An outcome-defining role for the triple-helical domain in regulating collagen-I assembly

Kathryn M. Yammine,[†,a] Rasia C. Li,[†,a] Isabella M. Borgula,[a] Sophia Mirda Abularach,[a] Andrew S. DiChiara,[a] Ronald T. Raines,[a,b,c] Matthew D. Shoulders[a,b,c,*]

[†]Equal contribution

Affiliations
[a]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA
[b]Koch Institute for Integrative Cancer Research at MIT, Massachusetts Institute of Technology, Cambridge, MA, USA
[c]Broad Institute of MIT and Harvard, Cambridge, MA, USA

*To Whom Correspondence May Be Addressed:
Matthew D. Shoulders
Department of Chemistry
Massachusetts Institute of Technology
77 Massachusetts Avenue, 16-573A
Cambridge, MA 02139
Tel: (617) 452-3525

Author Contributions: K.M.Y., R.C.L., A.S.D., and M.D.S. designed research. K.M.Y., R.C.L., I.M.B., S.M.A., and A.S.D. performed experiments. R.T.R and M.D.S. secured funding and supervised research. K.M.Y., R.C.L., and M.D.S. drafted the paper. All authors analyzed data and edited the manuscript.

Competing Interest Statement: The authors declare that they have no competing interests.

Major Classification: Biological Sciences
Minor Classification: Biochemistry

SIGNIFICANCE

Collagen is the molecular scaffold for animal life. By far the most prevalent collagen is type-I – the main proteinaceous component of skin and bone. Mature collagen-I is a 2:1 hetero-trimeric assembly of two distinctive procollagen strands. How the composition of this heterotri-mer is determined remains unknown. The longstanding paradigm is that assembly is specified exclusively by features within procollagen's ~30 kDa C-propeptide domain. Here, we show that the amino acid sequence near the C-terminus of the procollagen triple-helical domain also plays an outcome-defining role that is necessary for guiding proper assembly. These results provide critical insight into the molecular recognition and assembly challenge associated with producing the essential collagenous molecular scaffold in animals.

ABSTRACT

Collagens are the foundational component of diverse tissues, including skin, bone, cartilage, and basement membranes and are the most abundant protein class in animals. The fibrillar collagens are large, complex, multi-domain proteins, all containing the characteristic triple helix motif. The most prevalent collagens are heterotrimeric, meaning that cells express at least two distinctive procollagen polypeptides that must assemble into specific heterotrimer compositions. The molecular mechanisms ensuring correct heterotrimeric assemblies are poorly understood – even for the most common collagen, type-I. The longstanding paradigm is that assembly is controlled entirely by the ~30 kDa globular C-propeptide (C-Pro) domain. However, this dominating model for procollagen assembly has left many questions unanswered. Here, we show that the C-Pro paradigm is incomplete. In addition to the critical role of the C-Pro domain in templating assembly, we find that the amino acid sequence near the C-terminus of procollagen's triple-helical domain plays an essential role in defining procollagen assembly outcomes. These sequences near the C-terminus of the triple-helical domain encode conformationally stabilizing features that ensure only desirable C-Pro-mediated trimeric templates are committed to irreversible triple-helix folding. Incorrect C-Pro trimer assemblies avoid commitment to triple-helix formation thanks to destabilizing features in the amino acid sequences of their triple helix. Incorrect C-Pro assemblies are consequently able to dissociate and search for new binding partners. These findings provide a new perspective on the mechanism of procollagen assembly, revealing the molecular basis by which incorrect homotrimer assemblies are avoided and setting the stage for a deeper understanding of the biogenesis of this ubiquitous protein.

Keywords: procollagen assembly, extracellular matrix, endoplasmic reticulum, protein folding, macromolecular complex

INTRODUCTION

In humans, twenty-eight different types of collagen comprise the molecular scaffold for life. The defining feature of all collagens is the presence of a triple-helical structural motif, comprised of three individual polypeptide chains (1, 2). The highly abundant fibrillar collagens (types I, II, III, V, XI, XXIV, and XXVII) are synthesized as procollagen precursors, consisting of a single, lengthy triple-helical domain (THD; >1000 amino acids in length) bookended by short telopeptide and globular propeptide domains (**Fig. 1A**) (3). Following intracellular folding, the N-terminal (ranging from 6–50 kDa) and C-terminal (~30 kDa) propeptide domains (4) (termed the N-Pro and C-Pro, respectively) are cleaved, allowing the mature THD to oligomerize and crosslink into the fibrils that form the structural network of tissues, ranging from tendon to bone (1, 5).

While some fibrillar procollagen types are homotrimers comprised of three identical polypeptide chains, others assemble as heterotrimers of two or even three different polypeptide chains (1). For example, collagen-I, by far the most abundant of the fibrillar collagens and the key protein scaffold for skin and bone, is generally found in tissues as a 2:1 heterotrimer of collagen-α1(I):collagen-α2(I), although homotrimeric Colα1(I) is also observed in development and disease states (3). Homotrimeric Colα2(I) has not been observed in physiologic settings.

Decades of research have focused on elucidating the biogenesis of this long, complex, multidomain, and often heterotrimeric protein. Folding and assembly of the procollagen trimer begins at the C-Pro domain (6, 7). Individual C-Pro domains first fold separately, then recognize each other to form trimers (8), ultimately triggering THD folding into mature trimeric procollagen. This C-Pro-initiated process means that, unlike the vast majority of other multi-domain proteins, procollagen polypeptides must be entirely translated before folding can begin (9).

Early attempts to refold propeptide-cleaved, mature collagen THDs *in vitro* resulted in the formation of gelatin – disordered amalgamations of collagen molecules (10). In 1980, Bächinger, Engel, and colleagues demonstrated that procollagens begin folding from the C-Pro domain, in a “zipper-like” fashion (6). If still fused to their C-Pro, denatured collagen could be refolded into stable, protease-resistant triple helices (6). This seminal work sparked a flurry of investigations aimed at uncovering and defining the molecular features within the C-Pro domain

that govern procollagen assembly, from chain selection to registration and nucleation of the triple helix (6-8, 11-24).

In 1997, Bulleid and coworkers identified a discontinuous, 15-amino acid sequence in C-Proα1(III) that is sufficient to drive formation of protease-resistant triple helices of a heavily truncated "mini" procollagen construct (16). Residues within this "collagen recognition sequence" that may help to ensure type-specific procollagen assembly (i.e., that type-I procollagen does not incorrectly assemble with type-III when co-expressed) were later confirmed to be involved in salt-bridge formation between C-Proα1(III) monomers (20). However, these salt-bridges are not conserved amongst fibrillar collagens (21), suggesting that the chain recognition sequence may be primarily important for ensuring homotypic type-III C-Pro assembly.

In 2012 and later in 2017, Hulmes and coworkers determined high-resolution crystal structures of the C-Proα1(III) and homotrimeric C-Proα1(I) domains, respectively (20, 21), revealing that the C-Pro domain adopts a flower-like assembly consisting of three petal regions, a base, and a coiled-coil stalk. These structures identified specific interchain interactions proposed to be involved in assembly, unequivocally established the C-Pro domain's disulfide intra- and interchain disulfide pattern (**Fig. 1B**), and unveiled previously unknown Ca^{2+}-binding sites at the interfaces between C-Pro monomers.

Building on these structures, our group showed that Ca^{2+}-binding plays a critical role in collagen C-Pro assembly (22). Studying C-Pro domains in isolation (lacking a THD, like those employed in the structure studies), we discovered that Ca^{2+} coordination robustly mediates dynamic, non-covalent, and reversible assembly of all possible combinations of C-Pro trimers, including the formation of both desirable (e.g., 2:1 C-Proα1(I):C-Proα2(I) and homotrimeric C-Proα1(I)) and undesirable (e.g., 1:2 C-Proα1(I):C-Proα2(I) and homotrimeric C-Proα2(I)) assemblies (**Fig. 1C**, first step). The presence or absence of a single cysteine residue in the C-Pro domain that participates in an intermolecular disulfide bond determines which of these Ca^{2+}-mediated assemblies can then be irreversibly, covalently immortalized (**Fig. 1C**, second step). C-Proα2(I) lacks one of these cysteine residues that is required to form intermolecular disulfide bonds, and thus, undesirable 1:2 C-Proα1(I):C-Proα2(I) and homotrimeric C-Proα2(I) non-covalent assemblies cannot be covalently immortalized and instead disassemble until appropriate

partners are found (**Fig. 1C**, reversing the first step) (22). These features of C-Pro assembly are conserved across the fibrillar collagens, and the presence or absence of that key cysteine residue correlates perfectly with the capacity of a given procollagen to homotrimerize or not (22).

These observations are summarized in **Fig. 1C**. First, procollagen strands reversibly assemble in all possible combinations in a dynamic process mediated by Ca^{2+}. Productive, biologically relevant assemblies are then covalently immortalized via disulfide bonding (only possible when both cysteine residue 2 and cysteine residue 3 are present; **Fig. 1B**). These results stem directly from experimental observations. A reasonable speculation following these experimental findings is that subsequent THD folding occurs only after covalent immortalization locks in a specific C-Pro assembly (**Fig. 1C**, final step). This last prediction has never been experimentally tested and is founded on the current paradigm that the C-Pro domain itself fully encodes which collagens can or cannot homotrimerize.

Here, we explicitly test whether C-Pro domain interstrand disulfide bond formation is either necessary or sufficient to trigger THD folding. We find that it is neither. Thus, the longstanding C-Pro-only paradigm for procollagen assembly is misguided. Instead, we reveal that the C-terminal sequence of the THD has a decisive role in actually nucleating triple-helix formation of only certain desirable assemblies templated by the C-Pro domain. In this new model, both the C-Pro and the C-terminal THD sequence function in concert to guide productive trimer folding and to actively prevent undesirable outcomes.

RESULTS

Covalent immortalization of the C-Pro trimer is not necessary to trigger triple helix folding

The final step in the model for procollagen assembly presented in **Figure 1C** implies that disulfide-mediated covalent immortalization of the C-Pro domain is necessary to trigger triple-helix folding. To test this hypothesis, we expressed full-length procollagen constructs encoding either wild-type or C2S proColα1(I). In the latter construct, the cysteine at position 2 (C2; residue 1265) is replaced by a serine (i.e., C2S proColα1(I); **Fig. 1B**). This substitution prevents interchain disulfide bond formation without otherwise affecting C-Pro domain assembly (22). We examined the properties of these procollagen variants when secreted from HT-1080 cells, which are known to be capable of properly folding procollagen-I but, importantly, do not substantively express endogenous, potentially interfering fibrillar procollagens (11, 25-28).

We first evaluated the ability of these proColα1(I) constructs to form disulfide-linked trimers. As expected based on the model in **Fig. 1C**, wild-type proColα1(I) formed disulfide-linked trimers, while the C2S substitution eliminated disulfide-mediated covalent trimers (**Fig. 2A**). We note that proColα1(I) C2S did form some amount of an apparent disulfide-linked dimer. This dimer most likely results from free cysteine residues undergoing disulfide shuffling during denaturation, but may also result from a linkage between cysteines in the N-Pro or between free C3 residues.

We next assessed whether these two proColα1(I) constructs were able to form stably folded triple helices. If the final step in the procollagen assembly mechanism from **Fig. 1C**, which is founded on the current C-Pro-only model for procollagen assembly, is correct, then only the covalently immortalized (interstrand disulfide-linked) wild-type proColα1(I) variant would do so. To assess triple helix formation, we employed the classic protease digestion assay – properly folded fibrillar collagen triple helices are resistant to protease digestion, whereas unfolded THDs are not (6, 15, 28-34) (**Fig. 2B**). We precipitated collagen from the media of cells transfected with either wild-type or C2S proColα1(I), and then subjected it to trypsin/chymotrypsin digestion. Similar to primary fibroblasts, cells transfected with wild-type proColα1(I) secreted triple-helical, protease-resistant collagen (**Fig. 2C**). The observed shift in molecular

weight arises from cleavage and digestion of the non-triple-helical pro- and telopeptide domains, with the stable THD left behind. Surprisingly, C2S proColα1(I) was *also* able to assemble into protease-resistant triple helices (**Fig. 2C**), though we note that some proportion of this construct appears to be susceptible to digestion. This result shows that, contrary to predictions derived from a C-Pro paradigm, covalent immortalization is actually *not* required to trigger stable triple helix formation, at least for proColα1(I).

Covalent immortalization of the C-Pro domain is insufficient to trigger triple helix folding

The results in **Fig. 2** show that disulfide-mediated, covalent immortalization of a trimerized C-Pro domain is not *necessary* to trigger triple helix folding. This observation raises a puzzling question: Why are Colα2(I) homotrimeric THDs not observed in animals (7, 15)? C-Proα2(I), and all the other fibrillar C-Pro domains that do not trigger homotrimeric THD folding, lack one of the cysteine residues that is required for covalent immortalization of the C-Pro domain (22). Still, as illustrated in **Fig. 1C**, the C-Proα2(I) domain does reversibly homotrimerize (22). If covalent immortalization is not *necessary* to trigger THD folding, then why are Colα2(I) THD homotrimers not formed? To begin to address this question, we first asked whether installation of the missing cysteine residue in proColα2(I) (C2; residue 1169) is *sufficient* to drive Colα2(I) THD homotrimer formation, as the current C-Pro-only paradigm for procollagen assembly would suggest.

We expressed wild-type proColα2(I) and S2C proColα2(I), where the serine at position 1169 is replaced by a cysteine (**Fig. 1B**), in HT-1080 cells and evaluated the assembly of secreted procollagen. We observed that, as expected, the C2-lacking wild-type proColα2(I) does not form disulfide-linked homotrimeric assemblies. Also consistent with prior results derived from expression of the C-Proα2(I) domain in isolation (22), S2C proColα2(I) was indeed, and in contrast, able to form covalently immortalized, disulfide-linked homotrimers (**Fig. 3A**).

We next assessed whether a disulfide-linked proColα2(I) trimer triggers THD folding in S2C proColα2(I). As expected, we observed that the C2-lacking wild-type proColα2(I) variant did not form any detectable protease-resistant triple helices. Surprisingly, S2C proColα2(I) *also*

did not form protease-resistant triple helices (**Fig. 3B**), despite covalent immortalization of the homotrimerized C-Proα2(I) domain.

In concert with our findings for proColα1(I) in **Fig. 2**, these results show that covalent immortalization of the C-Pro domain is neither *necessary* nor *sufficient* to trigger stable, triple helix formation. They also falsify the hypothesized final step in the C-Pro-only procollagen assembly model shown in **Fig. 1C**.

The procollagen triple-helical domain plays an essential role in regulating trimerization

Our observations in **Figs. 2** and **3** suggest that there must either be major, unappreciated features within the C-Pro-only procollagen assembly mechanism that we do not understand, or that there are molecular features outside of the C-Pro domain (THD, telopeptides, or N-Pro) that partner with the C-Pro domain to determine whether a given THD can adopt a stable, homotrimeric triple-helical conformation. Given the extensive attention already given to the C-Pro domain itself (6-8, 12-22, 24), we looked outside the C-Pro domain for other key factors that could be involved – particularly within the THD.

To test whether or not the sequence of the THD has a key role in defining the capacity of procollagen strands to form homotrimeric triple helices, we engineered chimeric procollagen constructs in which we swapped the THDs of proColα1(I) and proColα2(I), while holding the rest of the sequence constant. Thus, the chimera proα1($\alpha2^{THD}$) consists of the Colα2(I) THD nestled between the α1(I) N- and C- telo- and propeptides. Similarly, the chimera proα2($\alpha1^{THD}$) consists of the Colα1(I) THD nestled between the α2(I) N- and C- telo- and propeptides (**Fig. 4A**). If the THD sequence plays a critical role in determining which procollagen strands form stable triple helices, we would predict the chimeras to behave consistently with their THDs, rather than their C-Pro domains. Alternatively, if the current paradigm that C-Pro domains autonomously govern procollagen trimerization is correct, then we would expect the chimeras' behavior to match the identity of their C-Pro domains.

We expressed proα1($\alpha2^{THD}$) or proα2($\alpha1^{THD}$) in HT-1080 cells, and examined the disulfide-mediated covalent assembly and triple helix formation of the resulting secreted procollagens. With respect to disulfide-mediated covalent trimer formation, as expected, proα1($\alpha2^{THD}$)

was able to form disulfide-linked trimers, consistent with the known behavior of its C-Pro domain. Proα2($\alpha1^{THD}$) formed no disulfide-linked trimers, again as expected based on the known behavior of its C-Pro domain (**Fig. 4B**).

Remarkably, however, protease digestion of the precipitated chimeras demonstrated that proα1($\alpha2^{THD}$), despite covalent immortalization and fusion to C-Proα1(I), still could not form protease-resistant triple helices. Equally striking, proα2($\alpha1^{THD}$) was able to assemble into protease-resistant triple helices, consistent with its THD behavior rather than the identity of the C-Pro domain (**Fig. 4C**). These data show that, beyond the C-Pro domain, the THD must also play an essential role in regulating procollagen's trimeric assembly.

C-Terminal triple-helical domain stability informs the composition of folded triple helices

To elucidate specific features of the THD that could be involved in regulating the capacity of a given procollagen strand to form homotrimeric triple helices, we turned our attention to the amino acid sequence near the C-terminal, triple helix-nucleating region of proColα1(I) and proColα2(I). One hypothesis to explain the lack of homotrimeric triple helix formation by proColα2(I), even when covalently stabilized by cysteine introduction or C-Pro domain swapping, could be that the C-terminal THD sequence is not sufficiently stable to commit the construct to formation of a homotrimeric triple helix, at least not before the non-covalently and reversibly assembled C-Proα2(I) domain dissociates. Such a mechanism would ensure that incorrect Colα2(I) homotrimeric triple helices are never formed to begin with. After the incorrect assemblies dissociate, the resulting monomeric C-Pro domains could then continue their search for appropriate partners that enable stable triple-helix folding. Because triple helix unfolding is slow (35-37), avoiding commitment to the triple helix would be quite important to avoid formation of proColα2(I) homotrimers. Meanwhile, the absent C2 in proColα2(I) ensures that disassembly is still possible. Following this hypothesis, the prediction would then be that the C-terminal region in the THD of proColα1(I) does form sufficiently stable triple helices to commit the domain to folding, even in the case of only transient C-Pro domain homotrimer assembly.

Triple-helix stability correlates strongly with total proline content (31, 38, 39). Moreover, proline residues in the Yaa position of the collagen Xaa-Yaa-Gly triplet are especially impactful,

as their 4*R*-hydroxylation (to form Hyp) markedly stabilizes the triple helix via preorganization (2, 40-47). Klein, Brodsky, Baum, and colleagues have shown that the presence of proline in the first few triplets is critical to initiate triple-helix folding (38, 39). Thus, we hypothesized that the more triple helix-inclined proColα1(I) would have higher C-terminal region proline content than proColα2(I). To evaluate this idea, we began by tallying prominent triple helix-stabilizing or destabilizing features in the C-terminal region of the THD. In the 10 most-C-terminal triplets of proColα1(I), there are 14 prolines, of which six are in the Yaa position, versus 11 prolines in proColα2(I), of which only four are in the Yaa position (**Fig. S1**). The higher proline content near the C-terminus of the proColα1(I) THD would likely kinetically and thermodynamically commit even a non-covalent C-Pro trimer complex to triple-helix formation and stability (2, 41). In contrast, even in the context of covalent trimerization, the lower proline content near the C-terminus of the proColα2(I) THD could avoid a commitment to triple helix formation, a process that is effectively irreversible (35), thereby enabling reversible monomerization. We note that the observation by others that Colα1(I) THD homotrimers are more stable than 2:1 Colα1(I):Colα2(I) heterotrimers (33) provides further support for the idea that Colα2(I) homotrimeric triple helices might be quite unstable.

To test the hypothesis that the C-terminal THD sequence is a critical component governing what THDs form homotrimeric triple helices, we first used a collagen triple-helix stability calculator to predict the melting temperature (T_m) of triple helices containing relevant Xaa-Yaa-Gly triplets (48). Examining proColα1(I) and proColα2(I)'s 15 most C-terminal triplets, we noted the existence of two striking wells of instability, where the predicted T_m of triple helices formed from the proColα2(I) THD sequence were predicted to be markedly lower than those formed from the proColα1(I) THD sequence (48) (**Fig. 5A**). This phenomenon is even further exacerbated if the final C-terminal Gly-Gly-Gly triplet is considered part of the proColα2(I) triple helix, as this extremely flexible polypeptide is dramatically destabilizing. Related to this point, in our design of THD chimeras for **Fig. 4**, we considered the C-terminal Gly-Gly-Gly triplet of proColα2(I) as part of the telopeptide to ensure equal THD length between proColα1(I) and proColα2(I) (1,014 triplets each), although as a Xaa-Yaa-Gly triplet, that first Gly-Gly-Gly

could be considered the initial triplet of the proColα2(I) THD. Despite this conservative decision, all the **Fig. 4** results are still consistent with the C-Pro/THD partnership model tested here.

Next, to experimentally test whether there is a substantial difference in stability between the C-terminal region of proColα1(I) and proColα2(I) triple helices, we designed host–guest collagen-mimetic peptides spanning the predicted wells of instability in the proColα2(I) THD (42, 49-52). Briefly, we introduced two or three 'guest triplets' from equivalent C-terminal regions of the proColα1(I) and proColα2(I) THDs into a triple helix-inducing host backbone of Gly-Pro-Hyp triplets (see list in **Table S1** and **Figs. S3–5**). We then experimentally measured the T_m of triple helices formed from these host–guest peptides using temperature-dependent circular dichroism (53). Consistent with the partnership model for regulation of THD formation, we observed a substantial difference in the T_m values for peptides containing proColα1(I) versus proColα2(I) triplets, with all of the proColα1(I) peptides forming considerably more stable triple helices (**Fig. 5B**). Notably, many of the proColα2(I) host–guest peptides exhibited a T_m far below physiologic temperature.

Thus, both theory and experiment strongly indicate that there is a substantial difference in the stability of triple-helical conformations formed from the C-terminal regions of the THDs of proColα1(I) versus proColα2(I). Building on these results, we wondered whether proColα2(I) strands that fail to form stable triple helices at 37 °C might nonetheless be able to stably fold at a lower temperature. However, even at 27 °C, and consistent with poor triple-helical stability in the C-terminal region of proColα2(I)'s THD strongly disfavoring committed triple helix folding, both wild-type proColα2(I) and S2C proColα2(I) could not form protease-resistant triple helices (**Fig. S2**).

Taken together with the results from our chimeric procollagen constructs in **Fig. 4**, these studies suggest a model in which the C-terminal THD sequence plays a critical role in defining the capacity of procollagens to form homotrimeric triple helices. Rather than the C-Pro domain autonomously governing procollagen-I assembly, we conclude that the C-Pro domain and THD act in concert to ensure or prevent irreversible trimer formation.

DISCUSSION

Our results show that, contrary to the current C-Pro-only paradigm for procollagen assembly, the C-terminal THD amino acid sequence also has a critical role governing stable procollagen folding. Specifically, we find that C-Pro covalent immortalization via disulfide bonds is neither *necessary* nor *sufficient* for stable THD folding. Rather, the C-Pro domain and THD must function in concert to both enable the formation of certain assemblies and prevent formation of undesired assemblies.

Critically, this new partnership model explains how non-homotrimerizing, heterotrimer-exclusive strands, such as proColα2(I), avoid becoming irreversibly trapped in unproductive assemblies through features of both the C-Pro domain and the THD (**Fig. 6A**). As illustrated in **Fig. 6B**, there are two ways a procollagen chain could become irreversibly trapped in an undesired assembly: by covalent bonding of the C-Pro domain or by THD folding after non-covalent C-Pro trimerization. The problem is exacerbated by the fact that all C-Pro domains can assemble non-covalently (and reversibly) in any stoichiometry and composition, including incorrect assemblies (22). Unfolding of the THD is known to be very slow (35-37), meaning that THD folding is functionally kinetically irreversible. In other words, once the first few triplets of the THD fold, the molecule likely becomes committed to that assembly. To avoid unproductive entrapment, a procollagen chain must therefore lack the ability to form undesirable disulfide bonds that immortalize an incorrect assembly ***and*** have a C-terminal THD sequence that is not prone to rapid and stable triple helix formation (38, 39). Evolution has ensured that non-homotrimerizing chains, such as proColα2(I), do not get trapped by *both* the absence of cysteine 2 and the inclusion of the unstable C-terminal THD sequence. As a result, proColα2(I) can ***only*** form heterotrimeric triple helices.

Meanwhile, the C2–C3 interchain disulfide bond is not actually *required* for driving triple-helix folding by proColα1(I). Both C2S proColα1(I) and the proα2(α1^{THD}) chimera could form protease-resistant triple helices, despite not being able to form immortalizing interchain disulfide bonds in the C-Pro domain. Non-covalent association of proColα1(I) appears to be sustained long enough, through Ca^{2+} binding and other non-covalent interactions, to nucleate protease-resistant folding of the proColα1(I) homotrimeric triple helix, which, by virtue in part of its

proline- and Hyp-rich C-terminal sequence, has a high propensity to form stable homotrimeric triple helices.

If the cysteine code is not necessary for proColα1(I) THD folding, then why is it so highly conserved amongst species and collagen types (22)? Although covalent assembly of the procollagen strands via disulfide bonds is dispensable for homotrimeric proColα1(I) folding, our data suggest that disulfide bonds could be critical to ensure successful incorporation of the less stable proColα2(I) strand into desired 2:1 heterotrimeric THDs. By covalently stapling a pro-Colα2(I) chain to a proColα1(I) chain, the C-Pro templated trimer is sustained stably and long enough to commit the heterotrimeric THD to folding. As an aside, previous work has shown that deletion of the 10 most C-terminal amino acids in proColα2(I) (which include a cysteine residue that forms an intrastrand disulfide) prevents the strand from associating into heterotrimers (17). However, more than simply disrupting the heterotrimeric assembly, this observation is likely due to gross misfolding of C-Proα2(I) owing to the lack of a key intrachain disulfide bond.

Beyond the cysteine code's likely essential role in heterotrimer formation, it may also offer an assembly advantage for homotrimer folding, as once the correct strands are transiently assembled via Ca^{2+}-binding the disulfide bond ensures that the trimer no longer dissociates. Moreover, the cysteines could be important for C-Pro domain folding (prior to assembly) by engaging in intermediate disulfide bonds during the folding process.

We also explored the partnership model's applicability beyond procollagen-I to the other heterotrimeric fibrillar collagens: types V and XI. Procollagen-XI assembles in a wide variety of stoichiometries, a process that appears to be heavily transcriptionally regulated (54-56). Procollagen-V, meanwhile, commonly assembles as either a 1:1:1 proColα1(V):proColα2(V):pro-Colα3(V) heterotrimer, or as a 2:1 proColα1(V):proColα2(V) (57). The C-terminal triple-helix stability of the proColα1(V) chain, which is able to form stable homotrimers, is substantially higher than that of the non-homotrimerizing proColα2(V) chain, especially across the first three THD triplets to fold (**Fig. S6**). While the stability difference is not as striking between pro-Colα1(V) and proColα3(V), the proColα3(V) THD is likely prevented from folding into a stable homotrimeric triple helix by the high positive charge of its telopeptide (**Fig. S6**), which would

create considerable electrostatic repulsion to prevent triple helix initiation. Thus, similar to collagen type-I, there are regions outside of the C-Pro domain that are critical in governing other heterotrimeric fibrillar collagen assemblies, further emphasizing the need to think beyond the C-Pro domain.

Altogether, the data presented here show that the ability of a given collagen strand to stably homotrimerize is not based solely on molecular features encoded in the C-Pro domain, as previously believed. Rather, the sequence of the THD plays a crucial role in trimer assembly. It is this cooperation between the different domains that ultimately guides procollagen-I into proper assemblies, a finding that redefines our understanding of procollagen assembly by shifting from a C-Pro-only paradigm to a partnership model that accounts for both the C-Pro domain and the THD. Notably, work on short, synthetic collagen-mimetic peptides has also shown that interstrand electrostatic interactions can be used to govern the assembly of short heterotrimeric triple helices (58, 59).

A key question still remaining for the field relates to how cells normally produce almost exclusively 2:1 proColα1(I):proColα2(I) heterotrimer despite the greater stability of the proColα1(I) homotrimer (11, 34). Predicted salt-bridges between the C-Proα1(I) and C-Proα2(I) have been proposed to stabilize heterotrimeric C-Pro assemblies (21); however, this non-covalent stabilization is likely small compared to that of both the covalently linked C-Proα1(I) homotrimer and the highly stable proColα1(I) triple helix. How do cells ensure that proColα1(I) strands are not sequestered in highly irreversible, covalently linked homotrimers that form very stable triple helices, and instead prefer to assemble with one strand of proColα2(I)? Answering this remaining question provides ample opportunity for future structural and biophysical work, as well as for exploring likely roles for elements of the cellular proteostasis network in biasing such assemblies (5, 25, 60).

MATERIALS AND METHODS

Chemicals

Commercial chemicals were of reagent grade or better and were used without further purification. Oxyma, Rink amide resin (low-loading), and Fmoc-protected amino acids except Fmoc-Hyp(O*t*Bu)-OH were obtained from CEM corporation (Matthews, NC). Fmoc-Hyp(O*t*Bu)-OH was obtained from AK Scientific (Union City, CA). *N*,*N*-Dimethylformamide (DMF), HPLC-grade acetonitrile (ACN), diethyl ether, and dichloromethane were obtained from Fischer Scientific (Hampton, NH). Trifluoroacetic acid (TFA) was obtained from Beantown Chemical (Hudson, NH). Diisopropylcarbodiimide (DIC) and 4-methylpiperidine were obtained from Oakwood Chemical (Tampa, FL).

Cell culture

HT-1080 cells (ATCC) were cultured in DMEM (Corning) supplemented with 10% fetal bovine serum (FBS), 100 IU penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine (Corning). Primary fibroblasts (GM05294; Coriell) were cultured in MEM (Corning) supplemented with 15% FBS, 100 IU penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine (Corning).

Vector construction

The proColα1(I) C2S plasmid was constructed from the wild-type HA-tagged proColα1(I) plasmid by site directed mutagenesis using the Q5 SDM kit, according to the manufacturer's instructions (New England Biolabs). Primers were designed using NEBaseChanger. Similarly, the proColα2(I) S2C plasmid was constructed by site directed mutagenesis of wild-type HA-proColα2(I). The RFP plasmid consisted of the RFP gene fused to a KDEL ER localization tag under a constitutive CMV promoter. Chimeras were designed to swap the THD of proColα1(I) and proColα2(I), defined as the continuous stretch of Gly-Xaa-Yaa triplets. For proColα2(I), the C-terminal Gly-Gly-Gly triplet was not considered to be part of the THD (and was instead considered to be part of the telopeptide), such that the total number of THD triplets between proColα1(I) and proColα2(I) were the same (1,014 triplets). Vectors were constructed by Genscript via modification of the wild-type HA-proColα1(I) and HA- proColα2(I) vectors. The

purchased DNA was transformed into competent cells, MIDI-prepped (E.N.Z.O), and sequence-confirmed using Primordium's full plasmid next-generation sequencing.

Procollagen construct expression

HT-1080 cells were plated at a density of 4×10^6 cells/10-cm dish the day before transfection. Cells were transfected using Transit-2020 transfection reagent (Mirus), according to the manufacturer's instructions. The next day, media was replaced with fresh complete DMEM supplemented with 50 μM sodium ascorbate (Amresco). Media was harvested for immunoblotting or precipitation ~24 h post-ascorbate addition. For primary fibroblast controls, cells were provided with fresh MEM and 200 μM sodium ascorbate 24 h prior to harvesting.

Immunoblotting

For non-reducing blots, media was treated with 100 mM iodoacetamide (VWR) to minimize disulfide shuffling. Samples were then treated with 6× gel loading buffer (300 mM Tris, pH 6.8, 15% glycerol, 6% SDS, and 10% (w/v) bromophenol blue). For reducing blots, samples were treated with 6× gel loading buffer supplemented with 200 mM DTT and boiled for 10 min. Samples were separated on homemade 4/8% polyacrylamide gels. Immunoblots were probed with an anti-HA primary antibody (ThermoFisher 26183; 1:10,000) followed by 800CW or 680LT secondary antibodies (LI-COR) or, for proteolysis experiments probed with anti-Colα1(I) (Abnova, H00001277-M01J clone 3G3; 1:2,000) or anti-Colα2(I) (Sigma, SAB4500363; 1:2,000) primary antibodies followed by 800CW secondary antibodies. Images were obtained using an Odyssey infrared imager (LI-COR).

Protease digestions

Media from transfected 10-cm dishes was split into two equal aliquots (0.9 mL) and precipitated with 176 mg/mL ammonium sulfate in 0.1 M Tris (pH 7.4) overnight at 4 °C, rotating end-over-end. Samples were centrifuged at max speed for 30 min to pellet the protein, and supernatant was removed. The pellets were resuspended in 30 μL of digestion buffer (100 mM sodium chloride, 150 mM Tris, pH 7.4) and either with or without proteases (0.1 mg/mL trypsin + 1 mg/mL chymotrypsin for final concentration). Samples were digested at rt for 25 min. The reaction was quenched by adding 1× gel loading buffer (300 mM Tris, pH 6.8, 15% glycerol, 6%

SDS, and 0.25% w/v bromophenol blue) supplemented with 200 mM DTT, and promptly boiling for 15 min. Samples were then separated on homemade 8% SDS-PAGE gels.

Peptide synthesis

Peptides were synthesized using a Liberty Blue™ Automated Microwave Peptide Synthesizer from CEM (Matthews, NC). All peptides were synthesized following CEM standard methods for both microwave and coupling cycles, as previously described (61). Standard solutions of DIC (0.5 M in DMF), Oxyma (1 M in DMF), 4-methylpiperidine (20% v/v in DMF), and Fmoc-protected amino acids (0.2 M in DMF) were prepared for each synthesis.

Standard microwave-assisted deprotection. The microwave was set to 155 W at 75 °C for 15 s, followed by 30 W at 90 °C for 50 s.

Standard microwave-assisted coupling. The microwave was set to 170 W at 75 °C for 15 s, followed by 30 W at 90 °C for 225 s.

Standard coupling cycle. FmocGly-loaded Wang resin (1 equiv) was added to the CEM reaction vessel, and the resin was allowed to swell for 5 min in DMF. The Fmoc group was removed using the standard deprotection solution and the microwave-assisted deprotection methods described above. The resin was then washed (4×), and Fmoc-AA-OH (5 equiv) was added, followed by DIC (20 equiv) and Oxyma (40 equiv). Standard microwave-assisted coupling was performed with additional Fmoc-protected amino acids, and the resin was washed (2×) and drained. When double-coupling was required, the cycle was repeated without the deprotection step.

Cleavage and precipitation. After the final deprotection step, the resin was removed from the reaction vessel into a cleavage filter, washed with DCM (4×), and air-dried crude peptides were then cleaved from the resin using a cleavage cocktail composed of 2.5:2.5:95 H_2O/TIS/TFA for 2 h. Peptide mixtures were then filtered and precipitated in ice-cold diethyl ether (10×). The peptides were collected by centrifugation, the supernatants were decanted, and the solid peptide was dissolved in 5 mL of 70:30 H_2O/ACN. The solutions were frozen and lyophilized using a FreeZone benchtop instrument from Labconco (Kansas City, MO). The crude peptide mixture was then subjected to purification.

Purification. The crude peptide products were purified by preparative reversed-phased HPLC using a XSelect Peptide CSH C18 OBD Prep Column 130 Å, 5 µm, 19 mm × 250 mm

from Waters (Milford, MA) and a 1260 Infinity II instrument (Agilent Technologies, Santa Clara, CA). Crude products were dissolved in the minimum amount of ACN and eluted with a linear gradient of 5–80% v/v ACN in H_2O containing TFA (0.1% v/v). After reviewing the initial chromatogram, the method was updated, if necessary. Chromatography fractions were analyzed by MALDI–TOF MS using a microflex LRF instrument and a CHCA matrix (Bruker, Billerica, MA). Alternatively, aliquots fractions were submitted to liquid chromatography/mass spectrometry (LCMS) using an Agilent 6125B mass spectrometer attached to an Agilent 1260 Infinity LC. Fractions containing purified peptide were pooled, lyophilized, and analyzed with reversed-phase HPLC using a 1260 Infinity II instrument (Agilent Technologies) and EC 250/4.6 Nucleosil 100-5 C18 column (Macherey–Nagel, Düren, Germany). Final purity was assessed using a XSelect CSH C18 5μm analytical HPLC column, 4.6 × 100 mm from Waters (Milford, MA).

Circular dichroism spectroscopy

Peptides were dried under vacuum for ≥48 h before being weighed and dissolved to 0.8 mM in 50 mM acetic acid (pH 3.0), as previously described (61). The resulting solutions were heated to 65 °C and cooled to 4 °C at a rate of 1 °C every 5 min. The solution was then incubated at ≤4 °C for ≥24 h before its CD spectrum was acquired with a Model J-1500 spectrometer (JASCO, Easton, MD) at the MIT Biophysical Instrumentation Facility. Spectra were measured with a bandpass of 1 nm. The signal was averaged for 3 s during the wavelength scan. To collect thermal denaturation curves, samples at 4 °C were warmed at a rate of 0.2 °C/min. CD signals at specified wavelengths were recorded every 3 °C. Values of T_m were determined by fitting the molar ellipticity at 225 nm to a four-parameter Hill equation.

ACKNOWLEDGMENTS

The authors would like to extend their gratitude to Anton V. Persikov for kindly assisting with and maintaining the online collagen stability calculator. We would also like to thank the staff at MIT's Biophysical Instrumentation Facility for their technical assistance. This work was supported by grants from the NSF (MCB 2236194 to M.D.S.) and NIH (R01AR071443 and R35GM136354 to M.D.S. and R35GM148220 to R.T.R.). K.M.Y. and A.S.D. were supported by NIH Ruth L. Kirschstein Predoctoral Fellowships (F31AR079263 and F31AR067615, respectively). R.C.L. was supported by an NSF Graduate Research Fellowship. Portions of the paper were developed from the thesis of R.C.L..

REFERENCES

1. S. Ricard-Blum, The collagen family. *Cold Spring Harb Perspect Biol* **3**, a004978 (2011).
2. M. D. Shoulders, R. T. Raines, Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929-958 (2009).
3. J. Bella, D. J. Hulmes, Fibrillar collagens. *Subcell Biochem* **82**, 457-490 (2017).
4. C. UniProt, UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531 (2023).
5. Y. Ishikawa, H. P. Bächinger, A molecular ensemble in the rER for procollagen maturation. *Biochimica et biophysica acta* **1833**, 2479-2491 (2013).
6. H. P. Bächinger, P. Bruckner, R. Timpl, D. J. Prockop, J. Engel, Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. *Eur J Biochem* **106**, 619-632 (1980).
7. D. J. S. Hulmes, Roles of the procollagen C-propeptides in health and disease. *Essays Biochem* **63**, 313-323 (2019).
8. K. J. Doege, J. H. Fessler, Folding of carboxyl domain and assembly of procollagen I. *J Biol Chem* **261**, 8924-8935 (1986).
9. J. Engel, D. J. Prockop, The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu Rev Biophys Biophys Chem* **20**, 137-152 (1991).
10. W. Traub, K. A. Piez, The chemistry and structure of collagen. *Adv Protein Chem* **25**, 243-352 (1971).
11. A. E. Geddis, D. J. Prockop, Expression of human *COL1A1* gene in stably transfected HT1080 cells: the production of a thermostable homotrimer of type I collagen in a recombinant system. *Matrix* **13**, 399-405 (1993).
12. J. F. Lees, N. J. Bulleid, The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J Biol Chem* **269**, 24354-24360 (1994).
13. N. J. Bulleid, R. Wilson, J. F. Lees, Type-III procollagen assembly in semi-intact cells: chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation. *Biochem J* **317 (Pt 1)**, 195-202 (1996).
14. J. E. Oliver, E. M. Thompson, F. M. Pope, A. C. Nicholls, Mutation in the carboxy-terminal propeptide of the Proα1(I) chain of type I collagen in a child with severe osteogenesis imperfecta (OI type III): possible implications for protein folding. *Hum Mutat* **7**, 318-326 (1996).
15. J. Myllyharju *et al.*, Expression of wild-type and modified proalpha chains of human type I procollagen in insect cells leads to the formation of stable [α1(I)]2α2(I) collagen heterotrimers and [α1(I)]3 homotrimers but not [α2(I)]3 homotrimers. *J Biol Chem* **272**, 21824-21830 (1997).
16. J. F. Lees, M. Tasab, N. J. Bulleid, Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *EMBO J* **16**, 908-916 (1997).
17. S. A. Doyle, B. D. Smith, Role of the pro-α2(I) COOH-terminal region in assembly of type I collagen: disruption of two intramolecular disulfide bonds in pro-α2(I) blocks assembly of type I collagen. *J Cell Biochem* **71**, 233-242 (1998).
18. J. P. Malone, K. Alvares, A. Veis, Structure and assembly of the heterotrimeric and homotrimeric C-propeptides of type I collagen: significance of the α2(I) chain. *Biochemistry* **44**, 15269-15279 (2005).
19. S. P. Boudko, J. Engel, H. P. Bächinger, The crucial role of trimerization domains in collagen folding. *Int J Biochem Cell Biol* **44**, 21-32 (2012).

20. J. M. Bourhis *et al.*, Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat Struct Mol Biol* **19**, 1031-1036 (2012).
21. U. Sharma *et al.*, Structural basis of homo- and heterotrimerization of collagen I. *Nat Commun* **8**, 14671 (2017).
22. A. S. DiChiara *et al.*, A cysteine-based molecular code informs collagen C-propeptide assembly. *Nat Commun* **9**, 4206 (2018).
23. A. M. Barnes *et al.*, COL1A1 C-propeptide mutations cause ER mislocalization of procollagen and impair C-terminal procollagen processing. *Biochim Biophys Acta Mol Basis Dis* **1865**, 2210-2223 (2019).
24. N. D. Doan *et al.*, Elucidation of proteostasis defects caused by osteogenesis imperfecta mutations in the collagen-alpha2(I) C-propeptide domain. *J Biol Chem* **295**, 9959-9973 (2020).
25. A. S. DiChiara *et al.*, Mapping and exploring the collagen-I proteostasis network. *ACS Chem Biol* **11**, 1408-1421 (2016).
26. R. C. Li, M. Y. Wong, A. S. DiChiara, A. S. Hosseini, M. D. Shoulders, Collagen's enigmatic, highly conserved *N*-glycan has an essential proteostatic function. *Proc Natl Acad Sci U S A* **118** (2021).
27. N. D. Doan, A. S. DiChiara, A. M. Del Rosario, R. P. Schiavoni, M. D. Shoulders, Mass Spectrometry-Based Proteomics to Define Intracellular Collagen Interactomes. *Methods Mol Biol* **1944**, 95-114 (2019).
28. K. M. Yammine *et al.*, ER procollagen storage defect without coupled unfolded protein response drives precocious arthritis. *Life Sci Alliance* **7**, in press (2024).
29. P. Bornstein, K. A. Piez, The nature of the intramolecular cross-links in collagen. The separation and characterization of peptides from the cross-link region of rat skin collagen. *Biochemistry* **5**, 3460-3473 (1966).
30. P. Bruckner, D. J. Prockop, Proteolytic enzymes as probes for the triple-helical conformation of procollagen. *Anal Biochem* **110**, 360-368 (1981).
31. H. P. Bächinger, J. M. Davis, Sequence specific thermal stability of the collagen triple helix. *Int J Biol Macromol* **13**, 152-156 (1991).
32. M. Raghunath, P. Bruckner, B. Steinmann, Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J Mol Biol* **236**, 940-949 (1994).
33. E. Makareeva *et al.*, Substitutions for arginine at position 780 in triple helical domain of the α1(I) chain alter folding of the type I procollagen molecule and cause osteogenesis imperfecta. *PLoS One* **13**, e0200264 (2018).
34. S. Han *et al.*, Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. *J Biol Chem* **285**, 22276-22281 (2010).
35. E. Makareeva *et al.*, Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J Biol Chem* **283**, 4787-4798 (2008).
36. E. Leikina, M. V. Mertts, N. Kuznetsova, S. Leikin, Type I collagen is thermally unstable at body temperature. *Proc Natl Acad Sci U S A* **99**, 1314-1318 (2002).
37. Y. Nishi *et al.*, Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of collagen triple helix. *Biochemistry* **44**, 6034-6042 (2005).
38. D. L. Bodian, B. Madhan, B. Brodsky, T. E. Klein, Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry* **47**, 5424-5432 (2008).
39. T. J. Hyde, M. A. Bryan, B. Brodsky, J. Baum, Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J Biol Chem* **281**, 36937-36943 (2006).
40. M. D. Shoulders, K. A. Satyshur, K. T. Forest, R. T. Raines, Stereoelectronic and steric effects in side chains preorganize a protein main chain. *Proc Natl Acad Sci U S A* **107**, 559-564 (2010).

41. R. A. Berg, D. J. Prockop, The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem Biophys Res Commun* **52**, 115-120 (1973).
42. J. A. Ramshaw, N. K. Shah, B. Brodsky, Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J Struct Biol* **122**, 86-91 (1998).
43. S. Sakakibara *et al.*, Synthesis of (Pro-Hyp-Gly) n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline. *Biochimica et biophysica acta* **303**, 198-202 (1973).
44. L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider, R. T. Raines, Conformational stability of collagen relies on a stereoelectronic effect. *J Am Chem Soc* **123**, 777-778 (2001).
45. S. K. Holmgren, K. M. Taylor, L. E. Bretscher, R. T. Raines, Code for collagen's stability deciphered. *Nature* **392**, 666-667 (1998).
46. M. D. Shoulders, J. A. Hodges, R. T. Raines, Reciprocity of steric and stereoelectronic effects in the collagen triple helix. *J Am Chem Soc* **128**, 8112-8113 (2006).
47. M. D. Shoulders, R. T. Raines, Modulating collagen triple-helix stability with 4-chloro, 4-fluoro, and 4-methylprolines. *Adv Exp Med Biol* **611**, 251-252 (2009).
48. A. V. Persikov, J. A. Ramshaw, B. Brodsky, Prediction of collagen stability from amino acid sequence. *J Biol Chem* **280**, 19343-19349 (2005).
49. N. K. Shah, J. A. Ramshaw, A. Kirkpatrick, C. Shah, B. Brodsky, A host-guest set of triple-helical peptides: stability of Gly-X-Y triplets containing common nonpolar residues. *Biochemistry* **35**, 10262-10268 (1996).
50. W. Yang, V. C. Chan, A. Kirkpatrick, J. A. Ramshaw, B. Brodsky, Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. *J Biol Chem* **272**, 28837-28840 (1997).
51. A. Choudhary, K. J. Kamer, M. D. Shoulders, R. T. Raines, 4-ketoproline: An electrophilic proline analog for bioconjugation. *Biopolymers* **104**, 110-115 (2015).
52. M. D. Shoulders, K. J. Kamer, R. T. Raines, Origin of the stability conferred upon collagen by fluorination. *Bioorg Med Chem Lett* **19**, 3859-3862 (2009).
53. F. H. Chu, A. Lukton, Collagenase induced changes in the circular dichroism spectrum of collagen. *Biopolymers* **13**, 1427-1434 (1974).
54. V. C. Lui, R. Y. Kong, J. Nicholls, A. N. Cheung, K. S. Cheah, The mRNAs for the three chains of human collagen type XI are widely distributed but not necessarily co-expressed: implications for homotrimeric, heterotrimeric and heterotypic collagen molecules. *Biochem J* **311 (Pt 2)**, 511-516 (1995).
55. M. Sun *et al.*, Collagen XI regulates the acquisition of collagen fibril structure, organization and functional properties in tendon. *Matrix Biol* **94**, 77-94 (2020).
56. J. J. Wu, M. A. Weis, L. S. Kim, B. G. Carter, D. R. Eyre, Differences in chain usage and cross-linking specificities of cartilage type V/XI collagen isoforms with age and tissue. *J Biol Chem* **284**, 5539-5545 (2009).
57. K. M. Mak, C. Y. Png, D. J. Lee, Type V collagen in death, disease, and fibrosis. *Anat Rec* **299**, 613-629 (2016).
58. V. Islami *et al.*, Self-Sorting Collagen Heterotrimers. *J Am Chem Soc* **146**, 1789-1793 (2024).
59. C. C. Cole *et al.*, Heterotrimeric collagen helix with high specificity of assembly results in a rapid rate of folding. *Nat Chem* **16**, in press (2024).
60. M. Y. Wong, M. D. Shoulders, Targeting defective proteostasis in the collagenopathies. *Curr Opin Chem Biol* **50**, 80-88 (2019).

61. J. M. Dones *et al.*, Optimization of interstrand interactions enables burn detection with a collagen-mimetic peptide. *Org Biomol Chem* **17**, 9906-9912 (2019).

Figure 1. (A) Procollagen-I consists of a lengthy (~1000 amino acids) triple-helical domain bookended by shorter N- and C- telo- and propeptides. **(B)** Cysteines and disulfide bonding patterns in the C-Pro of homo- versus non-homotrimerizing fibrillar procollagens. Non-homotrimerizing chains lack a cysteine at position 2 (or position 3 in some cases), preventing them from engaging in an interchain disulfide bond that covalently immortalizes C-Pro trimers. **(C)** C-Pro-only model for procollagen assembly based on the behavior of the C-Pro in isolation (22). Ca^{2+} mediates non-covalent, reversible assembly of all possible trimers. Disulfide bonds covalently immortalize assemblies whose composition is consistent with biologically relevant triple-helical assemblies. These steps, shown in the green box, have all been experimentally demonstrated using C-Pro-only constructs (22). The final step, shown in the grey box, in which only the disulfide-linked assemblies can drive triple-helix folding, is a prediction that is experimentally tested herein.

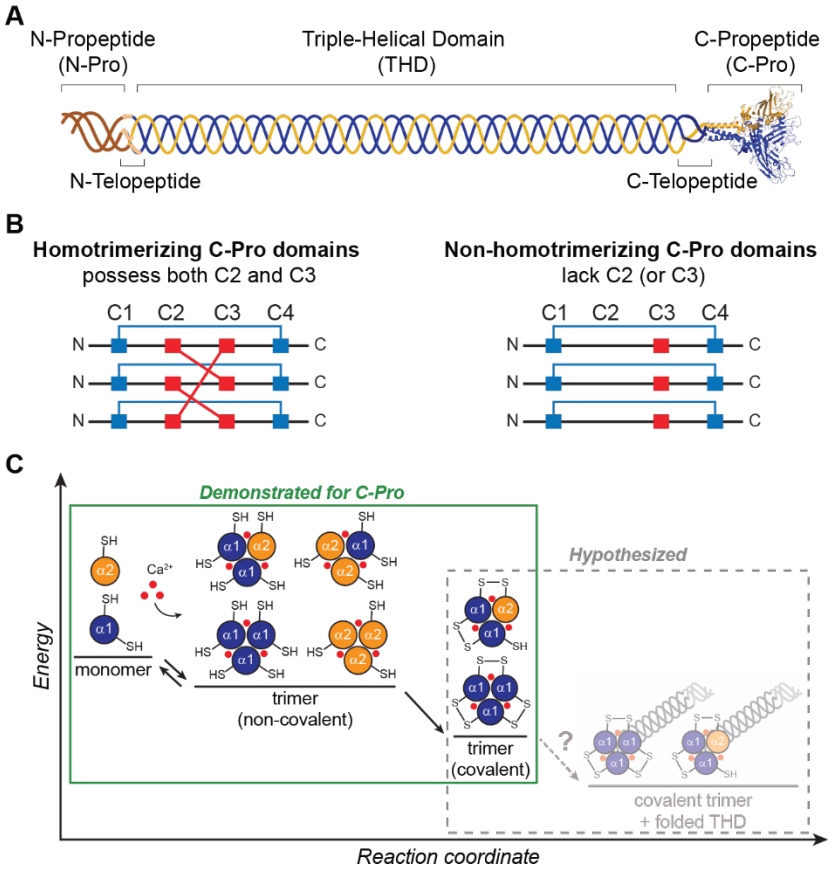
Figure 2. (A) Western blots of media from cells expressing HA-tagged wild-type or C2S pro-Colα1(I), or RFP as a negative control (ctrl). proColα1(I) was observed predominantly as a disulfide-linked trimer, whereas C2S proColα1(I) did not form disulfide-linked trimers. The blot was probed with an antibody against the HA epitope. **(B)** Properly folded collagen triple helices are resistant to protease digestion. Unfolded triple-helical domains, in contrast, are susceptible to digestion by proteases. **(C)** Western blot of procollagen precipitated from media of cells expressing proColα1(I), C2S proColα1(I), or RFP (negative control), or media from primary fibroblasts (fibro) that express endogenous procollagen-I (positive control), with or without protease treatment. Both primary fibroblasts and cells expressing wild-type proColα1(I) secrete protease-resistant, triple-helical collagen, as expected. Interestingly, C2S proColα1(I) can also form protease-resistant triple helices. The blot was probed with an antibody recognizing an epitope in the Colα1(I) THD.

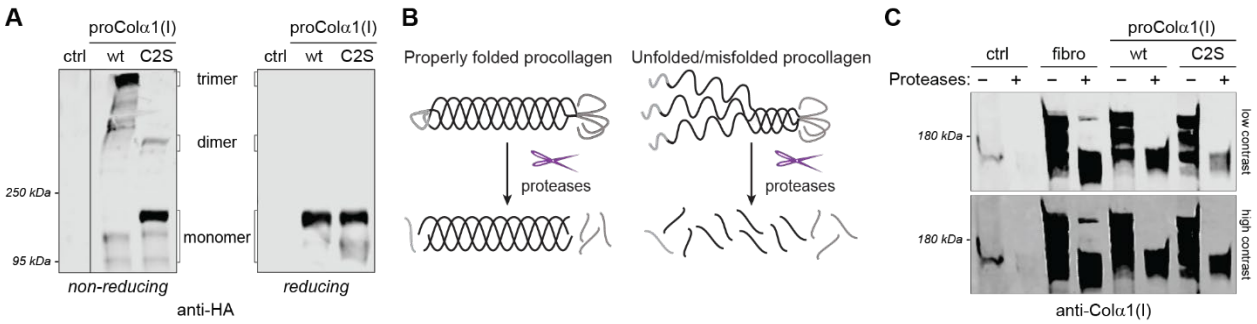
Figure 3. (A) Western blots of media from cells expressing HA-tagged wild-type or S2C pro-Colα2(I), or RFP as a negative control (ctrl). Wild-type proColα2(I) did not form any disulfide-linked assemblies, whereas S2C proColα2(I) was capable of forming disulfide-linked homotrimers. The blot was probed with an antibody against HA. **(B)** Western blot of procollagen precipitated from media of cells expressing wild-type proColα2(I), S2C proColα2(I), RFP as a negative control (ctrl), or media from primary fibroblasts (fibro) that express endogenous procollagen-I (positive control), with or without protease treatment. Whereas primary fibroblasts secrete protease-resistant proColα2(I), because it heterotrimerizes with endogenous proColα1(I), cells expressing just wild-type proColα2(I) or just S2C proColα2(I) fail to secrete protease-resistant, triple helical proColα2(I) homotrimers. The blot was probed with an antibody recognizing an epitope in the Colα2(I) THD.

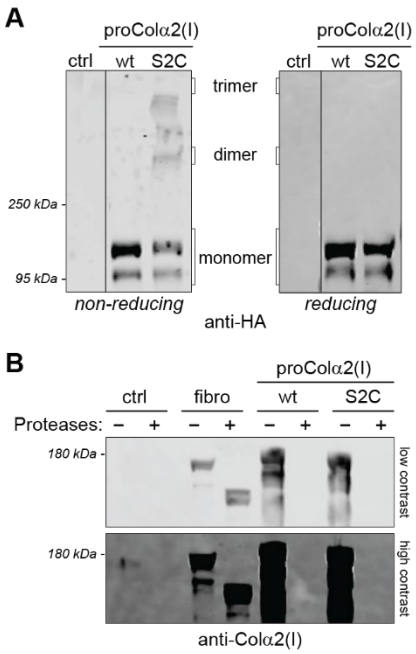
Figure 4. (A) Swapping the THD of proColα1(I) and proColα2(I) yields chimeras designated proα1(α2^{THD}) and proα2(α1^{THD}). The former is composed of the N- and C- telo- and propeptides of proColα1(I) flanking the proColα2(I) THD. The latter is composed of the N- and C- telo- and propeptides of proColα2(I) flanking the proColα1(I) THD. **(B)** Western blots of media from cells expressing the HA-tagged chimeras or RFP as a negative control (ctrl). proα2(α1^{THD}) did not form disulfide-linked homotrimers, consistent with its C2-lacking C-Proα2(I) domain. Proα1(α2^{THD}) could form disulfide-linked homotrimers, also consistent with its C2-containing C-Proα1(I) domain. The blot was probed with an antibody against HA. **(C)** Western blot of procollagen precipitated from media of cells expressing proColα1(I), proColα2(I), the indicated chimeras, or RFP (as a negative control; ctrl) and from primary fibroblasts that express endogenous procollagen-I (as a positive control; fibro), with or without protease treatment. Consistent with the sequence of its THD rather than its non-covalently homotrimerizing C-Pro domain, the proα2(α1^{THD}) chimera could form protease-resistant triple helices, similar to primary fibroblasts and wild-type proColα1(I). Also consistent with the sequence of its THD rather than its covalently homotrimerizing C-Pro domain, the proα1(α2^{THD}) chimera could not form protease-resistant triple helices. The blots were probed with antibodies against the Colα1(I) THD or Colα2(I) THD, respectively. Ctrl, fibroblast, and proColα2(I) lanes are reproduced from **Fig. 3B** for clarity.

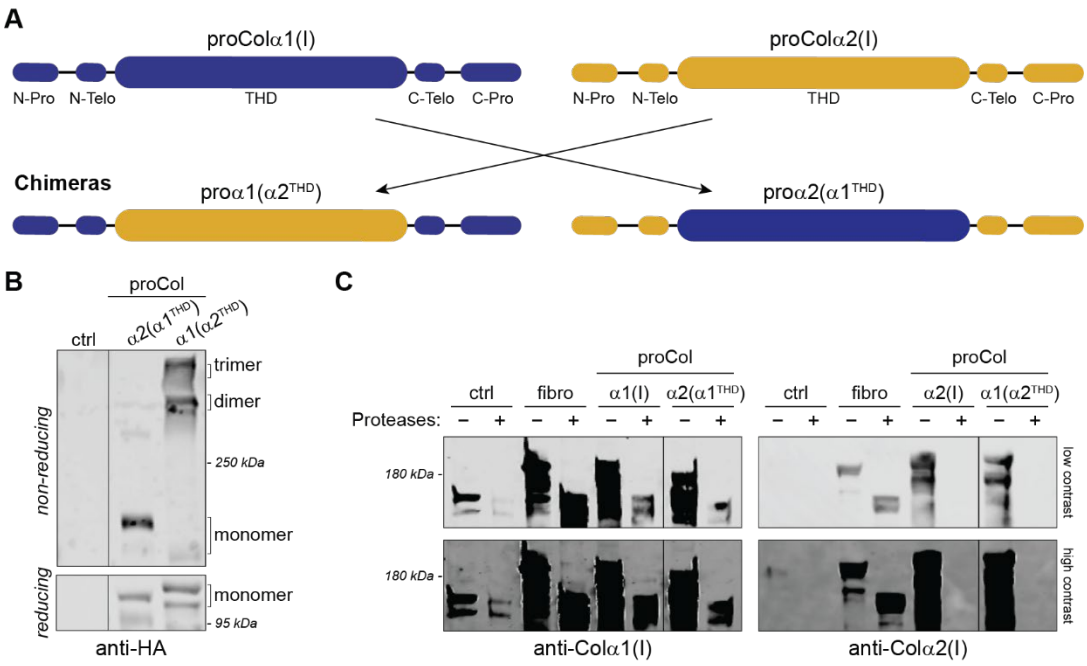
Figure 5. (A) Computational prediction of the triple-helix stability (T_m) encoded in the 15 most C-terminal THD triplets of proColα1(I) and proColα2(I) (with or without the GGG triplet included) (48). Note the presence of two wells of instability (highlighted in grey) where the predicted T_m for proColα2(I) is considerably lower than that of proColα1(I). **(B)** Experimentally measured T_m of paired host–guest peptides spanning the two wells of instability in the proColα2(I) THD. Triple helices formed from the proColα2(I) host–guest peptides all displayed substantially lower stability than the corresponding proColα1(I)-derived triple helices.

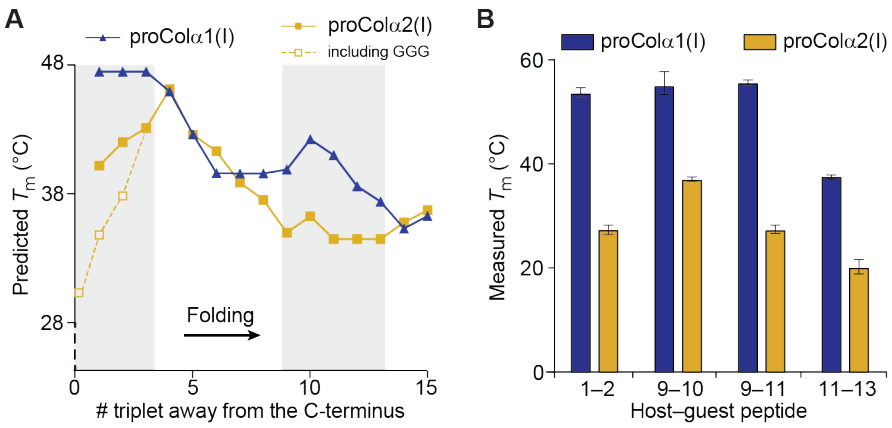
Figure 6. (A) Summary of the behavior of the different constructs studied in this work. Only pro-Colα2(I), which lacks the ability to both form a disulfide-linked trimer and a protease-resistant THD, results in a reversible outcome allowing the trimer to consistently dissociate and search for appropriate binding partners. **(B)** Beyond the C-Pro: A THD/C-Pro partnership model for procollagen-I assembly. Ca^{2+} mediates dynamic, reversible trimerization of all possible combinations of the C-Pro domain. Interstrand disulfide bonds enabled by the presence of key cysteine residues in the C-Pro domain covalently link those assemblies that are physiologically observed (i.e., 2:1 Colα1(I):Colα2(I) heterotrimers and Colα1(I) homotrimers). Covalent immortalization of undesired trimeric assemblies is precluded by missing Cys residues, and the presence of an unstable THD sequence avoids commitment to incorrect triple-helix folding, which would be functionally irreversible. Thus, a partnership between the THD and the C-Pro domain ensures that only desirable triple-helix compositions can be formed.


A
N-Propeptide
(N-Pro)
Triple-Helical Domain
(THD)
C-Propeptide
(C-Pro)
N-Telopeptide
C-Telopeptide
B
Homotrimerizing C-Pro domains
possess both C2 and C3
Non-homotrimerizing C-Pro domains
lack C2 (or C3)
C1 C2 C3 C4
N
C
C
Demonstrated for C-Pro
Hypothesized
monomer
trimer
(non-covalent)
trimer
(covalent)
covalent trimer
+ folded THD
Energy
Reaction coordinate

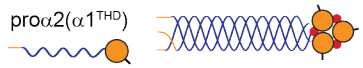

A
proColα1(I)
ctrl
wt
C2S
proColα1(I)
ctrl
wt
C2S
trimer
dimer
monomer
250 kDa
95 kDa
non-reducing
reducing
anti-HA
B
Properly folded procollagen
Unfolded/misfolded procollagen
proteases
proteases
C
proColα1(I)
ctrl
fibro
wt
C2S
Proteases: – + – + – + – +
180 kDa
low contrast
180 kDa
high contrast
anti-Colα1(I)


A
proColα2(I)
ctrl
wt
S2C
trimer
dimer
monomer
250 kDa
95 kDa
non-reducing
reducing
anti-HA
B
proColα2(I)
ctrl
fibro
wt
S2C
Proteases:
– + – + – + – +
180 kDa
low contrast
180 kDa
high contrast
anti-Colα2(I)

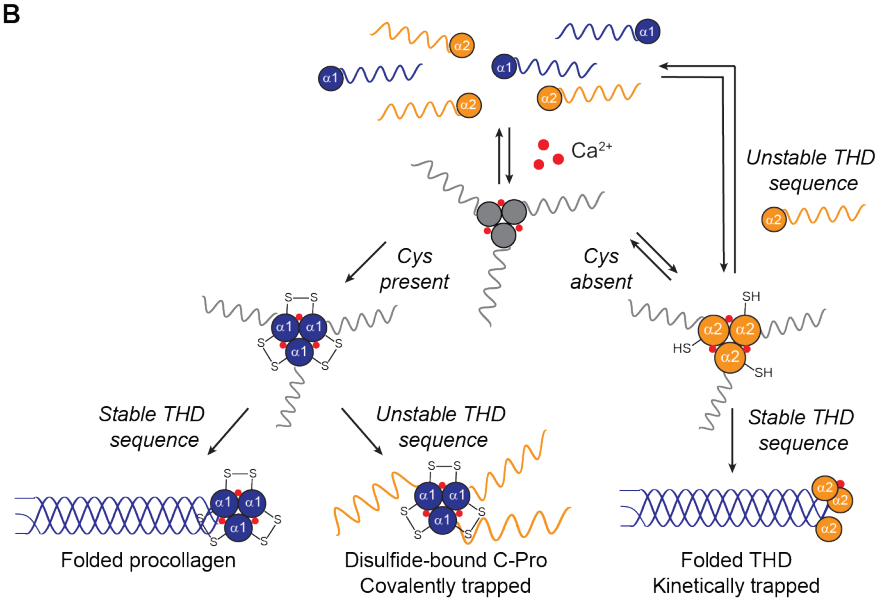

A
proColα1(I)
N-Pro
N-Telo
THD
C-Telo
C-Pro
proColα2(I)
N-Pro
N-Telo
THD
C-Telo
C-Pro
Chimeras
proα1(α2THD)
proα2(α1THD)
B
proCol
ctrl
α2(α1THD)
α1(α2THD)
trimer
dimer
250 kDa
monomer
non-reducing
reducing
monomer
95 kDa
anti-HA
C
proCol
ctrl
fibro
α1(I)
α2(α1THD)
proCol
ctrl
fibro
α2(I)
α1(α2THD)
Proteases:
– + – + – + – + – + – + – +
180 kDa
180 kDa
low contrast
high contrast
anti-Colα1(I)
anti-Colα2(I)


A
proColα1(I)
proColα2(I)
including GGG
48
38
28
Predicted T_m (°C)
Folding
0
5
10
15
triplet away from the C-terminus
B
proColα1(I)
proColα2(I)
60
40
20
0
Measured T_m (°C)
1–2
9–10
9–11
11–13
Host–guest peptide

A

Construct	Assembly	Disulfide-linked trimer?	Protease-resistant THD?	Outcome
proColα1(I)		✓	✓	Committed
C2S proColα1(I)		✗	✓	Committed
proα2(α1^{THD})		✗	✓	Committed
proColα2(I)		✗	✗	***Reversible***
S2C proColα2(I)		✓	✗	Committed
proα1(α2^{THD})		✓	✗	Committed

B



An outcome-defining role for the triple-helical domain in regulating collagen-I assembly

Kathryn M. Yammine, Rasia C. Li, Isabella M. Borgula, Sophia Mirda Abularach, Andrew S. DiChiara, Ronald T. Raines, and Matthew D. Shoulders

Page	Content
S1	Table of Contents
S2	Supporting Table
S3–S5	Amino Acid Sequences of Procollagen Constructs
S6–S12	Supporting Figures and Captions

SUPPORTING TABLE

Table S1 Host–guest peptide sequences and their measured melting temperatures. The guest peptides are bolded. O = (2*S*,4*R*)-4-hydroxyproline.

Triplets	Strand	Host–guest sequence	T_m (°C)
1 – 2	proColα1(I)	GPO-GPO-**GPO-GPO**-GPO-GPO-GPO	53.7
	proColα2(I)	GPO-GPO-**GPO-GVS**-GPO-GPO-GPO	27.2
9 – 10	proColα1(I)	GPO-GPO-GPO-**GPO-GPR**-GPO-GPO-GPO	55.0
	proColα2(I)	GPO-GPO-GPO-**GPA-GIR**-GPO-GPO-GPO	37.0
9 – 11	proColα1(I)	GPO-GPO-GPO-**GPI-GPO-GPR**-GPO-GPO-GPO	55.6
	proColα2(I)	GPO-GPO-GPO-**GTV-GPA-GIR**-GPO-GPO-GPO	27.3
11 – 13	proColα1(I)	GPO-GPO-GPO-**GLN-GLO-GPI**-GPO-GPO-GPO	37.5
	proColα2(I)	GPO-GPO-GPO-**GRT-GHO-GTV**-GPO-GPO-GPO	20.1
22 – 23	proColα1(I)	GPO-GPO-GPO-**GSO-GEQ**-GPO-GPO-GPO	35.6
	proColα2(I)	GPO-GPO-GPO-**GHH-GDQ**-GPO-GPO-GPO	17.2
24 – 26	proColα1(I)	GPO-GPO-**GLQ-GPO-GPO**-GPO-GPO-GPO	44.9
	proColα2(I)	GPO-GPO-**GLQ-GLO-GIA**-GPO-GPO-GPO	25.4

AMINO ACID SEQUENCES OF PROCOLLAGEN CONSTRUCTS

Legend:

Preprotrypsin leader, HA tag, *Telopeptide,* **Triple-helical domain**, C-propeptide Cys, Cys/Ser mutagenesis site

proColα1(I)

MSALLILALVGAAVAYPYDVPDYAAAQEEGQVEGQDEDIPPITCVQNGLRYHDRDVWK
PEPCRICVCDNGKVLCDDVICDETKNCPGAEVPEGECCPVCPDGSESPTDQETTGVEGP
KGDTGPRGPRGPAGPPGRDGIPGQPGLPGPPGPPGPPGPPGLGGNFAP*QLSYGYDEKST*
*GGISVP***GPMGPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGD**
DGEAGKPGRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPG
SPGENGAPGQMGPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAV
GAKGEAGPQGPRGSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGA
PGFPGARGPSGPQGPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEG
KRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEA
GRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGP
KGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQG
LPGPAGPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAP
GNDGAKGDAGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGV
RGLTGPIGPPGPAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADG
QPGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAA
GRVGPPGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGEVGPPGPPGPAGEKGSPGA
DGPAGAPGTPGPQGIAGQRGVVGLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPG
PMGPPGLAGPPGESGREGAPGAEGSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPV
GPAGKSGDRGETGPAGPAGPVGPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGL
QGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAG
PVGPPGPPGPPGPPGPPGP*SAGFDFSFLPQPPQEKAHDGGRYYRA*DDANVVRDRDLEVDTT
LKSLSQQIENIRSPEGSRKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNM
ETGETCVYPTQPSVAQKNWYISKNPKDKRHVWFGESMTDGFQFEYGGQGSDPADVAIQL
TFLRLMSTEASQNITYHCKNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVT
VDGCTSHTGAWGKTVIEYKTTKTSRLPIIDVAPLDVGAPDQEFGFDVGPVCFL

proColα2(I)

MSALLILALVGAAVAYPYDVPDYAAAQSLQEETVRKGPAGDRGPRGERGPPGPPGRDG
EDGPTGPPGPPGPPGPPGPPGLGGNFAA*QYDGKGVGLGP***GPMGLMGPRGPPGAAGAPGPQGF**
QGPAGEPGEPGQTGPAGARGPAGPPGKAGEDGHPGKPGRPGERGVVGPQGARGFPGTPG
LPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQTGARGLPGERGRVGAPGPA
GARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGAVGNAGPAGPAGPRGEVGLPGL
SGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGLVGEPG
PAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGPPGLRGSPGSRGLP

GADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGP
VGLPGIDGRPGPIGPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDG
NNGAQGPPGPQGVQGGKGEQGPPGPPGFQGLPGPSGPAGEVGKPGERGLHGEFGLPGPA
GPRGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAVGTAGPSGPSGLPGE
RGAAGIPGGKGEKGEPGLRGEIGNPGRDGARGAPGAVGAPGPAGATGDRGEAGAAGPAG
PAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGAKGPKGENGVVGPTGPVGAA
GPAGPNGPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGP
RGDQGPVGRTGEVGAVGPPGFAGEKGPSGEAGTAGPPGTPGPQGLLGAPGILGLPGSRG
ERGLPGVAGAVGEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRD
GQPGHKGERGYPGNIGPVGAAGAPGPHGPVGPAGKHGNRGETGPSGPVGPAGAVGPRGP
SGPQGIRGDKGEPGEKGPRGLPGLKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAG
PSGPAGKDGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPGVS*GGGYDFGYDGDF*
YRADQPRSAPSLRPKDYEVDATLKSLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSS
GYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIPAKNWYRSSKDKKHVWLGETINA
GSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCKNSIAYMDEETGNLKKAVILQ
GSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGA
DQEFFVDIGPVCFK

proα1(α2^{THD})

MSALLILALVGAAVAYPYDVPDYAAAQEEGQVEGQDEDIPPITCVQNGLRYHDRDVWK
PEPCRICVCDNGKVLCDDVICDETKNCPGAEVPEGECCPVCPDGSESPTDQETTGVEGP
KGDTGPRGPRGPAGPPGRDGIPGQPGLPGPPGPPGPPGPPGLGGNFAP***QLSYGYDEKST***
GGISVPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGARGPAGPPGKAGE
DGHPGKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPG
APGENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAP
GPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGA
PGLPGPRGIPGPVGAAGATGARGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEG
KRGPNGEAGSAGPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSRGASGPAGVRGPNGDA
GRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGEPGNIGFPGP
KGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGVQGGKGEQGPPGPPGFQG
LPGPSGPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPP
GPDGNKGEPGVVGAVGTAGPSGPSGLPGERGAAGIPGGKGEKGEPGLRGEIGNPGRDGA
RGAPGAVGAPGPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAG
QPGAKGERGAKGPKGENGVVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGMTGFPGAA
GRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTGEVGAVGPPGFAGEKGPSGE
AGTAGPPGTPGPQGLLGAPGILGLPGSRGERGLPGVAGAVGEPGPLGIAGPPGARGPPG
AVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPGNIGPVGAAGAPGPHGPV
GPAGKHGNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGPRGLPGLKGHNGL
QGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVGPAGIRGPQGHQG
PAGPPGPPGPPGPPGVS*SAGFDFSFLPQPPQEKAHDGGRYYRA*DDANVVRDRDLEVDTT
LKSLSQQIENIRSPEGSRKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNM
ETGETCVYPTQPSVAQKNWYISKNPKDKRHVWFGESMTDGFQFEYGGQGSDPADVAIQL
TFLRLMSTEASQNITYHCKNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVT
VDGCTSHTGAWGKTVIEYKTTKTSRLPIIDVAPLDVGAPDQEFGFDVGPVCFL

proα2(α1^{THD})

MSALLILALVGAAVAYPYDVPDYAAAAQSLQEETVRKGPAGDRGPRGERGPPGPPGRDG
EDGPTGPPGPPGPPGPPGPPGLGGNFAA*QYDGKGVGLGP***GPMGPSGPRGLPGPPGAPGPQGF**
QGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPQGARGLPGTAG
LPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRGLPGERGRPGAPGPA
GARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGPRGSEGPQGVRGEPGPPGP
AGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQGPGGPPGPKGNSGEPG
APGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFP
GADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGP
PGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDG
EAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPS
GARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGE
RGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDKGESGPSGPAG
PTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPP
GPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKEGGKGP
RGETGPAGRPGEVGPPGPPGPAGEKGSPGADGPAGAPGTPGPQGIAGQRGVVGLPGQRG
ERGFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAEGSPGRD
GSPGAKGDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGETGPAGPAGPVGPVGARGP
AGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPG
SAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP*GGGYDFGYDGDF*
*YRA*DQPRSAPSLRPKDYEVDATLKSLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSS
GYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIPAKNWYRSSKDKKHVWLGETINA
GSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCKNSIAYMDEETGNLKKAVILQ
GSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGA
DQEFFVDIGPVCFK

SUPPORTING FIGURES

FIGURE S1

Figure S1 C-terminal sequence of proColα1(I) and proColα2(I) highlighting the difference in proline content. The beginning of the C-telopeptide is denoted by |.

FIGURE S2

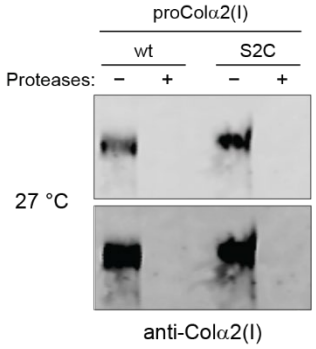


Figure S2 Protease digestion of media from cells expressing proColα2(I) and proColα2(I) S2C cultured at 27 °C. Even at the lower temperature, these constructs do not form detectable amounts of protease-resistant triple helices. The Western blot was probed with an antibody against Colα2(I).

FIGURE S3

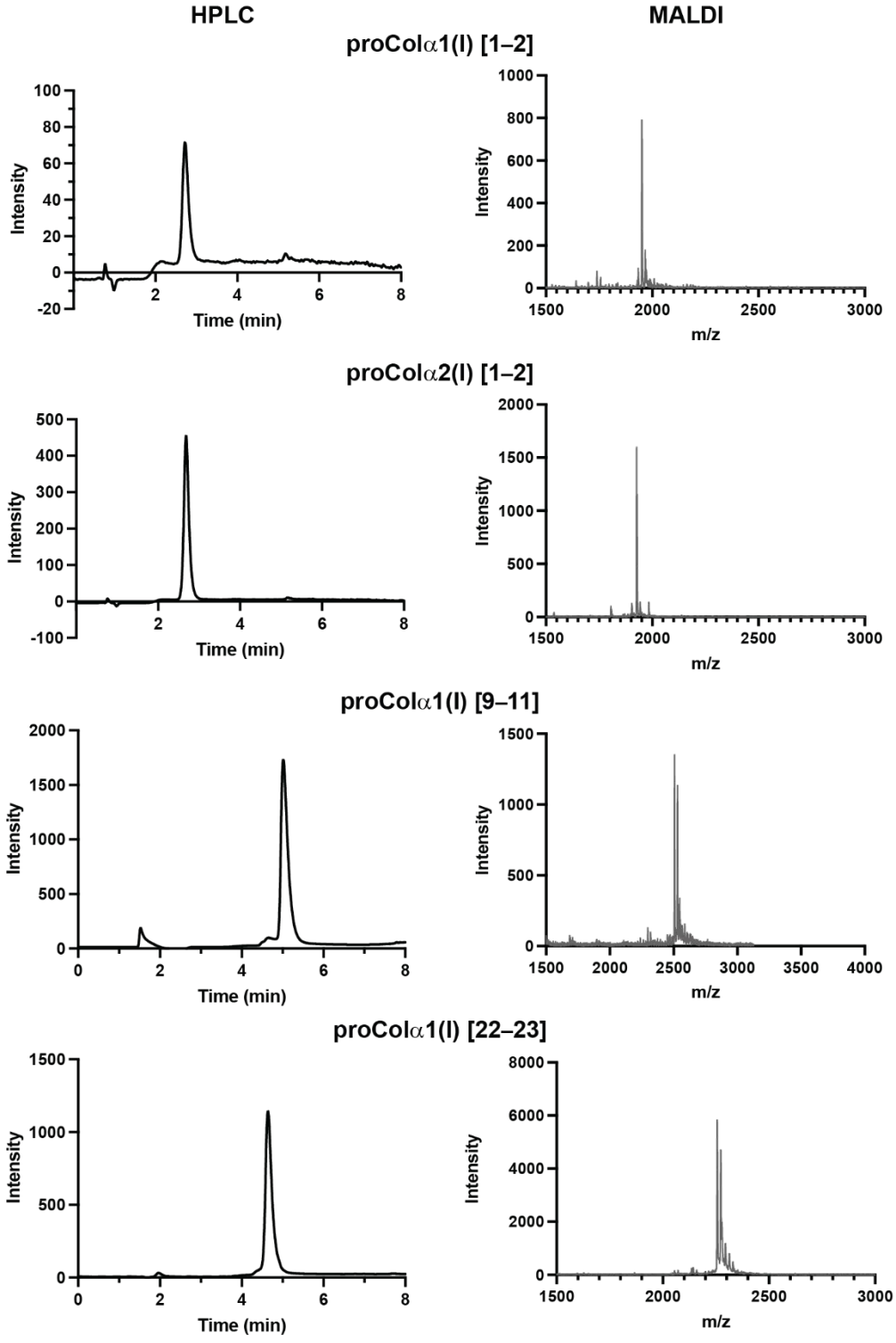

HPLC
MALDI
proColα1(I) [1–2]
Intensity
Time (min)
m/z
proColα2(I) [1–2]
proColα1(I) [9–11]
proColα1(I) [22–23]

Figure S3 HPLC and MALDI spectra of each synthesized peptide to confirm purity. HPLC Gradient: 0%–80% v/v acetonitrile in water + 0.1% v/v trifluoroacetic acid over 8 min. $[M + H]^+$ (Da): calculated 1927.891, found 1926.745; calculated 1905.890, found 1906.823; calculated 2507.224, found 2507.866; calculated 2235.322, found 2236.767. The numbers in brackets denote the collagen-I triplets serving as guests.

FIGURE S4

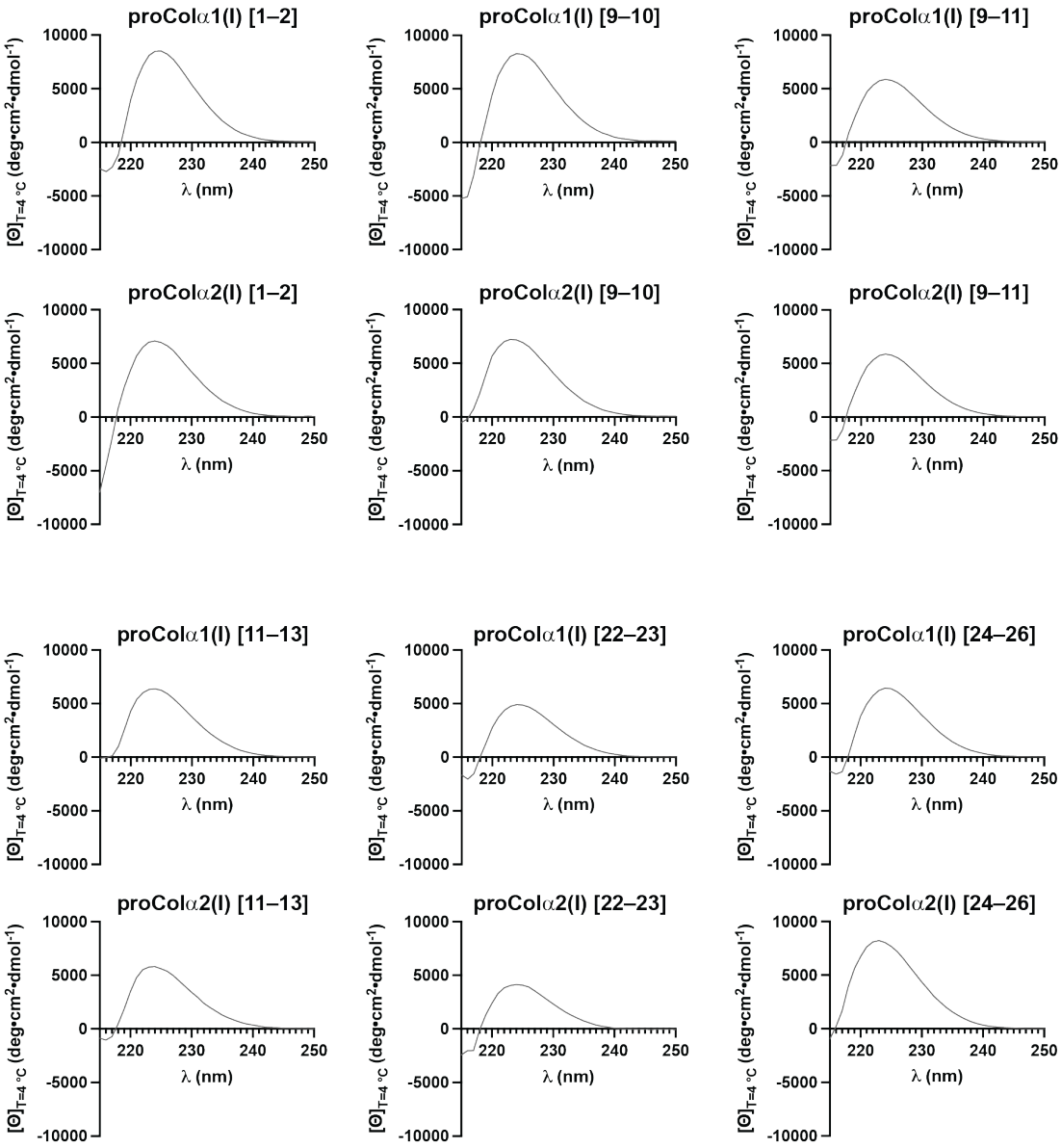


Figure S4 CD spectra of collagen-mimetic peptides (0.8 mM in 50 mM acetic acid, pH 3.0). The numbers in brackets denote the collagen-I triplets serving as guests.

FIGURE S5

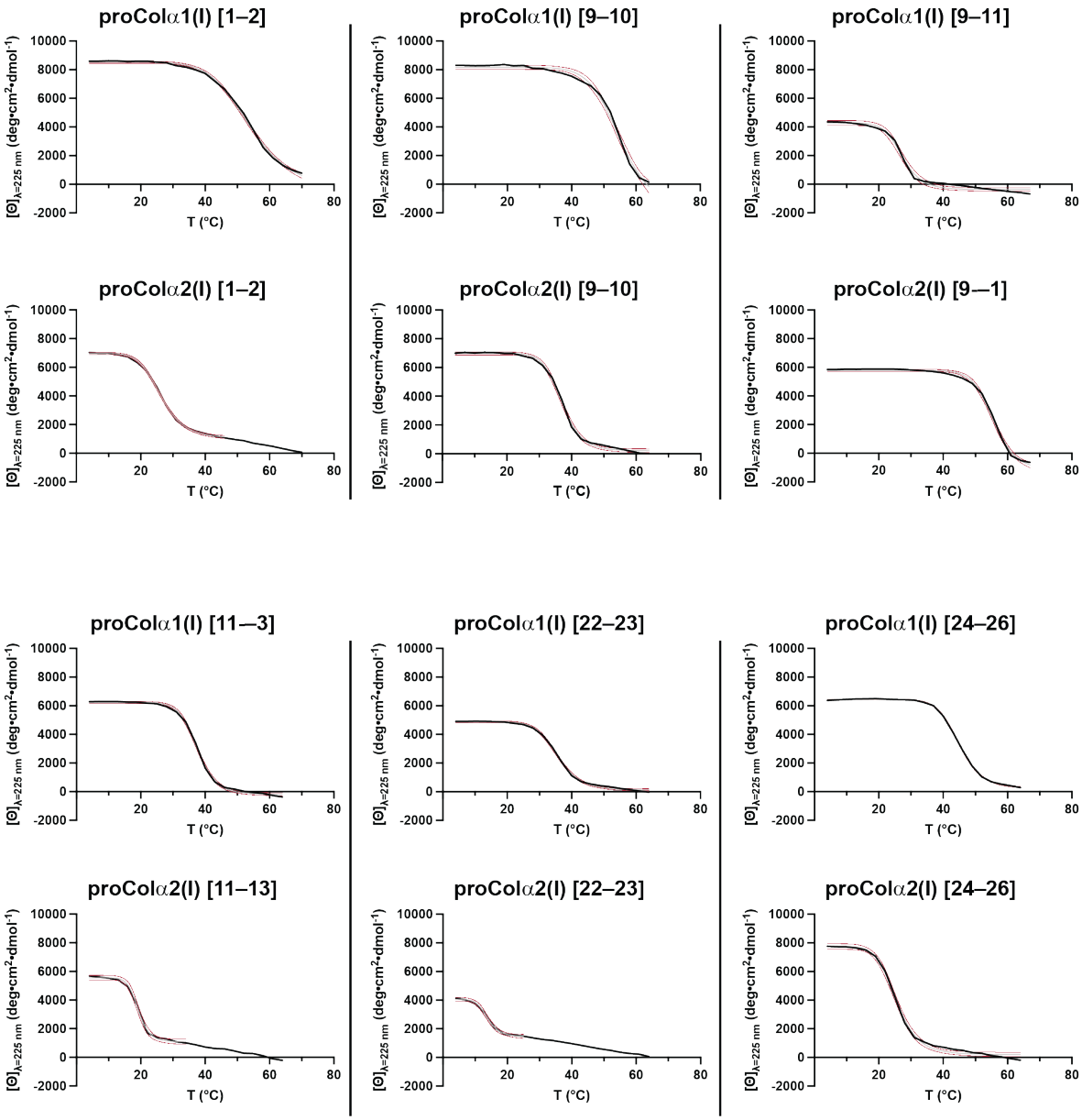


Figure S5 Thermal denaturation data of triple helices formed by collagen mimetic peptides (0.8 mM in 50 mM acetic acid, pH 3.0) as recorded at 225 nm, which has the maximum ellipticity in the CD spectra (**Figure S4**). Each of these triple helices undergoes cooperative denaturation, in which the trimers melt to form monomers. Values of T_m were calculated by fitting the temperature range containing the cooperative denaturation to a 4-parameter Hill equation. The 95% confidence interval for the best fit is shown in purple. The numbers in brackets denote the collagen-I triplets serving as guests.

FIGURE S6

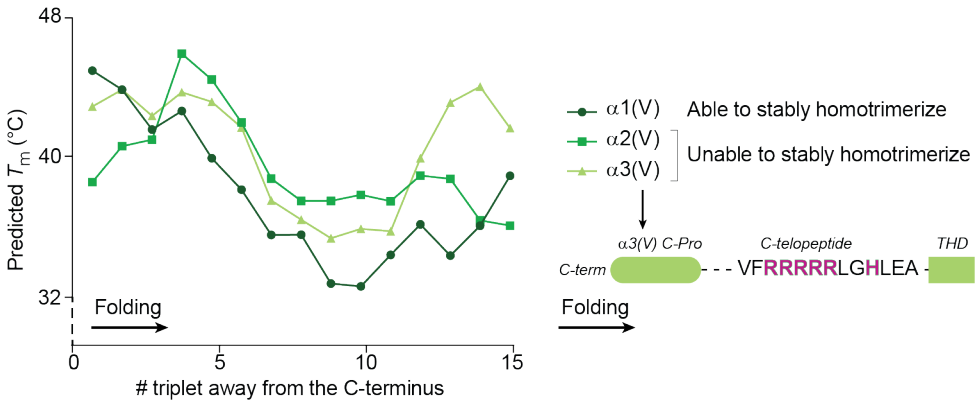


Figure S6 Computational prediction of the T_m values of the 15 most C-terminal THD triplets of procollagen-V. The low T_m of proColα2(V) across the first few triplets to fold likely prevents its forming stable homotrimers; similarly, the highly charged telopeptide of proColα3(V) likely prevents stable proColα3(V) homotrimerization. Charged residues are depicted in purple.