Structural Basis for Aza-Glycine Stabilization of Collagen

*Alexander J. Kasznel[1,2], Yitao Zhang[1], Yang Hai[1] and David M. Chenoweth[1]**

[1]Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104-6323, USA.
[2]Department of Bioengineering, University of Pennsylvania, 210 South 33rd Street, Philadelphia, Pennsylvania 19104-6323, USA.

Table of Contents

General Information

Throughout this document, the following abbreviations are used:

CD	Circular dichroism
CDI	1-1'-Carbonyldiimidazole
CHCA	α-Cyano-4-hydroxycinnamic acid
CMP	Collagen model peptide
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DIEA	*N,N*-Diisopropylethylamine
DMF	*N,N*-Dimethylformamide
EDT	Ethane-1,2-dithiol
Fmoc	9-Fluoroenylmethoxycarbonyl
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HOBt	Hydroxybenzotriazole
IBCF	Isobutyl chloroformate
HPLC	High-performance liquid chromatography
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
NMM	*N*-Methylmorpholine
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
PBS	Phosphate-buffered saline
SPPS	Solid-phase peptide synthesis
tBu	*tert*-Butyl
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
XRD	X-ray diffraction

Reagents:

All commercially available solvents reagents were used as received. Rink Amide AM resin was purchased from Novabiochem. HATU was purchased from Oakwood Chemical. DIEA and diethyl ether were purchased from Sigma-Aldrich. Fmoc-Arg(Pbf)-OH, Fmoc-L-Pro-OH, and CDI were purchased from Chem-Impex International. Fmoc-Gly-OH was purchased from Advanced ChemTech. DBU, IBCF, Phenol, and TFA were purchased from Acros Organics. HOBt was purchased from EMD Millipore. EDT was purchased from Aldrich Chemical Company. Lithium sulfate monohydrate ($Li_2SO_4 \cdot H_2O$) was purchased from Sigma Life Science. PEG4000 was purchased from Hampton Research. All other commercially available solvents and reagents were purchased from Fisher Scientific. The tripeptide synthon Pro-Hyp-Gly (Fmoc-PO(tBu)G-OH) was prepared in-house using solution-phase techniques and was used as the primary SPPS building block.[1-3] Fmoc-hydrazine was prepared in-house using a protocol established by Carpino and Han[4] and also utilized previously in our lab.[2]

Instrumentation:

HPLC was performed using a JASCO PU-2080 Plus Intelligent HPLC Pump and Phenomenex Luna C18(2) columns (5 μm particle size, 100 Å pore size). MALDI-TOF MS was performed using a Bruker MALDI-TOF Ultraflex III mass spectrometer. CHCA was used as the matrix for all MALDI-TOF MS measurements. Peptides were lyophilized using a Labconco FreeZone Plus 12 Liter Cascade Console Freeze Dry System. CD measurements were performed using a JASCO J-1500 Circular Dichroism Spectrometer. UV-vis measurements were performed using a JASCO V-650 UV-vis Spectrophotometer. XRD was performed using the 24-ID-E

undulator beamline operated by the Northeastern Collaborative Access Team (NE-CAT) at the Advanced Photon Source (APS) (Argonne National Laboratory, Argonne, IL).

Synthesis and Purification of Collagen Model Peptides

Synthesis

Control peptide **1**, $(POG)_3$-PRG-$(POG)_4$, and its azapeptide analog **2**, $(POG)_3$-PRazG-$(POG)_4$, were synthesized using manual SPPS. The peptides were prepared on Rink Amide AM resin (200-400 mesh, 0.69 mmol/g) using an Fmoc protecting group. The methods below were followed for both **1** and **2** with variations as noted. When it was necessary to stop the synthesis after completing a coupling, the post-coupling wash step was modified such that the resin was washed with DMF (5x) and DCM (2x); following the Arg coupling, additional washes with ethanol were also performed (see step 4 below). When beginning from a dry resin at any point during the synthesis, the resin was swelled for at least 30 min in DMF prior to initial deprotection.

1. Resin preparation

29 mg (0.02 mmol, 1 eq.) Rink Amide AM resin was added to a 10-mL SPPS vessel. The resin was swelled by stirring in DMF for 30 min. A deprotectant solution was prepared by combining 20 mL DMF, 200 mg HOBt, and 0.4 mL DBU. The Fmoc protecting group was removed from the resin by mixing with 1 mL (1% HOBt (w/v), 2% DBU (v/v) in DMF) of this deprotectant solution and then draining the solution (3 x 50 s).

2. POG synthon coupling

Following initial deprotection, the resin was washed with DMF (6x). A stock solution of HATU was prepared by dissolving 690 mg (1.81 mmol) HATU in 20 mL DMF. This solution was then used to prepare coupling solutions as noted. A coupling solution of 34 mg (0.06 mmol, 3 eq.) Fmoc-PO(tBu)G-OH, 0.67 mL (0.06 mmol, 3 eq.) HATU, and 20 µL DIEA (0.12 mmol, 6 eq.) was prepared and allowed to activate for ~10 min at ambient temperature. This solution was then added to the resin and stirred for 80 min. The coupling solution was drained from the vessel and the resin was washed with DMF (6x). These steps were repeated 4 times to couple a total of 4 POG trimers onto the resin.

3a. Glycine (Gly) coupling [CMP **1**]

The Fmoc protecting group was removed as described above, and the resin was washed with DMF (6x). A coupling solution of 18 mg (0.06 mmol, 3 eq.) Fmoc-Gly-OH, 0.67 mL (0.06 mmol, 3 eq.) HATU, and 20 µL DIEA (0.12 mmol, 6 eq.) was prepared and allowed to activate for ~10 min at ambient temperature. This solution was then added to the resin and stirred for 60 min. The coupling solution was drained from the vessel and the resin was washed with DMF (6x).

3b. Aza-glycine (azGly) coupling [CMP **2**]

The Fmoc protecting group was removed as described above, and the resin was washed with DMF (6x). A coupling solution of 10 mg (0.06 mmol, 3 eq.) CDI, 0.4 mL DMF, and 15.2 mg (0.06 mmol, 3 eq.) Fmoc-hydrazine (Fmoc-NH-NH_2) was prepared and allowed to activate for 5-10 min at ambient temperature. The coupling solution was then added to the resin and stirred overnight. The following day, a fresh coupling solution was prepared and allowed to activate as described above. The original coupling solution was drained from the vessel and the fresh solution was added to the resin and stirred for an additional 4 h. This second coupling solution was drained from the vessel and the resin was washed with DMF (6x).

4. Arginine (Arg) coupling

The Fmoc protecting group was removed as described above, and the resin was washed with DMF (5x) followed by THF (2x). A coupling solution of 39 mg (0.06 mmol, 3 eq.) Fmoc-Arg(Pbf)-OH, 8 µL IBCF (0.06 mmol, 3 eq.), 13 µL NMM (0.12 mmol, 6 eq.), and 0.65 mL THF was prepared and allowed to activate for ~10 min at

ambient temperature. This solution was then added to the resin and stirred for 5.5-6 h. The coupling solution was drained and the resin was washed with ethanol (2x) and DMF (5x).

5. Proline (Pro) coupling

The Fmoc protecting group was removed as described above, and the resin was washed with DMF (6x). A coupling solution of 20 mg (0.06 mmol, 3 eq.) Fmoc-L-Pro-OH, 0.67 mL (0.06 mmol, 3 eq.) HATU, and 20 μL DIEA (0.12 mmol, 6 eq.) was prepared and allowed to activate for ~10 min at ambient temperature. This solution was then added to the resin and stirred for 1 h. The coupling solution was drained from the vessel and the resin was washed with DMF (6x).

6. POG synthon coupling

The coupling protocol outlined in step 2 was repeated 3 more times to couple the final three POG trimers onto the peptide. When appropriate, some of these final coupling reactions were performed on a smaller scale to conserve reagents. At this scale, the coupling solutions consisted of 24 mg (0.04 mmol, 2 eq.) Fmoc-PO(tBu)G-OH, 0.47 mL HATU (0.04 mmol, 2 eq.), and 15 μL DIEA (0.08 mmol, 4 eq.).

7. Final deprotection and cleavage

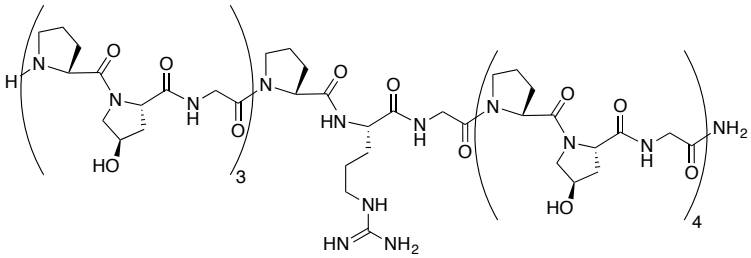
The Fmoc protecting group was removed as described above, and the resin was washed thoroughly with DMF (2x) followed by DCM (4x). The completed peptides were then cleaved from the resin by mixing with a 4-mL cleavage cocktail of TFA, phenol, H_2O, and EDT in an 87:5:5:3 (v/v) ratio for 1.5 h.

Purification

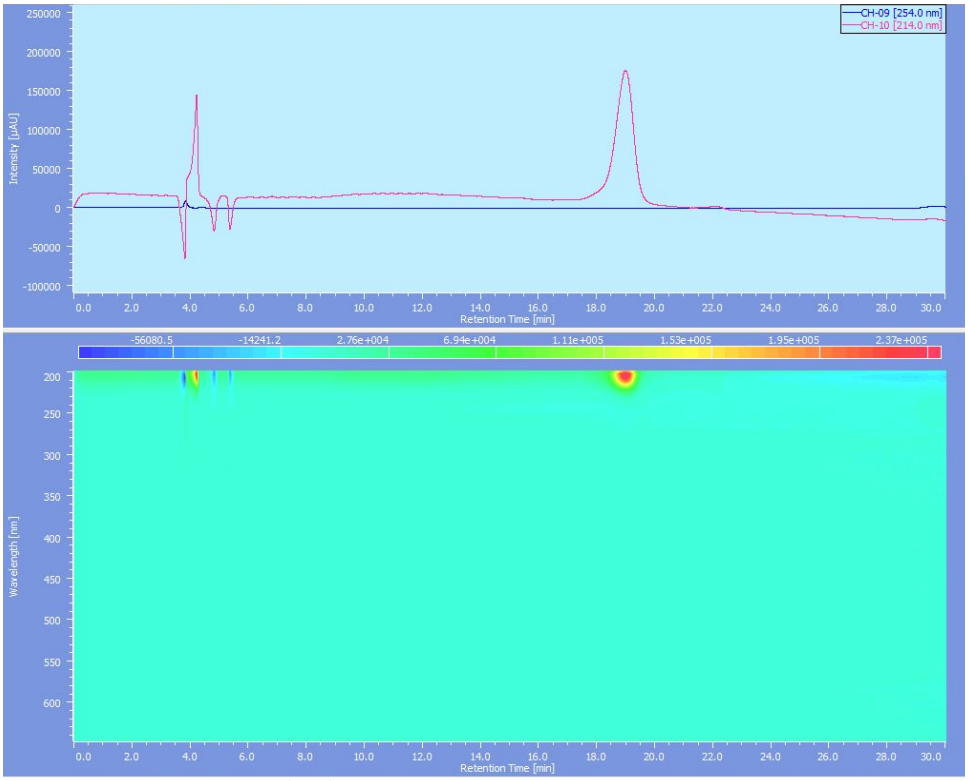
Following SPPS, the peptides were precipitated in cold diethyl ether. After initial precipitation, cleaved peptide solutions were centrifuged, the supernatants were decanted, and the solid peptide pellets were resuspended in ether. This process was repeated for a total of 3 resuspensions. After decanting the final supernatant, the solid peptide was dissolved in ~6 mL 18 MΩ H_2O and stored at 4 °C. Crude peptide stocks were then purified using preparative reverse-phase HPLC with a mobile phase gradient of 10-20% acetonitrile in H_2O (0.1% TFA). After reviewing initial chromatograms, some of the crude CMP **1** stock was also purified with a mobile phase gradient of 10-15% acetonitrile in H_2O (0.1% TFA) to enhance peak separation. Peptide solutions were heated for at least ~10 min at ~70 °C to ensure that peptides were in the single-stranded state prior to being loaded onto the column. The chromatographic fractions were analyzed by MALDI-TOF MS in positive ion mode. The fractions found to contain the desired product were pooled according to purity and lyophilized, after which stocks of **1** and **2** became fluffy white solids.

HPLC Traces and MALDI-TOF Mass Spectra

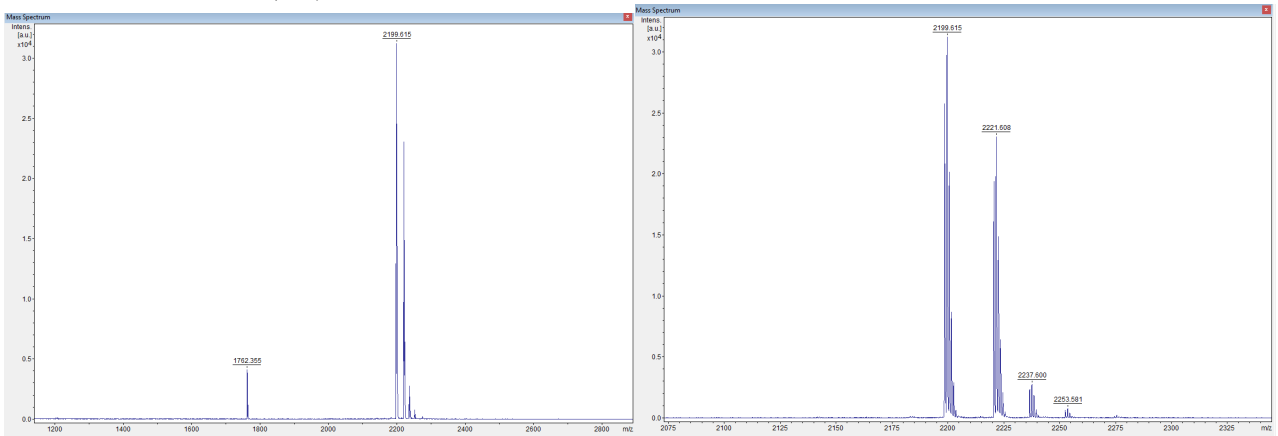
CMP 1: $(Pro\text{-}Hyp\text{-}Gly)_3\text{-}Pro\text{-}Arg\text{-}Gly\text{-}(Pro\text{-}Hyp\text{-}Gly)_4\text{-}NH_2$



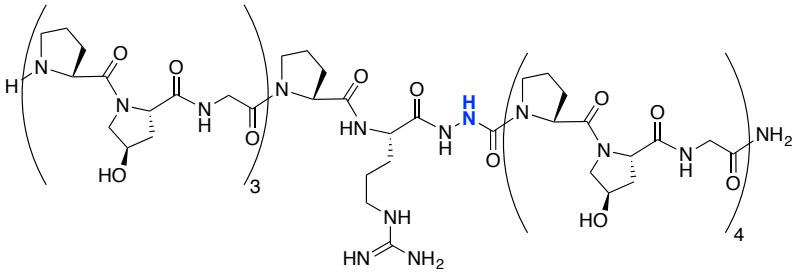
Analytical HPLC Gradient = 10-25% acetonitrile/H_2O (0.1% TFA), 30 min.



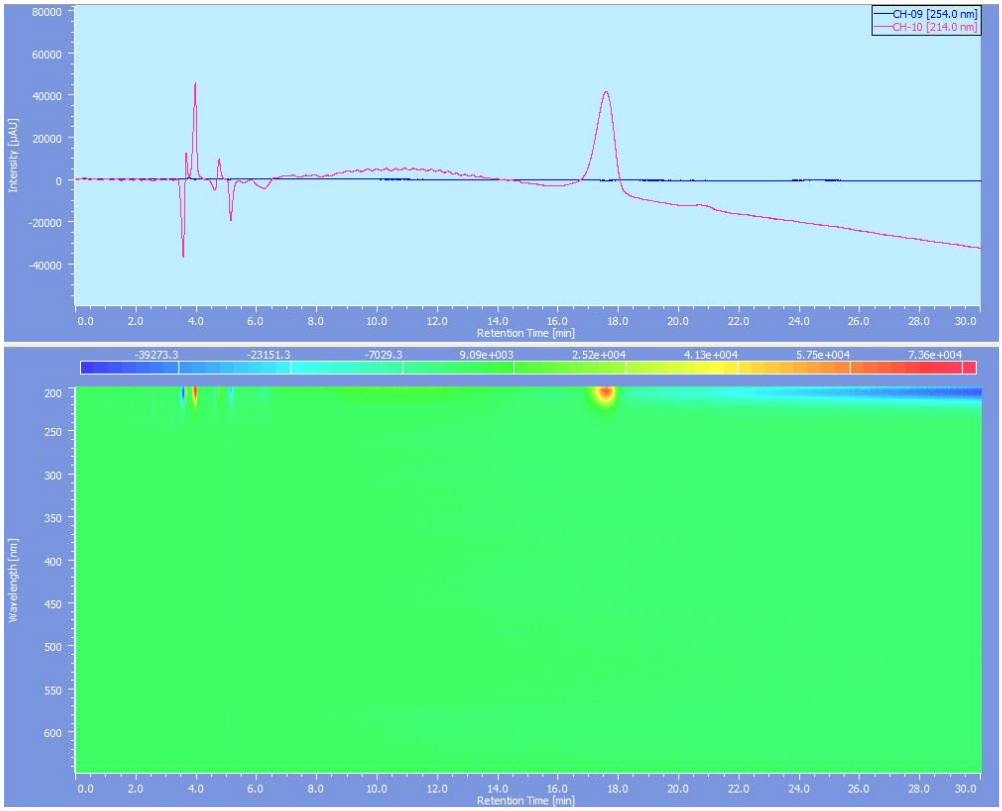
MALDI MS $[M+H^+]$ (Da): Calculated 2198.06; Found 2199.62.



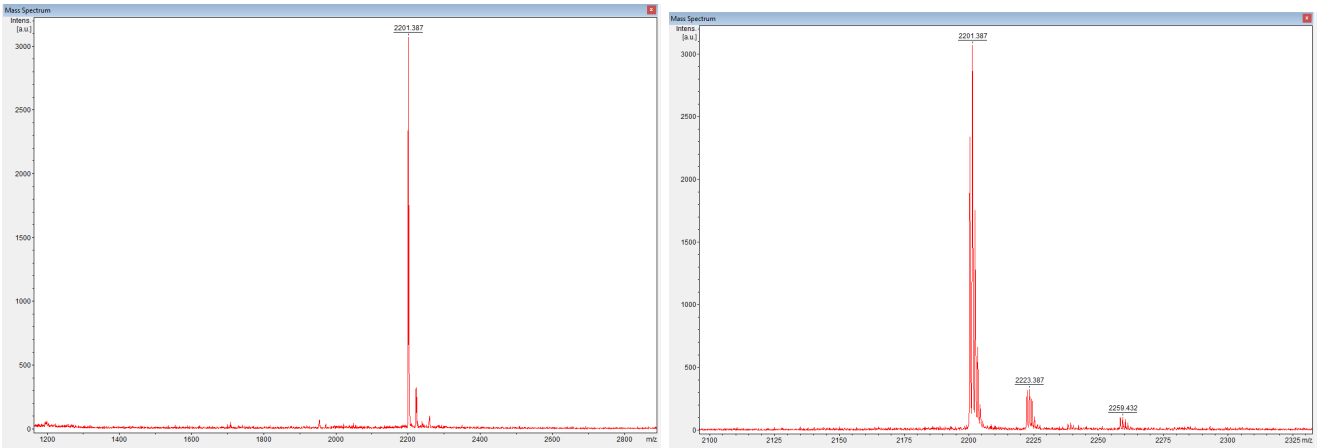
CMP 2: (Pro-Hyp-Gly)$_3$-Pro-Arg-**azGly**-(Pro-Hyp-Gly)$_4$-NH_2



Analytical HPLC Gradient: 10-25% acetonitrile/H_2O (0.1% TFA), 30 min.



MALDI MS [M+H$^+$] (Da): Calculated 2199.06; Found 2201.39.



CD Experimental Protocols

a. Peptide Solution Preparation

Small amounts of each purified solid peptide stock were dissolved in a few microliters of PBS (pH 7.4) to create concentrated aqueous solutions. These concentrated stocks were then diluted (e.g. 1:500, 1:1000). The absorbance of these dilute solutions was measured at 214 nm (A_{214}) in 1-cm quartz cuvettes. In some cases, multiple iterations of the same dilution were performed and an average of their absorbance values was used for the following calculations. Using a standard Beer's Law calculation with a molar extinction coefficient of 60 mM^{-1} cm^{-1}, the concentration of this stock was determined. The appropriate dilution factor was then applied to determine the concentration of the original concentrated stock. Once this value was determined, these concentrated stocks were used to prepare separate, diluted stocks to be used in the CD experiments (final concentration = 0.2 mM in PBS). These stocks were incubated overnight at 4 °C.

b. CD Wavelength Scan

Approximately 300 μL of each peptide solution was added to a 1-mm quartz cuvette. The ellipticity of each solution, θ, was measured from 260 to 190 nm at a temperature of 20 °C. This measurement was repeated in triplicate for **1** and **2**. The results of each trial were converted to mean residue ellipticity, $[\theta]_{MRW}$, and then averaged to generate the curves shown on pages S8-10.

c. Thermal Unfolding

Following the initial CD wavelength scan, solutions of **1** and **2** were heated at 12 °C/h to induce unfolding of the triple helix. The change in ellipticity of each peptide at 224 nm was monitored by CD spectroscopy. Their respective thermal transition temperatures, T_m, at which each peptide lost 50% of its initial ellipticity at this wavelength, were determined from the resulting melting curves (see Curve Fitting below). The overall values of T_m for **1** and **2** were calculated by averaging the T_m values generated from triplicate thermal melting measurements. The melting curves shown in this supporting info are expressed in terms of mean residue ellipticity, $[\theta]_{MRW}$, while the curves shown in the article itself (Figure 2) were normalized for ease of comparison.

d. Curve Fitting

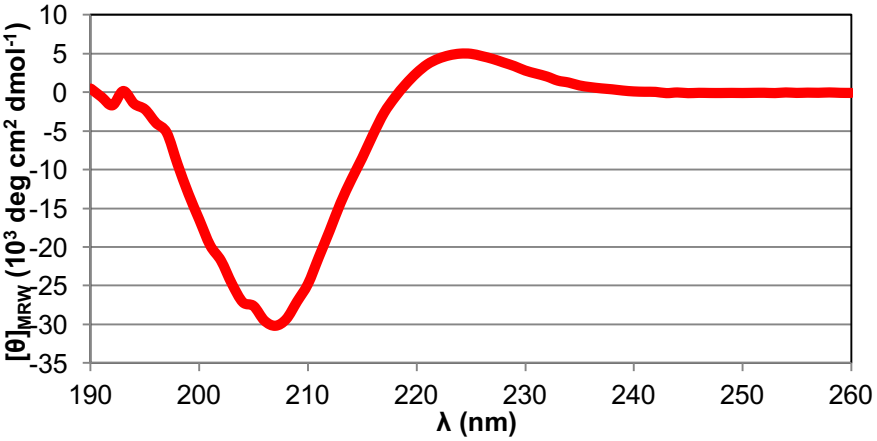
The thermal denaturation curves generated during these CD experiments were fitted using GraphPad Prism (version 6.0b for Mac OS X; GraphPad Software, San Diego, CA) as described previously.[2]

CD Spectra

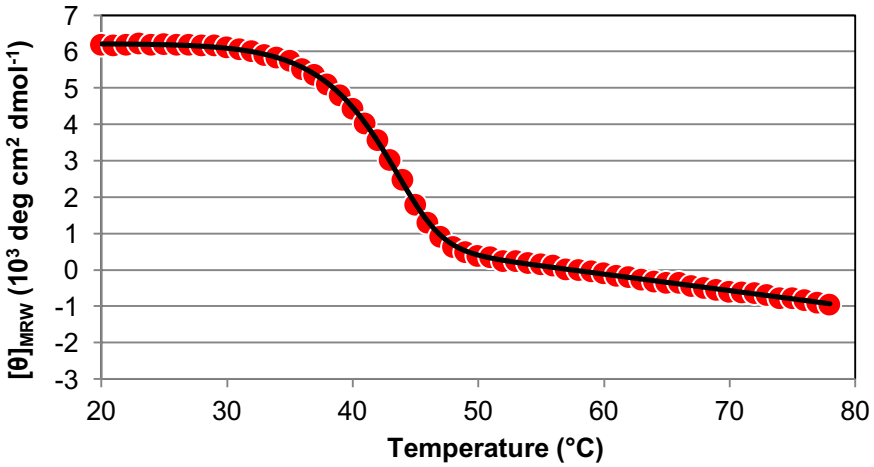
Thermal denaturation curves were collected at a 224-nm wavelength.

CMP 1: $(Pro\text{-}Hyp\text{-}Gly)_3\text{-}Pro\text{-}Arg\text{-}Gly\text{-}(Pro\text{-}Hyp\text{-}Gly)_4\text{-}NH_2$

a. CD Wavelength Scan

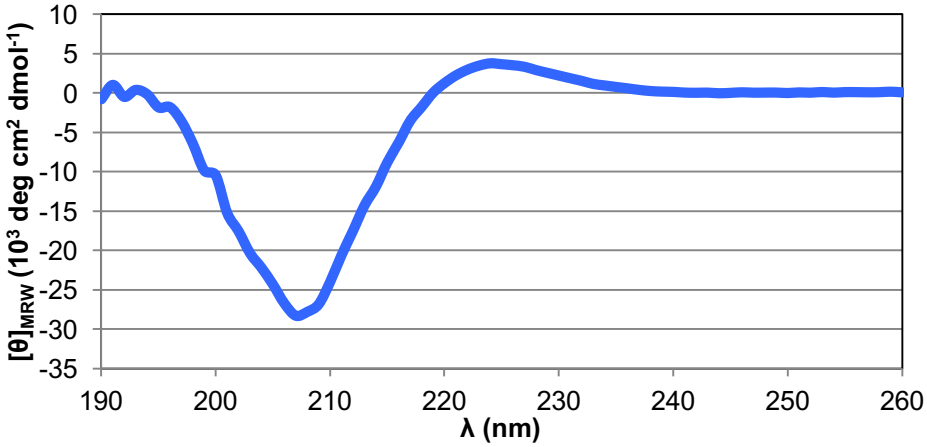


b. Thermal Unfolding Curve (T_m = 42.2 °C)

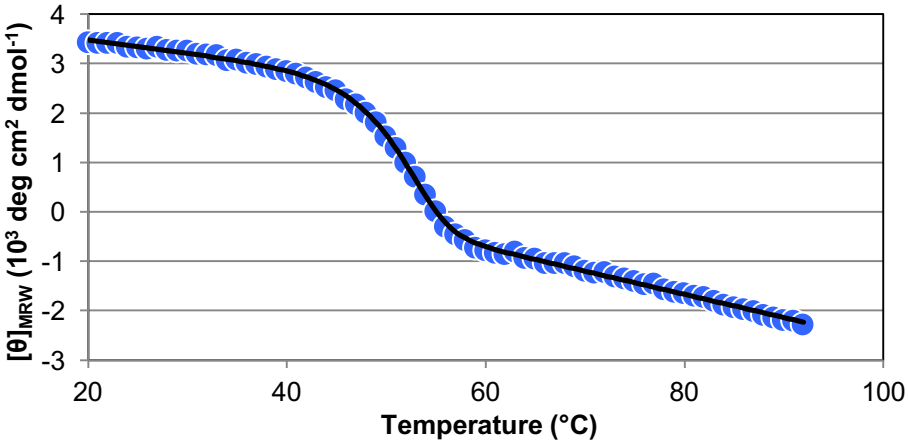


CMP 2: (Pro-Hyp-Gly)$_3$-Pro-Arg-azGly-(Pro-Hyp-Gly)$_4$-NH_2

a. CD Wavelength Scan

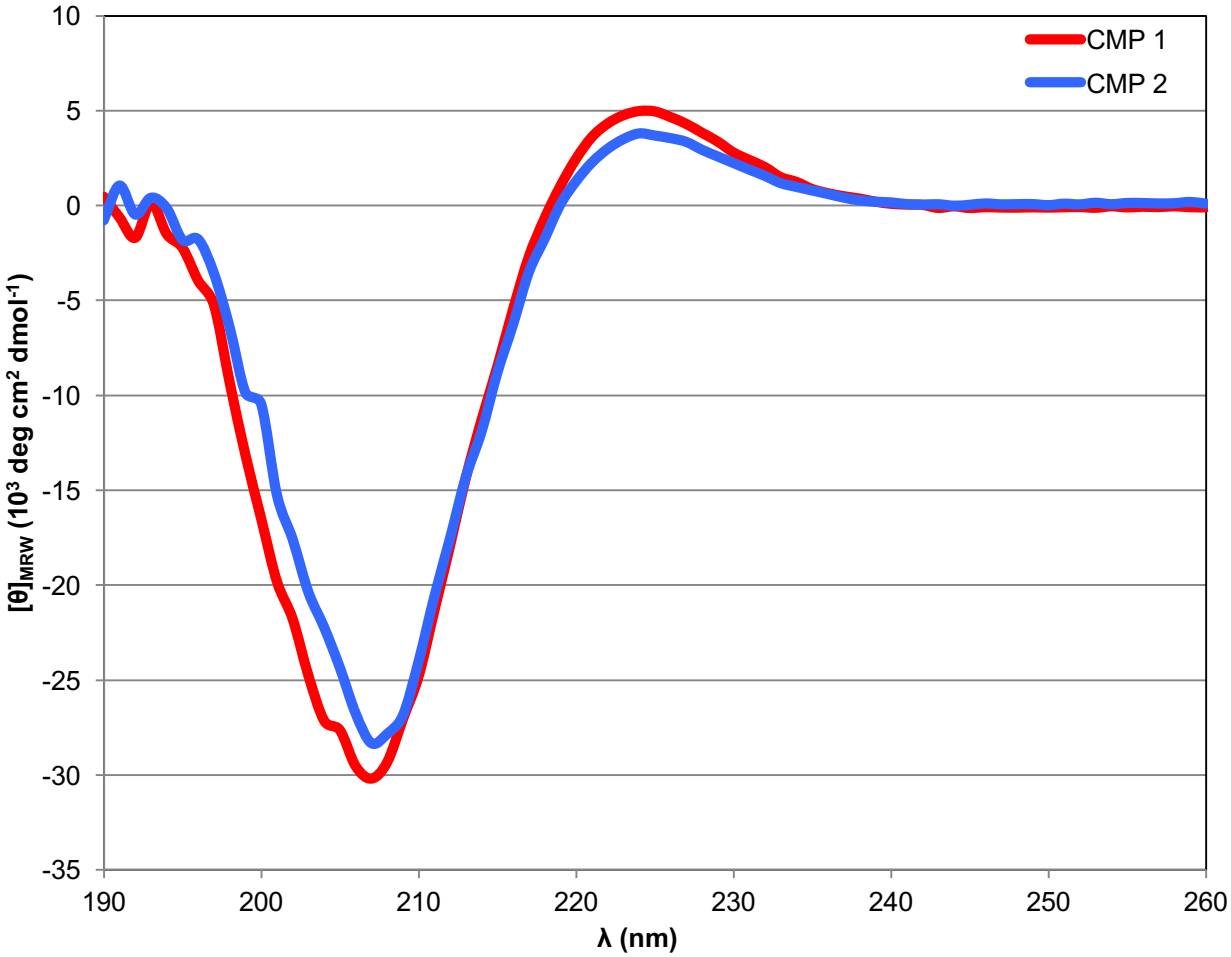


b. Thermal Unfolding Curve (T_m = 50.8 °C)



CMP 1 & 2: CD Wavelength Scan Comparison

In the overlaid CD spectra below, there is an apparent red shift in the CD spectrum of CMP **2** as compared to CMP **1**. Furthermore, the azGly substitution in CMP **2** yielded a slight decrease in the intensity of the positive peak at 224 nm. These results are consistent with our previous findings.[1-2]



Crystallization

Peptide **2** was crystallized using sitting-drop vapor diffusion under conditions adapted from Okuyama *et al.*[5] Peptide stock solutions were prepared by dissolving the purified solid product in 18 MΩ H_2O to a final concentration of 8.4 mg/mL using the solution preparation procedure outlined on page S7. Crystal trials were prepared by combining 1 µL of the peptide solution with 1 µL of a reservoir solution of 0.094 M Tris-HCl (pH 7.6), 30% (w/v) PEG4000, and 0.01 M $Li_2SO_4 \cdot H_2O$. This reservoir solution was prepared by first preparing a 0.94 M stock of Tris Base and adjusting its pH to 7.6 with concentrated HCl. Then, to prepare the reservoir solution, Tris-HCl was drawn from this stock and combined with the other listed reagents in the appropriate ratios to achieve the noted concentrations. Trays were sealed tightly with plastic tape to create a closed system and prevent solvent evaporation. Trays were incubated at 4 °C and monitored periodically using light microscopy; crystals became visible within approximately 1 week. Prior to beamline analysis, crystals were dipped in a drop of a cryoprotectant mixture containing equal volumes of the reservoir solution and 30% (w/v) PEG4000 and then frozen in liquid N_2.

XRD and Crystal Structure Refinement

The XRD data was refined using the CCP4 software suite,[6] including PHENIX (version dev_1370, structural refinement and phasing), iMOSFLM (data reduction), and SCALA (scaling). Further analysis was performed using MOLEMAN (X-UTIL package, Uppsala Software Factory).[7] The details of XRD data collection and refinement are presented in Table 1 below.

Data collection	
Beamline	APS 24-ID-E
Detector	ADSC Quantum 315
Wavelength (Å)	0.97918
Space group	$P\,1\,2_1\,1$
Resolution (Å)[a]	27.48-1.13 (1.15-1.13)
Unique reflections measured	18552 (897)
Unit cell parameters	
(a, b, c; Å)	27.49, 18.24, 48.84
(α, β, γ; °)	90.00, 91.66, 90.00
Completeness (%)	99.3 (97.7)
I/σ_I	7.1 (1.6)
R_{merge}[b]	0.088 (0.566)
R_{pim}[c]	0.063 (0.434)
Multiplicity	2.8 (2.4)
$CC_{1/2}$[d]	0.995 (0.607)
Refinement	
No. of reflections, refinement/test set	1790
R_{work} (%)	12.9
R_{free} (%)	15.9
No. of non-hydrogen atoms	
Protein	468
Solvent	134
Sulfate	5
RMSD	
Bonds (Å)	0.015
Angles (°)	1.688
Average B factors (Å^2)	
Protein	8.7
Solvent	16
Sulfate	10

Table 1. Data collection and refinement statistics. [a]Numbers in parentheses refer to the highest resolution shell. [b]$R_{merge} = \Sigma_h\Sigma_i|I(h)_i - \langle I(h)\rangle|/ /_h\Sigma_i I(h)_i$, where $I(h)$ is the intensity of reflection h, Σ_h is the sum over all reflections and Σ_i is the sum over i measurements of reflection h. [c]$R_{pim} = \Sigma sqrt(1/n\text{-}1)|I - \langle I\rangle|/)|I$ - ents$n$ is the number of observations (redundancy) and $\langle I\rangle$ is the average intensity calculated from replicate data. [d]$CC_{1/2} = s_t^2/(s_t^2 + s_e^2)$, where s_t^2 is the true measurement error variance and s_e^2 is the independent measurement error variance.

Collagen Backbone Analysis

The structures in this section were analyzed using Mercury (CCDC)[8-11] and the UCSF Chimera software package (Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco; supported by NIGMS P41-GM103311).[12]

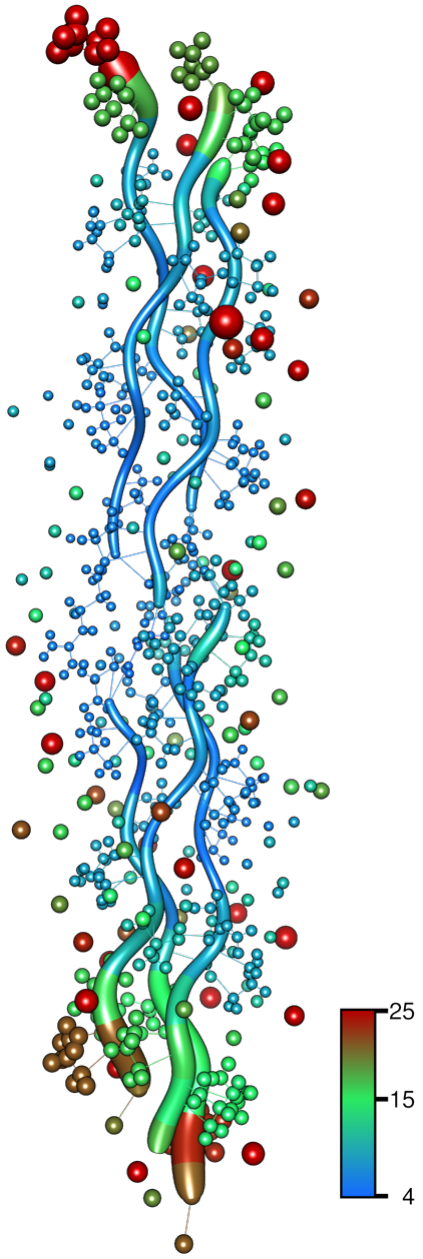
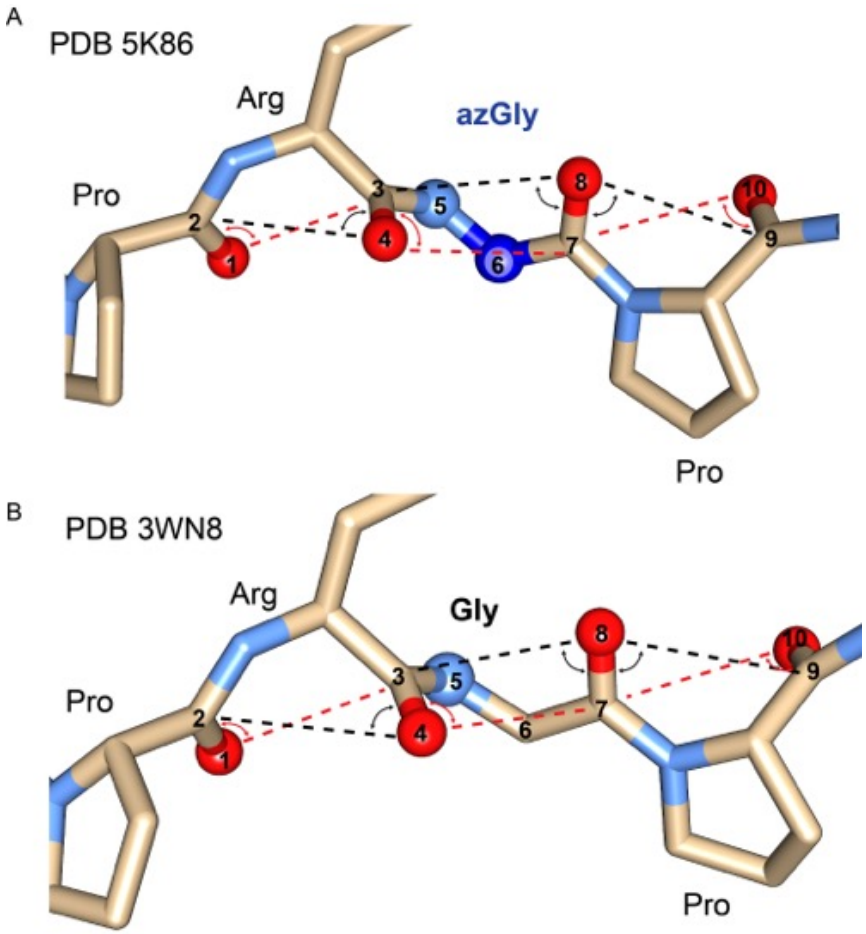


Figure S1. Illustration of average B factors in CMP 2. The average B factors are low (~4 Å^2) in and around the central –PRazG- triplet. This indicates limited atomic motion, likely due to added stability from the additional H-bonds integrated by the azGly residue. Going from the central triplet to the C and N termini, the average B factor increases substantially as the backbone becomes less ordered in these regions.



Interatomic Bond Distances (Å)					
Structure	**3-4**	**3-5**	**5-6**	**6-7**	**7-8**
PDB 5K86	1.23 ± 0.01	1.36 ± 0.00	1.39 ± 0.02	1.40 ± 0.00	1.22 ± 0.01
PDB 3WN8	1.24 ± 0.00	1.34 ± 0.02	1.44 ± 0.00	1.52 ± 0.00	1.23 ± 0.02

Interatomic Bond Angles (°)					
Structure	**∠3 5 6 7**	**∠4 3 5**	**∠3 5 6**	**∠5 6 7**	**∠6 7 8**
PDB 5K86	-72.45 ± 4.29	122.91 ± 0.16	120.82 ± 0.82	119.48 ± 1.42	121.63 ± 0.29
PDB 3WN8	-63.68 ± 4.63	121.98 ± 0.23	123.32 ± 2.14	114.13 ± 0.45	121.24 ± 2.37

n-π* Interaction Angles (°)						
Structure	**∠2 1 3**	**∠2 4 3**	**∠3 4 7**	**∠3 8 7**	**∠7 8 9**	**∠7 10 9**
PDB 5K86	78.04 ± 1.99	67.50 ± 3.05	76.92 ± 0.60	78.11 ± 0.84	78.65 ± 2.99	74.74 ± 3.79
PDB 3WN8	77.10 ± 0.47	69.75 ± 5.10	76.74 ± 3.10	78.35 ± 1.40	78.57 ± 3.44	72.81 ± 0.77

Figure S2. Interatomic bonding and potential n-π* interactions in the central arginine-containing triplet of PDB 5K86 and 3WN8. A. Numbering for atoms in PDB 5K86. B. Numbering for atoms in PDB 3WN8. Data: For each of the numbered atoms in the peptide backbone segments shown in A and B above, the interatomic bond distances, interatomic bond angles, and angles of potential n-π* interactions are presented. Values are listed as mean ± standard deviation.

n-π* Interaction Angles - PDB 5K86 (°)						
Chain	**∠2 1 3**	**∠2 4 3**	**∠3 4 7**	**∠3 8 7**	**∠7 8 9**	**∠7 10 9**
A	78.78	64.00	76.31	77.28	78.28	78.19
B	75.78	69.62	76.96	78.11	81.81	70.68
C	79.55	68.87	77.50	78.95	75.86	75.36
Mean	78.04	67.50	76.92	78.11	78.65	74.74
SD	1.99	3.05	0.60	0.84	2.99	3.79

n-π* Interaction Angles - PDB 3WN8 (°)						
Chain	**∠2 1 3**	**∠2 4 3**	**∠3 4 7**	**∠3 8 7**	**∠7 8 9**	**∠7 10 9**
A	76.82	64.96	79.36	79.96	74.59	73.14
B	77.65	75.11	73.32	77.46	80.62	73.37
C	76.84	69.17	77.53	77.63	80.49	71.93
Mean	77.10	69.75	76.74	78.35	78.57	72.81
SD	0.47	5.10	3.10	1.40	3.44	0.77

Interatomic Bond Angles - PDB 5K86 (°)					
Chain	**∠3 5 6 7**	**∠4 3 5**	**∠3 5 6**	**∠5 6 7**	**∠6 7 8**
A	-75.646	122.909	121.489	120.009	121.787
B	-74.125	122.752	121.070	120.566	121.811
C	-67.578	123.078	119.910	117.870	121.298
Mean	-72.45	122.91	120.82	119.48	121.63
SD	4.29	0.16	0.82	1.42	0.29

Interatomic Bond Angles - PDB 3WN8 (°)					
Chain	**∠3 5 6 7**	**∠4 3 5**	**∠3 5 6**	**∠5 6 7**	**∠6 7 8**
A	-60.005	121.967	121.245	114.562	118.611
B	-62.159	122.221	125.524	114.154	121.878
C	-68.873	121.758	123.186	113.671	123.217
Mean	-63.68	121.98	123.32	114.13	121.24
SD	4.63	0.23	2.14	0.45	2.37

azGly Interatomic Bond Distances - PDB 5K86 (Å)					
Chain	**3-4**	**3-5**	**5-6**	**6-7**	**7-8**
A	1.232	1.364	1.397	1.403	1.223
B	1.216	1.359	1.361	1.400	1.214
C	1.230	1.367	1.399	1.395	1.233
Mean	1.23	1.36	1.39	1.40	1.22
SD	0.01	0.00	0.02	0.00	0.01

Gly Interatomic Bond Distances - PDB 3WN8 (Å)					
Chain	**3-4**	**3-5**	**5-6**	**6-7**	**7-8**
A	1.231	1.359	1.442	1.519	1.253
B	1.234	1.322	1.435	1.519	1.225
C	1.240	1.337	1.437	1.519	1.225
Mean	1.24	1.34	1.44	1.52	1.23
SD	0.00	0.02	0.00	0.00	0.02

Figure S3. Interatomic bonding and potential n-π* interactions in the central arginine-containing triplet of PDB 5K86 and 3WN8 (full data). The measurements summarized previously in Figure S2 are shown here for each chain of PDB 3WN8 and PDB 5K86. Highlighted values represent an average due to the alternate conformation of the Arg residue in chain B of PDB 3WN8.[5]

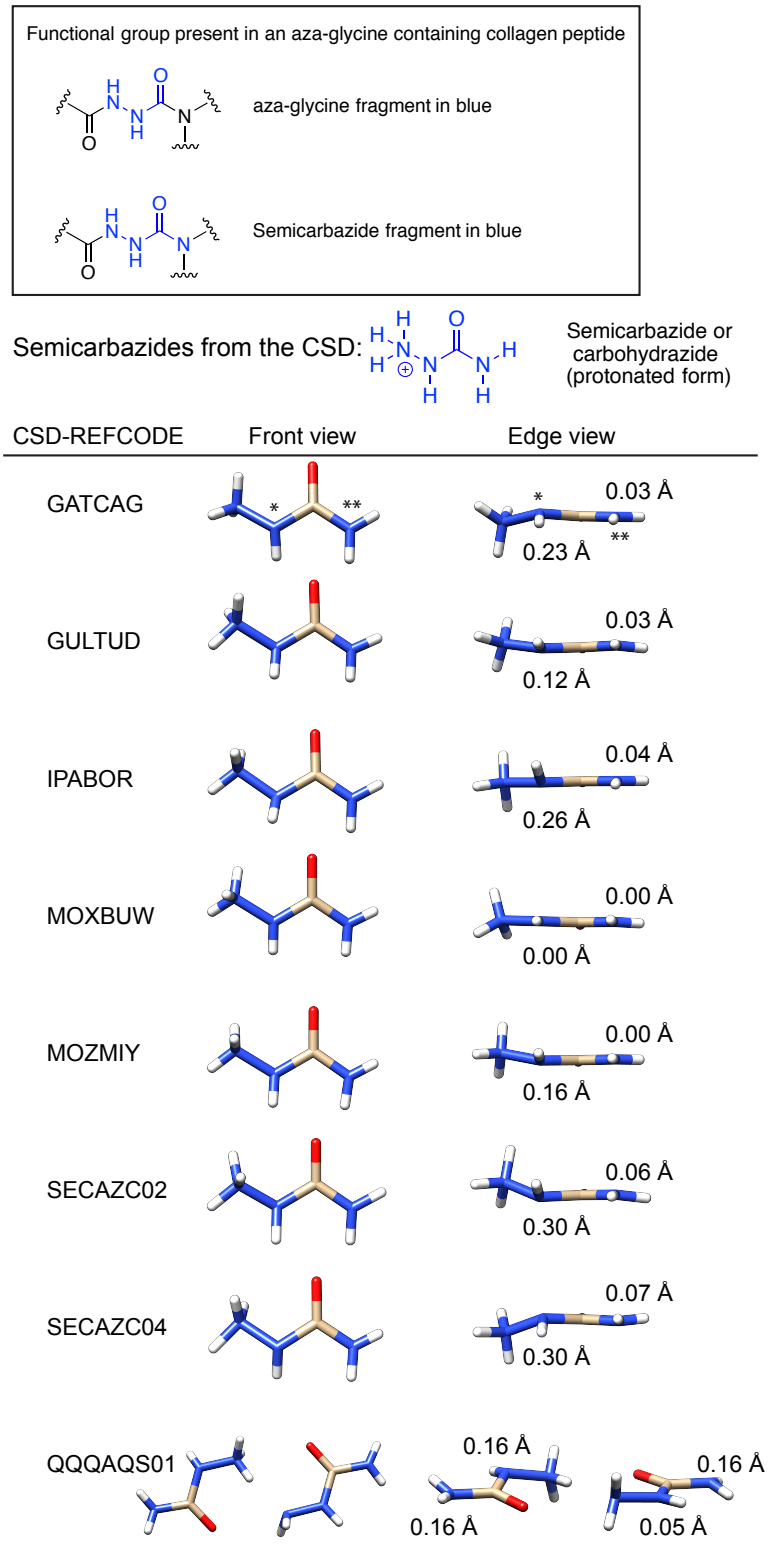
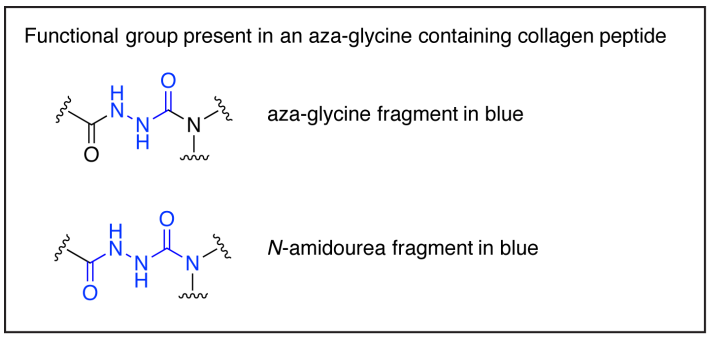


Figure S4. Structural comparison of backbone conformations in molecules containing azGly and azGly-like fragments. After analyzing the structures of several small molecules in the Cambridge Structural Database (CSD),[13] we posit that the nitrogen atoms in azGly adopt a planar sp^2-like geometry similar to that of *N*-amidourea small molecules. This contrasts with the sp^3-like pyramidal geometry adopted by the nitrogen atoms in some semicarbazide small molecules. Representative structures of each from the CSD are presented for comparison.



Structures from the CSD containing the N-amidourea functional group:

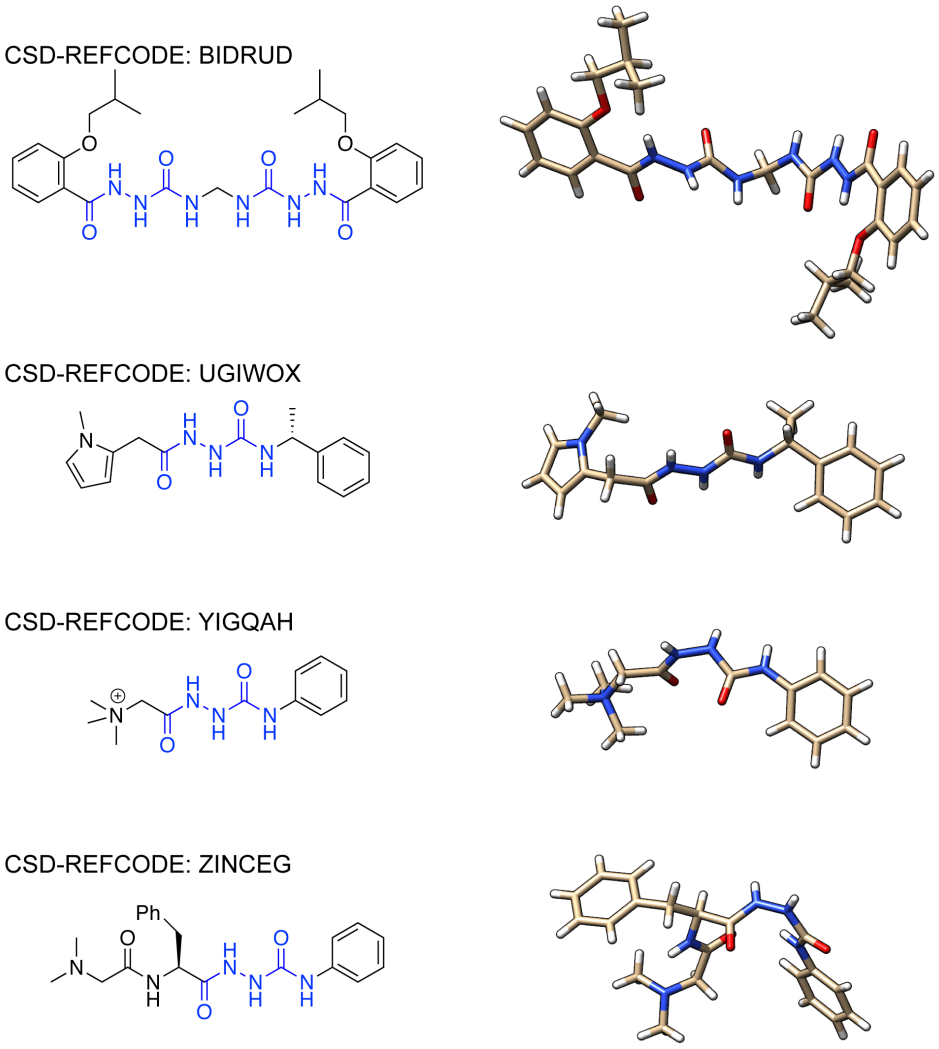
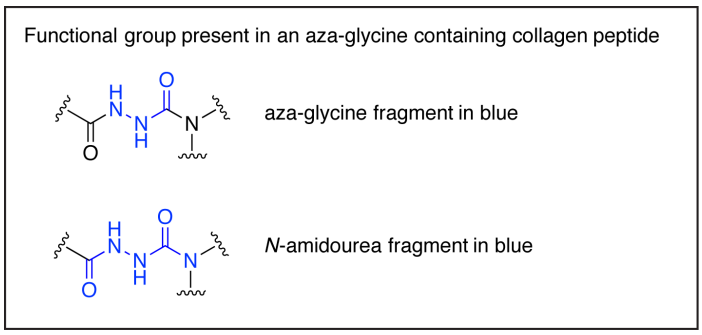
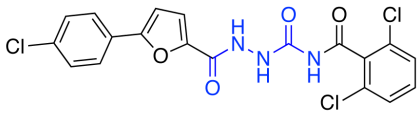


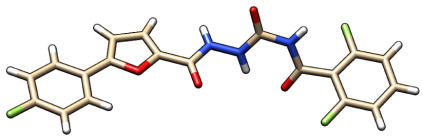
Figure S4, continued. Structural comparison of backbone conformations in molecules containing azGly and azGly-like fragments. After analyzing the structures of several small molecules in the Cambridge Structural Database (CSD),[13] we posit that the nitrogen atoms in azGly adopt a planar sp^2-like geometry similar to that of *N*-amidourea small molecules. This contrasts with the sp^3-like pyramidal geometry adopted by the nitrogen atoms in some semicarbazide small molecules. Representative structures of each from the CSD are presented for comparison.



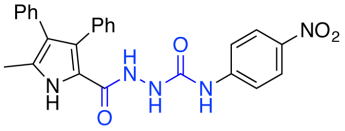
Structures from the CSD containing the N-amidourea functional group:

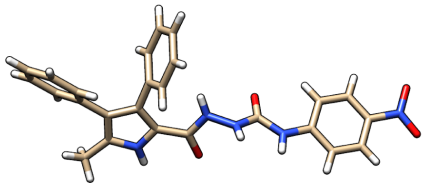
CSD-REFCODE: FUZQEX



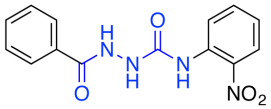


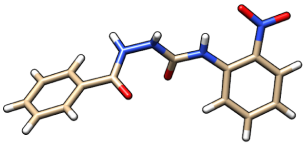
CSD-REFCODE: LECLEL



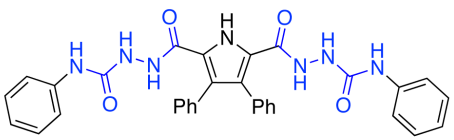


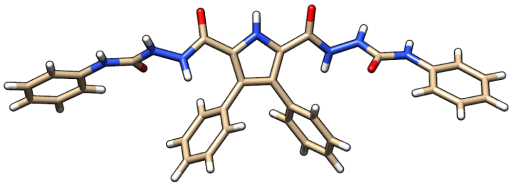
CSD-REFCODE: NUQFIP



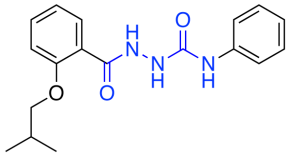


CSD-REFCODE: PEJKIZ





CSD-REFCODE: ULIXOD



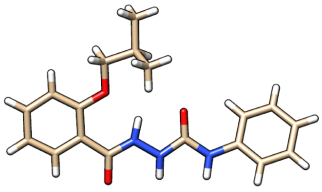
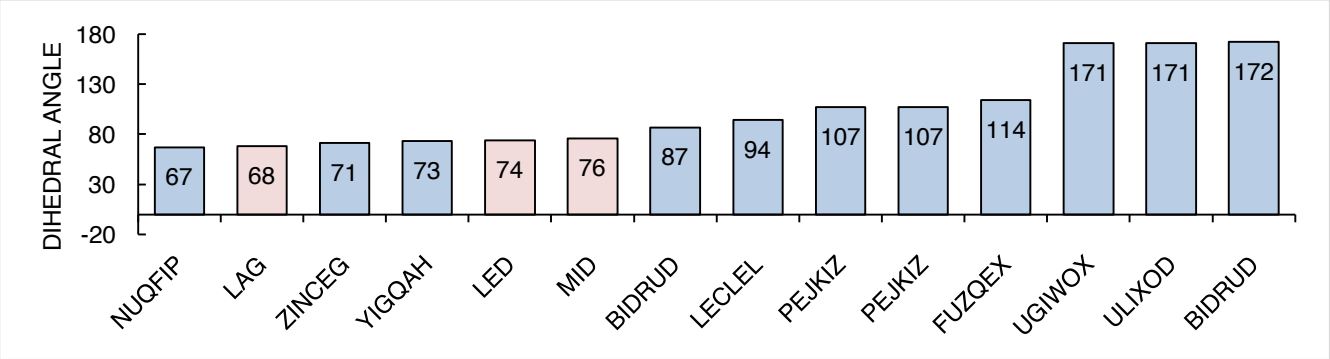


Figure S4, continued. Structural comparison of backbone conformations in molecules containing azGly and azGly-like fragments. After analyzing the structures of several small molecules in the Cambridge Structural Database (CSD),[13] we posit that the nitrogen atoms in azGly adopt a planar sp^2-like geometry similar to that of *N*-amidourea small molecules. This contrasts with the sp^3-like pyramidal geometry adopted by the nitrogen atoms in some semicarbazide small molecules. Representative structures of each from the CSD are presented for comparison.



		Pyramidalization	Pyramidalization	Pyramidalization
NAME	Dihedral	CO**N**NCONC	CON**N**CONC	CONNCO**N**C
NUQFIP	67	0.0	0.1	0.1
LAG	**68**	**0.0**	**0.0**	**0.0**
ZINCEG	71	0.0	0.0	0.0
YIGQAH	73	0.0	0.2	0.1
LED	**74**	**0.0**	**0.0**	**0.0**
MID	**76**	**0.0**	**0.0**	**0.0**
BIDRUD	87	0.0	0.0	0.0
LECLEL	94			
PEJKIZ	107			
PEJKIZ	107			
FUZQEX	114			
UGIWOX	171			
ULIXOD	171			
BIDRUD	172	0.0	0.0	0.0

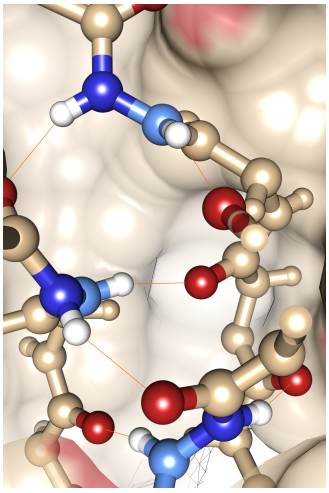


Figure S5. Comparison of dihedral angles and extent of pyramidalization in CMP 2 and *N*-amidourea-containing molecules. A) Dihedral angles. To further examine the backbone conformation of azGly-containing CMP **2**, the dihedral angles of several *N*-amidourea-containing molecules from Figure 2 are presented above for comparison with those of CMP **2**. The leading (LED), middle (MID), and lagging (LAG) strands of CMP **2** are in pink, while the *N*-amidourea-containing molecules are in blue. **B) Pyramidalization**. The extent of pyramidalization was determined by calculating the length of the normal vector between the plane of the dihedral angle and the peptide backbone (at the added nitrogen atom in azGly?). A comparison of the pyramidalization in the listed *N*-amidourea-containing molecules with those of the azGly-containing triplet in each of the three strands of CMP **2** reveals generally planar backbone geometry in all cases.

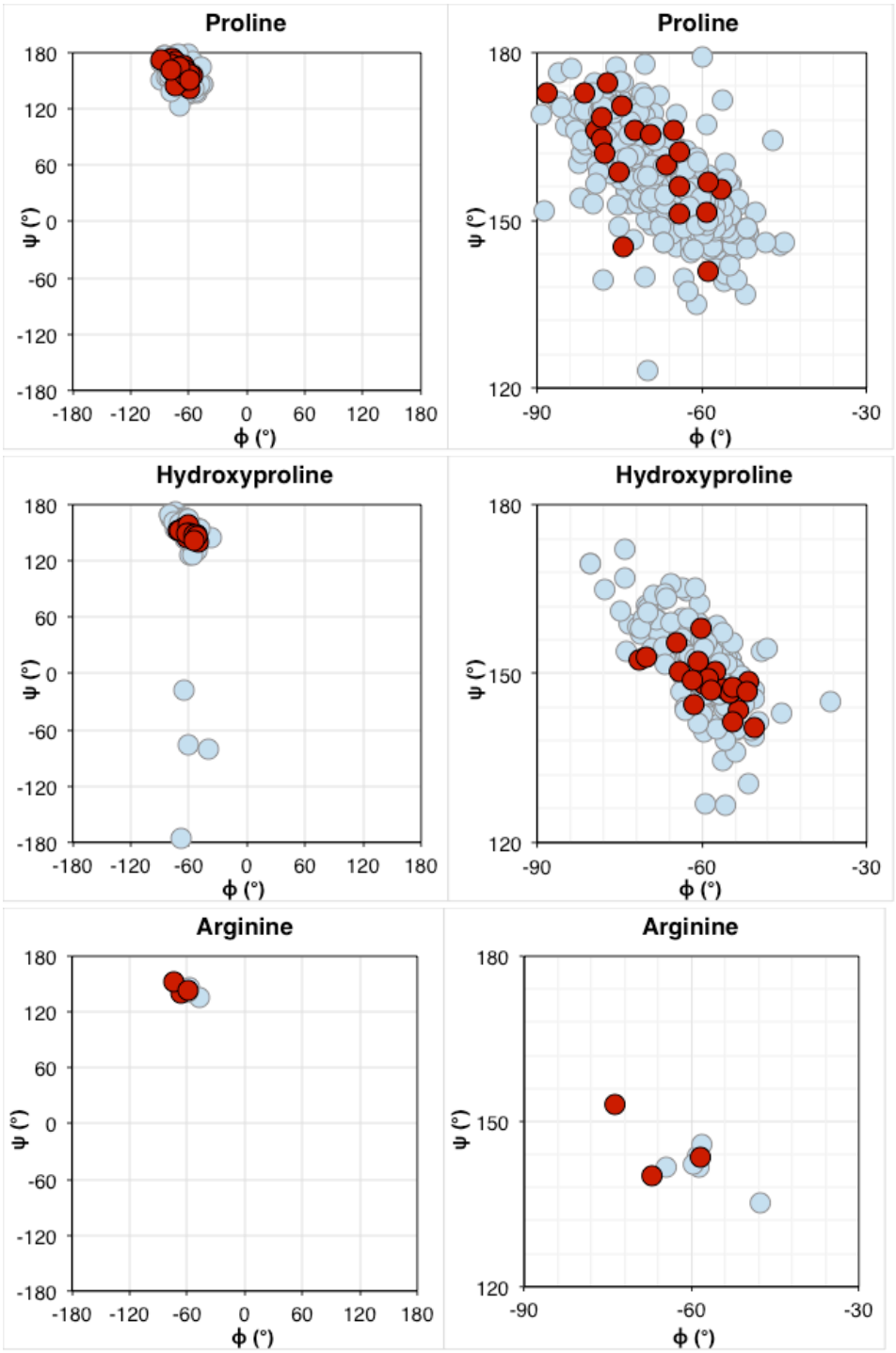


Figure S6. Dihedral angles of amino acid residues in collagen. Dihedral angles were calculated from structures in the PDB using UCSF Chimera.[12] Angles from aza-glycine-containing CMP **2** (PDB ID: 5K86) are shown in red, while data from all other collagen crystal structures are shown in light blue. Ramachandran plots on the right-hand side show a magnified view of the region of interest.

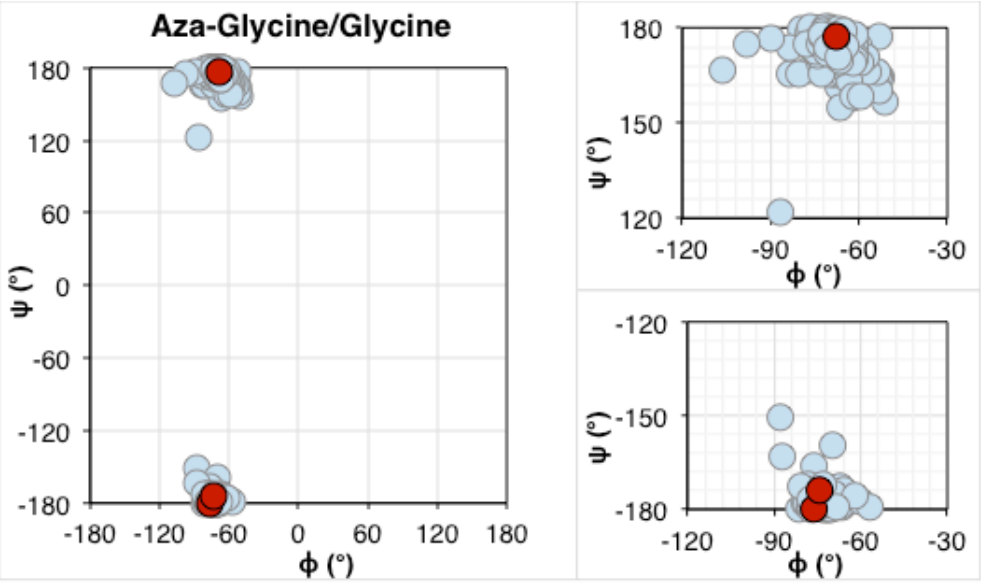


Figure S6, cont'd. Dihedral angles of amino acid residues in collagen. Dihedral angles were calculated from structures in the PDB using UCSF Chimera.[12] Angles from the aza-glycine residues in CMP **2** (PDB ID: 5K86) are shown in red, while data from the glycine residues in all other collagen crystal structures are shown in light blue. Ramachandran plots on the right-hand side show a magnified view of the regions of interest.

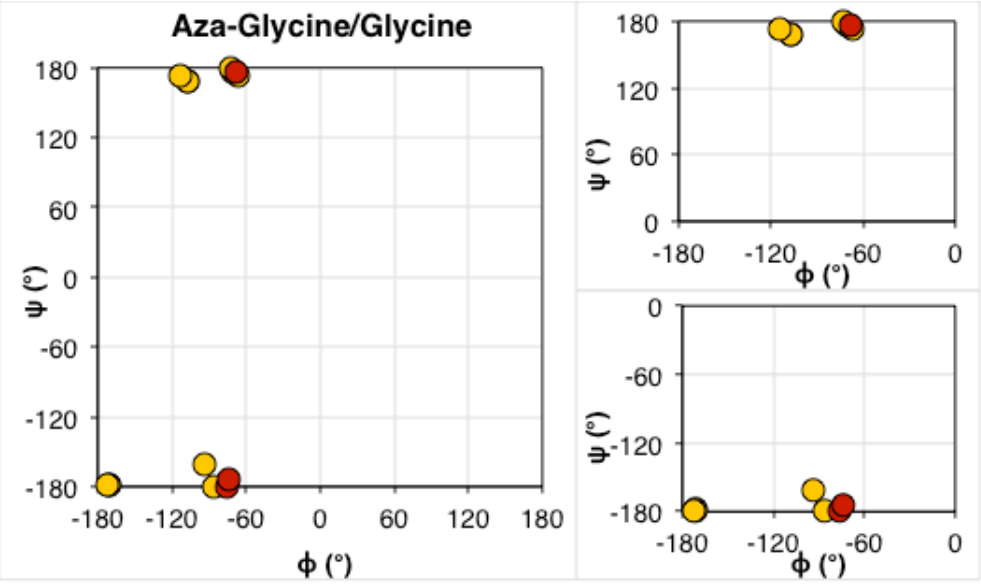


Figure S7. Comparison of dihedral angles in collagen and *N*-amidourea small molecules. Using structural data from the Cambridge Structural Database (CSD),[13] the dihedral angles of the aza-glycine residue in CMP **2** (red) were compared with those of the glycine residues in *N*-amidourea small molecules (yellow).

PDB ID	Reference
5K86	This Publication
3WN8	[5]
4Z1R	[14]
4AXY	[15]
3ADM	[16]
3ABN	[17]
3A1H	[16]
3A0A	To Be Published
3A0M	[16]
3A08	[18]
2DRX	[19]
1V4F	[20]
1V6Q	[20]
1V7H	[20]
1G9W	[21]
3A19	[18]
2DRT	[19]
1A3I	[22]
1A3J	[22]
1CGD	[23]
1CAG	[24]

Figure S8. PDB IDs for collagen crystal structures used in backbone analysis.

Hydrogen Bonding Analysis

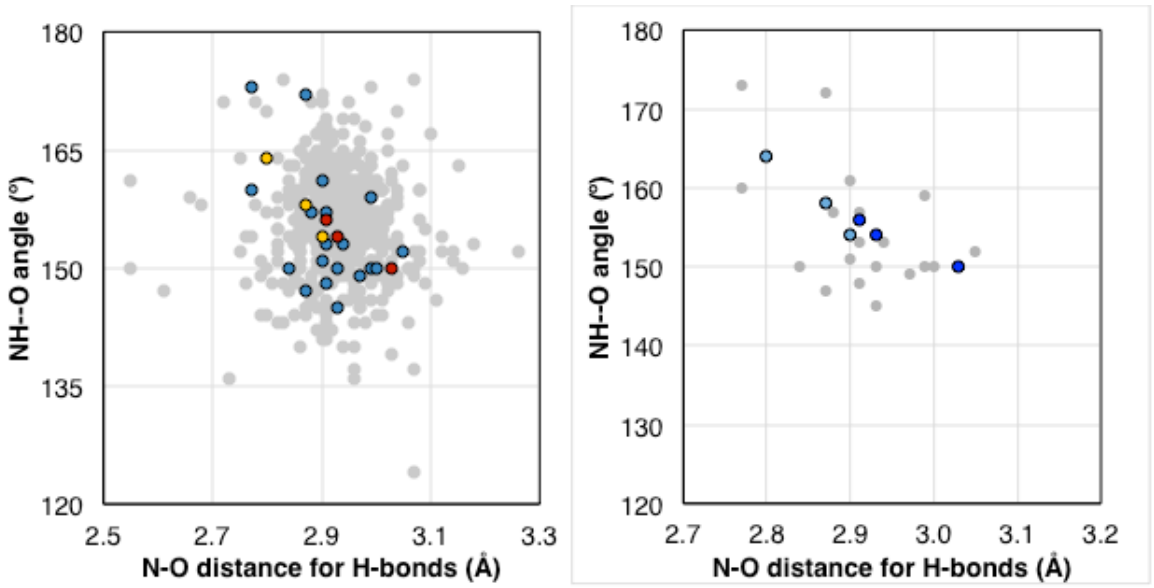


Figure S9. Hydrogen bond parameters in collagen crystal structures. Left) Plot of all H-bonds in CMP **2** in terms of N-H--O angle versus N-O distance. H-bond data for CMP **2** (colored) are overlaid on values for all collagen structures in the PDB (grey). **Right)** Plot of H-bonds in the aza-glycine residues of CMP **2** overlaid on H-bond data from the glycine residues for all collagen structures in the PDB. Colored markers represent data from CMP **2** (PDB 5K86) (Light blue = H-bond from nitrogen atoms already present in collagen, dark blue = H-bond from nitrogen atom added by azGly); grey markers represent H-bond data from all other non-azapeptide collagen crystal structures in the PDB.

C-terminus					
Chain B	C-terminus				
G	Chain A	C-terminus	H-Bond	C=O--H	N-H--O
O	G	Chain C	Distance (Å)	ω (°)	θ (°)
P	O	G	2.77	175.96	172.95
G	P	O	2.77	162.48	159.68
O	G	P	2.87	168.274	171.64
P	O	G	2.99	159.36	150.22
G	P	O	2.97	158.6	149.36
O	G	P	2.94	161.46	152.87
P	O	G	2.91	165.74	157.21
G	P	O	2.90	163.78	160.9
O	G	P	2.93	156.81	149.97
P	O	G	2.88	166.85	157.36
azG-1	P	O-1	2.93	147.51	144.65
R-1	**1-azG-2**	**P-1**	2.90	159.78	154.35
P-2	R-2-1	**2-azG-3**	2.80	155.81	164.5
G	P-3	R-3-2	2.87	160.18	158.28
O	G	P	2.91	156.39	148.01
P	O	G	3.05	161.15	151.83
G	P	O	2.91	159.67	152.72
O	G	P	3.00	156.88	150.17
P	O	G	2.84	154.24	149.71
G	P	O	2.90	159.45	150.87
O	G	P	2.87	155.57	147.14
P	O	G	2.99		
N-terminus	P	O	2.81		
	N-terminus	P			
		N-terminus			

Table 1. Characterization of hydrogen bonding in CMP 2. The characteristics of the backbone for each chain of the triple helix are presented in terms of H-bond distances and bond angles ω and θ. The pairs of amino acids in each chain are matched by color with their associated H-bond distances and Ψ and Θ angles; H-bond distances and angles associated with the central azGly-containing triplet are highlighted in red.

References

1. Zhang, Y.; Herling, M.; Chenoweth, D. M., General Solution for Stabilizing Triple Helical Collagen. *J. Am. Chem. Soc.* **2016,** *138* (31), 9751-9754.
2. Zhang, Y.; Malamakal, R. M.; Chenoweth, D. M., Aza-Glycine Induces Collagen Hyperstability. *J. Am. Chem. Soc.* **2015,** *137* (39), 12422-12425.
3. Zhang, Y.; Malamakal, R. M.; Chenoweth, D. M., A Single Stereodynamic Center Modulates the Rate of Self-Assembly in a Biomolecular System. *Angew. Chem. Int. Ed. Engl.* **2015,** *54* (37), 10826-10832.
4. Carpino, L. A.; Han, G. Y., 9-Fluorenylmethoxycarbonyl amino-protecting group. *J. Org. Chem.* **1972,** *37* (22), 3404-3409.
5. Okuyama, K.; Haga, M.; Noguchi, K.; Tanaka, T., Preferred Side-Chain Conformation of Arginine Residues in a Triple-Helical Structure. *Biopolymers* **2014,** *101* (10), 1000-1009.
6. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D. Struct. Biol.* **2011,** *67*, 235-242.
7. Kleywegt, G. J., Validation of protein models from Cα coordinates alone. *J. Mol. Biol.* **1997,** *273* (2), 371-376.
8. Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M. K.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R., New software for searching the Cambridge Structural Database and visualising crystal structures. *Acta Crystallogr. Sect. B: Struct. Sci.* **2002,** *58*, 389-397.
9. Taylor, R.; Macrae, C. F., Rules governing the crystal packing of mono- and di-alcohols. *Acta Crystallogr. Sect. B: Struct. Sci.* **2001,** *57*, 815-827.
10. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; Streek, J. v. d.; Wood, P. A., Mercury CSD 2.0 - New Features for the Visualization and Investigation of Crystal Structures *J. Appl. Crystallogr.* **2008,** *41*, 466-470.
11. Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; Streek, J. v. d., Mercury: visualization and analysis of crystal structures. *J. Appl. Crystallogr.* **2006,** *39*, 453-457.
12. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-1612.
13. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallogr. Sect. B: Struct. Sci.* **2016,** *72*, 171-179.
14. Plonska-Brzezinska, M. E.; Bobrowska, D. M.; Sharma, A.; Rodziewicz, P.; Tomczyk, M.; Czyrkoa, J.; Brzezinski, K., Triple helical collagen-like peptide interactions with selected polyphenolic compounds. *RSC Adv.* **2015,** *5*, 95443-95453.
15. Widmer, C.; Gebauer, J. M.; Brunstein, E.; Rosenbaum, S.; Zaucke, F.; Drögemüller, C.; Leeb, T.; Baumann, U., Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (33), 13243-13247.
16. Okuyama, K.; Miyama, K.; Morimoto, T.; Masakiyo, K.; Mizuno, K.; Bächinger, H. P., Stabilization of triple-helical structures of collagen peptides containing a Hyp-Thr-Gly, Hyp-Val-Gly, or Hyp-Ser-Gly sequence. *Biopolymers* **2011,** *95* (9), 628-640.
17. Okuyama, K.; Kawaguchi, T.; Shimura, M.; Noguchi, K.; Mizuno, K.; Bächinger, H. P., Crystal structure of the collagen model peptide (Pro-Pro-Gly)4-Hyp-Asp-Gly-(Pro-Pro-Gly)4 at 1.0 Å resolution. *Biopolymers* **2013,** *99* (7), 436-447.
18. Okuyama, K.; Morimoto, T.; Narita, H.; Kawaguchi, T.; Mizuno, K.; Bächinger, H. P.; Wu, G.; Noguchi, K., Two crystal modifications of (Pro-Pro-Gly)4-Hyp-Hyp-Gly-(Pro-Pro-Gly)4 reveal the puckering preference

of Hyp(X) in the Hyp(X):Hyp(Y) and Hyp(X):Pro(Y) stacking pairs in collagen helices. *Acta Crystallogr. Sect. D. Struct. Biol.* **2010,** *66*, 88-96.
19. Okuyama, K.; Narita, H.; Kawaguchi, T.; Noguchi, K.; Tanaka, Y.; Nishino, N., Unique Side Chain Conformation of a Leu Residue in a Triple-Helical Structure. *Biopolymers* **2007,** *86* (3), 212-221.
20. Okuyama, K.; Hongo, C.; Fukushima, R.; Wu, G.; Narita, H.; Noguchi, K.; Tanaka, Y.; Nishino, N., Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: Implications for proline ring puckering. *Biopolymers* **2004,** *76* (5), 367-377.
21. Vitagliano, L.; Berisio, R.; Mazzarella, L.; Zagari, A., Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **2001,** *58* (5), 459-464.
22. Kramer, R. Z.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H. M., X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* **1998,** *280* (4), 623-638.
23. Bella, J.; Brodsky, B.; Berman, H. M., Hydration structure of a collagen peptide. *Structure* **1995,** *3* (9), 893-906.
24. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M., Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **1994,** *266* (5182), 75-81.