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

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

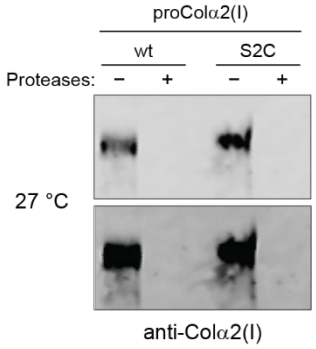

Figure S2 Protease digestion of media from cells expressing proColα2(I) and proColα2(I) S2C cultured at 27 °C. Even at the lower temperature, these constructs do not form detectable amounts of protease-resistant triple helices. The Western blot was probed with an antibody against Colα2(I).

FIGURE S3

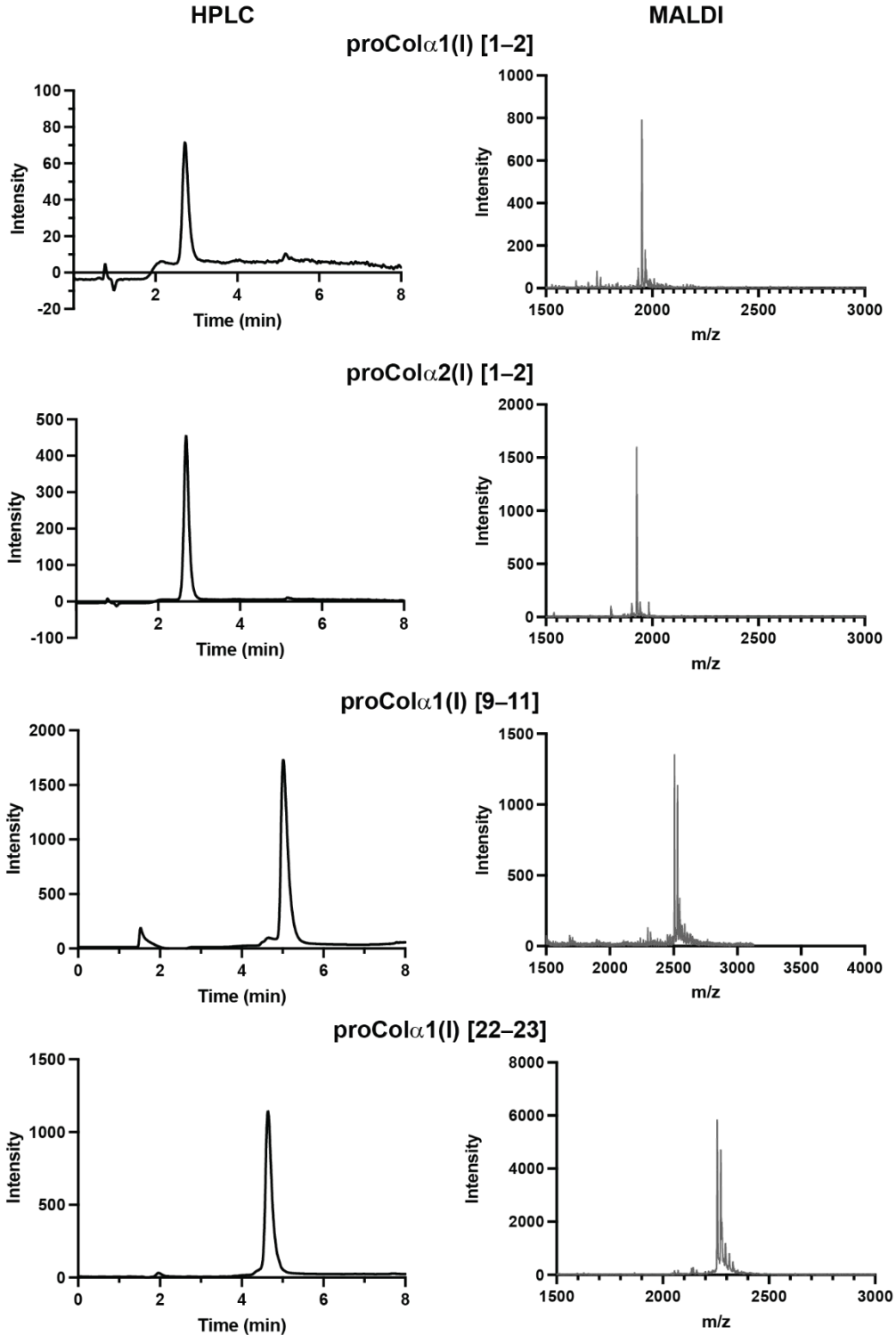
HPLC
MALDI
proColα1(I) [1–2]
Intensity
Time (min)
m/z
proColα2(I) [1–2]
proColα1(I) [9–11]
proColα1(I) [22–23]

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

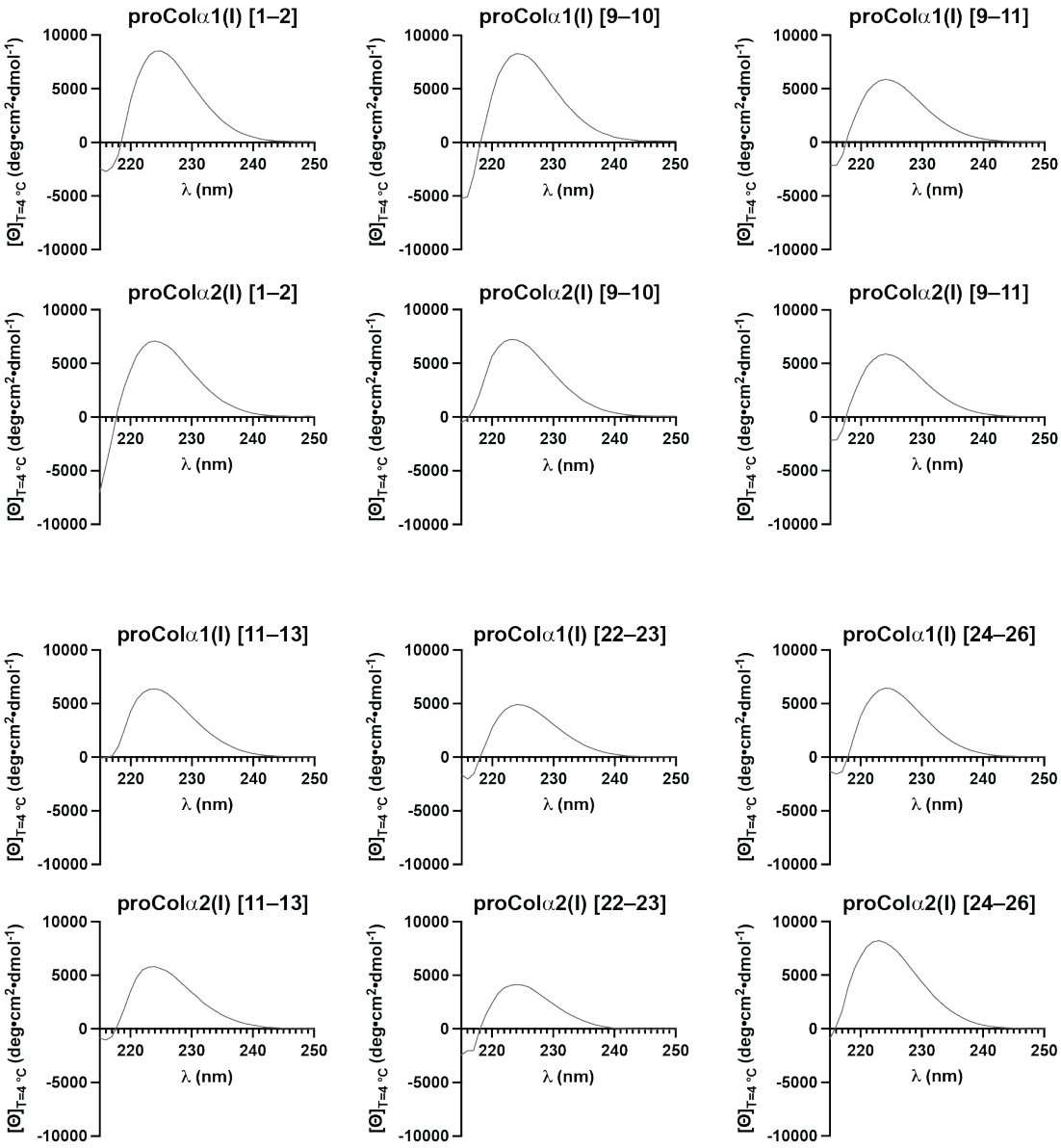

Figure S4 CD spectra of collagen-mimetic peptides (0.8 mM in 50 mM acetic acid, pH 3.0). The numbers in brackets denote the collagen-I triplets serving as guests.

FIGURE S5

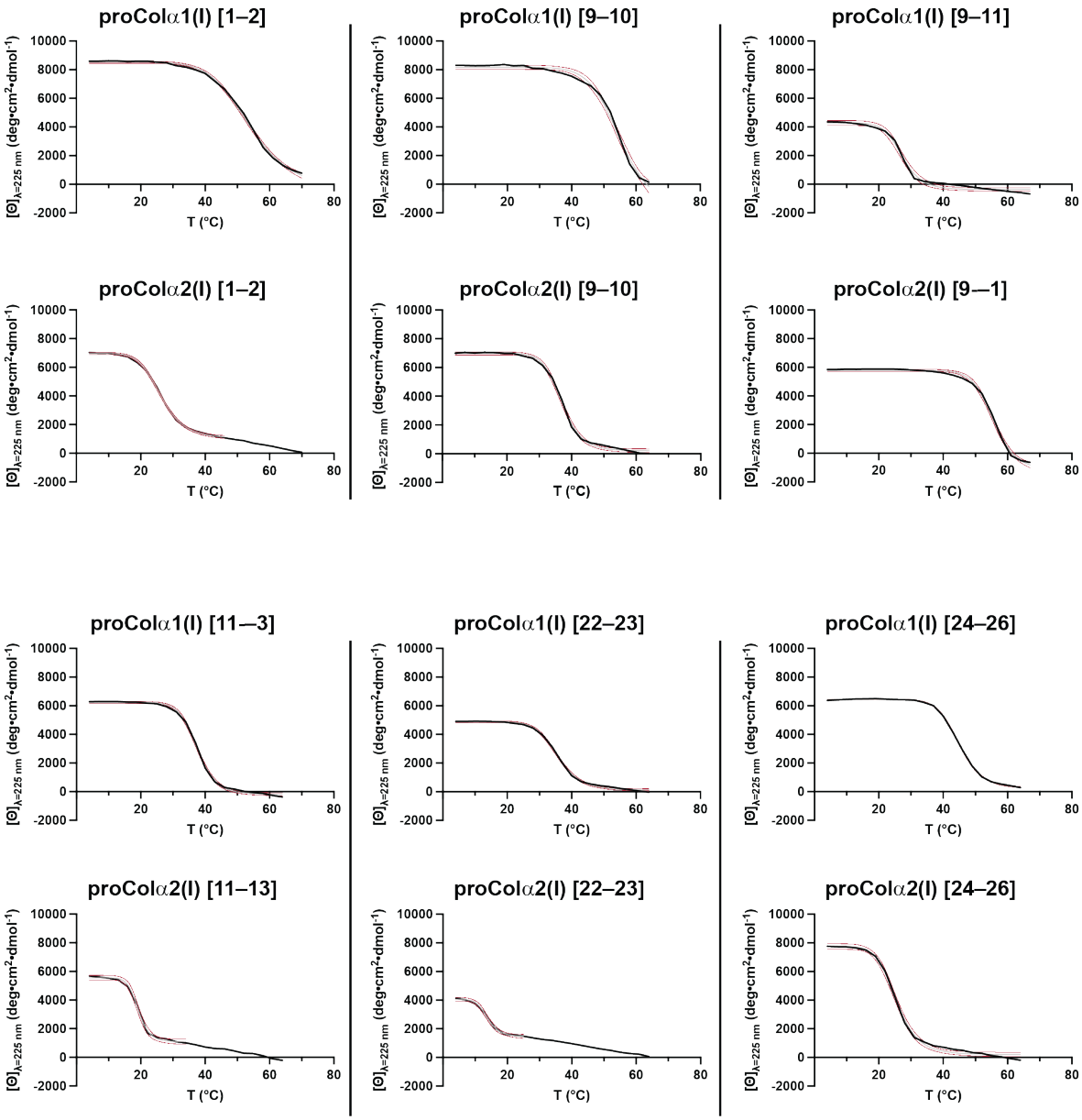

Figure S5 Thermal denaturation data of triple helices formed by collagen mimetic peptides (0.8 mM in 50 mM acetic acid, pH 3.0) as recorded at 225 nm, which has the maximum ellipticity in the CD spectra (**Figure S4**). Each of these triple helices undergoes cooperative denaturation, in which the trimers 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

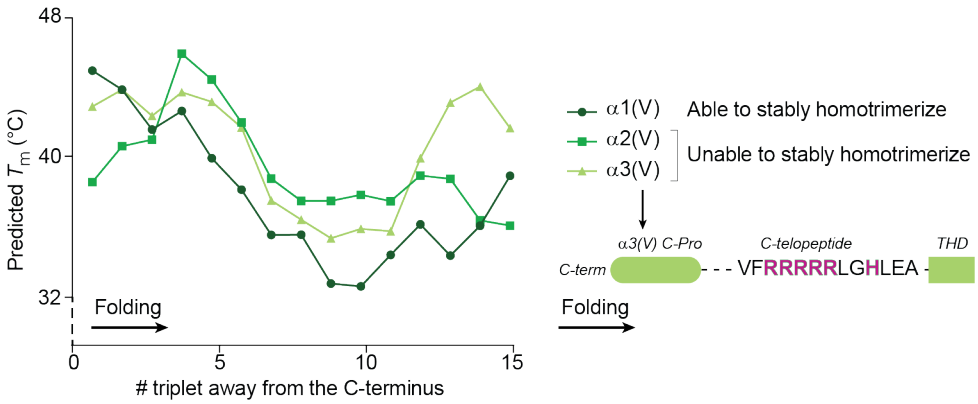

Figure S6 Computational prediction of the T_m values of the 15 most C-terminal THD triplets of procollagen-V. The low T_m of proColα2(V) across the first few triplets to fold likely prevents its forming stable homotrimers; similarly, the highly charged telopeptide of proColα3(V) likely prevents stable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