Academy of Sciences of the United States of America* **78**:4689-93
9. **Nilson BHK**, Frick IM, Akesson P, Forsen S, Bjorck L, Akerstrom B, Wikstrom M (1995) Structure and stability of protein-H and the M1 protein from Streptococcus-pyogenes - implications for other surface-proteins of Gram-positive bacteria. *Biochemistry* **34**:13688-13698 https://doi.org/10.1021/bi00041a051
10. **Sanderson-Smith M**, De Oliveira DMP, Guglielmini J, McMillan DJ, Vu T, Holien JK, Henningham A, Steer AC, Grp MPS (2014) A Systematic and Functional Classification of Streptococcus pyogenes That Serves as a New Tool for Molecular Typing and Vaccine Development. *Journal of Infectious Diseases* **210**:1325-1338 https://doi.org/10.1093/infdis/jiu260
11. **Macheboeuf P**, Buffalo C, Fu CY, Zinkernagel AS, Cole JN, Johnson JE, Nizet V, Ghosh P (2011) Streptococcal M1 protein constructs a pathological host fibrinogen network. *Nature* **472**:64-8 https://doi.org/10.1038/nature09967
12. **McNamara C**, Zinkernagel AS, Macheboeuf P, Cunningham MW, Nizet V, Ghosh P (2008) Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science* **319**:1405-8
13. **Stewart CM**, Buffalo CZ, Valderrama JA, Henningham A, Cole JN, Nizet V, Ghosh P (2016) . Coiled-coil destabilizing residues in the group A Streptococcus M1 protein are required for functional interaction. *Proceedings of the National Academy of Sciences of the United States of America* **113**:9515-9520 https://doi.org/10.1073/pnas.1606160113
14. **Buffalo CZ**, Bahn-Suh AJ, Hirakis SP, Biswas T, Amaro RE, Nizet V, Ghosh P (2016) Conserved patterns hidden within group A Streptococcus M protein hypervariability recognize human C4b-binding protein. *Nature Microbiology* **1** https://doi.org/10.1038/nmicrobiol.2016.155
15. **Fiedler T**, Koller T, Kreikemeyer B (2015) Streptococcus pyogenes biofilms-formation, biology, and clinical relevance. *Frontiers in Cellular and Infection Microbiology* **5**:15 https://doi.org/10.3389/fcimb.2015.00015
16. **Hoiby N**, Bjarnsholt T, Moser C, Bassi GL, Coenye T, Donelli G, Hall-Stoodley L, Hola V, Consulting External Expert Werner Z (2015) ESCMID guideline for the diagnosis and treatment of biofilm infections 2014. *Clinical Microbiology and Infection* **21**:S1-25 https://doi.org/10.1016/j.cmi.2014.10.024
17. **Ciofu O**, Moser C, Jensen PO, Hoiby N (2022) Tolerance and resistance of microbial biofilms. *Nature Reviews Microbiology* **20**:621-635 https://doi.org/10.1038/s41579-022-00682-4
18. **Siemens N**, Chakrakodi B, Shambat SM, Morgan M, Bergsten H, Hyldegaard O, Skrede S, Arnell P, Norrby-Teglund A (2016) Biofilm in group A streptococcal necrotizing soft tissue infections. *JCI Insight* **1**:e87882 https://doi.org/10.1172/jci.insight.87882
19. **Luca-Harari B**, Darenberg J, Neal S, Siljander T, Strakova L, Tanna A, Creti R, Ekelund K, Jasir A (2009) Clinical and microbiological characteristics of severe Streptococcus pyogenes disease in Europe. *Journal of Clinical Microbiology* **47**:1155-65 https://doi.org/10.1128/JCM.02155-08
20. **Bruun T**, Rath E, Madsen MB, Oppegaard O, Nekludov M, Arnell P, Karlsson Y, Babbar A, Group IS (2021) Risk Factors and Predictors of Mortality in Streptococcal Necrotizing Soft-tissue Infections: A Multicenter Prospective Study. *Clinical Infectious Diseases* **72**:293-300 https://doi.org/10.1093/cid/ciaa027

21. **Lembke C**, Podbielski A, Hidalgo-Grass C, Jonas L, Hanski E, Kreikemeyer B (2006) Characterization of biofilm formation by clinically relevant serotypes of group A streptococci. *Applied and Environmental Microbiology* **72**:2864-75 https://doi.org/10.1128/AEM.72.4.2864-2875.2006

22. **Koller T**, Manetti AGO, Kreikemeyer B, Lembke C, Margarit I, Grandi G, Podbielski A (2010) Typing of the pilus-protein-encoding FCT region and biofilm formation as novel parameters in epidemiological investigations of Streptococcus pyogenes isolates from various infection sites. *Journal of Medical Microbiology* **59**:442-452 https://doi.org/10.1099/jmm.0.013581-0

23. **Herwald H**, Cramer H, Mörgelin M, Russell W, Sollenberg U, Norrby-Teglund A, Flodgaard H, Lindbom L, Björck L (2004) M protein, a classical bacterial virulence determinant, forms complexes with fibrinogen that induce vascular leakage. *Cell* **116**:367-379 https://doi.org/10.1016/S0092-8674(04)00057-1

24. **Thern A**, Stenberg L, Dahlback B, Lindahl G (1995) Ig-Binding Surface-Proteins of Streptococcus-Pyogenes Also Bind Human C4b-Binding Protein (C4bp), a Regulatory Component of the Complement-System. *Journal of Immunology* **154**:375-386

25. **Sanderson-Smith M**, Batzloff M, Sriprakash KS, Dowton M, Ranson M, Walker MJ (2006) Divergence in the plasminogen-binding group A streptococcal M protein family - Functional conservation of binding site and potential role for immune selection of variants. *Journal of Biological Chemistry* **281**:3217-3226 https://doi.org/10.1074/jbc.M508758200

26. **Cue D**, Lam H, Cleary PP (2001) Genetic dissection of the M1 protein: regions involved in fibronectin binding and intracellular invasion. *Microbial Pathogenesis* **31**:231-242 https://doi.org/10.1006/mpat.2001.0467

27. **Dinkla K**, Rohde M, Jansen WT, Kaplan EL, Chhatwal GS, Talay SR (2003) Rheumatic fever-associated Streptococcus pyogenes isolates aggregate collagen. *Journal of Clinical Investigation* **111**:1905-12

28. **Bessen DE** (1994) Localization of Immunoglobulin A-Binding Sites within M or M-Like Proteins of Group-A Streptococci. *Infection and Immunity* **62**:1968-1974 https://doi.org/10.1128/Iai.62.5.1968-1974.1994

29. **Quek AJH**, Mazzitelli BA, Wu G, Leung EWW, Caradoc-Davies TT, Lloyd GJ, Jeevarajah D, Conroy PJ, Law RHP (2019) Structure and Function Characterization of the a1a2 Motifs of *Streptococcus pyogenes* M Protein in Human Plasminogen Binding. *Journal of Molecular Biology* **431**:3804-3813 https://doi.org/10.1016/j.jmb.2019.07.003

30. **Yuan Y**, Ayinuola YA, Singh D, Ayinuola O, Mayfield JA, Quek A, Whisstock JC, Law RHP, Castellino FJ (2019) Solution structural model of the complex of the binding regions of human plasminogen with its M-protein receptor from Streptococcus pyogenes. *Journal of Structural Biology* **208**:18-29 https://doi.org/10.1016/j.jsb.2019.07.005

31. **Bober M**, Morgelin M, Olin AI, von Pawel-Rammingen U, Collin M (2011) The Membrane Bound LRR Lipoprotein Slr, and the Cell Wall-Anchored M1 Protein from *Streptococcus pyogenes* Both Interact with Type I Collagen. *PLoS One* **6** https://doi.org/10.1371/journal.pone.0020345

32. **Barroso V**, Rohde M, Davies MR, Gillen CM, Nitsche-Schmitz DP, Dinkla K, Chhatwal GS (2009) Identification of active variants of PARF in human pathogenic group C and group G streptococci leads to an amended description of its consensus motif. *International Journal of Medical Microbiology* **299**:547-53

33. **Dinkla K**, Nitsche-Schmitz DP, Barroso V, Reissmann S, Johansson HM, Frick IM, Rohde M, Chhatwal GS (2007) Identification of a streptococcal octapeptide motif involved in acute rheumatic fever. *Journal of Biological Chemistry* **282**:18686-93

34. **Brandt CM**, Spellerberg B (2009) Human Infections Due to Streptococcus dysgalactiae Subspecies equisimilis. *Clinical Infectious Diseases* **49**:766-772 https://doi.org/10.1086/605085

35. **Xie O**, Zachreson C, Tonkin-Hill G, Price DJ, Lacey JA, Morris JM, McDonald MI, Bowen AC, Tong SYC (2024) Overlapping *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subspecies *equisimilis* household transmission and mobile genetic element exchange. *Nature Communications* **15** https://doi.org/10.1038/s41467-024-47816-1

36. **Singh B**, Fleury C, Jalalvand F, Riesbeck K (2012) Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. *FEMS Microbiology Reviews* **36**:1122-80 https://doi.org/10.1111/j.1574-6976.2012.00340.x

37. **Arora S**, Gordon J, Hook M (2021) Collagen Binding Proteins of Gram-Positive Pathogens. *Frontiers in Microbiology* **12** https://doi.org/10.3389/fmicb.2021.628798

38. **Zong Y**, Xu Y, Liang X, Keene DR, Hook A, Gurusiddappa S, Hook M, Narayana SV (2005) . A 'Collagen Hug' model for Staphylococcus aureus CNA binding to collagen. *EMBO Journal* **24**:4224-36 https://doi.org/10.1038/sj.emboj.7600888

39. **Leo JC**, Elovaara H, Bihan D, Pugh N, Kilpinen SK, Raynal N, Skurnik M, Farndale RW, Goldman A (2010) First analysis of a bacterial collagen-binding protein with collagen Toolkits: promiscuous binding of YadA to collagens may explain how YadA interferes with host processes. *Infection and Immunity* **78**:3226-36

40. **van Wieringen T**, Kalamajski S, Liden A, Bihan D, Guss B, Heinegard D, Farndale RW, Rubin K (2010) The streptococcal collagen-binding protein CNE specifically interferes with alphaVbeta3-mediated cellular interactions with triple helical collagen. *Journal of Biological Chemistry* **285**:35803-13 https://doi.org/10.1074/jbc.M110.146001

41. **Tandon R**, Sharma M, Chandrashekhar Y, Kotb M, Yacoub MH, Narula J (2013) Revisiting the pathogenesis of rheumatic fever and carditis. *Nature Reviews Cardiology* **10**:171-7 https://doi.org/10.1038/nrcardio.2012.197

42. **Nitsche DP**, Johansson HM, Frick IM, Mörgelin M (2006) Streptococcal protein FOG, a novel matrix adhesin interacting with collagen I in vivo. *Journal of Biological Chemistry* **281**:1670-1679 https://doi.org/10.1074/jbc.M506776200

43. **Delorenzi M**, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**:617-25 https://doi.org/10.1093/bioinformatics/18.4.617

44. **Ricard-Blum S** (2011) The Collagen Family. *Cold Spring Harbor Perspectives in Biology* **3** https://doi.org/10.1101/cshperspect.a004978

45. **Konitsiotis AD**, Raynal N, Bihan D, Hohenester E, Farndale RW, Leitinger B (2008) Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *Journal of Biological Chemistry* **283**:6861-8

46. **Zhou NE**, Kay CM, Hodges RS (1993) Disulfide bond contribution to protein stability: positional effects of substitution in the hydrophobic core of the two-stranded alpha-helical coiled-coil. *Biochemistry* **32**:3178-87 https://doi.org/10.1021/bi00063a033

47. **Howes JM**, Bihan D, Slatter DA, Hamaia SW, Packman LC, Knauper V, Visse R, Farndale RW (2014) The recognition of collagen and triple-helical Toolkit peptides by MMP-13: Sequence specificity for binding and cleavage. *Journal of Biological Chemistry* **289**:24091-24101 https://doi.org/10.1074/jbc.M114.583443

48. **Holm L**, Laiho A, Toronen P, Salgado M (2023) DALI shines a light on remote homologs: One hundred discoveries. *Protein Science* **32**:e4519 https://doi.org/10.1002/pro.4519

49. **Abramson J**, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Jumper JM (2024) Accurate structure prediction of biomolecular interactions with AlphaFold. *Nature* **3**:493-500 https://doi.org/10.1038/s41586-024-07487-w

50. **Hamaia SW**, Pugh N, Raynal N, Nemoz B, Stone R, Gullberg D, Bihan D, Farndale RW (2012) Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin alpha1beta1 using collagen toolkits II and III. *Journal of Biological Chemistry* **287**:26019-28 https://doi.org/10.1074/jbc.M112.353144

51. **Bergsten H**, Palma Medina LM, Morgan M, Moll K, Skutlaberg DH, Skrede S, Wajima T, Svensson M, Norrby-Teglund A (2021) Adjunctive Rifampicin Increases Antibiotic Efficacy in Group A Streptococcal Tissue Infection Models. *Antimicrobial Agents and Chemotherapy* **65**:e0065821 https://doi.org/10.1128/AAC.00658-21

52. **Siemens N**, Kittang BR, Chakrakodi B, Oppegaard O, Johansson L, Bruun T, Mylvaganam H, Group IS, Norrby-Teglund A (2015) Increased cytotoxicity and streptolysin O activity in group G streptococcal strains causing invasive tissue infections. *Scientific Reports* **5**:16945 https://doi.org/10.1038/srep16945

53. **Pedchenko V**, Kitching AR, Hudson BG (2018) Goodpasture's autoimmune disease - A collagen IV disorder. *Matrix Biology* **71-72**:240-249 https://doi.org/10.1016/j.matbio.2018.05.004

54. **Reissmann S**, Gillen CM, Fulde M, Bergmann R, Nerlich A, Rajkumari R, Brahmadathan KN, Chhatwal GS, Nitsche-Schmitz DP (2012) Region Specific and Worldwide Distribution of Collagen-Binding M Proteins with PARF Motifs among Human Pathogenic Streptococcal Isolates. *PLoS One* **7** https://doi.org/10.1371/journal.pone.0030122

55. **Tolken LA**, Neufend JV, Oppegaard O, Methling K, Moll K, Redanz S, Katsburg MMD, Ali MQ, Siemens N (2024) Streptokinase reduces Streptococcus dysgalactiae subsp. equisimilis biofilm formation. *BMC Microbiology* **24**:378 https://doi.org/10.1186/s12866-024-03540-w

56. **Davies MA**, de Gier B, Guy RL, Coelho J, van Dam AP, van Houdt R, Matamoros S, van den Berg M, van Sorge NM (2025) Streptococcus pyogenes emm Type 3.93 Emergence, the Netherlands and England. *Emerging Infectious Diseases* **31**:229-236 https://doi.org/10.3201/eid3102.240880

57. **Skutlaberg DH**, Wiker HG, Mylvaganam H, Group IS, Norrby-Teglund A, Skrede S (2022) Consistent Biofilm Formation by Streptococcus pyogenes emm *1* Isolated From Patients With Necrotizing Soft Tissue Infections. *Frontiers in Microbiology* **13**:822243 https://doi.org/10.3389/fmicb.2022.822243

58. **Liu H**, Naismith JH (2009) A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expression and Purification* **63**:102-11 https://doi.org/10.1016/j.pep.2008.09.008

59. **Dinkla K**, Talay SR, Morgelin M, Graham RMA, Rohde M, Nitsche-Schmitz DP, Chhatwal GS (2009) Crucial role of the CB3-region of collagen IV in PARF-induced acute rheumatic fever. *PLoS One* **4**:e4666

60. **Raynal N**, Hamaia SW, Siljander PR, Maddox B, Peachey AR, Fernandez R, Foley LJ, Slatter DA, Farndale RW (2006) Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *Journal of Biological Chemistry* **281**:3821-31

61. **Delaglio F**, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular Nmr* **6**:277-93 https://doi.org/10.1007/BF00197809

62. **Vranken WF**, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins-Structure Function and Genetics* **59**:687-96 https://doi.org/10.1002/prot.20449

63. **McCoy AJ**, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658-674 https://doi.org/10.1107/S0021889807021206

64. **Emsley P**, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D* **66**:486-501 https://doi.org/10.1107/S0907444910007493

65. **Murshudov GN**, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section* **67**:355-67 https://doi.org/10.1107/S0907444911001314

66. **Chen VB**, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) . MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**:12-21 https://doi.org/10.1107/S0907444909042073

67. **Kaul R**, McGeer A, Low DE, Green K, Schwartz B (1997) Population-based surveillance for group A streptococcal necrotizing fasciitis: Clinical features, prognostic indicators, and microbiologic analysis of seventy-seven cases. Ontario Group A Streptococcal Study. *American Journal of Medicine* **103**:18-24 https://doi.org/10.1016/s0002-9343(97)00160-5

68. **Johansson L**, Thulin P, Sendi P, Hertzen E, Linder A, Akesson P, Low DE, Agerberth B, Norrby-Teglund A (2008) Cathelicidin LL-37 in severe Streptococcus pyogenes soft tissue infections in humans. *Infection and Immunity* **76**:3399-404 https://doi.org/10.1128/IAI.01392-07

69. **Thenmozhi R**, Balaji K, Kumar R, Rao TS, Pandian SK (2011) Characterization of biofilms in different clinical M serotypes of Streptococcus pyogenes. *Journal of Basic Microbiology* **51**:196-204 https://doi.org/10.1002/jobm.201000006

70. **Bolte S**, Cordelieres FP (2006) A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* **224**:213-32 https://doi.org/10.1111/j.1365-2818.2006.01706.x

Peer reviews

Reviewer #1 (Public review):

Summary:

Wojnowska et al. report structural and functional studies of the interaction of Streptococcus pyogenes M3 protein with collagen. They show through X-ray crystallographic studies that the N-terminal hypervariable region of M3 protein forms a T-like structure, and that the T-like structure binds a three-stranded collagen-mimetic peptide. They indicate that the T-like structure is predicted by AlphaFold3 with moderate confidence level in other M proteins that have sequence similarity to M3 protein and M-like proteins from group C and G streptococci. For some, but not all, of these related M and M-like proteins, AlphaFold3 predicts, with moderate confidence level, complexes similar to the one observed for M3-collagen. Functionally, the authors show that emm3 strains form biofilms with more mass when surfaces are coated with collagen, and this effect can be blocked by an M3 protein fragment that contains the T-structure. They also show the co-occurrence of emm3 strains and collagen in patient biopsies and a skin tissue organoid. Puzzlingly, M1 protein has been reported to bind collagen, but collagen inhibits biofilm in a particular emm1 strain but that same emm1 strain colocalizes with collagen in a patient biopsy sample. The implications of the variable actions of collagen on biofilm formation are not clear.

Strengths:

The paper is well written and the results are presented in a logical fashion.

Weaknesses:

A major limitation of the paper is that it is almost entirely observational and lacks detailed molecular investigation. Insufficient details or controls are provided to establish the robustness of the data.

Comments on revisions:

The authors' response to this reviewer's Major issue #1 is inadequate. Their argument is essentially that if they denature the protein, then there is no activity. This does not address the specificity of the structure or its interactions.

They went only part way to addressing this reviewer's Major issue #2. While Figure 8 - supplement 3 shows 1D NMR spectra for M3 protein (what temperature?), it does not establish that stability is unaltered (to a significant degree).

This reviewer's Major issue #3 is one of the major reasons for considering this study to be observational. This reviewer agrees that structural biology is by its nature observational, but modern standards require validation of structural observations. The authors' response is that a mechanistic investigation involving mutant bacterial strains and validation involving mutated proteins is beyond their scope. Therefore, the study remains observational.

Major issue 4 was addressed suitably, but brings up the problematic point that the emm1 2006 strain colocalizes quite well with collagen in a patient biopsy sample but not in other assays. This calls into question the overall interpretability of the patient biopsy data.

The authors have not provided a point-by-point response. Issues that were indicated to be minor previously were deemed to be minor because this reviewer thought that they could easily be addressed in a revision. It appears that the authors have ignored many of these comments, and these issues are therefore now considered to be major issues. For example, no errors are given for Kd measurements, Table 2 is sloppy and lacks the requested information, negative controls are missing (Figure 10 - figure supplement 1), and there is no indication of how many independent times each experiment was done.

And "C4-binding protein" should be corrected to "C4b-binding protein."

https://doi.org/10.7554/eLife.105539.2.sa2

Reviewer #2 (Public review):

Streptococcus pyogenes, or group A streptococci (GAS) can cause diseases ranging skin and mucosal infections, plasma invasion, and post-infection autoimmune syndromes. M proteins are essential GAS virulence factors that include an N-terminal hypervariable region (HVR). M proteins are known to bind to numerous human proteins; a small subset of M proteins were reported to bind collagen, which is thought to promote tissue adherence. In this paper, authors characterize M3 interactions with collagen and its role in biofilm formation. Specifically, they screened different collagen type II and III variants for full-length M3 protein binding using an ELISA-like method, detecting anti-GST antibody signal. By statistical analysis, hydrophobic amino acids and hydroxyproline found to positively support binding, whereas acidic residues and proline negatively impacted binding. The authors applied X-ray crystallography to determine the structure of the N-terminal domain (42-151 amino acids) of M3 protein (M3-NTD). M3-NTD dimmer (PDB 8P6K) forms a T-shaped structure with three helices (H1, H2, H3), which are stabilized by a hydrophobic core, inter-chain salt bridges and hydrogen bonds on H1, H2 helices, and H3 coiled coil. The conserved Gly113 serves as the turning point between H2 and H3. The M3-NTD is co-crystalized with a 24-residue peptide, JDM238, to determine the structure of M3-collagen binding. The structure (PDB 8P6J) shows that two copies of collagen in parallel bind to H1 and H2 of M3-NTD. Among the residues involved binding, conserved Try96 is shown to play a critical role supported by structure and isothermal titration calorimetry (ITC). The authors also apply a crystal-violet assay and fluorescence microscopy to determine that M3 is involved in collagen type I binding, but not M1 or M28. Tissue biopsy staining indicates that M3 strains co-localize with collagen IV-containing tissue, while M1 strains do not. The authors provide generally compelling evidence to show that GAS M3 protein binds to collagen, and plays a critical role in forming biofilms, which contribute to disease pathology. This is a very well-executed study and a well-written report relevant to understanding GAS pathogenesis and approaches to combatting disease; data are also applicable to emerging human pathogen Streptococcus dysgalactiae. One caveat that was not entirely resolved is if/how different collagen types might impact M3 binding and function. Due to the technical constrains, the in vitro structure and other binding assays use type II collagen whereas in vivo, biofilm formation assays and tissue biopsy staining use type I and IV collagen; it was unclear if this difference is significant. One possibility is that M3 has an unbiased binding to all types of collagens, only the distribution of collagens leads to the finding that M3 binds to type IV (basement membrane) and type I (varies of tissue including skin), rather than type II (cartilage).

Comments on revisions:

We are glad to see that the authors addressed our prior comments on M3 binding to different types of collagens in discussion section; adding a prediction of M3 binding to type I collagen (Figure 8-figure supplement 1B and 1C) is helpful to fill in the gap. Although it would be nice to experimentally fill in the gap by putting all types of collagens into one experiment (For example, like Figure 9A, use different types of human collagens to test biofilm formation; or Figure 10, use different types of human collagens to compete for biofilm formation), this appears to be beyond the scope of this paper. Meanwhile, the changes they have made are constructive.

The authors have addressed the majority of our prior comments.

https://doi.org/10.7554/eLife.105539.2.sa1

Author response:

The following is the authors' response to the current reviews.

We thank the reviewers for their comments on the initial submission, which helped us improve and extend the paper. We would like to respond specifically to reviewer #1.

We disagree with the broad criticism of this study as being "almost entirely observational" and lacking "detailed molecular investigation". We report structures and binding data, show mechanistic detail, identify critical residues and structural features underlying biological activity, and present biologically meaningful data demonstrating a role of the interaction of the M3 protein with collagens. We disagree that insufficient details or controls are included. We agree that our report has limitations, such as an understanding of potential emm1 strain binding to collagen, which might play a role in host tissue colonization, but not in biofilm.

In response to issues raised in the initial review, we conducted several new experiments for the revised manuscript. We believe these strengthen what we report. Firstly, as the reviewer suggested, we conducted a binding experiment where the tertiary fold of M3-NTD was disrupted to confirm the T-shaped fold is indeed required for binding to collagen, as might be expected based on the crystal structure of the complex. To achieve this, we did not, as the reviewer states, use denatured protein in the ITC binding experiment. Instead, we used a monomeric form of M3-NTD, which does not adopt a well-defined tertiary structure, but retains all residues in the context of alpha helices. Secondly, we added more evidence for the importance of structural features (amino acid side chains defining the collagen binding site) by analysing the role of Trp103. Together, we provide clear evidence for the specific role of the T-shaped fold of M3-NTD for collagen binding.

Responding to a constructive criticism by reviewer #1 we characterised M3-NTD mutants to demonstrate conservation of overall structure. NMR is an exquisite tool for this as it is highly sensitive to structural changes. It is not clear why the reviewer suggested we should have measured the stability of the proteins, which is irrelevant here. What matters is that the fold is conserved between mutated variants at the chosen experimental temperature (now added to the Methods section), which NMR demonstrates.

We added errors for the ITC-derived dissociation constants.

In the submitted versions of the paper we did not include the negative control requested by reviewer #1 for experiments shown in Figure 10 - figure supplement 1B. In our view this does not add information supporting our findings. However, we have now added two negative controls, staining of emm1 and emm28 strains. As expected, no reactivity was found with the type-specific M3 HVR antiserum while the M3 BCW antiserum showed weak reactivity, in line with some sequence similarity of the C-terminal regions of M proteins.

Table 2 contains essential information, in line with what generally is shown in crystallographic tables in this journal. All other information can be found in the depositions of our data at the PDB. The structures have been scrutinised and checked by the PDB and passed all quality tests.

We stated how many times experiments were done where appropriate. We now added this information for CLC assays (as given in the previously published protocol, refs. 45, 47). ITC was carried out more than once for optimization but the results of single experiments are shown (as is common practice).

The following is the authors' response to the original reviews.

Many thanks for assessing our submission. We are grateful for the reviews that have informed a revised version of the paper, which includes additional data and modified text to take into account the reviewers' comments.

We addressed the major limitation identified by Reviewer #1 by including data to demonstrate that collagen binding is indeed dependent on the T-shaped fold (major issue 1). Reviewer #1 suggested this needs to be done through extensive mutational work. This in our view was neither feasible nor necessary. Instead, we used ITC to measure collagen peptide binding using a monomeric form of M3, which preserves all residues including the ones involved in binding, but cannot form the T-shaped structure. This achieves the same as unravelling the T fold through mutations, but without the risk of aJecting binding through altering residues that are involved in both binding and definition of the T fold. The experiment shows a very weak interaction, confirming the fold of the M3-NTD is required for binding activity.

Reviewer #1 finds the study limited for being "almost entirely observational". Structural biology is by its nature observational, which is not a limitation but the very purpose of this approach. Our study goes beyond observing structures. In the first version of our paper, we identified a critical residue within a previously mapped binding site, and demonstrated through mutagenesis a causal link between presence of this residue on a tertiary fold and collagen binding activity. However, we agree this analysis could have been strengthened by additional mutagenesis, which we carried out and describe in the revised manuscript. This identifies a second residue that is critical for collagen binding. We firmed up these mutational experiments with a characterisation of mutated forms of M3 by NMR spectroscopy to confirm that these mutations did not aJect the overall fold, addressing major issue no. 2 of reviewer #1. We further demonstrate that the interaction between M3 and collagen is the cause of greatly enhanced biofilm formation as observed in patient biopsies and a tissue model of infection. We show that other streptococci that do not possess a surface protein presenting collagen binding sites like M3 do not form collagen-dependent biofilm. We therefore do not think that criticising our study for being almost entirely observational is valid.

Major issue 3:

We agree with the reviewer that it would be useful to carry out experiments with k.o. and complemented strains. Such experiments go beyond the scope of our study, but might be carried out by us or others in the future. We disagree that emm1 is used "as a negative". Instead, we established that, in contrast to emm3 strains, emm1 strain biofilm formation is not enhanced by collagen.

We addressed major issue 4 by quantifying colocalizations in the patient biopsies and 3D tissue model experiments.

We thank Reviewer #2 for the thorough analysis of our reported findings. The main criticism here (issue 1) concerns the question of whether binding of emm3 streptococci would diJer to

diJerent types of collagen. Our collagen peptide binding assays together with the structural data identify the collagen triple helix as the binding site for M3. While collagen types diJer in their distribution, functions and morphology in diJerent tissues, they all have in common triple-helical (COL) regions with high sequence similarity that are non-specifically recognised by M3. Therefore, our data in conjunction with the body of published work showing binding to M3 to collagens I, II, III and IV suggest it is highly likely that emm3 streptococci will indeed bind to all types of collagen in the same manner. We added a statement to the manuscript to make this point more clearly. We also added a prediction of a complex between M3 and a collagen I triple-helical peptide, which supports the idea of conserved binding mechanism for all collagen types. Whether this means all collagen types in the various tissues where they occur are targeted by emm3 streptococci is a very interesting question, however one that goes beyond the scope of our study.

Minor issues identified by the reviewers were addressed through changes in the text and addition of figures.

Summary of changes:

(1) Two new authors have been added due to inclusion of additional data and analysis.

(2) New experimental data included in section "M3-NTD harbors the collagen binding site".

(3) Figure 3 panels A and B assigned and swapped.

(4) Figure 4 changed to include new data and move mutant M3-NTD ITC graphs to supplement.

(5) Table 2 corrected and amended.

(6) AlphaFold3 quality parameters ipTM and pTM added to all figures showing predicted structures.

(7) New supplementary figure added showing crystal packing of M3-NTD/collagen peptide complex.

(8) Figure supplement of predicted M-protein/collagen peptide complexes includes new panel for a type I collagen peptide bound to M3.

(9) New figure supplement showing mutant M3-NTD ITC data.

(10) New figure supplement showing 1D ^{1}H NMR spectra of M3-NTD mutants.

(11) Included data for additional M3-NTD mutants assessing role of Trp103 in collagen binding. Text extended to describe and place into context findings from ITC binding studies using these mutants.

(12) Added quantitative analysis of biopsy and tissue model data (Mander's overlap coeJicient).

(13) Corrected and extended table 3 to take into account new primers.

(14) Added experimental details for new NMR and ITC experiments as well as new quantitative image analysis.

(15) Minor adjustments to the text to improve clarity and correct errors.

https://doi.org/10.7554/eLife.105539.2.sa0