Templated Collagen "Double Helices" Maintain Their Structure

I. Caglar Tanrikulu,[†,‡] William M. Westler,[‡,§] Aubrey J. Ellison,[#] John L. Markley,[‡,§] and Ronald T. Raines*[,†,‡,#]

[†]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States
[‡]Department of Biochemistry, [#]Department of Chemistry, and [§]NMRFAM, University of Wisconsin–Madison, Madison, Wisconsin 53706, United States

Experimental Methodology and Results

Peptide Synthesis

All peptides were synthesized on glycine-preloaded Wang or polyethylene glycol-based resins using Fmoc-protected monomers from Chem-Impex (Wood Dale, IL). Synthesis of unlabeled $(PPG)_{10}$ and $(PPG)_{10}$ variants has been described previously.[1] ^{15}N-labeled $(PPG)_{10}$-variants were synthesized on a Liberty Blue automated microwave-assisted peptide synthesizer from CEM (Matthews, NC). Fmoc removal was achieved in piperidine (20% v/v in DMF), and peptide building blocks (5 equiv), activated through treatment with DCC and Oxyma, were coupled to the free amine on the growing chain. Peptides were cleaved from the resin and deprotected in 1.5–2.0 mL of reagent R, which was 90:5:3:2 TFA/thioanisole/ethanedithiol/anisole, precipitated from methyl *t*-butyl ether below 0 °C, and isolated by centrifugation. Dried crude peptides were dissolved in TFA (0.1% v/v) and filtered and purified by preparative HPLC at 45 °C using a linear gradient of CH_3CN/water containing TFA (0.1% v/v) with a Shimadzu Prominence unit equipped with a VarioPrep 250/21 C18 column from Macherey–Nagel (Düren, Germany). All peptides were >90% pure according to analytical HPLC and MALDI–TOF mass spectrometry (MS). MALDI–TOF analyses were carried out on an Voyager DE-Pro mass spectrometer from Applied Biosystems (Foster City, CA) at the University of Wisconsin–Madison Biophysics Instrumentation Facility (BIF), where single peptides and linked-dimers were analyzed by using a 1:10 matrix mixture of 2-(4-hydroxyphenylazo)benzoic acid:α-cyano-4-hydroxycinnamic acid to suppress the reduction of disulfide bonds due to in-source decay.[2]

Linked-Dimer Synthesis

Unlabeled disulfide-linked strand dimers, as well as the unproductive linked-dimers $(s1c)_2$ and h_x–c were produced as described previously.[1] The ^{15}N-labeled linked-dimer h^n–c^n (*i.e.*, $s1h^n$–$s3c^n$) was synthesized similarly, with minor modifications. Briefly, dried $s3c^n$ peptide (1.2–1.4 mM) and 2,2′-dithiobis(5-nitropyridine) (DTNP; 5 equiv) were dissolved in 3:1 HOAc/water (which was degassed and saturated with Ar(g)), and the resulting solution was stirred vigorously for ≥6 h. The reaction was stopped by the addition of 1.5 reaction-volumes of HOAc, freeze-dried, sonicated in aqueous TFA (0.1% v/v), and filtered through a 0.2-μm membrane. Peptide bearing the activated thiol, $s3c^n$-Npys, was isolated by HPLC, freeze-dried, and weighed. Coupling of $s1h^n$ and $s3c^n$-Npys was initiated by mixing equimolar amounts of the components (2.2 mM final concentration) dissolved previously in degassed and Ar(g)-saturated 50 mM NH_4OAc buffer, pH 5.3. Reactions were stirred under Ar(g) for 2–4 h, after which the solution was acidified and the solvent was removed by lyophilization. Disulfide-linked h^n–c^n was isolated by HPLC and analyzed by MALDI–TOF mass spectrometry as described above. For ^{15}N-labeled peptide constructs, (m/z) $[M + H]^+$ calcd 5088.8, found 5090.1 for h^n–c^n; calcd 2532.8, found 2533.3 for $s2^n$. For unproductive linked-dimers, (m/z) $[M + H]^+$ calcd 5114.8, found 5114.7 for h_x–c; calcd 5072.8, found 5072.4 for $(s1c)_2$. Analytical HPLC results for purified linked-dimers in 50 mM HOAc are shown in Figure S1.

Circular Dichroism Spectroscopy

Samples for circular dichroism (CD) spectroscopy and analytical ultracentrifugation (AUC) were prepared as reported previously.[1] All peptides and linked-dimers were dissolved in 50 mM HOAc to a 0.7 mM concentration based on weight. To ensure 1:1 equimolar composition for

x–y·s2 and *(s1c)₂+s2* linked-dimer:*s2* mixtures, relative concentrations were determined by monitoring the absorbance at 214 nm by HPLC. Integrations assumed identical extinction coefficients for all strands, and the concentrations of all components were matched to that of 60 μM *s2* peptide. CD samples were prepared at 180 μM for *s2,* 60 μM:60 μM for linked-dimer:*s2* mixtures, and 60 μM for linked-dimers alone. To facilitate the formation of the thermodynamic product for peptide association, mixtures were heated to ≥55 °C and allowed to cool to 4 °C slowly (over ~4 h). Samples were incubated at 4 °C for at least 48 h before data acquisition.

CD spectra of all samples were acquired at 4 °C with a 1-nm band-pass in quartz cuvettes with a 0.1-cm pathlength, using an averaging time of 3 s. For thermal denaturation experiments, the CD signal was monitored at 226 nm while the sample was heated from 4 to 64 °C in 3-°C steps over 4.5 h. CD data from denaturation experiments were converted to fraction folded, and data near 50% folded were used to obtain the T_m values for each sample, which are reported in Table S1.

CD data were acquired on three CD spectrometers, all from Aviv Biomedical (Lakewood, NJ) and each equipped with a 5-cell thermoelectric sample changer. These spectrometers were an Aviv 202SF and 420 in the BIF, and an Aviv 420 in the Gellman laboratory in the Department of Chemistry at the University of Wisconsin–Madison. Data on 180 μM *s2* were collected in every experiment to allow comparison of instrument performance and to account for differences in wavelength calibration, and results were well reproducible between instruments (T_m = 36.5 ± 0.6 °C for 180 μM *s2*; n = 5). All CD spectra and thermal denaturation curves used in this study are shown in Figures S2 and S3.

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed at the BIF with a Beckman XL-A analytical ultracentrifuge equipped with an An-60 Ti rotor. Samples were prepared at a strand concentration of 120 μM, but were diluted to 90 μM before the experiment. Sample (100 μL) and buffer (110 μL) were placed in a cell with a 12-mm double-sector charcoal-filled centerpiece (Epon). Experiments were run at 4 °C at speeds of 12, 22, and 32 k rpm, and gradients recorded at 231 nm were monitored until they were superimposable when recorded 4 h apart. A buffer density of 1.00037 g/mL and a partial specific volume of 0.7275 mL/g calculated based on $(PPG)_{10}$ amino acid content was used. Equilibrium gradients at 4 °C were modeled as single and multiple non-interacting species through nonlinear least-squares fits to gradient data. Analysis was performed with programs written for IGOR PRO (WaveMetrics, Lake Oswego, OR) by Dr. D. R. McCaslin (University of Wisconsin–Madison). Non-sedimenting baselines of 0.02–0.07 OD were applied for all samples. Plots of gradients and fits are shown in Figure S4A.

Overall, the data collected on *x–y* variants fitted best to a four-stranded, dimer of linked-dimers model (single species with MW = 10.2 kDa). For *c–c*, a model that includes monomers in addition to dimers better explains the data, especially at the top of the sample cell. In contrast, *(s1c)₂* exhibits a gradient that is significantly shallower than any of the *x–y* linked-dimers, closer to what is expected from a monomer. Still, inclusion of a higher molecular weight species is mandatory for better fits to gradients forming at the bottom of the sample cell. Among models entertained, a monomer + trimer model performs the best. This model also makes chemical sense: *(s1c)₂* strands cannot fold with each other (Figure S3A), but that does not exclude them from forming homotrimers with strands from two other *(s1c)₂* molecules. Nevertheless, this higher molecular weight species is a minor component of the *(s1c)₂* ensemble (Figure S4B).

Nuclear Magnetic Resonance Spectroscopy

Samples of *s2*^{n}, *h*^{n}–*c*^{n} and *h*^{n}–*c*^{n}·*s2* for analysis by nuclear magnetic resonance (NMR) spectroscopy were prepared in 45 mM HOAc with 10% D_2O and at ^{15}N-label concentrations of 1.2, 0.8, and 0.8 mM, respectively. Samples were heated, annealed, and incubated at either 4 or 5 °C for two days prior to analysis, as described previously.[1]

All NMR spectroscopy experiments were conducted at the National Magnetic Resonance Facility at Madison (NMRFAM). Heteronuclear Single Quantum Correlation spectra (HSQC) (Bruker pulse program: hsqcfpf3gpphwg) were collected on a Bruker Avance III 600 MHz equipped with a 1.7 mM ^{1}H(^{13}C/^{15}N) cryogenic probe and a SampleJet. Spectra were collected with a 1-s repetition delay, 8 steady-state transits, 4 acquisition transits, 256 increments with an acquisition time of 0.106 s for ^{1}H and 0.105 s for ^{15}N. Sweep widths for the ^{1}H and ^{15}N dimensions were 16 and 20 ppm, respectively. The spectra were referenced to DSS. Spectra at different temperatures were collected using NMRbot data-acquisition software.[3] The temperature in the probe was calibrated using neat ethylene glycol from 293–313 K and the NMR Temperature Measurements calculator.[4]

1-D NMR spectra shown as part of the denaturation series in Figures 5A–C are projections of 2-D spectra acquired at different temperatures. Although all 2-D spectra were processed identically, the projection process increases noise and results in the loss of broad low-intensity peaks. Samples that are able to form triple helices, such as those in Figures 5A and 5B, exhibit strong triple-helix peaks and maintain a high signal-to-noise ratio after processing. The triple-helix peak in Figure 5C, however, is lost to noise and does not appear in the 5-ºC projection. At higher temperatures (10–28 ºC) peak intensities increase, peak widths narrow, and triple-helix peaks become apparent in the projections (Figure 5C).

Computational Methodology and Results

Computational Evaluation of Disulfide Bridges Between (PPG)$_{10}$ Strands

Evaluations of disulfide linker strain on *x*–*y* linked-dimers were performed as was reported previously for *x*–*y*·*s2* heterotrimers, but with a few modifications.[1] All calculations were performed on Intel Xeon 2.33-GHz processors at the Materials and Process Simulation Center at the California Institute of Technology (Pasadena, CA). Computational models were built on the crystal structure of the (PPG)$_{10}$ trimer (PDB entry 1kf6).[5] Hydrogens were added using Reduce[6] (ver.3.03), and the model was minimized fully. All minimizations were carried out to a 0.2 kcal/mol/Å RMS-force convergence criterion using conjugate gradient minimization on MPSim[7] without solvation. The forces on the model peptides were described by the DREIDING force-field[8] without atomic charges.

For the analysis of disulfide linkers on linked-dimers, proline residues at neighboring Xaa and Yaa sites were replaced with Cys or Hcy on a collagen "double helix" obtained from a previously-prepared triple-helix model. The energy of the complete model, including the neighboring pair of free thiolate side chains, was then minimized fully. This structure was set as the reference state for each *x*–*y*; its backbone coordinates and proline side-chains were made immutable for all remaining calculations, which yielded the energy $E(x{\cdot}y)$. Free-thiolate side-chains were then connected to form a disulfide, and the linker was optimized by minimization followed by multiple rounds of

simulated annealing to reveal the optimized *x–y* with energy $E(x–y)$. The energy change due to disulfide formation was calculated as $E_{strain} = E(x–y) - E(x{\cdot}y)$. This metric omits solvation and Coulombic contributions to reduce noise and focuses on changes in bonding and steric interactions, which allow comparison across all linkers.

Strain Energy for Linkers on Linked-Dimers versus Triple Helices

E_{strain} calculations on *x–y·s2* triple helices were carried out on an optimized but immutable triple-helix model based on the PDB entry 1kf6.[5] Although our study suggests a triple-helix-like structure for the linked-dimers, there is no standard "CMP dimer" structure. Therefore, each *x–y* linked-dimer was allowed to minimize independently as a strand dimer prior to the formation of the disulfide linkage. This procedure allowed the remaining strands to optimize in the absence of *s2* and provided additional room for further optimization of disulfide linkers. The deviations that arise from this optimization are small (0.04 Å and 0.05Å backbone heavy-atom RMSD for *c–c*/*h–c* and *c–h*/*h–h*, respectively) and are localized mostly to Gly residues on the *s1x* strand, all of which have lost their hydrogen-bonding partners by the removal of *s2*. An exception to this general rule is the thiol-bearing residue on the *s1x* strand, which displays a high RMSD on the strained *c–c*, *c–h*, and *h–h* systems.

Despite these deviations, E_{strain} values for disulfide bridges on linked-dimers do not deviate far from their values calculated on a triple-helical background (Table S2 and Figure S6A). As anticipated, linked-dimers generally produce slightly improved E_{strain} values than do triple helices. This change is small for *c–c*, as it is highly strained on both backgrounds due to its short linker. The improvement is also minimal for *h–c*, which exhibits low strain on both backgrounds. In contrast, *c–h* and *h–h* both feature longer, but more strained linkers that can take advantage of the small deviations in backbone coordinates. The impact is most visible for the *c–h* linker, which, unlike other linkers, adopts a different conformation on a linked-dimer background than on a triple helix (Figure S6B). This lower E_{strain} correlates with and might be responsible for the minimal drop in T_m that the *c–h* linked-dimer experiences upon removal of the *s2* strand. The agreement between these computational models, which assume structured linked-dimers, with experimental data gives further credence to the double-helical association model of linked-dimer dimerization.

Construction of Molecular Models for Figures and Displays

All molecular models were created using coordinates from PDB entry 1k6f,[5] with the exception of models displayed in Figure 1B–E, which are based on coordinates generated by Gencollagen (UCSF).[9] Coordinates for the unfolded sections of damaged collagen were generated through the use of PEP-FOLD3.[10] All models shown in Figures 1 and S5 were rendered with PyMOL v1.3 (Schödinger), while the displays in Figures 2C–D were generated by VMD v1.9 (University of Illinois at Urbana–Champaign).

Possible Dimerization Modes for Linked-Dimers

Although the association of *x–y* and *s2* into *x–y·s2* can be explained easily by triple helix formation, there is no obvious model for the association of two *x–y* dimes into an *x–y* dimer of dimers. The data, however, indicate that certain association modes are likely.

The removal of *s2* from *x–y·s2* leaves numerous unpaired molecular contacts available on *x–y* for further association. As free linked-dimers are not present in significant amounts in our samples,

we believe *x–y* dimerization is robust. It is unlikely for van der Waals forces to support robust dimerization because these interactions are generally weak unless *x–y* dimers are well-packed. Interactions between terminal charges are also weak in 50 mM HOAc at pH 3. Thus, dimerization is most likely driven by hydrogen bonds forming between donors and acceptors exposed on each linked-dimer.

The inadequacy of a triple-helical association model (Figure S5A) has already been discussed in the main text. For double-helical association, since there is no canonical *x–y* dimer structure, multiple dimerization routes could be possible. One such possibility is dimerization relying on full or partial triple-helix formation between two otherwise-structured linked-dimers. (Figure S5B) Indeed, h^n–c^n shows a weak triple-helix NMR signal just above noise (Figure 5C), and "triple helices" that produce this signal might also be responsible for the dimerization of *x–y*. Although this model predicts a monomeric linked-dimer population in samples, AUC does not agree. The linked-dimers, *h–c*, *c–h* and *h–h* do not present a monomer population in their gradients. Furthermore, data for *c–c* also contradict this theory. The linked-homodimer $(s1c)_2$ contains the same cystine linker as does *c–c*, but does not produce a significant triple-helix population. This incompatibility between a cystine bridge and strand exchange should significantly deter the dimerization of *c–c* linked strands. Yet, *c–c* linked strands exist predominantly as a dimer of linked strands. Thus, we do not think this is the dominant association model for *x–y* pairs.

Given that all *x–y* exist almost exclusively as dimers in solution according to AUC, it is likely that the interactions enabling dimerization are relatively specific. Due to their strength and the availability of donors and acceptors on linked-dimer surfaces, hydrogen bonds are likely critical in this process. Each strand of a linked-dimer presents either only hydrogen-bond donors or only hydrogen-bond acceptors. All of these functionalities face the region previously occupied by the third strand, and this region is likely the binding region (*BR*) responsible for maintaining *x–y* dimerization. When one *x–y* associates with another in parallel, hydrogen-bond donors on each linked-dimer can be oriented towards the acceptors on the opposing strand. (Figure S5C) Even though steric factors prevents long runs of hydrogen bonds to form this way, associations between sections of the *BR* near the termini might be utilizing this mode of association. A terminal interaction, however, would allow linear self-assembly, or "polymerization", of linked-dimers, which is not observed.

Any steric complementarity between two linked-dimers would improve their dimerization. The *BR* runs the length of the linked-dimer, twisting along the path left vacant by the third strand, forming a "major groove" akin to that observed on B-form double helices of DNA. Because linked-dimers, themselves, are extended and helical, their major grooves can associate both in parallel and anti-parallel orientations, albeit with a slight angle. (Figure S5D). Thus, *BRs* on both linked-dimers could be in contact for about half a turn of the major groove. This orientation would not support the type of regular H-bond interactions highlighted above (Figure S5C), but sufficient H-bonds and good packing can possibly allow a good dimerization interface.

Although a single association state is apparent for triple-helical association (Figure S5A), none of the dimerization modes we have discussed here (Figure S5B–D) specify a unique association state. Linked-dimer sequence and structure are repetitive. Thus, multiple dimerization interfaces will appear on each *x–y*, and the choice of dimerization interface by each *x–y* in the dimer will likely influence the overall structure of the *x–y* dimer. Thus, although there might be unique association mode for *x–y* dimerization, a specific *x–y*-dimer structure might not exist. Rather, an ensemble of

x–y dimers, interacting through the same association mode, but at different sites on their *BR*s, could better explain these elusive structures.

References

(1) Tanrikulu, I. C.; Raines, R. T. Optimal interstrand bridges for collagen-like biomaterials. *J. Am. Chem. Soc.* **2014**, *136*, 13490–13493.

(2) Huwiler, K. G.; Mosher, D. F.; Vestling, M. M. Optimizing the MALDI–TOF–MS observation of peptides containing disulfide bonds. *J. Biomol. Technol.* **2003**, *14*, 289–297.

(3) Clos, L. J.; Jofre, M. F.; Ellinger, J. J.; Westler, W. M.; Markley, J. L. NMRbot: Python scripts enable high-throughput data collection on current Bruker BioSpin NMR spectrometers. *Metabolomics* **2013**, *9*, 558–563.

(4) Chemical Shift Thermometers for Variable Temperature NMR. https://www.spectroscopynow.com/userfiles/sepspec/file/specNOW/HTML%20files /NMR_temperature_measurement.htm (accessed July 16, 2019).

(5) Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal structure of the collagen triple helix model $[(\text{Pro-Pro-Gly})_{10}]_3$. *Protein Sci.* **2002**, *11*, 262–270.

(6) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.

(7) Lim, K. T.; Brunett, S.; Iotov, M.; McClurg, R. B.; Vaidehi, N.; Dasgupta, S.; Taylor, S.; Goddard, W. A. Molecular dynamics for very large systems on massively parallel computers: The MPSim program. *J. Comput. Chem.* **1997**, *18*, 501–521.

(8) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING—A generic force-field for molecular simularions. *J. Phys. Chem.* **1990**, *94*, 8897–8909.

(9) Gencollagen. https://www.rbvi.ucsf.edu/cgi-bin/gencollagen.py (accessed December 29, 2017).

(10) Lamiable, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuffery, P. PEP-FOLD3: Faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44*, W449–454.

Table S1. Sequences and T_m Values for Linked-Dimers and Their 1:1 Mixtures with *s2*[a]

Linked-dimer (*x–y*)	T_m (°C)		Amino acid sequence of *x–y*
	x–y	*x–y*·*s2*	
c–c	24	28[b]	PPGPPGPPGPPGPPG-Cys-PGPPGPPGPPGPPG PPGPPGPPGPPGP-Cys-GPPGPPGPPGPPGPPG
h–c	32	35[b]	PPGPPGPPGPPGPPG-Hcy-PGPPGPPGPPGPPG PPGPPGPPGPPGP-Cys-GPPGPPGPPGPPGPPG
c–h	27	28[b]	PPGPPGPPGPPGPPG-Cys-PGPPGPPGPPGPPG PPGPPGPPGPPGP-Hcy-GPPGPPGPPGPPGPPG
h–h	25	27[b]	PPGPPGPPGPPGPPG-Hcy-PGPPGPPGPPGPPG PPGPPGPPGPPGP-Hcy-GPPGPPGPPGPPGPPG
$(s1c)_2$	22	29	PPGPPGPPGPPGPPG-Cys-PGPPGPPGPPGPPG PPGPPGPPGPPGPPG-Cys-PGPPGPPGPPGPPG
h_x–*c*	n.o.[c]	n.d.[d]	PPGPP-Sar-PPGPPGPPG-Hcy-PGPPGPP-Sar-PPGPPG PPG-Pro-PGPPGPPGP-Cys-GPPGPPG-Pro-PGPPGPPG

[a]Strand *s2* is $(PPG)_{10}$.
[b]Values from ref 1.
[c]Thermal transition not observed.
[d]T_m could not be determined due to shallow transition.

Table S2. Breakdown of E_{strain} into Force-Field Energy Components for Linked-dimers

Linker type	CH_2 groups on linker, $n_{Xaa}+n_{Yaa}$	E_{strain} components (kcal/mol)				
		Total	Bonds	Angles	Torsions	van der Waals
h–c	3	3.1	0.1	0.3	2.3	0.4
c–h	3	6.9	0.1	0.1	5.3	1.5
h–h	4	7.3	0.1	0.8	5.2	1.3
c–c	2	14.6	1.1	8.4	4.5	0.9

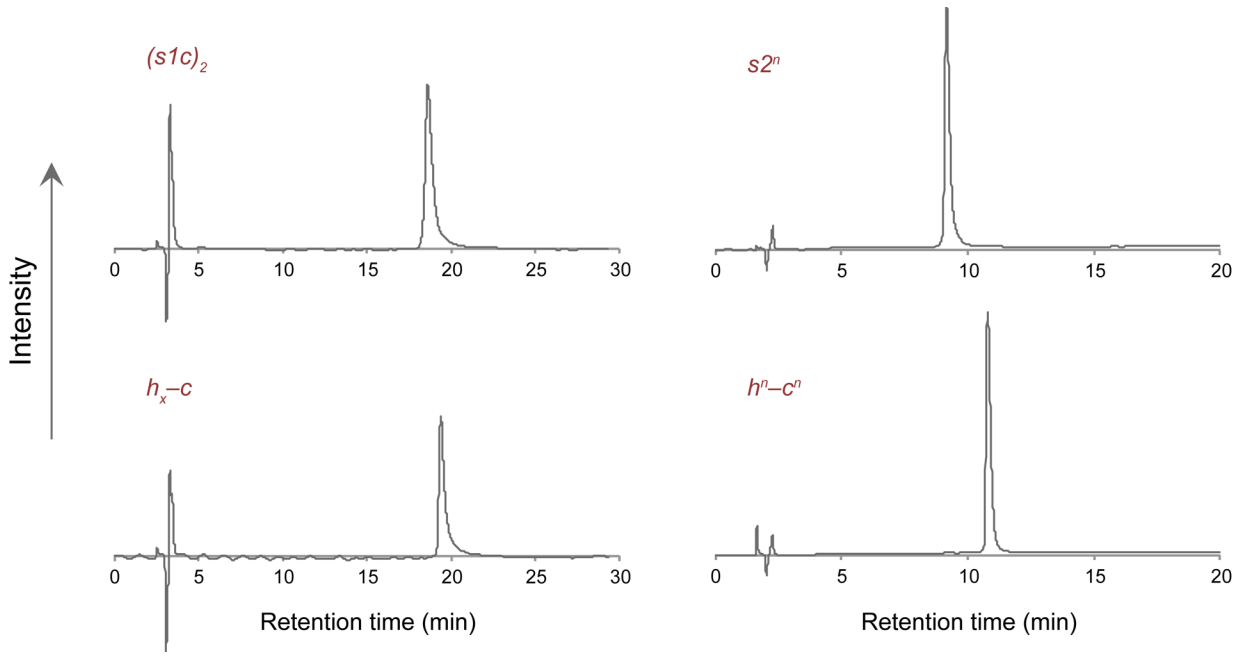

Figure S1. Chromatograms from analytical HPLC of purified linked-dimers and the *s2* peptide acquired on a VarioPrep 250/4.6 C18 column from Macherey–Nagel. Chromatograms for $(s1c)_2$ and h_x–c feature a 20–45% v/v gradient of acetonitrile in water over 25 min at 1 mL/min, whereas those for $s2^n$ and h^n–c^n feature a 20–50% v/v gradient of acetonitrile in water over 20 min at 1.5 mL/min flow rate. All HPLC runs feature solvents containing TFA (0.1% v/v).

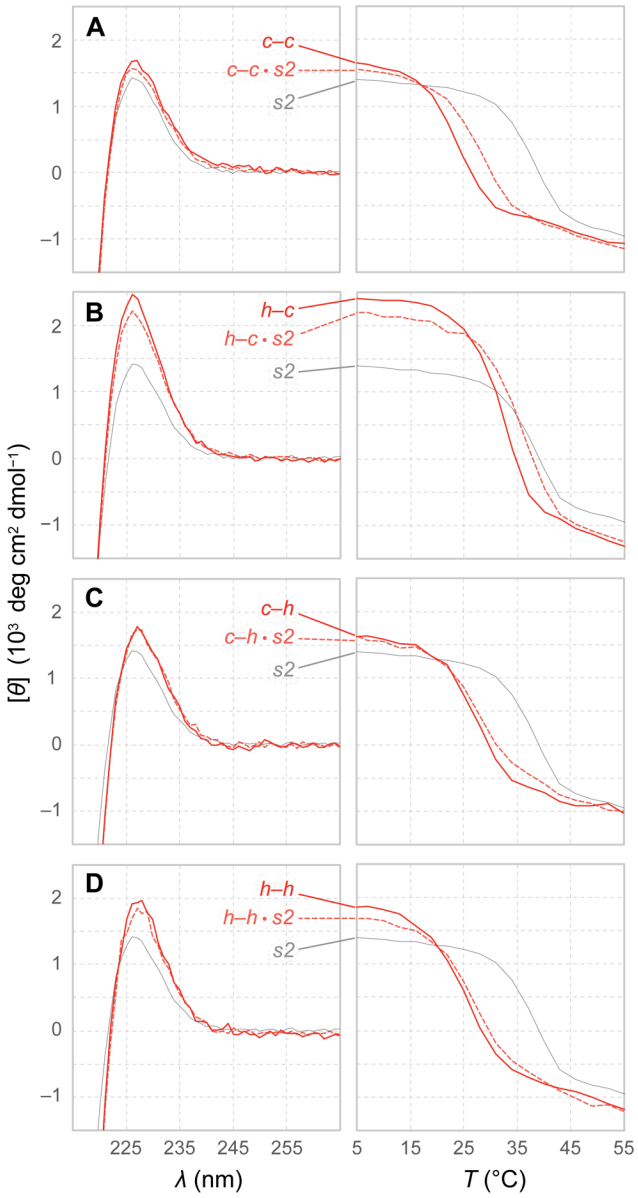

Figure S2. Circular dichroism spectra and thermal denaturation profiles for linked-dimers in comparison to their mixtures with *s2*. Data are shown for (A) *c–c*, (B) *h–c*, (C) *c–h*, and (D) *h–h*. For reference, data for the *s2* homotrimer are displayed in gray in each panel.

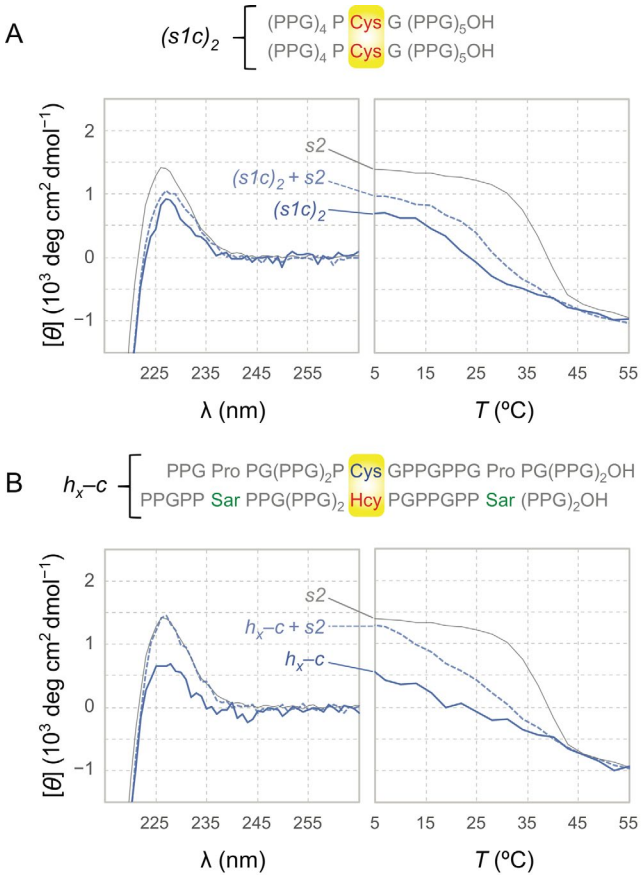

Figure S3. Circular dichroism spectra and thermal denaturation profiles for linked-dimers whose strands cannot fold together into collagen-like structures. Data are shown for (A) *(s1c)₂* and (B) h_x–*c*, as well as the sequences for each construct. Such linked-dimers do not display a strong collagen signal, unless *s2* is introduced. For reference, data for the *s2* homotrimer are displayed in gray in each panel.

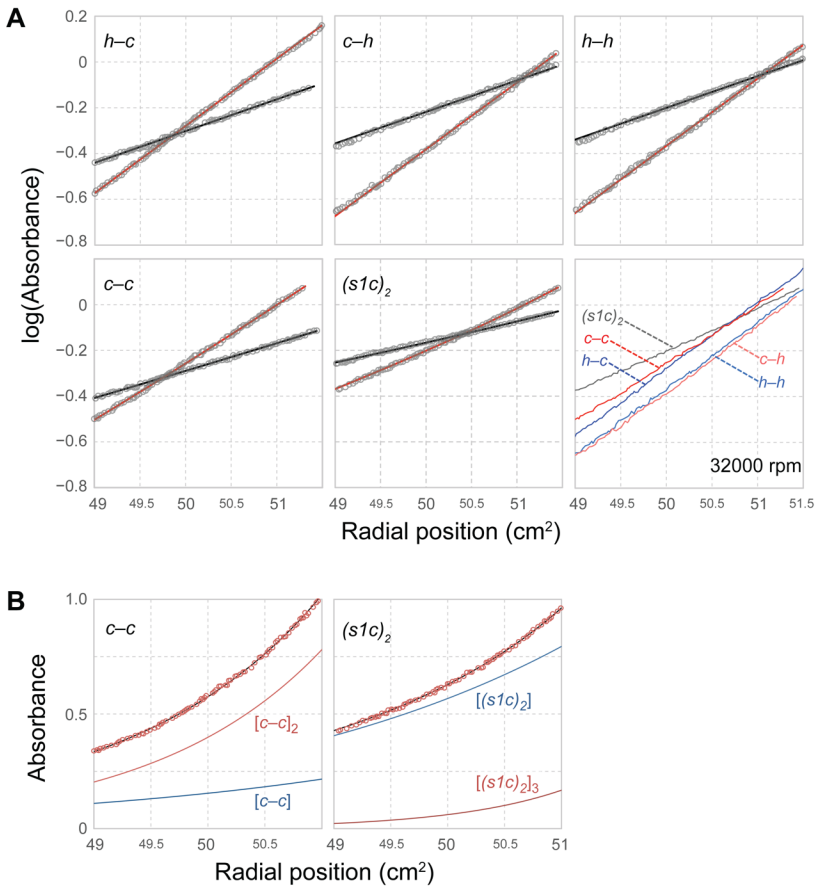

Figure S4. Sedimentation equilibrium analysis of linked-dimers. Equilibrium gradients (circles) are shown with models (lines) that provide optimal fits. (A) Gradients at 22k and 32k rpm (gray circles) and models that best fit the data (black and red lines for 22k and 32k rpm, respectively). The *h–c*, *c–h*, and *h–h* linked-dimers agree well with a four-stranded model, indicating an *x–y* dimer. The *c–c* and *(s1c)*$_2$ gradients necessitate the use of a model featuring both monomers and multimers. Equilibrium gradients of all species at 32k rpm (colored lines) are compared in the final panel and their varying slopes highlight the different oligomerization characteristics of *c–c* and *(s1c)*$_2$ with respect to other linked-dimers. (B) Two-species models for the isomers, *c–c* and *(s1c)*$_2$, display very different oligomerization patterns. The gradient at 32 k rpm (red circles) and the model (gray line) is shown, in addition to the calculated gradients of each species included in the model (teal and red lines for monomer and multimer, respectively). Whereas a small amount of monomer accompanies the *c–c* dimer, *(s1c)*$_2$ appears mostly as a monomer with a small population of *(s1c)*$_2$ trimers.

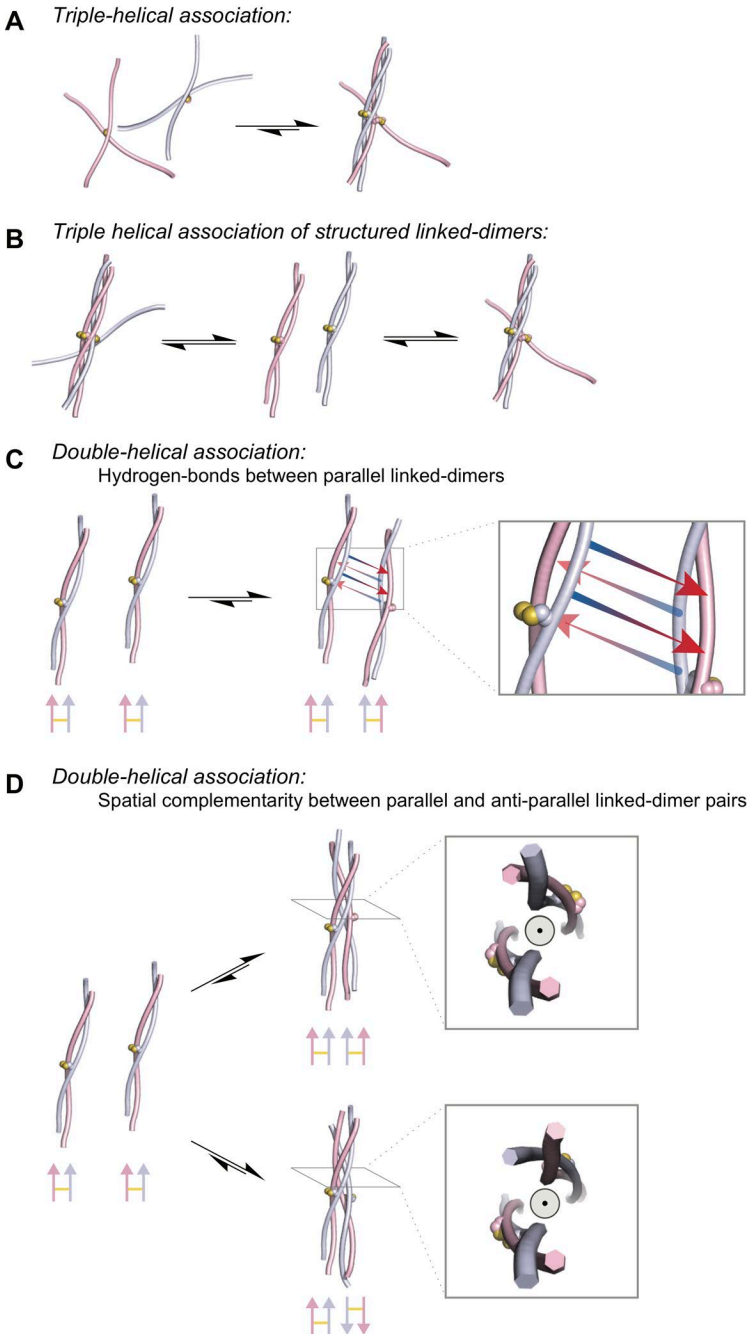

Figure S5. Possible models for the association of two linked-dimers. (A) Two unstructured linked-dimers could associate by forming a triple helix. (B) A pair of structured linked-dimers could form dimers via strand exchange. (C) Hydrogen-bond donors (on gray strands) and acceptors (on pink strands) on a pair of structured linked-dimers that face each other are aligned, which could assist dimerization. Arrows indicate hypothetical hydrogen bonds. (D) The absence of a third strand leaves a "major" groove that draws a helical trace on each structured linked-dimer. Structured linked-dimers could interface at a slight angle, bringing sections of their major grooves in contact. Top-down views show two linked-dimers (parallel and anti-parallel) with their major grooves aligned around a vertical central line (grey circle).

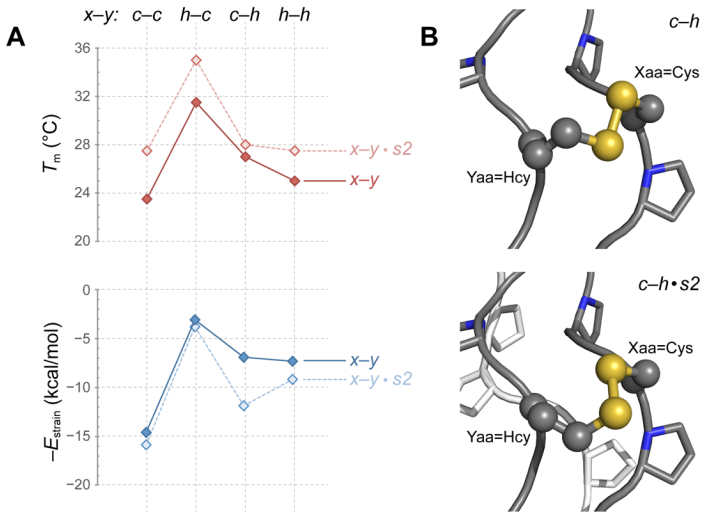

Figure S6. Effect of the *s2* strand on disulfide-linker strain energy in linked-dimers. (A) Linker-strain energy, $E_{strain} = E(x–y) - E(x \cdot y)$, computed on linked-dimers agrees well with E_{strain} for *x–y·s2* triple helices, and reproduces the trend of T_m values determined experimentally. The absence of the *s2* strand allows for additional relaxation of the *x–y* linked-dimer backbone and generally relieves E_{strain}; this effect is most prominent for the *c–h* linked-dimer. In addition, the *c–h* linked-dimer does not experience the drop in T_m observed for other linked-dimers when the *s2* strand is removed. The calculated strain energy is plotted as $-E_{strain}$ (rather than E_{strain}) to simplify comparisons with experimental data. (B) The conformation of the *c–h* linker changes in the absence of the *s2* strand. This change in conformation could be relevant to the minimal change in T_m observed for *c–h*. Models present the CMP backbone as cartoons with proline side chains visible. The *c–h* linker conformation is displayed in balls-and-sticks with yellow sulfur atoms forming the disulfide bond. The *s2* strand is shown in white.