nature chemistry

Article

https://doi.org/10.1038/s41557-024-01573-2

Heterotrimeric collagen helix with high specificity of assembly results in a rapid rate of folding

Received: 28 September 2023

Accepted: 6 June 2024

Published online: 15 July 2024

Check for updates

Carson C. Cole[1], Douglas R. Walker[1], Sarah A. H. Hulgan[1], Brett H. Pogostin[2], Joseph W. R. Swain[1], Mitchell D. Miller[3], Weijun Xu[3], Ryan Duella[1], Mikita Misiura[1], Xu Wang[4], Anatoly B. Kolomeisky[1,5,6], George N. Philips Jr[1,3] & Jeffrey D. Hartgerink[1,2]

The most abundant natural collagens form heterotrimeric triple helices. Synthetic mimics of collagen heterotrimers have been found to fold slowly, even compared to the already slow rates of homotrimeric helices. These prolonged folding rates are not understood. Here we compare the stabilities, specificities and folding rates of three heterotrimeric collagen mimics designed through a computationally assisted approach. The crystal structure of one ABC-type heterotrimer verified a well-controlled composition and register and elucidated the geometry of pairwise cation–π and axial and lateral salt bridges in the assembly. This collagen heterotrimer folds much faster (hours versus days) than comparable, well-designed systems. Circular dichroism and NMR data suggest the folding is frustrated by unproductive, competing heterotrimer species and these species must unwind before refolding into the thermodynamically favoured assembly. The heterotrimeric collagen folding rate is inhibited by the introduction of preformed competing triple-helical assemblies, which suggests that slow heterotrimer folding kinetics are dominated by the frustration of the energy landscape caused by competing triple helices.

Collagen-like proteins are found in all domains of life, including in prokaryotes and archaea. Their unifying structural feature is the triple helix. Prokaryotes use collagen-like proteins to bind to hosts during pathogenesis and for structural functions[1]. In humans, collagen triple helices form the basis of more than fifty different types of collagen-like proteins[2,3]. Fibre-forming collagens can be used as templates for biomineralization processes that give rise to rigid structures[4]. The proteins C1q and SPA, which contain collagen-like triple-helical domains, play essential roles in innate immunity[5,6]. Most natural collagen-like proteins exist as heterotrimers, where not all strands in the triple helix are equivalent.

Synthetic, self-assembled triple helices of collagen mimetic peptides have served as models to study the structure and the stability of natural protein triple helices. Triple helices are the tertiary structure formed from three left-handed PPII monomers that fold into a supercoiled, right-handed helix[7,8]. The common sequential motif of collagen is X-Y-Gly where X and Y are most frequently the amino acids proline (P) and 4-hydroxyproline (O), respectively[9]. The three peptide strands in a triple helix are offset by a single amino acid, which differentiates the leading, middle and trailing peptide strands. Therefore, spatial registration differentiates an ABC register from ACB, BAC, BCA, CAB or CBA registers. Two key inter-strand pairwise interactions are termed axial

[1]Department of Chemistry, Rice University, Houston, TX, USA. [2]Department of Bioengineering, Rice University, Houston, TX, USA. [3]Department of Biosciences, Rice University, Houston, TX, USA. [4]Shared Equipment Authority, Rice University, Houston, TX, USA. [5]Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, USA. [6]Department of Physics and Astronomy, Rice University, Houston, TX, USA. e-mail: jdh@rice.edu

https://doi.org/10.1038/s41557-024-01573-2

Table 1 | Heterotrimer sequences of HT1–HT3 and affiliated thermostability of binary and ternary mixtures

Monomer[a]	Sequence[b]	CD T_m (°C)			
HT1 A	KPGYPGDRGPOGQRGKRGPPGDRGFPGFOG	AB	AC	BC	ABC
HT1 B	PRGLKGLOGAQGPOGAKGLKGPRGYKGLOG	DNF[d]	DNF[d]	12.5	23.5
HT1 C	PKGYOGDOGIAGPDGPKGDQGDRGAOGDOG	Specificity = 11.0 °C			
HT2 A[c]	PKGPKGFOGPOGFKGFKGPKGPOGFKGPOG	AB	AC	BC	ABC
HT2 B[c]	PKGDOGDKGPOGPOGDKGDOGDKGPKGDOG	20.0	19.5	21.0	39.5
HT2 C[c]	PRGEPGPRGERGPPGPPGERGPPGEPGEPG	Specificity = 18.5 °C			
HT3 A	PKGROGPKGFOGYOGPRGROGKKGPRGPOG	AB	AC	BC	ABC
HT3 B	SKGDOGPOGDRGPKGPOGYKGPOGDKGFRG	17.5	17.0	14.0	44.0
HT3 C	PDGDRGPRGPOGYOGDDGPEGDOGPPGDOG	Specificity = 26.5 °C			

[a]All samples were prepared in a ratio of 1:1 (binary mixtures) or 1:1:1 (ternary mixtures) for a final concentration of 0.3 mM. Specificity = T_m (most stable) − T_m (second most stable). [b]Underlined amino acids are ^{15}N isotopically labelled. [c]Sequence is adapted from ref. 32. [d]No thermal transition observed. DNF, did not fold.

and lateral[10] and are essential for designing and controlling the register and composition of triple helices. Lys–Asp (KD), Arg–Tyr (RY) and Gln–Phe (QF) axial interactions have all been used to design triple helices and understand the chemical basis for collagen folding[11–14]. Increasing amino acid diversity beyond these interactions is hampered by the fragility of the assemblies and is substantially more challenging when attempting to control the registration of a heterotrimeric triple helix.

While much of the literature has focused on understanding the collagen triple helix's stability upon assembly, less attention has been paid to studying the folding pathway from free monomer to helix, especially in heterotrimers. The folding process for traditional collagens is believed to be driven by hydrophilic, rather than hydrophobic, interactions[15–17]. Collagen mimetic peptide folding starts with a nucleation event at the C terminus[18–20]. After nucleation, molecular dynamics and experimental studies with NMR have shown that *cis–trans* isomerization is the rate-limiting step for helical propagation down the axis[21–23]. This mechanism has led to the hypothesis that the rate of collagen folding is effectively concentration independent[23]. For natural collagen folding, isomerases catalyse the *cis–trans* isomerization and once in the necessary *trans* state, folding can commence[24–26]. Alternative theories suggest that the leading, middle and trailing strands' folding propagate at different rates[27]. Finally, it has also been proposed that two strands can associate first, building a template for the third strand, or all three strands can simultaneously fold[21].

The kinetics of folding have been modelled extensively in simple homotrimers with the sequence $(POG)_n$ (ref. 19). The rate of folding for homotrimers has been found to vary according to the position along the axis, with C-terminal folding being second order, while the central amino acids are best modelled as first order[27]. In other studies, covalently tethered homodimers have been used to enhance the kinetic rate of assembly, similar to disulfide bond formation in natural collagens[28,29].

Folding rates have yet to be studied in collagen heterotrimers for several reasons. First, the design of well-controlled, synthetic heterotrimers has only recently been accomplished[30–32]. Second, quantifying the folding pathway and delineating if there are intermediates forming is challenging and requires the confirmation of both composition and register throughout the folding process[33–35]. Finally, the heterotrimeric collagen helix is slow to form, frequently requiring days or even months[6,36], and the pathway by which a competing heterotrimer can interchange with the most stable triple helix is unknown. This slow formation is contrary to the typical rapid formation of globular proteins such as lysozyme, which can fold on the order of microseconds to milliseconds[37,38]. Triple-helical homotrimers are much slower (on the order of minutes), but still orders of magnitude faster than heterotrimers[27].

Herein we compare a series of synthetic ABC heterotrimeric collagen helices, heterotrimer 1 (HT1), heterotrimer 2 (HT2) and heterotrimer 3 (HT3), to investigate the kinetics and mechanism of folding. These were designed with the assistance of our scoring function, SCEPTTr (ref. 32). Of these, HT3 is notable because it possesses a higher specificity over other compositions and registers in the folding landscape, relative to previously published triple helices. By using circular dichroism (CD) and NMR, specific folding to only the ABC register was confirmed. A successful crystal structure of HT3 reveals molecular-level insight into the geometry and interactions of cation–π and axial and lateral salt bridges. Folding rates were compared and HT3 was found to fold the most rapidly. The mechanism of the much-faster-folding HT3 was investigated in detail. Notably, when competing species are in solution, they hinder the folding process because off-target triple helices must first unwind into free monomer and can only then fold into the ABC register. This shows how an off-target folding event is difficult to remediate for heterotrimers. Therefore, we conclude that the energy landscape of competing species governs the overall rate at which the triple-helix assembly occurs.

Results

Designing ABC-type heterotrimers with varying specificities

The collagen scoring algorithm, SCEPTTr, was used to aid in the design of three ABC-type heterotrimers, as shown in Table 1 (ref. 32). In the current work, HT1–HT3 are named from the lowest to highest melting temperature and specificity of their ABC-type assembly (Table 1). Melting temperature (T_m) is assigned to the minimum of the derivative of a thermal melting curve as determined by CD, and specificity is defined as the T_m of the most stable species minus the T_m of the second most stable species. HT2 is a previously published ABC-type heterotrimer and was designed with axial charge pair and cation–π interactions[32]. The design of HT2 follows a straightforward geometric stabilization with the strongest pairwise charge pairing (K–D axial interactions) being used from strands A to B. Interactions between B and C strands used K–E axial charge pair interactions, while the C to A strands used R–F axial cation–π interactions. Additionally, prolines were incorporated at select Y positions instead of hydroxyproline as a negative design to destabilize C homotrimers. Previous study[32] of HT2 has shown that it forms the correct ABC register in solution and with good thermal stability and specificity (T_m = 39.5 °C, ΔT_m = 18.5 °C).

The novel heterotrimers, HT1 and HT3, use these stabilizing design elements while also incorporating other stabilizing interactions and non-interacting, destabilizing amino acids. Hydrophobic amino acids such as leucine, isoleucine, alanine and Y-substituted prolines were incorporated into HT1's sequence to limit the number of viable competing species (specificity, ΔT_m = 11.0 °C; Table 1). However, these amino acids also destabilize the ABC-type assembly (T_m = 23.5 °C). The design of HT3 incorporates the most diverse set of stabilizing and destabilizing

https://doi.org/10.1038/s41557-024-01573-2

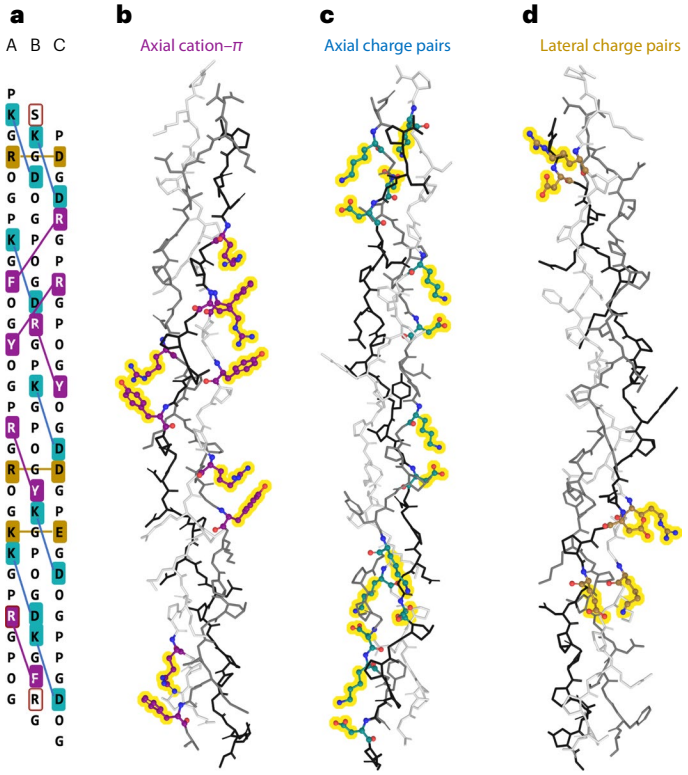

Fig. 1 | Sequence and crystal structure of HT3. a, The sequence of HT3 shows a variety of pairwise interactions with axial charge pairs (blue) and lateral charge pairs (gold), axial cation–π interactions (purple) and non-interacting residues (brown box). A single amino acid stagger arises between each strand in the triple helix. **b**, Crystal structure of HT3 where the A strand is white, the B strand is grey and the C strand is black. The cation–π interactions between arginine and tyrosine or phenylalanine are highlighted. **c**, Axial charge pairs constituted of lysine and aspartate amino acids. **d**, Lateral charge pairs between lysine or arginine with glutamate or aspartate.

elements. HT3 uses non-interacting amino acids (B-Ser1, B-Arg29 and C-Pro26) to promote the designed ABC register. In HT3's A to B strands, there are three KD axial charge pairs and two axial cation–π interactions (RY and RF). The B to C strand interaction has three axial KD charge pairs. Finally, from the C to A strands, there are three lateral salt bridges with two axial cation–π interactions. This design yielded not only a high stability of T_m = 44.0 °C but also a high specificity of ΔT_m = 26.5 °C. All pairwise interaction geometries that direct higher-order assembly are highlighted for HT1–HT3 in Supplementary Fig. 10.

Seven mixtures of all components of the heterotrimers were studied with CD to determine melting temperatures of the various compositions (Fig. 2a–c). NMR was then used to confirm a distinct composition and ABC register in solution for HT3, which is presented in the Supplementary Information (Supplementary Figs. 14 and 15). Subsequent crystallization of HT3 (Protein Data Bank (PDB) entry 8TW0) also confirmed the correctly predicted composition and register for the triple helix in addition to confirmation of the theorized geometries of the pairwise interactions employed.

Crystal structure of HT3 at 1.53 Å resolution

The crystal structure of HT3 revealed a well-ordered collagen triple helix with a backbone confirmation as expected. The triple helix was confirmed to have an ABC composition and register. In the crystal structure, pairwise axial cation–π interactions have an average interaction distance of 4.2 Å between the centre of the aromatic ring of Tyr (Y) or Phe (F) with the δ-guanidinium cation's central ζ-carbon of Arg (R; Fig. 1b). The crystal structure also demonstrates that even in cases where a possible cation–π lateral geometry can occur, such as between B-Lys14 and C-Tyr13, the axial conformer is preferred (Supplementary Fig. 12). The two theorized geometries, planar (between C-Arg5 and A-Phe10) and T-shaped (C-Arg8 and A-Tyr13)[39], were also observed for the arginine-containing pairwise interactions.

The KD axial charge pairs, highlighted in Fig. 1c, showed the closest distance for any pairwise interaction with the average distance of 3.6 Å and near ideal N–O hydrogen bonding distance of 2.7 Å, as measured between the ζ-nitrogen of Lys (K) and the γ-carbon of Asp (D). This once again demonstrates the robustness of the K–D axial interaction in the collagen de novo design. Reverse-lateral charge pairs were measured from the ζ-nitrogen of Lys (K) or the ζ-carbon of the δ-guanidinium cation of Arg (R) to the γ-carbon of Asp (D) or the δ-carbon of Glu (E) and had an average interaction distance of 3.9 Å and a corresponding hydrogen bonding distance of 2.8 Å (Fig. 1d). The N and C termini of the helix have weaker density and higher B-factors, indicating some disorder. The final model contains a major conformer for all residues except for the side chain of B-Arg29, all of B-Gly30 and the C termini of C-Gly30, which are too disordered to model their major conformation. B-Arg29, in particular, was intentionally designed not to have a pairwise interaction partner; the lack of order in this region supports its non-interacting state. All pairwise interactions are summarized in Supplementary Table 1.

The PPII secondary structure yields a maximum CD signal at around 225 nm, and the mean residual ellipticity is correlated to triple-helix assembly. The maximum mean residual ellipticity for each heterotrimer was determined by measuring the CD of the folded ABC mixture after one month of folding at 4 °C. For all experiments, each sample's observed mean residual ellipticity was then divided by this maximum value at 5 °C to yield a 'fraction of maxima'. Monitoring the triple-helix formation in the case of ABC-type heterotrimer formation does not guarantee that solely the ABC-type register is observed; however, we did observe relative folding rates for HT1, HT2 and HT3. The folding of HT3 reached completion in two days (Fig. 2d). HT1 and HT2, because of their slower folding nature, were monitored for thirty days. The maximum signal was then used to normalize the kinetic plots and are presented in Table 2. To fold to 50% of the observed fraction, HT1 and HT2 took 147 h and 27 h, respectively, considerably slower than HT3, which took only 0.50 h.

Monitoring the formation of HT3 at various concentrations and temperatures, without denaturant, yields further insight into heterotrimer folding. In Fig. 2e, we see that the folding kinetics are concentration independent, in the range of 150–400 µM (Supplementary Table 3). The concentration independence supports the notion that the nucleation step in collagen folding does not govern the folding paradigm. Therefore, similar to homotrimer folding, heterotrimer folding could be rate limited by isomerization[26].

A temperature variability study shown in Fig. 2f appeared to show a decrease in folding rate as temperature increased (Supplementary Table 4). Thermal melting derivative curves of HT3 in Fig. 2c show a loss in CD signal starting at 20 °C. Therefore, we correlated the maxima for each temperature (Supplementary Fig. 15a) and used those maxima to plot the fraction folded for each temperature. At temperatures below 20 °C, the heterotrimer folding rate was consistent and agrees with previous studies on homotrimers[23]. However, at temperatures greater than 20 °C, the folding rate was reduced, presumably due to increasing denaturation[40].

Heterotrimer folding kinetics with NMR

The CD experiments described above provided relative folding rates but are limited in their ability to distinguish systems with mixtures of compositions and/or registers (Supplementary Fig. 17b). Using NMR, folding specific to particular compositions and registers could be investigated. The formation of HT3 was observed with heteronuclear single quantum coherence (HSQC) NMR by labelling peptide strands

https://doi.org/10.1038/s41557-024-01573-2

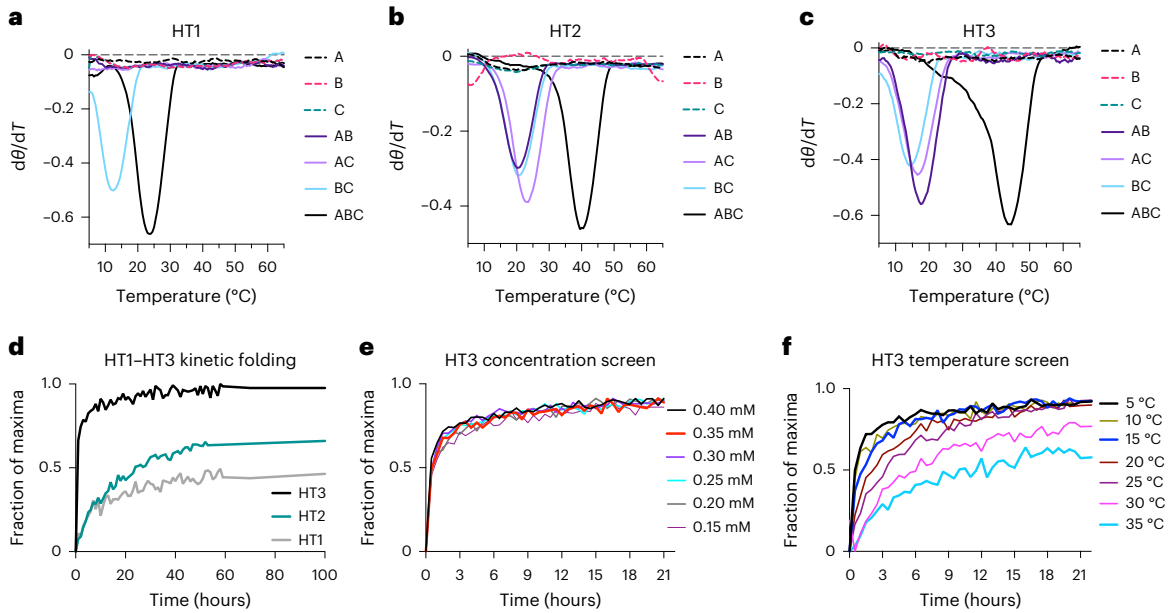

Fig. 2 | Monitoring heterotrimer folding with CD. a, HT1 melt derivative curves as determined by CD. dθ is change in mean residue ellipticity (MRE) and dT is the change in temperature (°C). The minima of the curves signify the thermal melt transition temperature (T_m) and the curves of HT1 and its unary (A, B and C), binary (AB, AC and BC) and ternary (ABC) mixtures yields a specificity of 11.0 °C. **b**, HT2 melt derivative curves from each constituting mixture, where the specificity is 18.5 °C. **c**, HT3 melt derivative curves from each constituting mixture, where the specificity of the mixture is 26.5 °C. **d**, Kinetic folding curves of HT1–HT3 as observed by CD. Curves are normalized by a maximum value for each heterotrimer after folding for days (λ_{max}). **e**, CD kinetic folding curves of the HT3 ABC solution at variable total peptide concentration folding values at 5 °C. **f**, Kinetic folding curves of a temperature screen of HT3 folding at 0.3 mM.

Table 2 | CD kinetics of HT1, HT2 and HT3

ABC mixture	Percent folded[a]	Time (h)	λ_{max} (mDeg)
HT1			
	25	6.6	
	50	147	
	75	410	
	95	691	83.0
HT2			
	25	6.6	
	50	27	
	75	262	
	95	610	60.5
HT3			
	25	0.04	
	50	0.50	
	75	2.12	
	95	43	99.5

[a]Percent folded values are reported from an observed circular dichroism signal, in mDeg, divided by the signal observed after 30 days of folding (λ_{max}). The first 55 hours of each folding experiment were continuously monitored and then extrapolated to the λ_{max} signal. All experimental values were determined at 0.30 mM and 5 °C.

A and B with a ^{15}N-labelled glycine at the 21 position, and strand C with a ^{15}N–^{13}C glycine at the 21 position (Table 1). Strand B also incorporated an amide ^{15}N–^{13}C lysine at the 2 position. We correlated each peak to the free monomer or assembled ABC heterotrimer (Supplementary Figs. 18 and 19). These monomer and trimer peaks are in equilibrium with each other in solution[27]. In the HSQC of spectra as shown in Fig. 3e,f, the monomer strands (M_x, where x is A, B or C) are distinct in their chemical shifts from the three trimer strands (T_x).

Samples were heated to 55 °C, well above the 44.0 °C melting temperature, to ensure the system was fully denatured, and then cooled back to 5 °C to monitor folding. All NMR kinetics experiments were conducted in triplicate. Figure 3b plots the time course of the raw HSQC signal intensities of the A, B and C strands upon assembling into the ABC heterotrimer. The intensity of each observed peptide strand of the triple helix is statistically similar. The sum of these peak intensities, shown in black, gives the rate of triple-helix formation, which, unlike CD, is specific to both composition and register (Supplementary Fig. 18). As expected, we observed that as the ABC heterotrimer is formed, the monomer signals from all peptides decreased, as seen in Fig. 3c.

These kinetic data were found to be best fit with a two-phase logarithmic function. The folding curve profile possesses an initial fast folding rate and then a long plateau correlated with a slower rate that leads to full folding after approximately two days. Our hypothesis is that free monomer in solution can relatively quickly form into the target ABC-type triple helix. However, small amounts of alternative triple helices also fold, and the second, slow rate is based on the unfolding of these species followed by the refolding to the target ABC system.

To investigate the possibility of the folding pathway being frustrated by competing triple helices, several additional NMR studies were performed in which the ABC triple helix was forced to fold from prefolded alternate compositions. Binary mixtures of AB, AC and BC were separately folded into their preferred triple-helical structure by annealing followed by one-week equilibration at 4 °C. The AB, AC and BC mixtures form a number of competing triple helices with melting temperatures of 17.5, 17.0 and 14.0 °C, respectively (Table 1 and Fig. 2d). However, it is important to note that the 'AB' mixture after equilibrium contains A monomer and B monomer as well as triple helices of one or more different compositions and registers, for example, AAB, ABA, BAA, ABB, BAB and/or BBA. The AC and BC mixtures are similarly complex. The missing peptide strand was then added to the preformed heterotrimer solution to result in a final molar mixing ratio of 1:1:1 A/B/C. In a fourth experiment, all three prefolded systems, AB, AC and BC, were mixed in a 1:1:1 ratio. Folding of these four systems was

https://doi.org/10.1038/s41557-024-01573-2

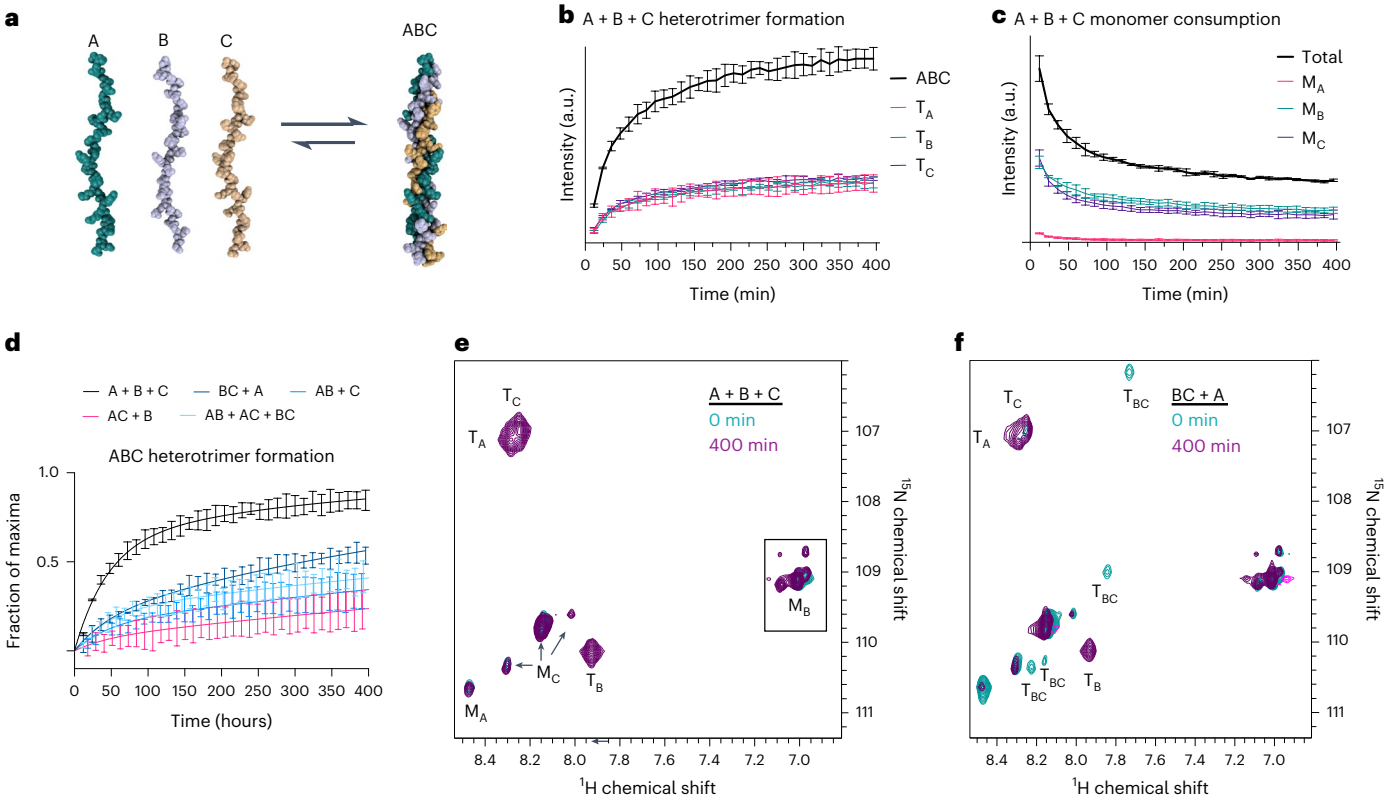

Fig. 3 | Pathway of free monomer folding into HT3. a, Scheme of monomers folding to a trimer in an uninhibited system. **b**, The kinetic folding curve of each individual trimer strand (T_x) of HT3 and their total intensity (black). **c**, The curve for the consumption of free monomers (M_x) and their total decrease (black). **d**, Comparative rates of folding as observed by NMR for the uninhibited system and four different inhibited systems. All plots were fit with a constrained two-phase plateau function. **e**, HSQC spectra from the initial and final time-points of the heated and then immediately cooled A + B + C HT3 system with the free monomers assembling. **f**, The same HSQC spectra but with a BC heterotrimer (T_{BC}) in solution. The BC heterotrimer peaks disappear after HT3 assembles. Spectra were obtained at 5 °C. All experiments were performed in triplicate, and data are presented as mean values ± s.d.

monitored as before by HSQC (Supplementary Fig. 21). The rates of ABC triple-helix formation are plotted in Fig. 3d for the simple A + B + C folding as well as the four inhibited-folding experiments AB + C, AC + B, BC + A and AB + AC + BC.

We observed that the presence of preformed non-target triple helices substantially slowed the rate of folding as compared to the mixture that initially contained only monomers. Of these, preformed AC resulted in the most dramatic rate reduction while preformed BC resulted in the least impact. All four inhibited-folding experiments could be fit by a two-phase fit in which the fast rate from the uninhibited A + B + C folding experiment was used in all cases, while the fractions of fast rate and slow rate were varied. (The Supplementary Information has a detailed discussion of the data workup and fitting.)

The BC + A solution (dark blue) possessed the best signal-to-noise ratio among the prefolded mixtures and has the lowest stability (T = 14.0 °C). This made it especially useful to follow the details of the system as it evolved and to compare this to the simpler A + B + C folding system. In the A + B + C folding experiment, the formation of ABC triple helix is inversely proportional to the consumption of A, B and C monomers (Fig. 3b,c), but for the BC + A system, we instead observe a much slower consumption of B and C monomer (almost flat) over time while the ABC triple helix is formed. Instead, the competing BC heterotrimer and A monomer are consumed (Fig. 4b,c). As shown in Fig. 3f, the BC trimer peaks are no longer detectable after 400 min. These data support our hypothesis that preformed heterotrimers must unfold before they can refold to a more stable triple helix.

The general mechanism for heterotrimer formation is outlined in Fig. 4a. First, available monomers can associate to form the most stable ABC triple helix. However, in the various inhibited systems, this concentration is comparatively low as some peptides are in preformed, alternative triple helices. As monomer is consumed in the formation of the ABC triple helix, the equilibrium no longer favours the concentration of preformed alternate triple helices, which unwind, regenerating the monomer concentration and allowing the additional ABC triple-helix formation. Figure 4c shows that as A peptide is consumed, the BC heterotrimer also disappears. Figure 4d demonstrates that monomer consumption is markedly slower for the BC + A system compared to a free A + B + C system. The same trend was observed by monitoring the N-terminal labelled residues in each system (Supplementary Fig. 23). This demonstrates that the unwinding event of the competing heterotrimer species is necessary for the formation of the ABC assembly and shows the path-dependent folding for heterotrimeric collagens.

Off-target association of monomer strands in solution greatly impacts the rate of triple-helix assembly. This explains why, in the cases of HT1 and HT2, where competing species are much closer to the stability of the target helix, months have been required for heterotrimers to reach their fully folded state. Natural collagens presumably fold more readily due to interactions in their pro-peptide regions, through the assistance of chaperonins and the formation of disulfide and other covalent bonds, all of which help to guide and maintain the formation of the appropriate assembly[3,27].

Conclusion

The computationally aided design of HT3 generated a heterotrimer with a high specificity. To achieve this, HT3 uses a wide variety of stabilizing charge pairs, cation–π interactions and destabilizing,

https://doi.org/10.1038/s41557-024-01573-2

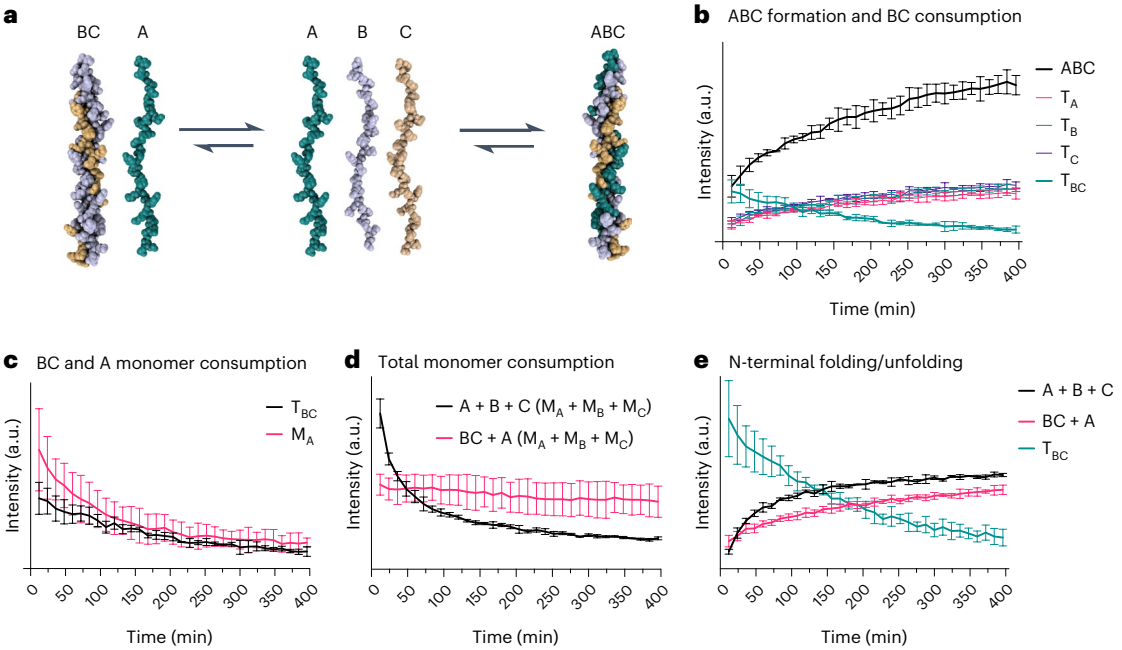

Fig. 4 | BC folding impacts HT3 formation. a, Illustration of the slow folding pathway of HT3 formation with preformed heterotrimer. Competing species (T_{BC}) unwind before forming the correct ABC-type heterotrimer. **b**, The formation of HT3 ABC (black) is inversely proportional to the loss of the competing T_{BC} heterotrimer. **c**, T_{BC} heterotrimer's disappearance correlates to the disappearance of the A monomer (M_A). **d**, Comparison of free monomer for A + B + C and BC + A. The preformed system with T_{BC} shows a relatively slower rate of monomer consumption. **e**, The N-terminal folding of each mixture as observed from the 2-Lys in strand B for each experimental set. Similar to the other experimental sets, the folding of ABC is more rapid when the sample has free monomers (A + B + C) compared with the frustrated preformed system. Spectra were obtained at 5 °C. All experiments were performed in triplicate, and data are presented as mean values ± s.d.

non-interacting amino acids. The design was structurally validated by X-ray crystallography, which demonstrated each of the intended pairwise amino acid interactions in the fully folded helix. The high specificity correlated to a much faster rate of folding compared to heterotrimers HT1 and HT2. As has been reported for homotrimeric helices, CD demonstrated that the rate of folding is largely independent of concentration. Monitoring rates by ^{1}H,^{15}N HSQC NMR allowed us to see the disappearance of starting materials and production of products with a precision that differentiates between alternative compositions and registrations of the helix and is impossible by simpler techniques such as CD. NMR experiments that begin with prefolded, off-path triple helices show dramatically slower rates of folding, supporting the conclusion that HT3's fast folding is due to the system's ability to avoid these kinetic traps. This complex de novo design success shows how a detailed supramolecular understanding of pairwise interactions can motivate the investigation of biologically relevant collagen sequences and the next generation of collagen mimetic biomaterials.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-024-01573-2.

References

1. Lukomski, S., Bachert, B. A., Squeglia, F. & Berisio, R. Collagen-like proteins of pathogenic streptococci. *Mol. Microbiol.* **103**, 919–930 (2017).
2. Thomas, A. H., Edelman, E. R. & Stultz, C. M. Collagen fragments modulate innate immunity. *Exp. Biol. Med.* **232**, 406–411 (2007).
3. Gordon, M. K. & Hahn, R. A. Collagens. *Cell Tissue Res.* **339**, 247–257 (2009).
4. Fratzl, P. & Weinkamer, R. Nature's hierarchical materials. *Prog. Mater. Sci.* **52**, 1263–1334 (2007).
5. Abendstein, L. et al. Complement is activated by elevated IgG3 hexameric platforms and deposits C4b onto distinct antibody domains. *Nat. Commun.* **14**, 4027 (2023).
6. Yu, L. T. et al. Hollow octadecameric self-assembly of collagen-like peptides. *J. Am. Chem. Soc.* **145**, 5285–5296 (2023).
7. Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958 (2009).
8. Brodsky, B. & Persikov, A. V. *Molecular Structure of the Collagen Triple Helix* Vol. 70 of *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers* (Academic Press, 2005). https://www.sciencedirect.com/science/article/pii/S0065323305700097
9. Shoulders, M. D., Guzei, I. A. & Raines, R. T. 4-Chloroprolines: synthesis, conformational analysis, and effect on the collagen triple helix. *Biopolymers* **89**, 443–454 (2008).
10. Persikov, A. V., Ramshaw, J. A. M. & Brodsky, B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280**, 19343–19349 (2005).
11. Cole, C. C. et al. Cation–π interactions and their role in assembling collagen triple helices. *Biomacromolecules* **23**, 4645–4654 (2022).
12. Cole, C. C. et al. Stabilization of synthetic collagen triple helices: charge pairs and covalent capture. *Biomacromolecules* **24**, 5083–5090 (2023).
13. Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A. & Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960–14967 (2000).
14. Hentzen, N. B., Islami, V., Köhler, M., Zenobi, R. & Wennemers, H. A lateral salt bridge for the specific assembly of an ABC-type collagen heterotrimer. *J. Am. Chem. Soc.* **142**, 2208–2212 (2020).

15. Canty, E. G. & Kadler, K. E. Collagen fibril biosynthesis in tendon: a review and recent insights. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **133**, 979–985 (2002).
16. Leikin, S., Rau, D. & Parsegian, V. Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions. *Nat. Struct. Biol.* **2**, 205–210 (1995).
17. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
18. Stultz, C. M. The folding mechanism of collagen-like model peptides explored through detailed molecular simulations. *Protein Sci.* **15**, 2166–2177 (2006).
19. Baum, J. & Brodsky, B. Folding of peptide models of collagen and misfolding in disease. *Curr. Opin. Struct. Biol.* **9**, 122–128 (1999).
20. Park, S., Klein, T. E. & Pande, V. S. Folding and misfolding of the collagen triple helix: Markov analysis of molecular dynamics simulations. *Biophys. J.* **93**, 4108–4115 (2007).
21. Hartmann, J. & Zacharias, M. Mechanism of collagen folding propagation studied by molecular dynamics simulations. *PLoS Comp. Biol.* **17**, e1009079 (2021).
22. Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L. & Raines, R. T. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **123**, 777–778 (2001).
23. Bachmann, A., Kiefhaber, T., Boudko, S., Engel, J. & Bächinger, H. P. Collagen triple-helix formation in *all-trans* chains proceeds by a nucleation/growth mechanism with a purely entropic barrier. *Proc. Natl Acad. Sci. USA* **102**, 13897–13902 (2005).
24. Favretto, F. et al. Catalysis of proline isomerization and molecular chaperone activity in a tug-of-war. *Nat. Commun.* **11**, 6046 (2020).
25. Chen, J., Edwards, S. A., Grater, F. & Baldauf, C. On the cis to trans isomerization of prolyl–peptide bonds under tension. *J. Phys. Chem. B* **116**, 9346–9351 (2012).
26. Buevich, A. V., Dai, Q.-H., Liu, X., Brodsky, B. & Baum, J. Site-specific NMR monitoring of cis–trans isomerization in the folding of the proline-rich collagen triple helix. *Biochemistry* **39**, 4299–4308 (2000).
27. Baum, J. & Brodsky, B. Real-time NMR investigations of triple-helix folding and collagen folding diseases. *Fold. Des.* **2**, R53–R60 (1997).
28. Tanrikulu, I. C., Westler, W. M., Ellison, A. J., Markley, J. L. & Raines, R. T. Templated collagen "double helices" maintain their structure. *J. Am. Chem. Soc.* **142**, 1137–1141 (2020).
29. Tanrikulu, I. C. & Raines, R. T. Optimal interstrand bridges for collagen-like biomaterials. *J. Am. Chem. Soc.* **136**, 13490–13493 (2014).
30. Fallas, J. A., Lee, M. A., Jalan, A. A. & Hartgerink, J. D. Rational design of single-composition ABC collagen heterotrimers. *J. Am. Chem. Soc.* **134**, 1430–1433 (2012).
31. Zheng, H. et al. How electrostatic networks modulate specificity and stability of collagen. *Proc. Natl Acad. Sci. USA* **115**, 6207–6212 (2018).
32. Walker, D. R. et al. Predicting the stability of homotrimeric and heterotrimeric collagen helices. *Nat. Chem.* **13**, 260–269 (2021).
33. Madhan, B., Xiao, J., Thiagarajan, G., Baum, J. & Brodsky, B. NMR monitoring of chain-specific stability in heterotrimeric collagen peptides. *J. Am. Chem. Soc.* **130**, 13520–13521 (2008).
34. Buevich, A. & Baum, J. Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases. *Philos. Trans. R. Soc. B* **356**, 159–168 (2001).
35. Saccà, B., Renner, C. & Moroder, L. The chain register in heterotrimeric collagen peptides affects triple helix stability and folding kinetics. *J. Mol. Biol.* **324**, 309–318 (2002).
36. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–2535 (2006).
37. Matagne, A., Radford, S. E. & Dobson, C. M. Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J. Mol. Biol.* **267**, 1068–1074 (1997).
38. Chaffotte, A. F., Guillou, Y. & Goldberg, M. E. Kinetic resolution of peptide bond and side chain far-UV circular dichroism during the folding of hen egg white lysozyme. *Biochemistry* **31**, 9694–9702 (1992).
39. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
40. Simpson, R. B. & Kauzmann, W. The kinetics of protein denaturation. I. The behavior of the optical rotation of ovalbumin in urea solutions. *J. Am. Chem. Soc.* **75**, 5139–5152 (1953).

Methods

X-ray crystallography

HT3 crystallization. The assembled HT3 heterotrimer was diluted to 20 mg ml^{-1} in MES buffer (10 mM, pH 6.1). HT3 was then annealed at 85 °C for 15 min before being cooled to room temperature. A diluted aliquot of HT3 was then confirmed to fold, using CD. Subsequently, high-throughput crystal screening was performed using a Mosquito LCP robot (SPT LabTech) to set up vapour diffusion sitting drop experiments with commercial screens from Hampton Research (IndexHT and PegRx-HT) and Rigaku Reagents (WizardHT screens 1–2 and 3–4). Drops formed by mixing 200 nl peptide and 200 nl reservoir solution. These were equilibrated against a reservoir of 200 µl at 20 °C. Of the initial 384 conditions screened, over 35 conditions produced crystals within one week, and more than 25 conditions had crystals large enough to mount.

Crystal data collection and structure refinement. A crystal looped from the Index-G8 condition (0.2 M ammonium acetate, 0.1 M HEPES buffer pH 7.5, 25% (w/v) polyethylene glycol 3350) was flash-cooled in a 100 K nitrogen stream (Oxford Cryosystems, Cryostream 800). Data were collected to 1.53 Å using a Rigaku XtaLAB Synergy-S-DW diffractometer equipped with a HyPix-Arc 150° hybrid photon counting detector and Cu microfocus sealed X-ray source. Data were integrated, scaled and merged within the CrysAllis Pro software package. Phasing was accomplished with Phaser (ref. 41) and Shelxe (ref. 42) within the Arcimboldo pipeline[43], which placed three copies of a blunt-end starting model derived from PDB entry 3t4f (ref. 44; residues A, 8–16; B, 7–15; C, 6–14 with Lys and Glu residues truncated at the Cβ position). Shelxe auto-traced 75 residues with a final correlation coefficient of 41% between the structure factor amplitudes calculated from the poly-alanine model with the measured amplitudes. ARP/wARP software (ref. 45) was used for building side chains and built 76 residues. The HT3 model was inspected and rebuilt in Coot (ref. 46) and refined with PHENIX (ref. 47). PyMOL was used for structure visualization and analysis. The structure was visualized using a collaborative three-dimensional graphics system[48]. Structural biology software applications used in this project were compiled and configured by SBGrid (ref. 49). Data processing and refinement statistics are summarized in Supplementary Table 2. Atomic coordinates of HT3 (PDB entry 8TW0) crystal structure have been deposited with the PDB.

NMR

Structural studies. All NMR experiments were carried out on a Bruker NEO 600 MHz NMR spectrometer equipped with a 5 mm TCI cryoprobe. Isotope-labelled collagen peptide was prepared at 2.7 mM in an NMR buffer containing 9 mM phosphate, 50 mM KCl, 10 µM trimethylsilyl propanoic acid and 90% H_2O/10% D_2O. The hydrogen and nitrogen carrier frequency were set to 4.7 ppm and 117 ppm, respectively. For the ^{1}H,^{1}H,^{15}N three-dimensional nuclear Overhauser effect spectroscopy (NOESY) HSQC and ^{1}H,^{15}N two-dimensional HSQC, the spectral width of 25 ppm was used in the nitrogen dimension. The proton carrier frequency was set at the water signal for all experiments. In the ^{15}N NOESY–HSQC, the first and second proton dimension values were identical to those used in the HSQC experiments. HT3 spectra for the ^{1}H,^{15}N HSQC were collected at 5 °C. For the HSQC experiments, 32 dummy scans were run before each experiment, and then eight scans were run for each transient in the experiments. For all HSQC experiments, 256 transients were collected in the nitrogen dimension. For the NOESY–HSQC experiments, 2,048 × 1 × 700 complex data points, with eight scans per data point, and 128 dummy scans were run. Three-dimensional ^{15}N NOESY–HSQC was performed with a mixing time of 100 ms at 25 °C. Regarding chemical shift referencing, ^{1}H chemical shifts were referenced to trimethylsilyl propanoic acid, and ^{15}N chemical shifts were referenced indirectly according to the gyromagnetic ratio[50]. NMR data were processed using Bruker TopSpin software. PyMOL and DLPacker were used to render energy-minimized structures to determine the correct register[51–53].

Kinetics. To quantify the kinetics of HT3 formation with NMR, HSQC experiments were optimized at 5 °C. Briefly, three independent solutions were prepared to have a final total peptide concentration of 2.7 mM in a 10% solution of D_2O with trimethylsilyl propanoic acid. For the formation of HT3 from free monomers (sample A + B + C), the sample was loaded and parameterized at 5 °C; then the sample was heated for one hour at 55 °C. After one hour, the sample was cooled to 5 °C and HSQC spectra were immediately collected with the same parameters as above. For all HSQC experiments, 256 transients were collected in the nitrogen dimension. Samples were collected for roughly 400 minutes, where the experimental time stamp was used to determine the step time between each collection. For the competing heterotrimer studies, such as BC + A, the heterotrimer and monomer strands were annealed and folded for one week. After folding, the offset frequency in parts per million (ppm), was determined on a sample prepared in parallel, and then both samples were kept at 5 °C. Samples were combined by first adding one component to the bottom of a 5 mm NMR tube, and the next component(s) was (were) held above the first sample mixture by capillary action. The tube was then capped and mixed before being immediately loaded into the magnet. After shimming data acquisition commenced after about 15 minutes. Peak intensities were compiled using Bruker's Dynamic Center and then were plotted in GraphPad Prism 10 software. All equations are reported and rationalized in the Supplementary Information.

Data availability

The authors declare that all the data supporting the findings of this research are available within the article and its Supplementary Information. HT3 crystal structures have been deposited with the PDB (entry 8TW0). Source data are provided with this paper.

References

41. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
42. Usón, I. & Sheldrick, G. M. An introduction to experimental phasing of macromolecules illustrated by *SHELX*; new autotracing features. *Acta Crystallogr. D Struct. Biol.* **74**, 106–116 (2018).
43. Caballero, I. et al. *ARCIMBOLDO* on coiled coils. *Acta Crystallogr. D Struct. Biol.* **74**, 194–204 (2018).
44. Fallas, J. A., Dong, J., Tao, Y. J. & Hartgerink, J. D. Structural insights into charge pair interactions in triple helical collagen-like proteins. *J. Biol. Chem.* **287**, 8039–8047 (2012).
45. Langer, G. G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Struct. Biol.* **66**, 486–501 (2010).
47. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
48. Yennamalli, R., Arangarasan, R., Bryden, A., Gleicher, M. & Phillips, G. N. Jr. Using a commodity high-definition television for collaborative structural biology. *J. Appl. Crystallogr.* **47**, 1153–1157 (2014).
49. Morin, A. et al. Cutting edge: collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
50. Wishart, D. S. et al. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140 (1995).
51. Misiura, M., Shroff, R., Thyer, R. & Kolomeisky, A. B. DLPacker: deep learning for prediction of amino acid side chain conformations in proteins. *Proteins Struct. Funct. Bioinform.* **90**, 1278–1290 (2022).
52. Gauba, V. & Hartgerink, J. D. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.* **129**, 2683–2690 (2007).

53. Kawahara, K. et al. Polymorphism of collagen triple helix revealed by ^{19}F NMR of model peptide [Pro-4(*R*)-hydroxyprolyl-Gly]$_3$-[Pro-4(*R*)-fluoroprolyl-Gly]-[Pro-4(*R*)-hydroxyprolyl-Gly]$_3$. *J. Phys. Chem. B* **116**, 6908–6915 (2012).

Acknowledgements

J.D.H. acknowledges funding for this work that was provided by the US National Science Foundation (CHE 2203937) and the Robert A. Welch Foundation Grant (C-2141). G.N.P. Jr's funding was provided by the US National Science Foundation BioXFEL STC (1231306) and the Robert A. Welch Foundation Grant (C-2118). C.C.C. acknowledges funding from the US National Science Foundation Graduate Research Fellowship Program (grant no. 1842494). Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US National Science Foundation. This work was partly done using resources of the Shared Equipment Authority at Rice University.

Author contributions

C.C.C. contributed to the conceptualization, data curation, methodology, investigation and writing. D.R.W. contributed to the conceptualization, methodology, writing and investigation. S.A.H.H. contributed to the methodology and investigation. B.H.P. contributed to the investigation and data curation. J.W.R.S. contributed to the investigation and data curation. M.D.M. contributed to the methodology (crystallography) and investigation. W.X. contributed to the methodology (crystallography) and investigation. R.D. contributed to the methodology and investigation. M.M. contributed to the investigation and methodology. X.W. contributed to the methodology (NMR) and investigation. A.B.K. contributed to the methodology (kinetics) and investigation. G.N.P. Jr contributed to the methodology (crystallography) and investigation. J.D.H. contributed to the conceptualization, investigation, editing, writing, supervision and funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41557-024-01573-2.

Correspondence and requests for materials should be addressed to Jeffrey D. Hartgerink.

Peer review information *Nature Chemistry* thanks Fei Xu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.