

## **A Cysteine-Based Molecular Code Informs Collagen C-Propeptide Assembly**

DiChiara, A.S. and co-workers

<b>Page</b>	<b>Contents</b>
S1	Table of Contents
S2–S10	Supplementary Figures
S11–12	Supplementary Tables
S13–16	Supplementary Note

## SUPPLEMENTARY FIGURES

		C1	C2	
sponge_ColF1	-----VNLGSVADVIELHKKLQHL--KSPTGT-KDSPARS	C	HDLFLE	382
Hydra_colα1(I)	-----EKNKDTV1-EKLIRLEDITQGA--QKPDGS-EYFPAKS	C	KDLKMC	1262
Hydra_col1	-----TEIENNFNRRVKILKSSVEAY--KKPNGS-KEFPART	C	RDIIYAF	1213
Hydra_col2	-----EKNKDTV1-EKLIRLEDITQGA--QKPDGS-EYFPAKS	C	KDLKMC	1286
Hydra_col3	-----LVKLIKRR--TQENLRNFNKVWNTVLDLVLKNELG-T-RMHPAIT	C	RHFVFN	1690
Hydra_col5	-----VDEIIKR--NIDGFTMYRVFKT-ADKV-FLGGDGT-RDNPARS	C	LELFW	1194
Red fire ant_α1(I)	-----ENLSSFKL--VKPDG-EKNSPAKT	C	RDLFV	29
Mollusk_α1(I)	-----QYE--NLNRVREAIVRI--GGTRLGSRTPSGKN	C	RDIKLS	734
Sea urchin_Col1Pa	-----KIQDTELLGAIISALGQQIELI--KAPQKAKTNPARS	C	KDVFLN	3362
Sea urchin_Col2Pa	-----DRTOFQIYLAKFESEILSL--IEPLGS-RDQPIRS	C	KDLFKC	3008
lancelet_α1(I)	-----IGKDTKVSDALEISAQIESL--KKPTGT-RKNPART	C	VLDLAL	217
lancelet_α2(I)	STDIDHLKKLISRVPAAEWTDYFEGEMDRLQGI V K T I -VGGPWGT-AEYPART	C	KDLMFA	2093
Tunicate_α1(I)	-----PEGLEEIIYAAMETLKQOLEMM--KEPMGRTOQDNPGRS	C	KDIWLC	1177
Tunicate_α2(I)	-----QPVTS--VGRDPEMLVLKELTSSVEDI--KAPRGVSRKTPARS	C	LDIYLA	1225
Lamprey_α1(I)	-----LPYGLVLDASIKALNAQVESI--TSPDGS-RKHPARS	C	RDLMLC	1078
Lamprey_α2(I)	-----GERNLEASISSLNVRVENM--VSPDGS-QKNPART	C	RNLKLC	992
Shark_α1(I)	-----RNRDLEVDSTLKSLSQIENI--RSPEGT-KKNPART	C	RDLMKC	1256
Shark_α2(I)	-----SAKDFEVDATLKLQSSQIETL--LFPEGS-KKNPART	C	RDRLLS	1168
Coelacanth_α1(I)	-----RDRHAEVETTLKSLTKQIDNI--RSPEGT-RKNPART	C	RDLMKC	1262
Coelacanth_α2(I)	-----TEKDLEVDSTLKVLTNQIESI--RTPEGS-RKNPARS	C	RDRLLS	1169
Frog_α1(I)	-----RDRDHEVDSTLKSLSQIENI--QSPEGT-RKNPART	C	RDLMKC	1250
Frog_α2(I)	-----RPKDYEV DATLKSLSLNQIETI--LTPEGS-KKNPART	C	RDRLLS	1156
Human_α1(I)	-----RDRDLEVDSTLKSLSQIENI--RSPEGS-RKNPART	C	RDLMKC	1265
Human_α2(I)	-----RPKDYEV DATLKSLSLNQIETL--LTPEGS-RKNPART	C	RDRLLS	1169
		C3	C4	C5
sponge_ColF1	--DNST--SD--GYYWIDPNGGIGDAVKVFCNFTGG--VQQTCISATKNAG--			426
Hydra_colα1(I)	--HPNV--IS--GEYYIDPSLGNVNDKIKVTC E F Y --PTASETCIKPTVSMF--			1306
Hydra_col1	--YPDS--SS--GMYIDPNKGGIDDAIYVHCNFTKSHDEETKITTCVYPEKTMSS--			1262
Hydra_col2	--HPNV--IS--GEYYIDPSLGNVNDKIKVTC E F Y --PTASETCIKPTVSMF--			1330
Hydra_col3	--NNNS--KS--GDYWIDPNEGSPVDAFLVYCNAS--TLETCTIPFKQPLV--			1732
Hydra_col5	--HLGA--NT--GDYWIDPNEGSPDSDVLVHCNKL--TNETCVYSKNTQI--			1236
Red fire ant_α1(I)	A-YPD--KLSGEYWIDPNEGIRDAILVYC--DAE--KRA--TCILPNP--S--RSP			73
Mollusk_α1(I)	--NPDF--KD--GDYWIDPNGSALDAVKVFCRM--ETLETCTVCKPKIEY--			776
Sea urchin_Col1Pa	--NVEA--ES--GYYWVDPNLGQKDAIQVYCEA--ETGATCVPTNNVV--			3404
Sea urchin_Col2Pa	--YPEA--ED--GNYWIDPNEGSKDAFLAHCVKRGE--SGSPETCITPRVDEI--			3054
lancelet_α1(I)	--HPTW--PS--GNYWIDPNQGTVDIAIEVWCDM--KTLETCTVYKPAKV--			259
lancelet_α2(I)	--QPGL--KD--GYYFIDPDGGIENAFQAF CNFT--AGGLTCFQPTNDTL--			2136
Tunicate_α1(I)	--HPDF--PS--GNYWIDPNGGCSADAIIEVFCDFE--AEGDTCISPVERTASVSWL			1225
Tunicate_α2(I)	--EQQGTVPKSGVRWIDPNGGQADGLEVYCNF--HTMETCVVYPTNRNI--			1271
Lamprey_α1(I)	--HPEY--KSDTGEYWIDPNEGCRADAIKVWCNM--ETGESCVNPGMPSL--			1122
Lamprey_α2(I)	--YDDL--PS--GLYWIDPNEGSKSDAIQVWCNM--ETGESCMNADNPSI--			1034
Shark_α1(I)	--HPEW--KS--GDYWIDPNQGSALDAIRVYCNL--ETGETCVVSSPQSI--			1304
Shark_α2(I)	--HPEW--KS--GYYWIDPNQGTQDAIRVFCDF--TTGETCVHASPDI--			1211
Coelacanth_α1(I)	--HPEW--KS--GDYWIDPNQGMMLDAIKVYCNM--ETGETCVVYPTQSSV--			1298
Coelacanth_α2(I)	--HPEW--SS--GFYWIDPNQGTMDAIKVYCDF--TNGETCINANPETI--			1210
Frog_α1(I)	--HSDW--KS--GEYWIDPNQGIILDAIKVYCNM--ETGETCVVYPTQNSI--			1292
Frog_α2(I)	--HPDW--SS--GFYWIDPNQGTSDSIRVFCDF--STGETCIIHANPDSI--			1198
Human_α1(I)	--HSDW--KS--GEYWIDPNQGNLDAIKVYCNM--ETGETCVVYPTQPSV--			1307
Human_α2(I)	--HPEW--SS--GYYWIDPNQGTMDAIKVYCDF--STGETCIIAQPENI--			1211
sponge_ColF1	DLKS--WSGH-----SIWFSDM-LGGFKLTYDIS-----			452
Hydra_colα1(I)	EKKK--WVSQST--DQWKWFNGEISSDSIFTYASQ-----			1337
Hydra_col1	VEKD--SWPTKL-HTKAQRWFVEDH-ELGKLSYAAD-----			1294
Hydra_col2	EKKK--WVSQST--DQWKWFNGEISSDSIFTYASQ-----			1361
Hydra_col3	EKAD--WFTGK--DHLMWAYKDLAEGGITYSSD-----			1762
Hydra_col5	SKNN--YFKDNT--DSYKWLMEVA--NNFDYAME-----			1265
Red fire ant_α1(I)	NITH-ITE----QOET--WLSE-IDNG--MK-----I-T			96
Mollusk_α1(I)	RRDR--WTKDTT--SGQYFMYDVPFGKVKQFYDID-----			807
Sea urchin_Col1Pa	SNMT--WYVGKT--KR--AFFSSMHGGDKFAYIED-----			3433
Sea urchin_Col2Pa	SRAR--WYEGASG--SRYI--TEMGLEKFSYEAS-----			3082
lancelet_α1(I)	PKAS--WYNGP--AKHVWFSESIKGGYQFGYTAD-----			289
lancelet_α2(I)	ASPR--SPYK--VDNFTWFSEV-EGGFEIEYEGD-----			2165
Tunicate_α1(I)	TSKR--WPKA-----QPGDWFSY-RMGDRFEYNTS-----			1253
Tunicate_α2(I)	ENGT--HYTGE--PGHTYIYGEEMTRVEHADY-----			1298
Lamprey_α1(I)	PRKN--WWRSA-ADKKHVWLGETMSPGSPQTYGDG-----			1155
Lamprey_α2(I)	ARKN--WWLKPS-GSKKHVWFGVTMSPDAQTYGED-----			1067
Shark_α1(I)	PQKN--WYTSKNPKKKHWFGEISMNGGFQFGYDAD-----			1339
Shark_α2(I)	ERRN--WWISNDAQEKHWFGETIKDGAQTYNNEE-----			1245
Coelacanth_α1(I)	PQKN--WYVSKNPKDKKHWFGETMNEGFPQFEYGE-----			1332
Coelacanth_α2(I)	PSKT--WYISKNPMDKKHWFGETINGGTQFEYNDE-----			1244
Frog_α1(I)	AQKN--WYTSKNPREKKHVWFGAAMSDFQFEYGE-----			1326
Frog_α2(I)	PQKN--WYTSKT-QQRKHVWFGETINGGTQIEYNNEE-----			1231
Human_α1(I)	AQKN--WYISKNPDKRHVWFGESMTDGFQFEYGGQ-----			1341
Human_α2(I)	PAKN--WYRSS--KDKKHVWLGETINAGSQFEYNVE-----			1243

**C6**

sponge_ColF1	-----RSQLQFIRAASRHAVQSFTYKCRNSAAAVIFR	484
Hydra_colα1(I)	-----VQLRFLRLNSQFVRQNLTYHCLNSHAHNSDF	1368
Hydra_col1	-----QSQLTFLGLYSREAYQNVTFHCQNTLVVYDKQ	1326
Hydra_col2	-----VQLRFLRLNSQFVRQNLTYHCLNSHAHNSDF	1392
Hydra_col3	-----MVQLKMMQLLSAKCRQNTYFCKNSYSNITIK	1794
Hydra_col5	-----TPQMKILSLLSTTVRQTITFHCKNIPLLRGA-	1296
Red fire ant_α1(I)	YK---A-D-----SNQISF---LQLLSKHANQNTYHCKNSVAYFDY-	131
Mollusk_α1(I)	-----AYQLKVLQFDSQAARQGVYHCLNSHTYGRTRF	839
Sea urchin_Col1Pa	-----STQMTFLRLLSTARSQTVTYFCKNVQ-----	3459
Sea urchin_Col2Pa	-----EVQLTFLRLLSTKAHQNVTYHCKNSVAVRDRQ	3114
lancelet_α1(I)	-----DIQMQLRLLSSTSARQNVTYHCKNSVAYYDAV	321
lancelet_α2(I)	-----PIQLNYIQALSTRATQTFTECSSAVAWYNNW	2197
Tunicate_α1(I)	-----IPQFNFLRLLSSQAKQRFYTKCVNSIGWENQQ	1285
Tunicate_α2(I)	-----ASQLTFLRLLSSKAKQVTFRCRNMVAYYDAS	1330
Lamprey_α1(I)	-----SPNMEVQLTFLRLLSTDASQKITTYHCKNSVAYLDSR	1191
Lamprey_α2(I)	-----SHSTEIQTLFLRFLSTEASQKITTYHCKNSVAYQDGA	1103
Shark_α1(I)	-----GVLASDVAIQMTFLRLLSSEATQNVTYHCKNSIAYMDEE	1378
Shark_α2(I)	-----HITPLVMATQLTFLQLLSNEASONVTTYCKNSIAYMDEE	1284
Coelacanth_α1(I)	-----GSDPADVAIQTLFLRLLSSEASONITYHCKNSIAYMDQQ	1371
Coelacanth_α2(I)	-----SITPKIMATQLAFMRLLANQASONITYHCKNSIAYMDEQ	1283
Frog_α1(I)	-----GSDPADVAIQTLFLRLLMATEASONITYHCKNSVAYMDQA	1365
Frog_α2(I)	-----GVTSKDMATQFAPMRLLANHASONITYHCKNSIAYMDGQ	1270
Human_α1(I)	-----GSDPADVAIQTLFLRLLSSEASONITYHCKNSVAYMDQQ	1380
Human_α2(I)	-----GVTSKEMATQLAFMRLLANASONITYHCKNSIAYMDEE	1282

**C7**

sponge_ColF1	TQDN-----KEIA-----ANKVTYDGC-----KSRPS	506
Hydra_colα1(I)	GNNRPYVKI-----MSSDDIEIH---TGSHMKNRL--KVLQDQC-----NKKDN	1407
Hydra_col1	NNDY----KKAMKFKGTEDQEFA---YSEDKQSMFMHAVNDEC-----SNMSK	1368
Hydra_col2	GNNRPYVKI-----MSSDDIEIH---TGSHMKNRL--KVLQDQC-----NKKDN	1431
Hydra_col3	TDENVHLHIHNGKISINPDETYIR---HGIALKLNL--MAIKDDC-----KVKDD	1839
Hydra_col5	K-NTIKFKLDNG-----AIY---HKKMRGVVL--KVTDDC-----TV-EN	1330
Red fire ant_α1(I)	-----GNPFLSSSDVELV---DDQDSKFNYY--RTLEDGC-----SSSSS	3494
Mollusk_α1(I)	TGST----EQALRLMTSDVELS---LDAPSQEQY--EVIDEGC-----QERSA	3154
Sea urchin_Col1Pa	-E-RKTY--RRSLKLLAWNDAE-LSPR-GN-QRLRY-EMI-MDEC--LHQ-----N	172
Sea urchin_Col2Pa	IT-----DSGDELDSA---EGRFKRSTYI--DILEGECEN-----VSSKDN	875
lancelet_α1(I)	SSNY----KKGLMLMGNGETELG---AQGPKKYQL--YAVEDGC-----QTPTG	361
lancelet_α2(I)	TDGY----DQAVRLLSNNENVLV---YGTGPV---KTIYDGC-----QFASP	2234
Tunicate_α1(I)	TGSF----DQAIHLLAANDEVLT---YGSEHL---TVIEDNC-----KT-GH	1321
Tunicate_α2(I)	ADNK----AQALKLRGFGDAEFT---AEGAVGTTY--RVLHDGC-----STRPT	1370
Lamprey_α1(I)	AGNL----KKALLMQGSSDVEIR---AEGNSRFTY--SVLEDGC-----THTG	1231
Lamprey_α2(I)	VGNL----QQALLQGSNEMEIR---AEGNSKLAY--SVLEDGC-----TMHTG	1143
Shark_α1(I)	AGNL----KKAVLLQGSNEIEIR---AEGNSRFTY--SVLEDGC-----TRHTG	1418
Shark_α2(I)	SGSL----KKAVLLQGSNDVELR---AEGNSRFTY--SVLEDGC-----TKHTG	1324
Coelacanth_α1(I)	AGNL----KKSLLQGSNEIEIR---AEGNSRFTY--SVLEDGC-----TRHTG	1411
Coelacanth_α2(I)	TGNL----KKAVMLQGSNDVELR---AEGNSRFTF--SVLEDGC-----TKHTS	1323
Frog_α1(I)	TGNL----KKALLQGSNEIEIR---AEGNSRFTY--SVLEDGC-----TQHTG	1405
Frog_α2(I)	TGNL----KKALLQGSNDVELR---AEGNSRFTF--SVLEDGC-----TQHTG	1310
Human_α1(I)	TGNL----KKALLQGSNEIEIR---AEGNSRFTY--SVTVDGC-----TSHTG	1420
Human_α2(I)	TGNL----KKAVILQGSNDVELV---AEGNSRFTY--TVLVDGC-----SKKTN	1322

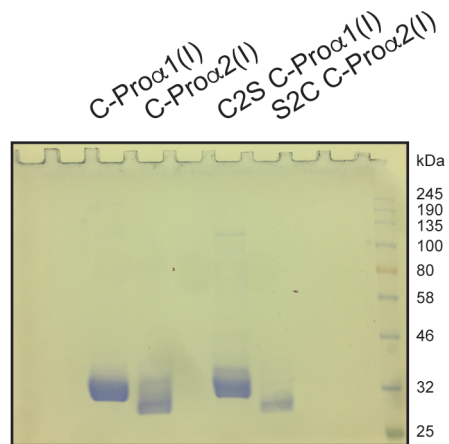
**C8**

sponge_ColF1	VPDAAFVAVETKRVEQLPIRDFASSDIAGQHQEFQFEMGPACFY-	550
Hydra_colα1(I)	QWHKTVFEFSSKITSRLPIVDVAVFDVANVGEQFGIELGPVCFY-	1451
Hydra_col1	EWRTTTLRFTSRKYIRLPIDFAPISSEDNNAMFVELGPVCFM-	1412
Hydra_col2	QWHKTVFEFSSKITSRLPIVDVAVFDVANVGEQFGIELGPVCFY-	1475
Hydra_col3	KWRESVFEITSNRLENLPIQDIGVTDIADGGEQFGLLIGPVCFS-	1883
Hydra_col5	KWQKLVLELSTSKNEILPVIDLGVHNGITSSIEIGVDIGPACFS-	1374
Red fire ant_α1(I)	QWGSVVEYETKKTTRLPIDVFAPGEVGSQFMGLEMPVCFFS-	3538
Mollusk_α1(I)	EWSQTVINYSRRNTRLPIDVAPSDIGGEGQFEGITLGPVCFFS-	3198
Sea urchin_Col1Pa	HWGKTVVSJETDKPVRLPIDVALRDIGNPDQSFSEIGAACY--	215
Sea urchin_Col2Pa	QWHTNRVEVRTNKSELLPLVDVLLFDIGGENQQFVIDVGEVCFFS-	919
lancelet_α1(I)	KWSSTVLEYKTKKTRLPFTDIAPYD-----	387
lancelet_α2(I)	QLDLTVIEINTTATECPVVRDFGVFELDENGQEFQFVSVGVCFQ-	2278
Tunicate_α1(I)	GNGQVVLELRTREVDLLPLFDYKAFDFGTRSQRHGYQLDRVCFSG	1366
Tunicate_α2(I)	QWDRTEIEFETRLVGRMPTIDIAFPDIDGADQFQAKFGPVCFK-	1414
Lamprey_α1(I)	VWGTKVIEYRTQKTSRLPFMDIAPMDVGGSDQEFQFVGVGPVCF-	1275
Lamprey_α2(I)	QWGTKVIEYRTPKTSRLPIVDIAPKDVGGPDQEFQFVGVGPVCF-	1187
Shark_α1(I)	EWGRTVIEYKTKTSRLPIIDVAPMDVGGADQEFQFVIEGPVCF-	1462
Shark_α2(I)	EWSKTIIEYRTQKTSRLPFVMDIAPMDIGGPEQEFGLDIGPVCFK-	1368
Coelacanth_α1(I)	AWGRTVIDYKTKTSRLPVIDIAPMDVGGADQEFQFVIEGPVCF-	1455
Coelacanth_α2(I)	QWGTKVIEYRTNKPTRLPILDIAPMDIGGADQEFGLDIGPVCFK-	1367
Frog_α1(I)	EWGKTVIDYKTKTSRLPITDVAPMDIGADQEFQFVIEGPVCFV-	1449
Frog_α2(I)	EWGRTVIEYRTNKPSRLPILDIAPLDIGDDQEFQFVIEGPVCFK-	1354
Human_α1(I)	AWGKTVIEYKTKTSRLPIIDVAPLDVGGADQEFQFVIEGPVCF-	1464
Human_α2(I)	EWGKTIVIEYKTKPSRLPFLDIAPLDIGGADQEFQFVIEGPVCFK-	1366

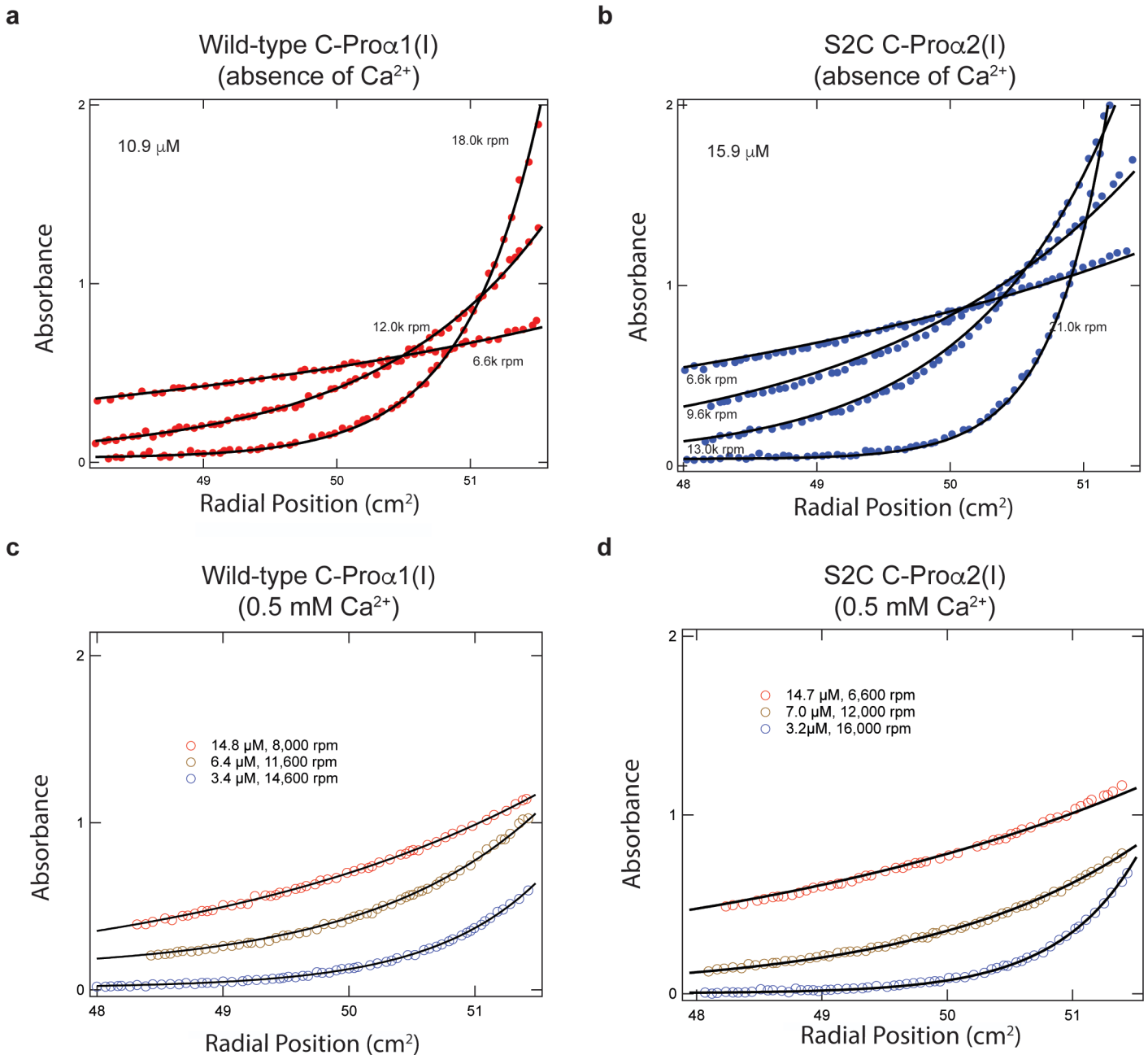
**Supplementary Figure 1 | Full alignments of ancestral collagen-I C-Pro domains.**

Alignment of collagen-I C-Pro domains within the chordate lineage. Cys residues are highlighted in yellow. Amino acids previously proposed to be important for salt bridge formation are bolded in red. Amino acids other than cysteine in the conserved cysteine positions are bolded. The collagen recognition sequence is outlined by a black box in sequences where it is present. Amino acids involved in the Ca<sup>2+</sup> coordination network are marked with a “•” below the alignment. The accession numbers for each sequence are as follows: sponge\_ColF1 P18856, hydra\_α1(I) T2MHG7, hydra\_col1 AF525468, hydra\_col2 A1XVT1, hydra\_col3 A1XVT2, hydra\_col5 A1XVT3, sea urchin\_Col1Pa M92040, sea urchin\_ColP2a M92041, red fire ant\_α1(I) UniRef90\_UPI0005960200, mollusk\_α1(I) UniRef90\_UPI00084B9720, tunicate\_α1(I) H2YGA7, tunicate\_α2(I) H2YJN4, lancelet\_α1(I) AB193827.1, lancelet\_α2(I) XM\_019787639.1, lamprey\_α1(I) S4R787, lamprey\_α2(I) S4RDA0, shark\_α1(I) XM\_007910713.1, shark\_α2(I) XM\_007909256.1, coelacanth\_α1(I) XM\_005992184.2, coelacanth\_α2(I) XM\_006011624.2, frog\_α1(I) NM\_001011005.1, frog\_α2(I) NM\_001079250.1, human\_α1(I) NM\_000088.3, and human\_α2(I) NM\_000089.3.

## Reducing Coomassie SDS-PAGE

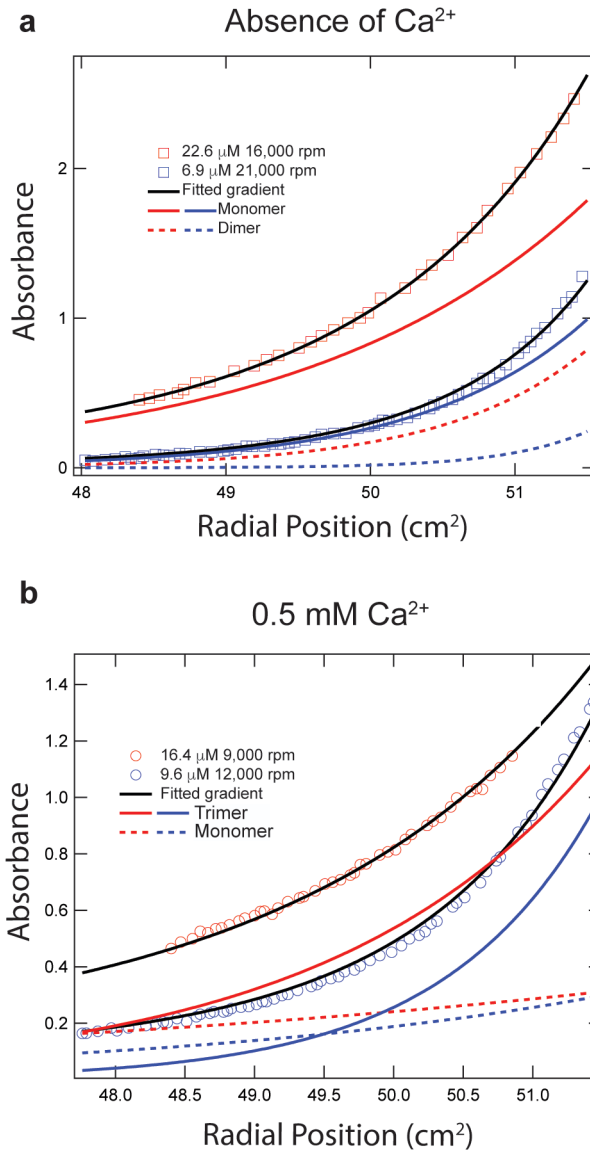
**Supplementary Figure 2 | Collagen-I C-Pro purification.**

Coomassie staining of a reducing SDS-PAGE gel showing recombinantly prepared, purified samples of wild-type C-Pro $\alpha$ 1(I), wild-type C-Pro $\alpha$ 2(I), C2S C-Pro $\alpha$ 1(I), and S2C C-Pro $\alpha$ 2(I).



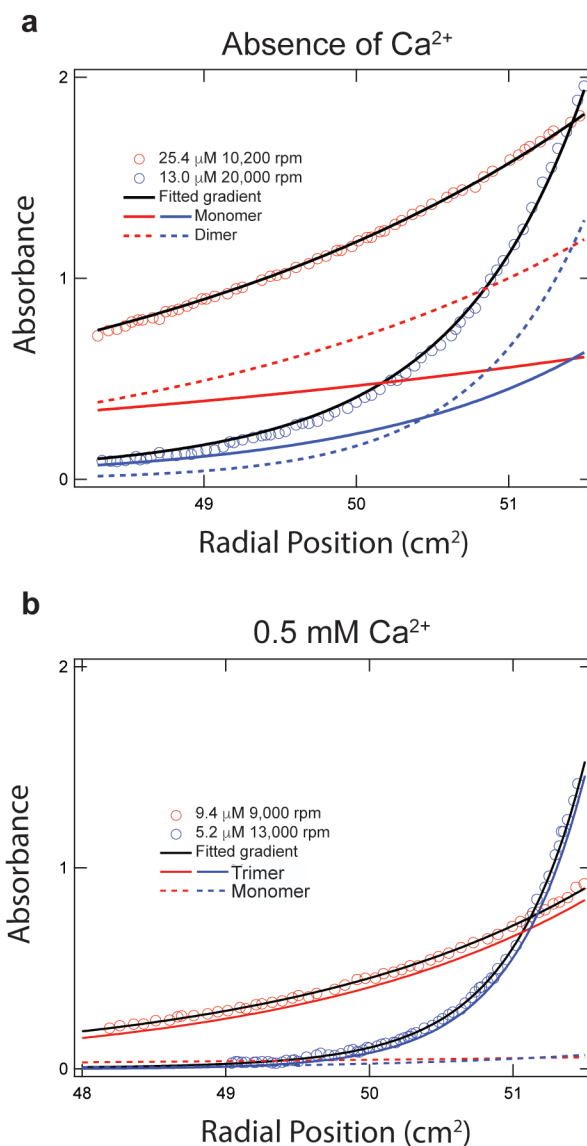
**Supplementary Figure 3 | Examples of global fits of the data collected for disulfide-linked homotrimeric C-Pro variants during sedimentation equilibrium analyses in the absence and presence of  $\text{Ca}^{2+}$ .**

- (a–b) Single species fits for wild-type C-Pro $\alpha$ 1(I) and S2C C-Pro $\alpha$ 2(I) at a single initial concentration and multiple speeds in the absence of  $\text{Ca}^{2+}$ . The fitted single species molecular weights based on a global fit to all the data available for each variant were 88,500 Da and 90,900 Da, respectively (showing only every second data point for clarity).
- (c–d) Examples of global single species fits of the data for wild-type C-Pro $\alpha$ 1(I) (c) and S2C C-Pro $\alpha$ 2(I) (d) at several concentrations and speeds (showing only every second data point for clarity) in the presence of  $\text{Ca}^{2+}$ .



**Supplementary Figure 4 | Global fits for wild type C-Pro $\alpha$ 2(I) in the absence and presence of  $\text{Ca}^{2+}$ .**

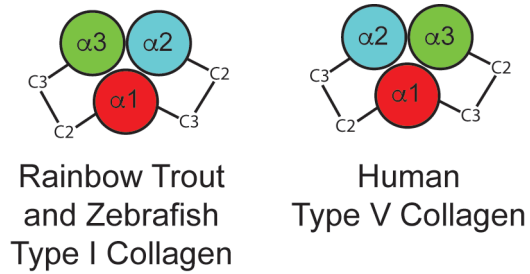
- (a) Example global fit to a monomer–dimer equilibrium model for wild-type C-Pro $\alpha$ 2(I) in the absence of  $\text{Ca}^{2+}$ . The fitted gradient superimposes on the measured data (showing only every second data point for clarity). The contributions of the monomer and dimer species to the total gradient are also shown, based on a fitted equilibrium constant of  $6500 \text{ M}^{-1}$ . This distribution could be replaced by a single gradient with a molecular weight intermediate to the monomer and dimer, as shown in Fig. 4d.
- (b) Example global fit to a monomer–trimer equilibrium model for wild-type C-Pro $\alpha$ 2(I) in the presence of 0.5 mM  $\text{Ca}^{2+}$ . The fitted gradient superimposes on the measured data (showing only every second data point for clarity). The contributions of the monomer and trimer species to the total gradient are also shown.



**Supplementary Figure 5 | Global fits for C2S C-Pro $\alpha$ 1(I) in the absence and presence of Ca $^{2+}$ .**

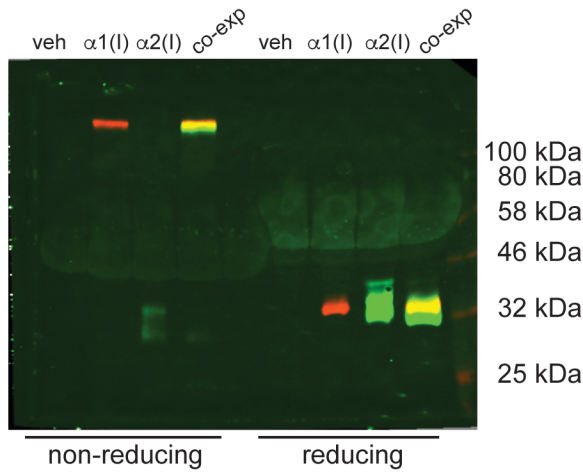
- (a) Example global fit to a monomer–dimer equilibrium model for C2S C-Pro $\alpha$ 1(I) in the absence of Ca $^{2+}$ . The fitted gradient superimposes on the measured data (showing only every second data point for clarity). The contributions of the monomer and dimer species to the total gradient are also shown, based on a fitted equilibrium constant of 76,300 M $^{-1}$ . This distribution could be replaced by a single gradient with a molecular weight intermediate to the monomer and dimer, as shown in Fig. 4e.
- (b) Example global fit to a monomer–trimer equilibrium model for C2S C-Pro $\alpha$ 1(I) in the presence of 0.5 mM Ca $^{2+}$ . The fitted gradient superimposes on the measured data (showing only every second data point for clarity). The contributions of the monomer and trimer species to the total gradient are also shown.



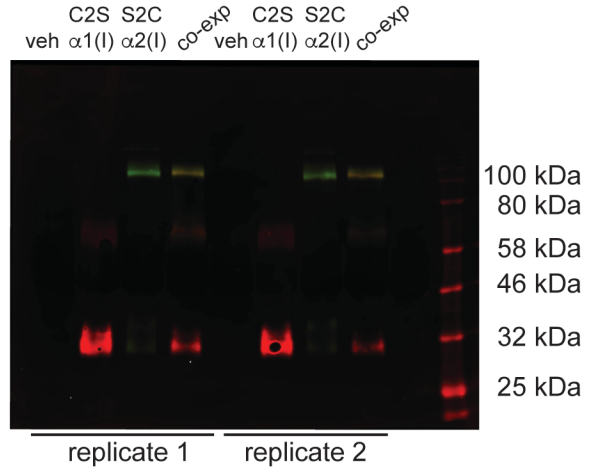
**Supplementary Figure 6 | Topology of 1:1:1 heterotrimers.**

The cysteine-based code provides insight into the topology of the 1:1:1 heterotrimers that can form. Given the difference in cysteine patterns between Col $\alpha 2$ (I) and Col $\alpha 3$ (I) of rainbow trout and zebrafish versus Col $\alpha 2$ (V) and Col $\alpha 3$ (V) in humans, the  $\alpha 2$  and  $\alpha 3$  chains are likely in different orientations relative to the  $\alpha 1$  chain, which in all three species maintains cysteine residues in both the C2 and C3 positions.

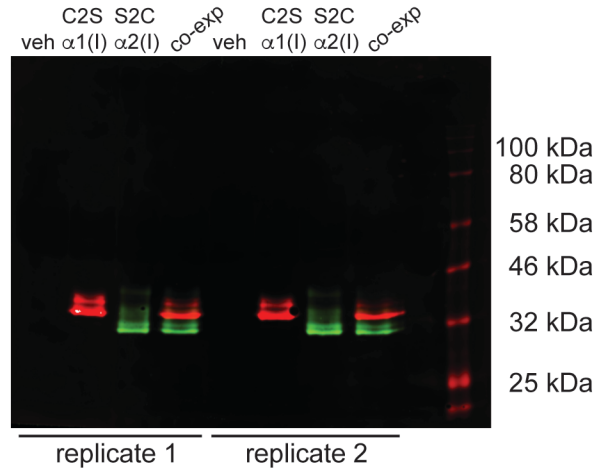
**a Wild-type collagen-I C-Pro domains**



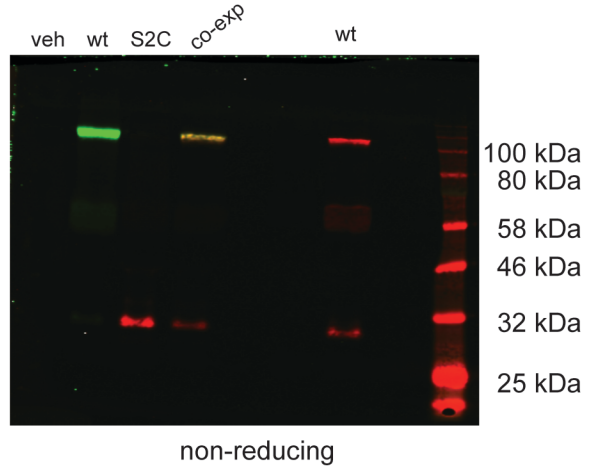
**b Cys/Ser variants of collagen-I C-Pro domains**



**c Cys/Ser variants of collagen-I C-Pro domains**



**d Wild-type and Ser variant of Col $\alpha$ 1(III)**



**Supplementary Figure 7 | Uncropped immunoblots for select results**

Select immunoblots are shown in the raw, uncropped form for transparency. The immunoblots shown correspond to Fig. 3a (a), Fig. 3b (b), Fig. 3c (c), and Fig. 7d (d).

## SUPPLEMENTARY TABLES

**Supplementary Table 1.** Disulfide bonding network of wild-type C-Pro $\alpha$ 1(I).

Disulfide linkage	Digestion condition	Expected mass (charge)	Observed mass (charge)	Error (ppm)
C1 + C4	GluC + LysC	957.4489 (+2)	957.4457 (+2)	3.33
		638.6352 (+3)	638.6330 (+3)	3.44
		479.2284 (+4)	479.2269 (+4)	3.04
		383.5843 (+5)	383.5823 (+5)	5.08
C2 + C3	GluC + LysC	1327.5892 (+2)	1327.5896 (+2)	0.31
		885.3954 (+3)	885.3944 (+3)	1.13
		664.2985 (+4)	664.2977 (+4)	1.21
C5 + C8	GluC + LysC	531.6404 (+5)	531.6400 (+5)	0.70
		1310.6386 (+2)	1310.6331 (+2)	4.23
		874.0950 (+3)	874.0915 (+3)	4.04
C6 + C7	GluC + LysC	655.8232 (+4)	655.8195 (+4)	5.69
		1198.8800 (+3)	1198.8787 (+3)	1.11
		899.4120 (+4)	899.4109 (+4)	1.20
C6 + C7	GluC + LysC	719.7312 (+5)	719.7332 (+5)	2.85
		599.9440 (+6)	599.9439 (+6)	0.05
		514.3816 (+7)	514.3812 (+7)	0.83

**Supplementary Table 2.** Disulfide bonding network of S2C C-Pro $\alpha$ 2(I).

Disulfide linkage	Digestion condition	Expected mass (charge)	Observed mass (charge)	Error (ppm)
C1 + C4	GluC	724.0294 (+3)	724.0243 (+3)	7.00
	GluC	543.2740 (+4)	543.2710 (+4)	5.49
	GluC	434.8208 (+5)	434.8181 (+5)	6.09
C2 + C3	GluC	577.5895 (+3)	577.5933 (+3)	6.64
	GluC	433.4441 (+4)	433.4469 (+4)	6.56
C5 + C8	GluC	839.9301 (+2)	839.928 (+2)	2.49
	GluC	560.2893 (+3)	560.2884 (+3)	1.67
	GluC	420.4690 (+4)	420.4677 (+4)	2.99
C6 + C7	trypsin	1024.1642 (+3)	1024.1644 (+3)	0.23
	trypsin	768.3751 (+4)	768.3765 (+4)	1.85
	trypsin	614.9016 (+5)	614.9039 (+5)	3.69

**Supplementary Table 3.** Primers used for site-directed mutagenesis of the indicated C-Pro domains.

Primer Name	Sequence
<i>COL1A1</i> C1265S F	5'- CCGTGACCTCAAGATGTCCCACTCTGACTGGAAG -3'
<i>COL1A1</i> C1265S R	5'- CTTCCAGTCAGAGTGGGACATCTTGAGGTCACGG -3'
<i>COL1A2</i> S1169C F	5'- CGTGACTTGAGACTCUGCCACCCAGAGTGGAG -3'
<i>COL1A2</i> S1169C R	5'- CTCCACTCTGGGTGGCAGAGTCTCAAGTCACG -3'
<i>COL2A1</i> C1306S F	5'-CCCCAACCAAGGCAGCACCTTGGACGC-3'
<i>COL2A1</i> C1306S R	5'-GCGTCCAAGGTGCTGCCTTGGTTGGGG-3'
<i>COL3A1</i> C1268S F	5'-TTGAGTTCAGGATGGCTGAATTCAGGTCTCTGC-3'
<i>COL3A1</i> C1268S R	5'-GCAGAGACCTGAAATTCAGCCATCCTGAACTCAA-3'
<i>COL5A1</i> C1645S F	5'-AGTCGGGGTGGCTGAGCTGCAGGTC-3'
<i>COL5A1</i> C1645S R	5'-GACCTGCAGCTCAGCCACCCCGACT-3'
<i>COL5A2</i> S1319C F	5'-TTGATTGCATCTTCAACACATCCTTGGTTAGGATCAATCC-3'
<i>COL5A2</i> S1319C R	5'-GGATTGATCCTAACCAAGGATGTGTTGAAGATGCAATCAA-3'
<i>COL5A3</i> N1550C F	5'-CGAGCTGCACCGCTGCCACCCGCACCTG-3'
<i>COL5A3</i> N1550C R	5'-CAGGTGCGGGTGGCAGCGGTGCAGCTCG-3'
<i>COL11A1</i> S1613C F	5'-GGGAAGTCAGGATGGCAGAGTTGCAGGTCTTTA-3'
<i>COL11A1</i> S1613C R	5'-TAAAGACCTGCAACTCTGCCATCCTGACTTCCC-3'
<i>COL11A2</i> C1577S F	5'-AGCTCTGGGTGGCTCAGCTTCAGGTCC-3'
<i>COL11A2</i> C1577S R	5'-GGACCTGAAGCTGAGCCACCCAGAGCT-3'
<i>COL24A1</i> C1551S F	5'-TCCATCTGATACTTTTGTTCCTGTTAAGTAAATCTTTGCAGATTC-3'
<i>COL24A1</i> C1551S R	5'-GAATCTGCAAAGATTTACTTAACAGTGAACAAAAAGTATCAGATGGA-3'
<i>COL27A1</i> C1696S F	5'-CCATCTTCTGCTCACTGTCCATGAGGTCCCT-3'
<i>COL27A1</i> C1696S R	5'-AGGGACCTCATGGACAGTGAGCAGAAGATGG-3'

**Supplementary Table 4.** Molecular properties (calculated and measured) of collagen-I C-Pro variants.

	wild-type C-Pro $\alpha$ 1(I)	Cys1265Ser C-Pro $\alpha$ 1(I)	wild-type C-Pro $\alpha$ 2(I)	Ser1169Cys C-Pro $\alpha$ 2(I)
residues	251	251	252	252
$M_s$ (Da)	27,909	27,893	28,225	28,241
$\bar{v}$ (mL/g)	0.723	0.723	0.728	0.728
$\epsilon$ (M $^{-1}$ cm $^{-1}$ )	39,420	39,420	43,890	43,890
MALDI-TOF (Da)	29,183	29,174	29,371	29,476

## SUPPLEMENTARY NOTE 1

## Sequences of Open Reading Frames for C-Pro Domains

Color legend for all plasmids: Pre-protrypsin leader sequence Epitope tag HRV-3C protease cleavage site C-Pro domain

Wild-type PPT.His.C-Pro $\alpha$ 1(I):

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TCACCATCACCATCACCATCTTGAAGTTCTTTTTCAAGGTCCTGATGATGCCAATGTGGTTCGTGACCGTG  
 ACCTCGAGGTGGACACCACCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGGGC  
 AGCCGCAAGAACCCCGCCCGCACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAG  
 TACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTG  
 AGACCTGCGTGTACCCCACTCAGCCAGTGTGGCCAGAAAGTGGTACATCAGCAAGAACCCCAAGG  
 ACAAGAGGCATGTCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGGCGGCCAGGGCT  
 CCGACCCTGCCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTCCACCGAGGCCTCCAGAACA  
 TCACCTACCACTGCAAGAACAGCGTGGCCTACATGGACCAGCAGACTGGCAACCTCAAGAAGGCCCTGC  
 TCCTCCAGGGCTCCAACGAGATCGAGATCCGCGCCGAGGGCAACAGCCGCTTCACCTACAGCGTCACTG  
 TCGATGGCTGCACGAGTCACACCGAGCCTGGGGCAAGACAGTGATTGAATACAAAACCACCAAGACCT  
 CCCGCTGCCCATCATCGATGTGGCCCCCTTGACGTTGGTGCCCCAGACCAGGAATTCGGCTTCGACG  
 TTGGCCCTGTCTGCTTCCTGTAA

Wild-type PPT.His.C-Pro $\alpha$ 2(I):

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTGACTACAAAGACGATGACGACAAGC  
 ACCATCACCATCACCATCTTGAAGTTCTTTTTCAAGGTCCTGCTGACCAGCCTCGCTCAGCACCTTCTCTC  
 AGACCAAGGACTATGAAGTTGATGCTACTCTGAAGTCTCTCAACAACCAGATTGAGACCCTTCTTACTCC  
 TGAAGGCTCTAGAAAGAACCCAGCTCGCACATGCCGTGACTTGAGACTCAGCCACCCAGAGTGGAGCAG  
 TGTTACTACTGGATTGACCCTAACCAAGGATGCACTATGGATGCTATCAAAGTATACTGTGATTTCTCTA  
 CTGGCGAAACCTGTATCCGGGCCCAACCTGAAAACATCCCAGCCAAGAACTGGTATAGGAGCTCCAAGG  
 ACAAGAAACACGTCTGGCTAGGAGAACTATCAATGCTGGCAGCCAGTTTGAATATAATGTAGAAGGAGT  
 GACTTCCAAGGAAATGGCTACCCAACCTTGCCTTCATGCGCCTGCTGGCCAACCTATGCCTCTCAGAACATC  
 ACCTACCACTGCAAGAACAGCATTGCATACATGGATGAGGAGACTGGCAACCTGAAAAGGCTGTCATTC  
 TACAGGGCTCTAATGATGTTGAACTTGTGCTGAGGGCAACAGCAGGTTCACTTACACTGTTCTTGTAGAT  
 GGCTGCTCTAAAAGACAAATGAATGGGGAAAGACAATCATTGAATACAAAACAAATAAGCCATCACGCCT  
 GCCCTTCCTTGATATTGCACCTTTGGACATCGGTGGTGCTGACCAGGAATTCTTTGTGGACATTGGCCCA  
 GTCTGTTTCAAATAA

PPT.HA C-Pro $\alpha$ 1(II):

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TGCGGGCCGACAGACCAGGCAGCCGGTGGCCTGAGACAGCATGACGCCGAGGTGGATGCCCACTCAAGT  
 CCCTCAACAACCAGATTGAGAGCATCCGCAGCCCCGAGGGCTCCCGCAAGAACCCTGCTCGCACCTGCA  
 GAGACCTGAACTCTGCCACCCTGAGTGGAAAGAGTGGAGACTACTGGATTGACCCCAACCAAGGCTGCA  
 CTTGGACGCCATGAAGTTTTCTGCAACATGGAGACTGGCGAGACTTGCCTACCCCAATCCAGCAAA  
 CGTTCCCAAGAAGAACTGGTGGAGCAGCAAGAGCAAGGAGAAGAAACACATCTGGTTTTGGAGAAACCAT  
 CAATGGTGGCTTCCATTTTCACTATGGAGATGACAATCTGGCTCCCAACACTGCCAACGTCCAGATGACC  
 TTCCTACGCCTGCTGTCCACGGAAGGCTCCAGAACATCACCTACCACTGCAAGAACAGCATTGCCTATC  
 TGGACGAAGCAGCTGGCAACCTCAAGAAGGCCCTGCTCATCCAGGGCTCCAATGACGTGGAGATCCGGG  
 CAGAGGGCAATAGCAGGTTACGTACACTGCCCTGAAGGATGGCTGCACGAAACATACCGGTAAGTGGG

GCAAGACTGTTATCGAGTACCGGTCACAGAAGACCTCACGCCTCCCATCATTGACATTGCACCCATGGA  
CATAGGAGGGCCCGAGCAGGAATTCGGTGTGGACATAGGGCCGGTCTGCTTCTTGTA

**PPT.HA C-Pro $\alpha$ 1(III):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
TGGCGCCGCAGATGAACCAATGGATTTCAAATCAACACCGATGAGATTATGACTTCACTCAAGTCTGTTA  
ATGGACAAATAGAAAGCCTCATTAGTCTGATGGTTCTCGTAAAACCCCGCTAGAAACTGCAGAGACCT  
GAAATTCTGCCATCCTGAACTCAAGAGTGGAGAATACTGGGTTGACCTAACCAAGGATGCAAATTGGAT  
GCTATCAAGGTATTCTGTAATATGGAACTGGGGAAACATGCATAAGTGCCAATCCTTTGAATGTTCCAG  
GAAACTGTTGGACAGATTCTAGTCTGAGAAGAAACACGTTTGGTTTGGAGAGTCCATGGATGGTGGT  
TTTCAGTTTAGCTACGGCAATCCTGAACTTCTGAAGATGTCCTTGATGTGCAGCTGGCATTCTTCGACT  
TCTCTCCAGCCGAGCTTCCCAGAACATCACATATCACTGCAAAAATAGCATTGCATACATGGATCAGGCCA  
GTGGAAATGTAAAGAAGGCCCTGAAGCTGATGGGGTCAAATGAAGGTGAATTCAAGGCTGAAGGAAATAG  
CAAATTCACCTACACAGTTCTGGAGGATGGTTGCACGAAACACACTGGGGAATGGAGCAAACAGTCTTT  
GAATATCGAACACGCAAGGCTGTGAGACTACCTATTGTAGATATTGCACCCTATGACATTGGTGGTCTGA  
TCAAGAATTTGGTGTGGACGTTGGCCCTGTTTGCTTTTTATAA

**PPT.HA C-Pro $\alpha$ 1(V):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
TGGCGCCGCAGACGGCATGGAAGAGATCTTCGGCTCTCTCAACTCTCTGAAGCTGGAGATTGAGCAGAT  
GAAACGGCCCTGGGCACGCAGCAGAACCCCGCCCGCACCTGCAAGGACCTGCAGCTCTGCCACCCCG  
ACTTCCCAGATGGTGAATACTGGGTCGATCCTAACCAAGGATGCTCCAGGGATTCTTCAAGGTTTACTG  
CAACTTCACAGCCGGGGGTCGACATGCGTCTTCCCTGACAAGAAGTCCGAAGGGGCCAGAATCACTTC  
TTGGCCCAAAGAAAACCCGGGCTCCTGGTTCAAGCGTGGGAACTGCTCTCCTATGTGGAC  
GCCGAGGGCAACCCTGTGGGTGTGGTACAGATGACCTTCTGCGGCTGCTGAGCGCCTCTGCCACCA  
GAACGTCACCTACCACTGCTACCAAGTCAGTGGCCTGGCAGGACGCAGCCACGGGCAGCTACGACAAGG  
CCCTCCGCTTCTGGGCTCCAACGACGAGGAGATGTCTATGACAACAACCCCTACATCCGCGCCCTGG  
TGGACGGCTGTGCTACCAAGAAAGGCTACCAGAAGACGTTCTGGAGATCGACACCCCAAAGTGGAGC  
AGGTGCCCATCGTGGACATCATGTTCAATGACTTCGGTGAAGCGTCACAGAAATTTGGATTTGAAGTGGG  
GCCGGCTTGCTTCATGGGCTAG

**PPT.HA C-Pro $\alpha$ 2(V):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
TGGCGCCGCAACAGCTGCTCTTGGGGATATCATGGGGCACTATGATGAAAGCATGCCAGATCCACTTCT  
GAGTTTACTGAAGATCAGGCGGCTCCTGATGACAAAACAAAACGGACCCAGGGGTTTCACTACCCTGA  
AGTCACTCAGTAGTCAGATTGAAACCATGCGCAGCCCGATGGCTCGAAAAGCACCCAGCCCGCACGT  
GTGATGACCTAAAGCTTTGCCATTCCGCAAAGCAGAGTGGTGAATACTGGATTGATCCTAACCAAGGATCT  
GTTGAAGATGCAATCAAAGTTTACTGCAACATGGAAACAGGAGAAACATGTATTTAGCAAACCCATCCAG  
TGTAACACGTAAAACCTGGTGGGCCAGTAAATCTCCTGACAATAAACCTGTTTGGTATGGTCTTGATATGA  
ACAGAGGGTCTCAGTTCGCTTATGGAGACCACCAATCACCTAATACAGCCATTACTCAGATGACTTTTTTG  
CGCCTTTTATCAAAGAAGCCTCCCAGAACATCACTTACATCTGTAAAACAGTGTAGGATACATGGACGA  
TCAAGCTAAGAACCTCAAAAAGCTGTGGTTCTCAAAGGGGCAAATGACTTAGATATCAAAGCAGAGGGA  
AATATTAGATTCGGTATATCGTTCTTCAAGACACTTGTCTAAGCGGAATGGAAATGTGGGCAAGACTGT  
CTTTGAATATAGAACACAGAATGTGGCACGCTTGCCCATCATAGATCTTGCTCCTGTGGATGTTGGCGGC  
ACAGACCAGGAATTCGGCGTTGAAATTGGGCCAGTTTGTGTTGTGTA

**PPT.HA C-Pro $\alpha$ 3(V):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TGGCGCCGCAGAGGAGGTGCTGGCCTCGCTCACATCGCTGAGCTTGGAGCTGGAGCAGCTGCGGCGTC  
 CTCCCGGCACTGCGGAGCGCCCGGGCCTCGTGTGCCACGAGCTGCACCGCAACCACCCGCACCTGCCT  
 GATGGGGAATACTGGATTGACCCCAACCAGGGCTGCGCGCGGGACTCGTTCAGGGTTTTTTGCAACTTC  
 ACGGCGGGAGGAGAGACCTGCCTCTATCCCGACAAGAAGTTTGAGATCGTGAAATTGGCCTCCTGGTCC  
 AAGGAAAAGCCTGGAGGCTGGTATAGCACATTCCGTCGAGGGGAAGAAGTTCTCCTACGTGGACGCCGAC  
 GGGTCCCAGTGAATGTCGTGCAGCTGAACTTCTGAAACTGCTGAGTGCCACAGCTCGCCAGAATTCA  
 CCTACTCCTGCCAGAATGCAGCTGCCTGGCTGGACGAAGCCACGGGTGACTACAGCCACTCCGCCCGCT  
 TCCTTGGCACCAATGGAGAGGAGCTGTCTTTCAACCAGACGACAGCAGCCACTGTCAGCGTCCCCAGG  
 ATGGCTGCCGGCTCCGAAAGGACAGACGAAGACCCTTTTCGAATTCAGCTCTTCTCGAGCGGGATTTCT  
 GCCCCTGTGGGATGTGGCGGCCACTGACTTTGGCCAGACGAACCAAAAGTTTGGGTTTGAAGTGGGCC  
 CGTCTGCTTCAGCAGCTGA

**PPT.HA C-Pro $\alpha$ 1(XI):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TGGCGCCGCAGATGCAGATGATAATATTCTTGATTACTCGGATGGAATGGAAGAAATATTTGGTTCCCTCA  
 ATCCCTGAAACAAGACATTGAGCATATGAAATTTCCAATGGTACTCAGACCAATCCAGCCCGAACTTGT  
 AAAGACCTGCAACTCAGCCATCCTGACTTCCCAGATGGTGAATATTGGATTGATCCTAACCAAGGTTGCTC  
 AGGAGATTCCTTCAAAGTTTACTGTAATTTACATCTGGTGGTGAGACTTGCATTTATCCAGACAAAAAATC  
 TGAGGGAGTAAGAATTTATCATGGCCAAAGGAGAAACCAGGAAGTTGGTTTAGTGAATTTAAGAGGGGA  
 AAAGTCTTTTACTACTTAGATGTTGAAGGAAATTCATCAATATGGTGCAATGACATTCCTGAAACTTCTG  
 ACTGCCTCTGCTCGGCAAAATTTACCTACCCTGTCATCAGTCAGCAGCCTGGTATGATGTGTCATCAG  
 GAAGTTATGACAAAGCACTTCGCTTCTGGGATCAAATGATGAGGAGATGTCCTATGACAATAATCCTTTT  
 ATCAAAACACTGTATGATGGTTGTGCGTCCAGAAAAGGCTATGAAAAGACTGTCATTGAAATCAATACACC  
 AAAAATTGATCAAGTACCTATTGTTGATGTCATGATCAATGACTTTGGTGATCAGAATCAGAAGTTCGGATT  
 TGAAGTTGGTCTGTTTGTCTTGGCTAA

**PPT.HA C-Pro $\alpha$ 2(XI):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TGGCGCCGCAATCCAGCCACTGCCATTCAGATGCCAAGAAGACTCGGCGCTCGGTGGATGGAAGCCG  
 TCTGATGCAGGAAGATGAGGCCATACCGACCGGGGGAGCCCCGGCAGTCTGGGGGGCTGGAGGAGA  
 TCTTTGGCTCACTCGACTCCCTGCGGGAGGAGATCGAGCAGATGAGGCGGCCAACAGGGACCCAGGAC  
 AGCCCTGCTCGCACCTGCCAGGACCTGAAGCTGTGCCACCCAGAGCTTCCCGATGGAGAGTACTGGGTC  
 GACCCCAACCAGGGCTGTGCTCGGGATGCCTTCCGAGTTTTCTGCAACTTCACAGCAGGGGGTGAGACC  
 TGTGTGACGCCTAGGGATGACGTCACGCAGTTCTTACGTGGACTCAGAGGGCTCCCCAGTGGGTGTG  
 GTCCAGCTCACCTTCTGCGGCTGCTCAGCGTCTCAGCCCACCAGGACGTCTCCTACCCCTGCTCTGGA  
 GCAGCCCGTGACGGTCCCCTGAGACTCCGTGGGGCCAATGAGGATGAGCTGAGCCCGGAGACTAGCCC  
 CTATGTCAAAGAATTCAGAGATGGCTGCCAGACACAGCAAGGCCGGACGGTGTGGAGGTGCGAACGCC  
 TGTGCTGGAGCAGCTGCCAGTGTGATGCCTCCTTCTCAGACCTGGGAGCCCCACCGAGGCGGGGAG  
 GGGTGTGCTGGGGCCTGTCTGCTTCATGGGATAG

**PPT.HA C-Pro $\alpha$ 1(XXIV):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TGGCGCCGCAGAAGAGATATCAAACCCTGAACTACCTTAGCAATTTATTGCACAGCATCAAGAATCCTC  
 TTGGCACACGAGATAACCCAGCACGAATCTGCAAAGATTTACTTAACTGTGAACAAAAAGTATCAGATGGA  
 AAATACTGGATTGACCCAAATCTTGGCTGTCTTCAGATGCCATTGAGGTTTTCTGCAATTTTCAGTGCTGG

TGGCCAGACATGCTTACCTCCTGTTTCTGTAACAAAGTTGGAGTTTGGAGTTGGGAAAGTCCAGATGAACT  
 TCCTTCATTTACTGAGTTCGGAAGCCACCCATATCATCACCATTCACTGTCTAAACACCCCAAGGTGGACA  
 AGCACACAAACAAGTGGCCAGGATTGCCTATTGTTTTCAAGGGATGGAATGGCCAGATTTTTAAAGTAAA  
 CACTCTACTTTGAACCTAAAGTGCTTTTCAAGATGACTGCAAGATTCAAGATGGCAGCTGGCATAAGGCAACAT  
 TTCTTTTTTACACCCAGGAACCTAATCAACTTCCAGTGATTGAAGTACAAAACTTCTCATCTCAAACTG  
 AACGAAAGTATTACATTGACAGCAGTTCTGTATGCTTTTCTGTAA

**PPT.HA C-Pro $\alpha$ 1(XXVII):**

ATGTCTGCACTTCTGATCCTAGCTCTTGTGGAGCTGCAGTTGCTTACCCATACGATGTTCCAGATTACGC  
 TGGCGCCGCAATCCAATTGCAACAAGATGATCTTGGGGCAGCTTTCCAGACGTGGATGGACACCAGTGG  
 AGCACTCAGGCCAGAGAGTTACAGCTATCCAGACCGGCTGGTGCTGGACCAGGGAGGAGAGATCTTTAA  
 AACCTTACACTACCTCAGCAACCTCATCCAGAGCATTAAAGACGCCCTGGGCACCAAAGAGAACCCCGCC  
 CGGGTCTGCAGGGACCTCATGGACTGTGAGCAGAAGATGGTGGATGGTACCTACTGGGTGGATCCAAAC  
 CTTGGCTGCTCCTCTGACACCATCGAGGTCTCCTGCAACTTCACTCATGGTGGACAGACGTGTCTCAAGC  
 CCATCACGGCCTCCAAGGTCGAGTTTGCCATCAGCCGGGTCCAGATGAATTTCTGCACCTGCTAAGCTC  
 CGAGGTGACCCAGCACATCACCATCCACTGCCTTAACATGACCGTGTGGCAGGAGGGCACTGGGCAGAC  
 CCCAGCCAAGCAGGCCGTACGCTTCCGGGCCTGGAATGGACAGATTTTTGAAGCTGGGGGTGAGTTCCG  
 GCCCGAGGTGTCCATGGATGGCTGCAAGGTCCAAGATGGCCGCTGGCATCAGACACTCTTACCTTCCG  
 GACCCAAGACCCCAACAGCTGCCATCATCAGTGTGGACAACCTCCCTCCTGCCTCATCAGGGAAGCA  
 GTACCGCCTGGAAGTTGGACCTGCGTGCTTCCTCTGA