General Solution for Stabilizing Triple Helical Collagen

Yitao Zhang, Madison Herling, and David M. Chenoweth*

Department of Chemistry, University of Pennsylvania, 231 South 34th Street,

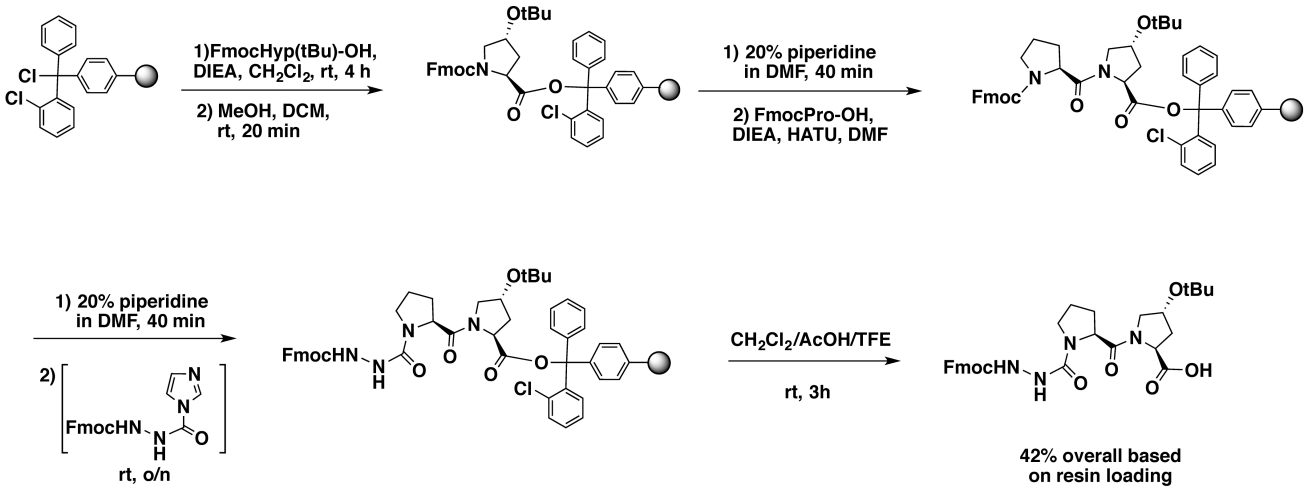
Philadelphia, Pennsylvania 19104-6323, United States

Table of contents

General information

All commercial reagents and solvents were used as received. Fmoc-Pro-OH, Fmoc-Gly-OH, HATU and Rink Amide AM Resin (100-200 mesh) were purchased from Novabiochem. Fmoc-Hyp(tBu)-OH was purchased from Advanced Chemtech. Piperidine was purchased from American Bioanalytical. All remaining chemicals were purchased from Sigma Aldrich. Flash column chromatography was performed using Silicycle silica gel (55–65 Å pore diameter) or using an ISCO flash chromatography system. Thin-layer chromatography was performed on Sorbent Technologies silica plates (250 μm thickness). High-resolution mass spectra were obtained at the University of Pennsylvania's Mass Spectrometry Service Center on a Micromass AutoSpec electrospray/chemical ionization spectrometer. Molecular masses of long peptides were obtained via a Bruker Ultraflex III Matrix-assisted laser desorption/ionization (MALDI) mass spectrometer. Ultraviolet absorption spectrophotometry was performed on a JASCO V-650 spectrophotometer with a PAC-743R multichannel Peltier using quartz cells with a 1 cm cell path length. High performance liquid chromatography analysis was performed using a Jasco HPLC instrument equipped with a Phenomenex column (Luna 5u C18(2) 100A; 250 × 4.60 mm, 5 μm). Circular dichroism experiments were performed with a Jasco J-1500 CD Spectrometer with a 6-cell holder.

Experimental procedures on small molecule



Scheme S1. Synthesis of Fmoc-azGly-Pro-Hyp(tBu)-OH.

Fmoc-azGly-Pro-Hyp(tBu)-OH

For the synthesis of Fmoc-azGly-Pro-Hyp(tBu)-OH, 2-chlorotrityl chloride resin was used. To a solution of Fmoc-Hyp(tBu)-OH (5 g, 12 mmol) in anhydrous CH_2Cl_2 (62 mL), 2-chlorotrityl chloride resin (5 g, 8.5 mmol, 1.7 mmol/g) and DIEA (2.12 mL, 12.2 mmol) were added under nitrogen. After agitating the mixture for 10 minutes, additional DIEA (3.19 mL, 18.4 mmol) was added. After agitating for ~4 hours, HPLC grade methanol (18 mL) was added to cap any remaining reactive trityl groups. After 15-20 minutes, the reaction mixture was filtered, and the resin was washed with CH_2Cl_2 (6 x 50 mL) and left to air-dry overnight. The loading was measured according to a known protocol[1] and found to be 0.75 mmol/g. The total mass obtained from the batch was 5.27 g (3.94 mmol).

The Hyp(tBu)-loaded 2-chlorotrityl chloride resin (3.73 g, 2.85 mmol, 1 equiv) was suspended in DMF with stirring (~30 min). After draining the DMF used to swell the resin, the base labile Fmoc-protecting group was removed with 20% piperidine in DMF (30 mL, 20 min, twice). The resin was washed with DMF (30 mL x 6). After deprotection, a solution of Fmoc-Pro-OH (2.88 g, 8.55 mmol),

HATU (3.25 g, 8.55 mmol), and DIEA (3.0 ml, 18.31 mmol) in DMF (31 mL) was added to the resin after 5 minutes of activation. After 4 hours, the solution was drained, and the resin was washed with DMF (6x). The base labile Fmoc-protecting group was cleanly removed with 20% piperidine in DMF (30 mL, 20 min, twice). The solution was drained, and the resin was washed with DMF (6x).

After deprotection, a solution of FmocNH-NH_2 (synthesized following a literature precedent[2]) (2.17 g, 8.55 mmol) and carbonyldiimidazole (1.39 g, 8.55 mmol) in 30 mL of DMF was prepared (5-10 minutes of stirring FmocNH-NH_2 with carbonyldiimidazole at room temperature is sufficient to convert FmocNH-NH_2 to the acyl-activated form, which is the electrophile in the next coupling on solid phase). This solution was added to the aforementioned Pro-Hyp(tBu)-Resin, and the mixture was allowed to stir overnight at room temperature. The solution was drained, and another fresh batch of carbonyldiimidazole-activated Fmoc-hydrazine in DMF was added to the solid phase vessel, and the stirring was kept for another 6-12 hours at room temperature.

The resin was then collected via filtration, rinsed with DMF (1 X), CH_2Cl_2 (6 X), and dried under vacuum for 6 hours. The dried resin was treated with an 80% CH_2Cl_2 (150 mL), 10% AcOH (15 mL), and 10% TFE (15 mL) cocktail solution for 3 hours at room temperature. The mixture was filtered through cotton, and the filtrate was concentrated *in vacuo*. AcOH was removed by azeotroping with C_6H_6 (3 x 120 mL). The resulting foamy solid residue was purified by silica gel column chromatography (0% - 5% MeOH in CH_2Cl_2). The isolated yield for this compound was 670 mg, which corresponds to a 42% yield based on determined Hyp(tBu)-loaded trityl resin active-sites.

TLC (10% MeOH in CH_2Cl_2) R*f* = 0.09.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.77-7.00 (m, 10H), 4.62-4.22 (m, 5H), 4.22-3.31 (m, 5H), 2.65-1.77 (m, 6H), 1.11 (s, 9H).

^{13}C NMR (125 MHz, $CDCl_3$) δ 173.8, 173.0, 157.6, 156.9, 143.7, 141.1, 127.6, 127.0, 125.3, 125.2, 119.8, 74.2, 69.5, 67.7, 58.4, 58.0, 53.1, 46.8, 45.9, 36.0, 28.1, 24.6.

HRMS (ESI) calculated for $C_{30}H_{36}N_4O_7$ $[M+Na]^+$ 587.2476, found 587.2485.

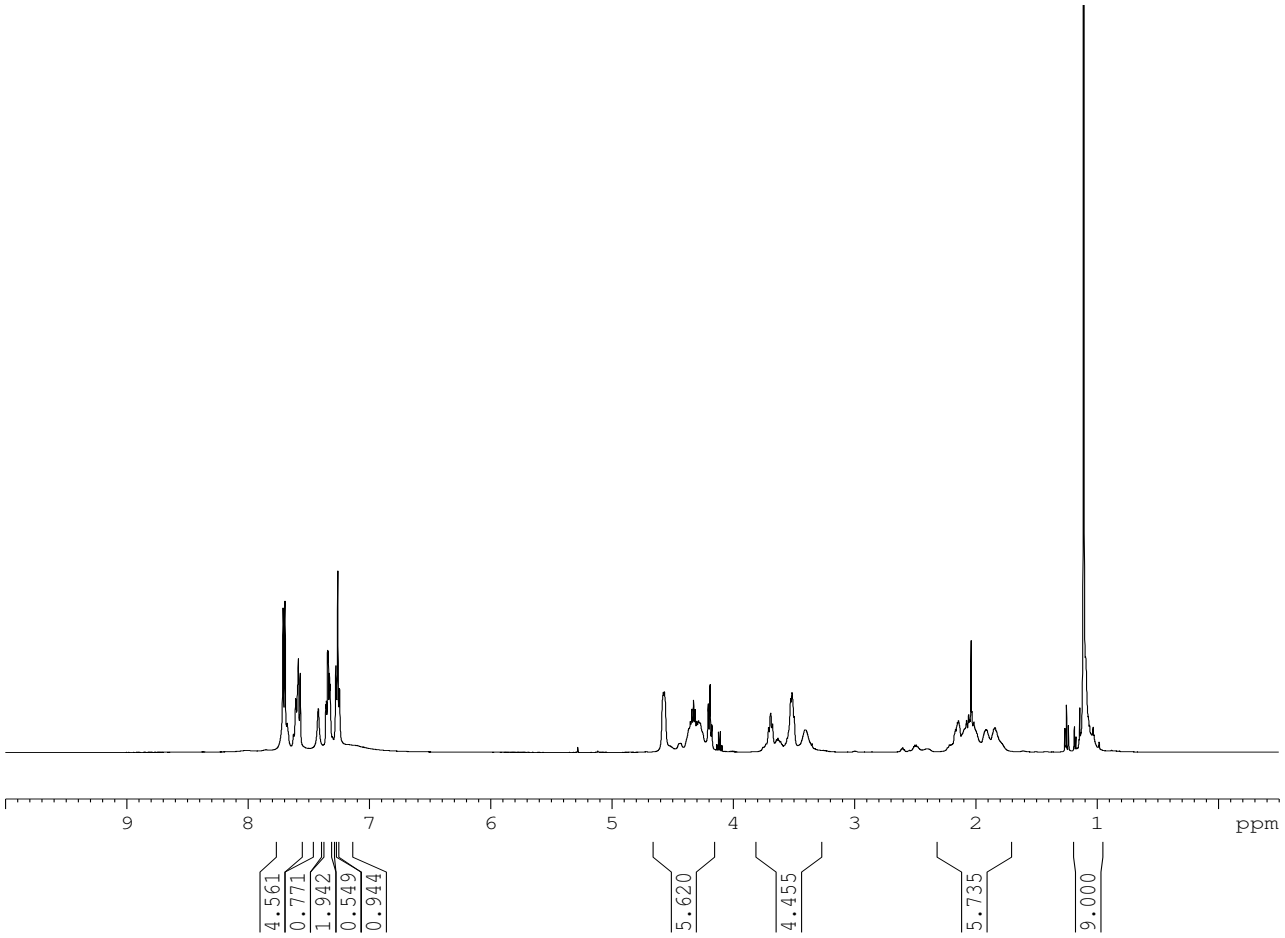


Figure S1 ^{1}H NMR of Fmoc-azGlyProHyp(tBu)-OH

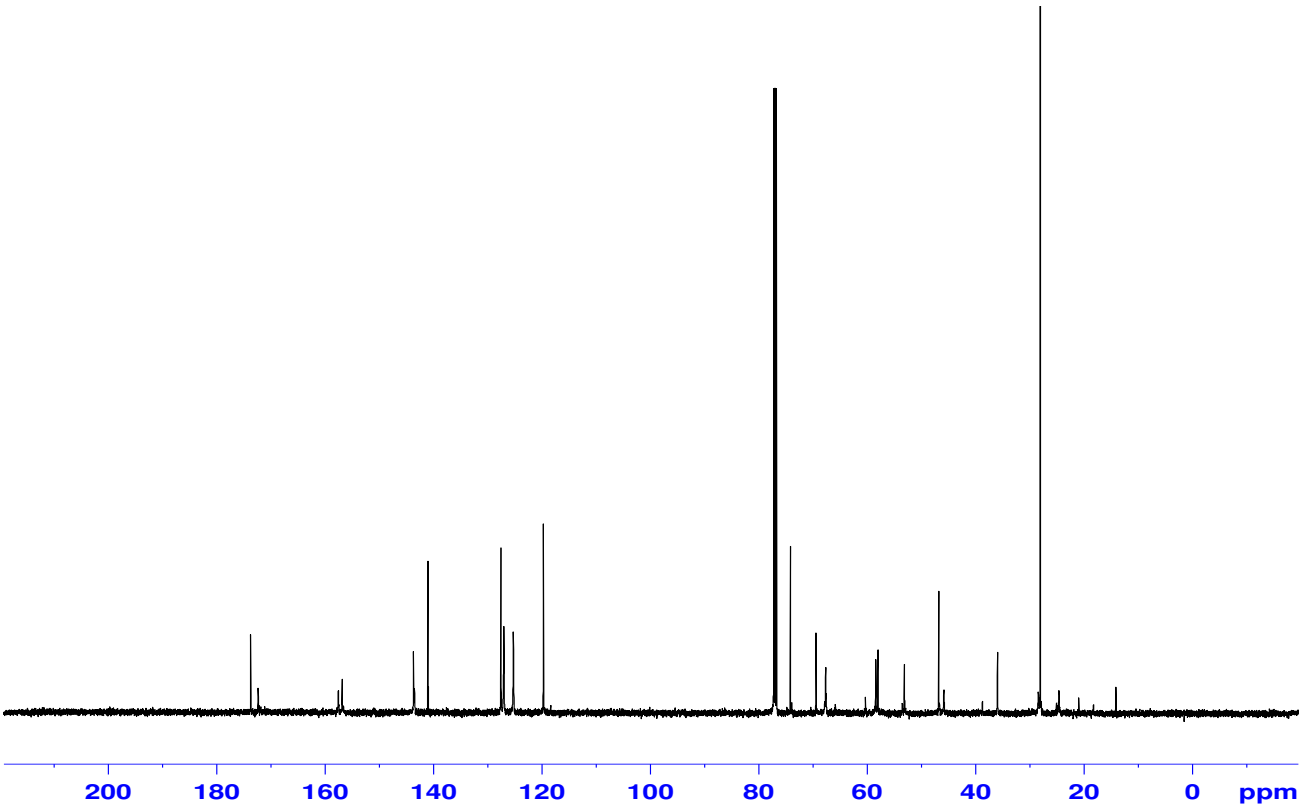


Figure S2 ^{13}C NMR of Fmoc-azGlyProHyp(tBu)-OH

Synthesis and Purification of Collagen Model Peptides

General protocols

Protocol A – Resin preparation followed by Fmoc-deprotection

The peptides used in this paper were synthesized by manual SPPS method with Rink Amide AM resin (0.54 mmol/g) on a 0.02-mmol scale. Rink Amide resin (37 mg) was carefully weighed out and transferred to a 5 mL solid phase synthesis vessel. The resin was then swelled in 4 mL of DMF for 30 min. After draining, the resin was treated with piperidine in DMF (20% v/v) (1 mL) at ambient temperature with stirring for 15 minutes (2x). The solution was drained, and the resin was washed with DMF (6x). An alternative deprotection protocol is shown at the end of this section (see below).

Protocol B – Oligopeptide building block coupling followed by Fmoc-deprotection

Fmoc-Xaa-Yaa-Gly-OH (3 equiv) and HATU (3 equiv) were dissolved in DMF (0.67 mL). DIEA (6 equiv) was subsequently added to the mixture, and the entire solution was transferred to the reaction vessel containing the resin. The mixture was stirred for 60-80 min, drained, and washed with DMF (6x). The resin was treated with piperidine in DMF (20% v/v) (1 mL) at ambient temperature with stirring for 15 minutes (2x). The solution was drained, and the resin was thoroughly washed with DMF (6x).

Protocol B' – Single amino acid coupling followed by Fmoc-deprotection

Fmoc-Aa-OH (3 equiv) and HATU (3 equiv) were dissolved in DMF (0.67 mL). DIEA (6 equiv) was subsequently added to the mixture, and the entire solution was transferred to the vessel containing amino-functionalized resin. The mixture was stirred for 45-60 min, drained, and washed with DMF (6x). The resin was treated with piperidine in DMF (20% v/v) (1 mL) at ambient temperature with stirring for 15 minutes (2x). The solution was drained and the resin was washed with DMF (6x).

Protocol B" – Special coupling of FmocNH-NH_2 followed by Fmoc-deprotection

Carbonyldiimidazole (3 equiv) was dissolved in DMF (0.67 mL). FmocNH-NH_2 (3 equiv) was subsequently added to the solution. After 5-10 minutes activation at room temperature, the mixture was transferred to the vessel containing amino-functionalized resin. The mixture was stirred for overnight at room temperature, drained, and "recharged" with a fresh batch of in-situ generated FmocNH-NH_2/CDI solution. After another 6-10 hours, the solution was drained, and the resin was washed with DMF (6x). The resin was treated with piperidine in DMF (20% v/v) (1 mL) at ambient temperature with stirring for 15 minutes (2x). The solution was drained, and the resin was washed with DMF (6x).

Protocol B"' – Special coupling of FmocHyp(tBu)-OH (or Fmoc-azGly-Pro-Hyp(tBu)-OH) onto azGly followed by Fmoc-deprotection

Fmoc-Hyp(tBu)-OH (3 equiv) was dissolved in THF (0.67 mL). NMM (6 equiv) and isobutyl-chloroformate (3 equiv) were subsequently added to the solution. After 5-10 minutes at room

temperature, the mixture was transferred to the vessel containing amino-functionalized resin. The mixture was stirred for 5-6 h at room temperature, drained, and washed with DMF (6x). The resin was treated with piperidine in DMF (20% v/v) (1 mL) at ambient temperature with stirring for 15 minutes (2x). The solution was drained and the resin was washed with DMF (6x).

Protocol C– Acylation

A solution of NMM (0.15 mL) and Ac_2O (0.25 mL) in DMF (2.1 mL) were added to the amino-functionalized resin. This mixture was stirred for 20 minutes at room temperature (2x). The solution was drained and thoroughly washed with CH_2Cl_2 (6x).

Protocol D– Cleavage off the resin and collection of the crude product

The resin was suspended for 35 min in a 4-mL mixture of TFA/H_2O/TIPS (95:2.5:2.5) at room temperature. Argon was used to push the mixture through the filter of the vessel, and the filtrate was collected and dropwise added to cold Et_2O (~11 mL). The sample was cooled to 4 °C for approximately 1 h, during which white solid precipitated. The resulted mixture was centrifuged and the supernatant was decanted. The white solid was dissolved in water (HPLC grade) and refrigerated at 4 °C before HPLC purification.

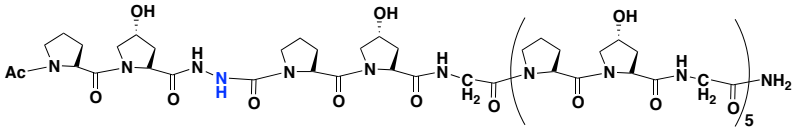
Protocol E– HPLC purification

MeCN (B) and water containing 0.1% TFA (A) were used as eluents. The flow rates used for semi-preparative HPLC and analytical HPLC were 4 mL/min and 1 mL/min, respectively. Crude samples were heated to 80 °C for 10-15 min before injection to prevent early triple helix formation.

An alternative Fmoc-deprotection method was used by treating the resin with a solution of 2% (v/v) DBU, 1% HOBt (m/v) in DMF. The resin is immersed in 1 mL of the above-mentioned solution for 1 minute followed by draining, and this process is repeated twice more (a total 3 doses of DBU/HOBt solution). When the growing peptide sequence was longer than nine amino acids, this method provided a slightly cleaner deprotection result compared to the traditional double treatment with piperidine/DMF solution.

Characterization of Collagen Model Peptides (CMPs).

CMP **3** Ac-(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_6$-NH_2

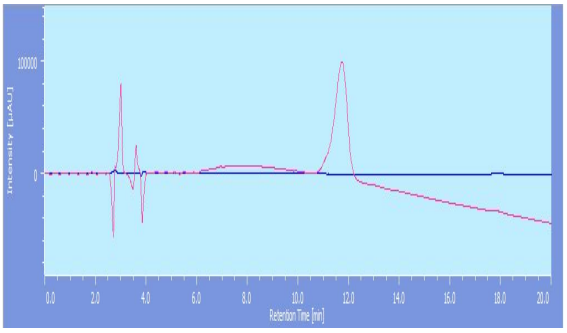


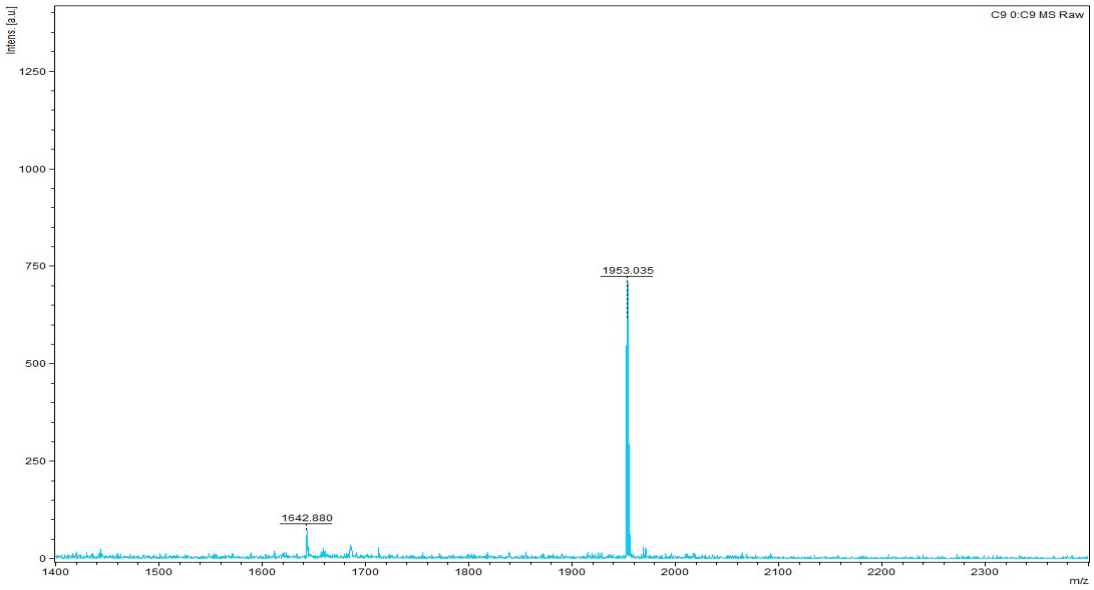
Synthesis: A + B(POG) + B(POG) + B(POG) + B(POG) + B(POG) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + C + D + E

Note: POG refers to the tripeptide building block FmocProHyp(tBu)Gly-OH. Its synthesis was previously described by Moroder et al.[3]

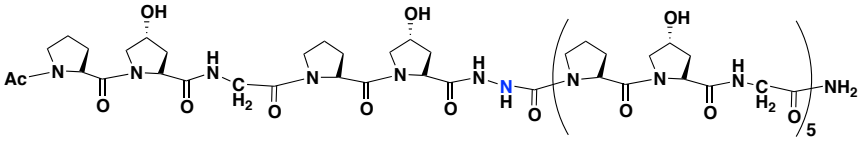
The synthesis of Fmoc-hydrazine, FmocNH-NH_2, was previously reported.[2]

MALDI-MS calculated $[M+Na]^+$ 1952.89, found 1953.03



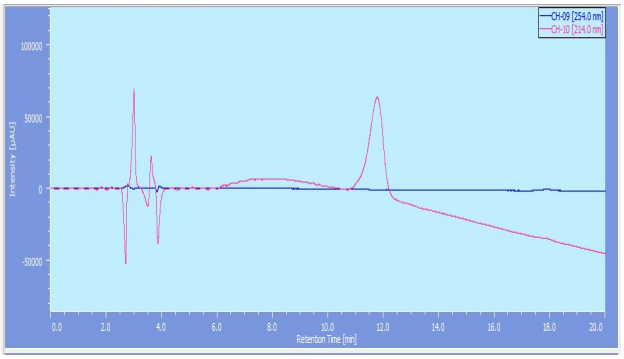


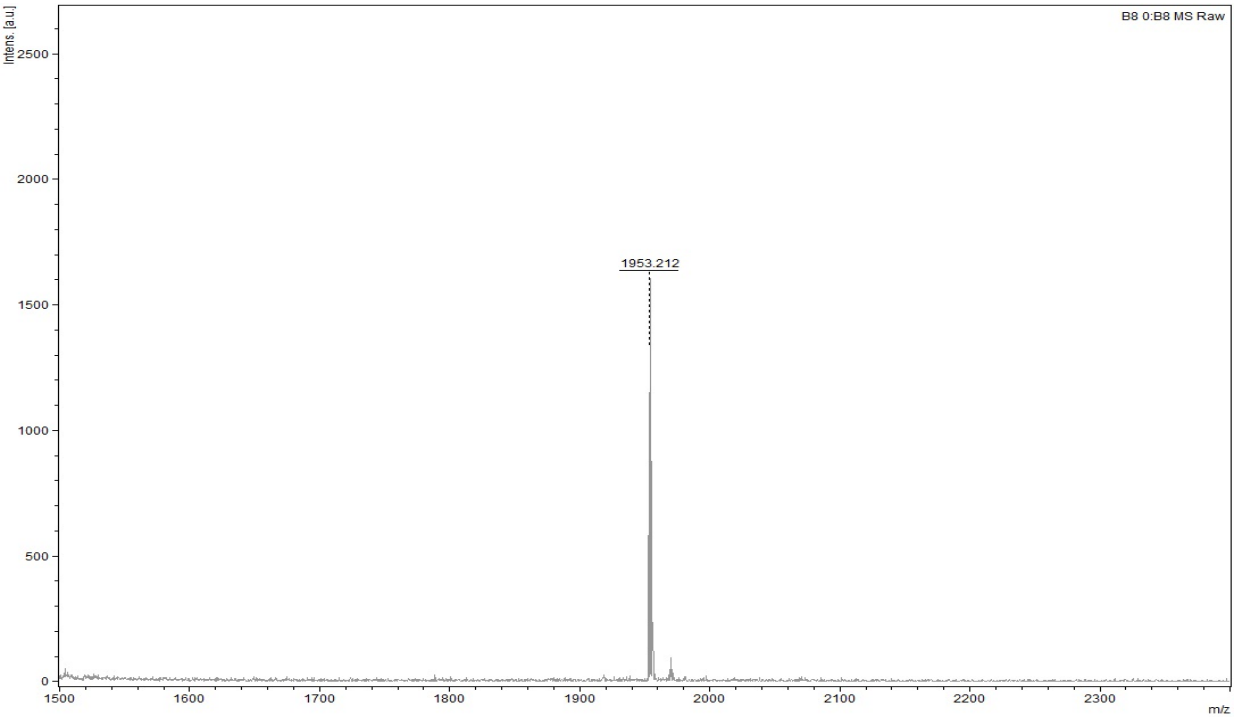
CMP **4** Ac-(Pro-Hyp-Gly)(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_5$-NH$_2$



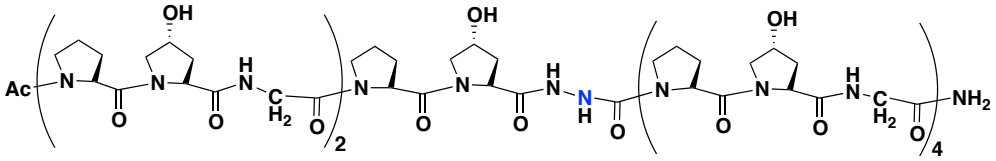
Synthesis: A + B(POG) + B(POG) + B(POG) + B(POG) + B(POG) + B''(FmocNH-NH$_2$) + B'''(O) + B'(P) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1952.89, found 1953.21



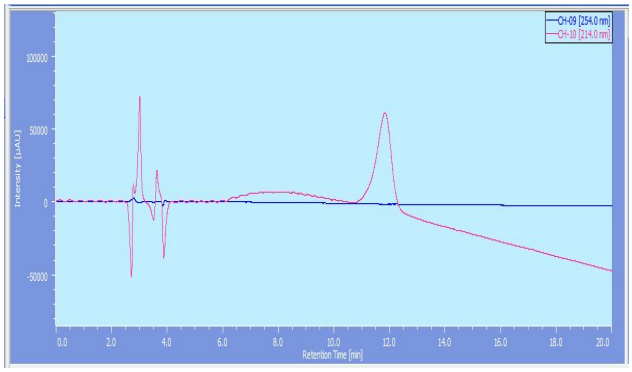


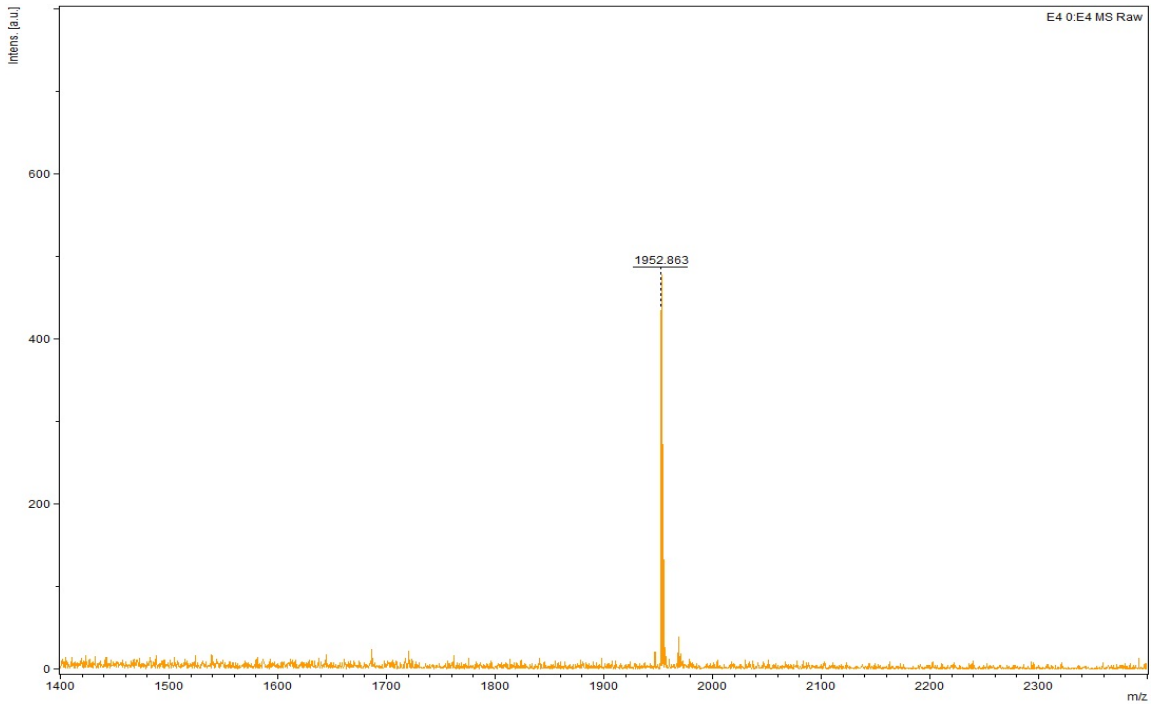
CMP **5** Ac-(Pro-Hyp-Gly)$_2$(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_4$-NH$_2$



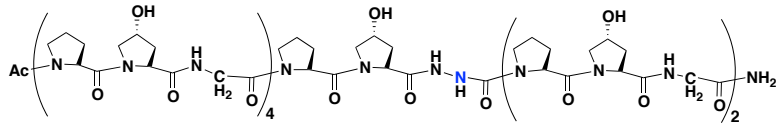
Synthesis: A + B(POG) + B (POG) + B(POG) + B(POG) + B''(FmocNH-NH$_2$) + B'''(O) + B'(P) + B(POG) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1952.88, found 1952.86



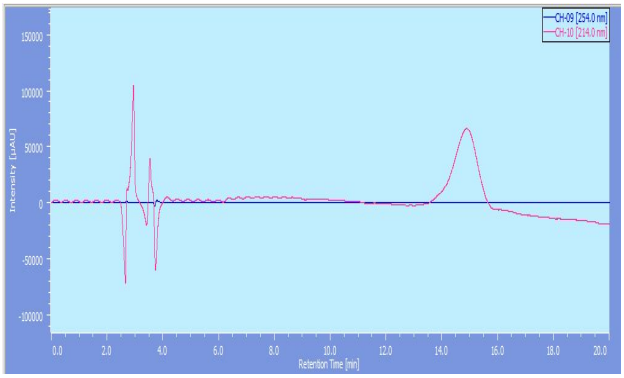


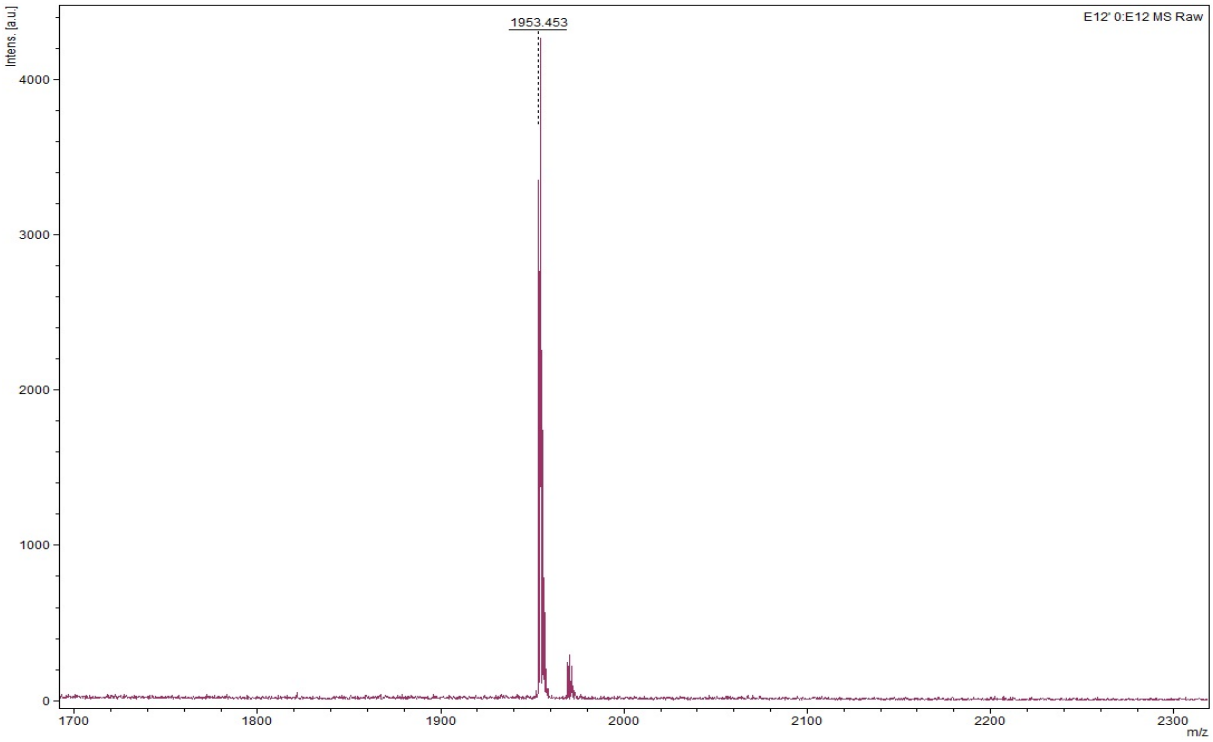
CMP **6** Ac-(Pro-Hyp-Gly)$_4$(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_2$-NH$_2$



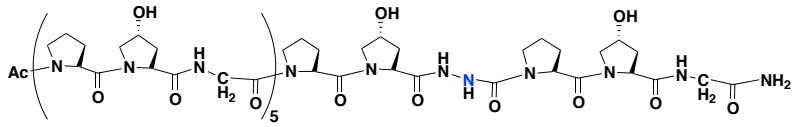
Synthesis: A + B(POG) + B(POG) + B''(FmocNH-NH$_2$) + B'''(O) + B'(P) + B(POG) + B(POG) + B(POG) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1952.89, found 1953.45



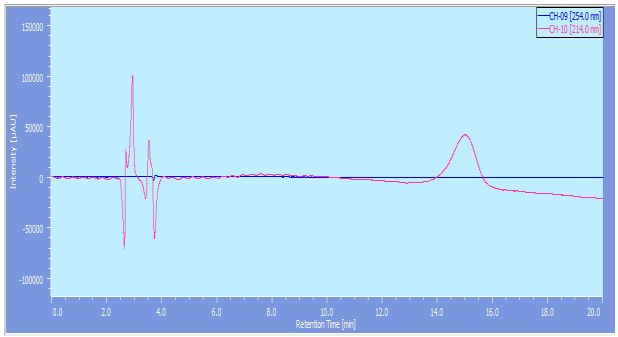


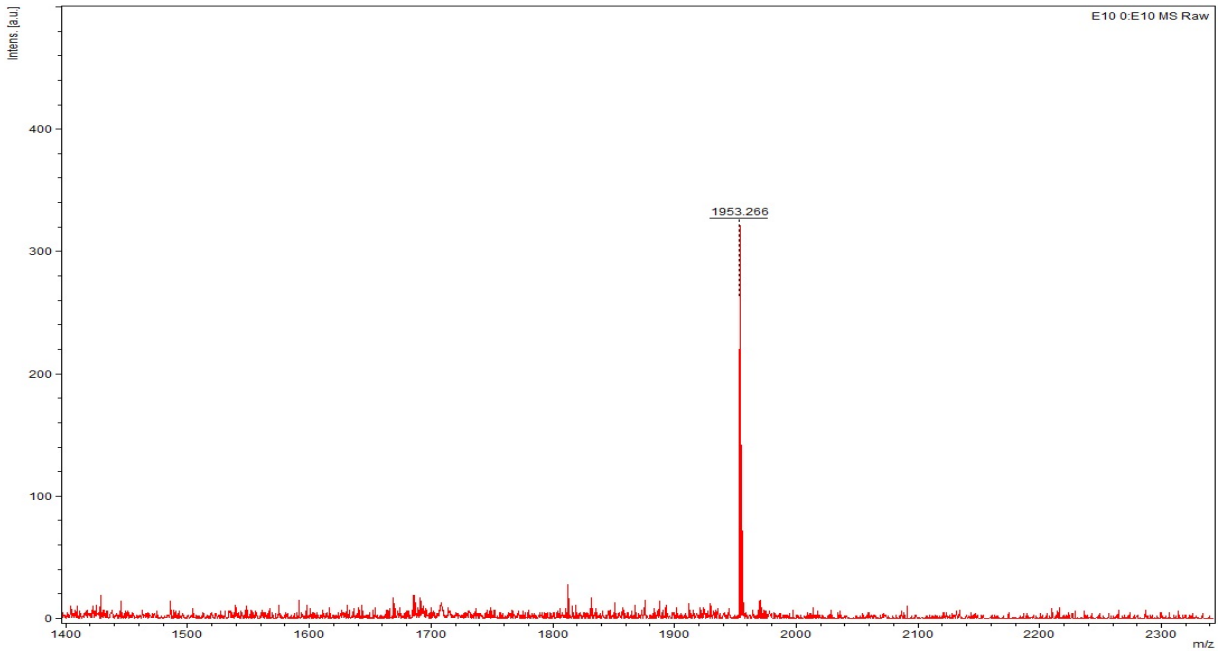
CMP **7** Ac-(Pro-Hyp-Gly)$_5$(Pro-Hyp-azGly)(Pro-Hyp-Gly)-NH_2



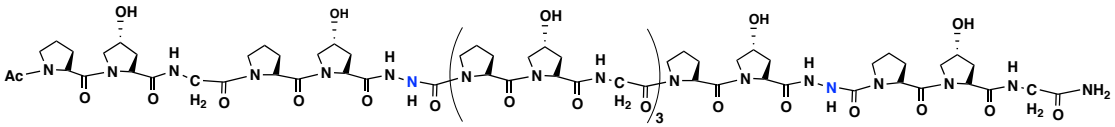
Synthesis: A + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B(POG) + B(POG) + B(POG) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1952.89, found 1953.27



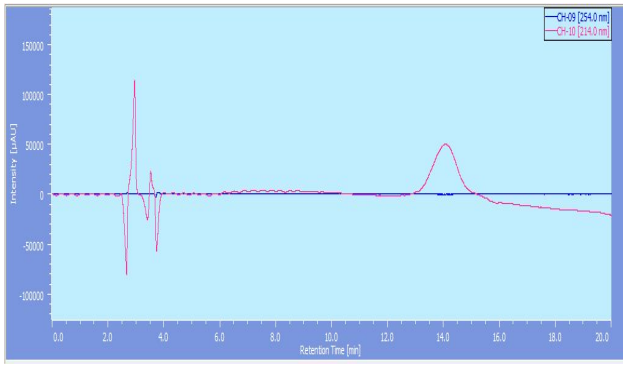


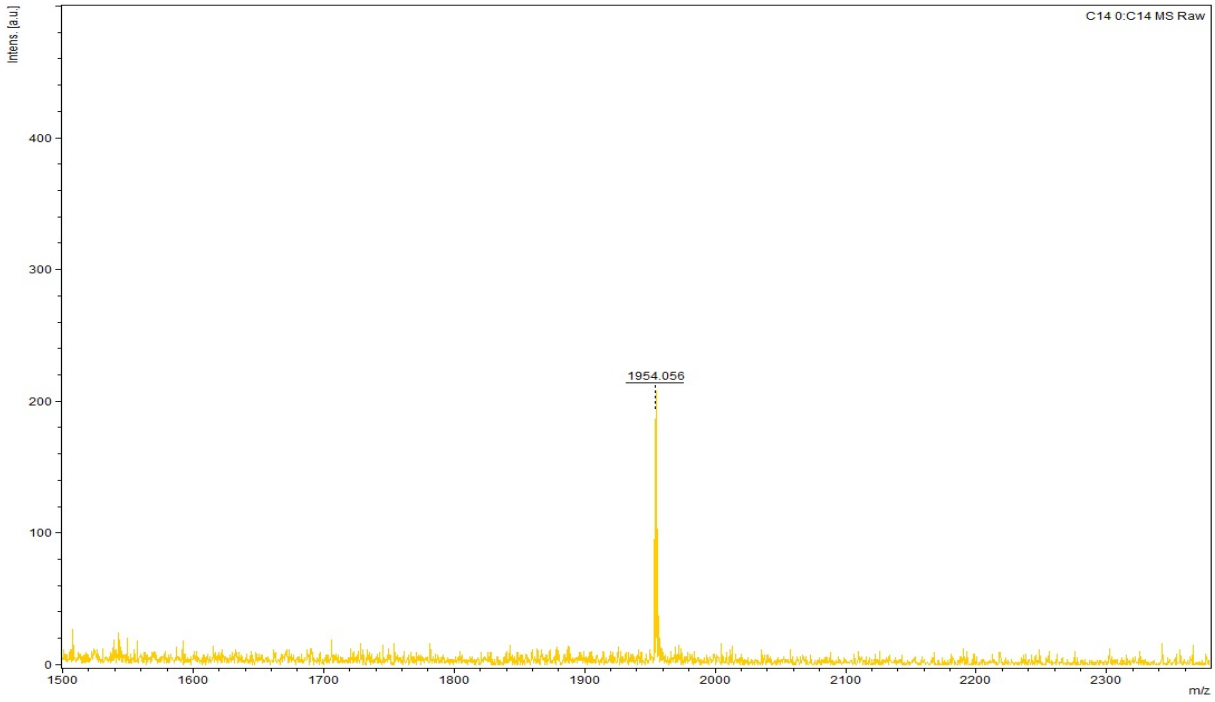
CMP **8** Ac-(Pro-Hyp-Gly)(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_3$(Pro-Hyp-azGly)(Pro-Hyp-Gly)-NH_2



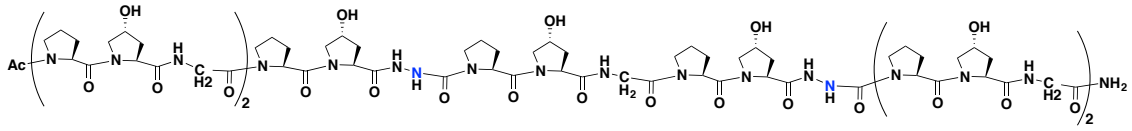
Synthesis: A + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B(POG) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1953.88, found 1954.06



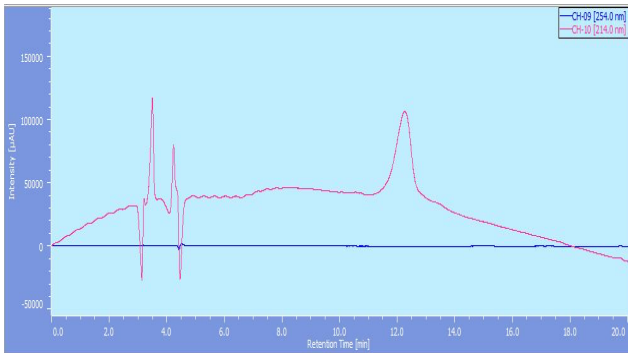


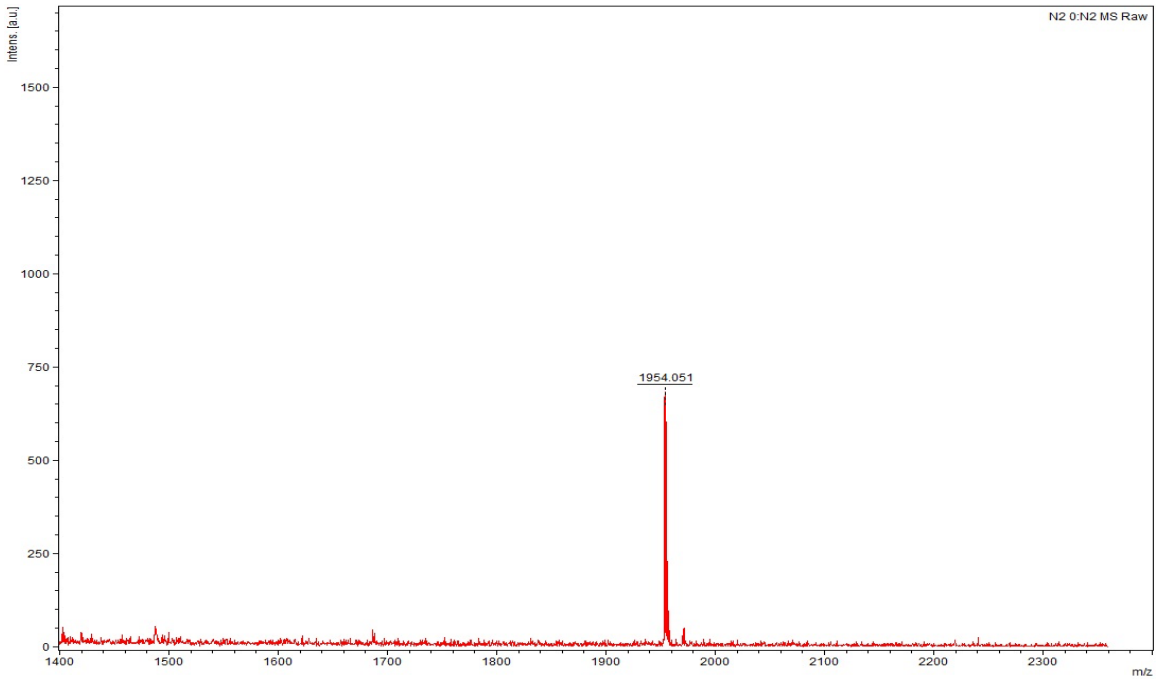
CMP **9** Ac-(Pro-Hyp-Gly)$_2$(Pro-Hyp-azGly)(Pro-Hyp-Gly)(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_2$-NH_2



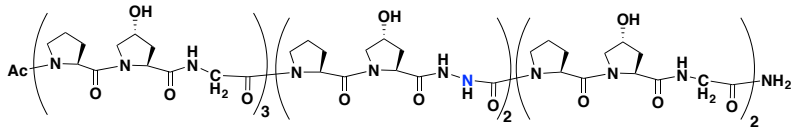
Synthesis: A + B(POG) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1953.87, found 1954.05



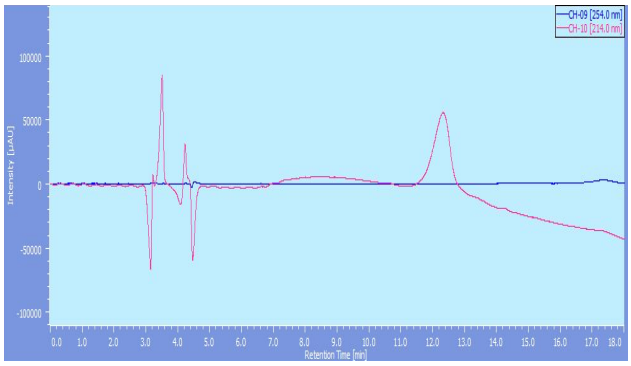


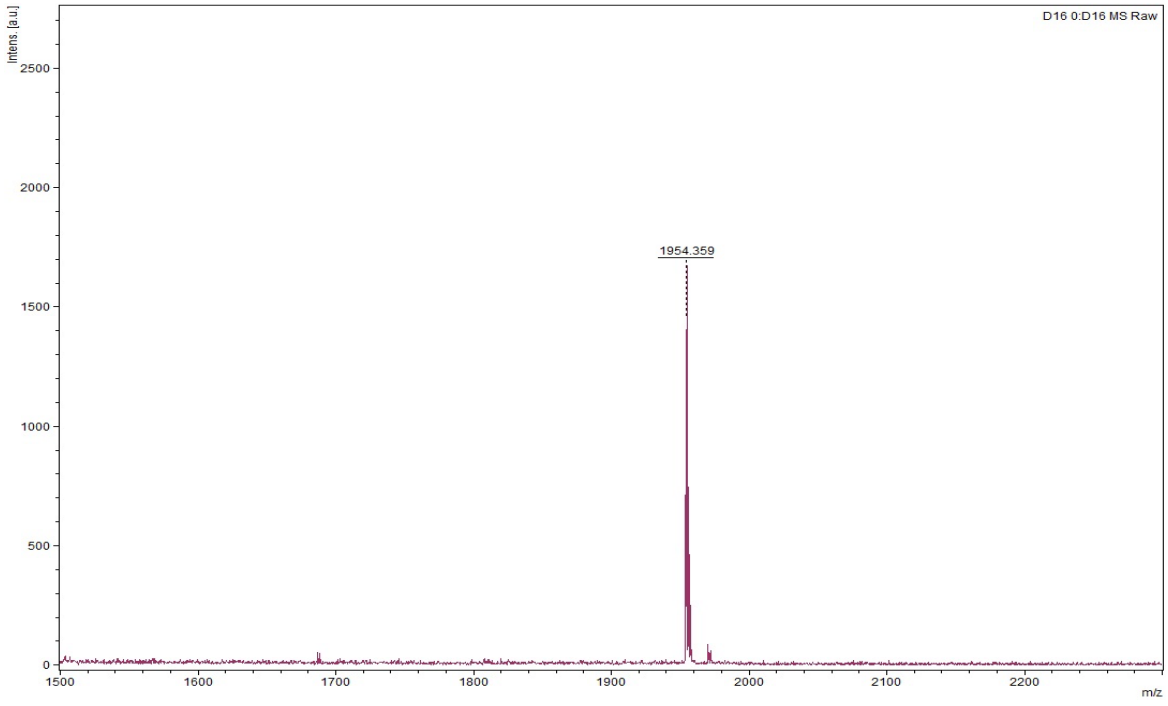
CMP **10** Ac-(Pro-Hyp-Gly)$_3$(Pro-Hyp-azGly)$_2$(Pro-Hyp-Gly)$_2$-NH_2



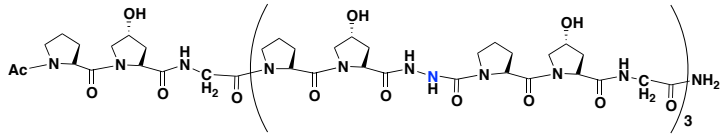
Synthesis: A + B(POG) +B(POG) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1953.87, found 1954.36



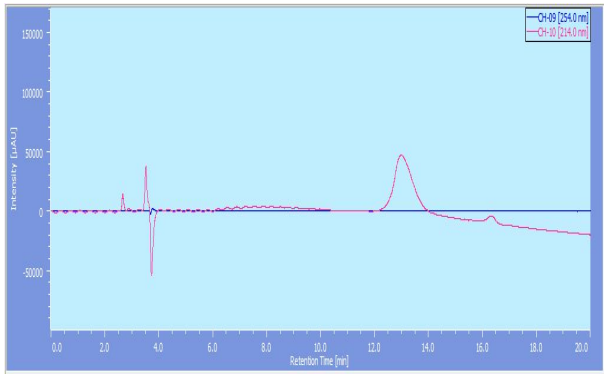


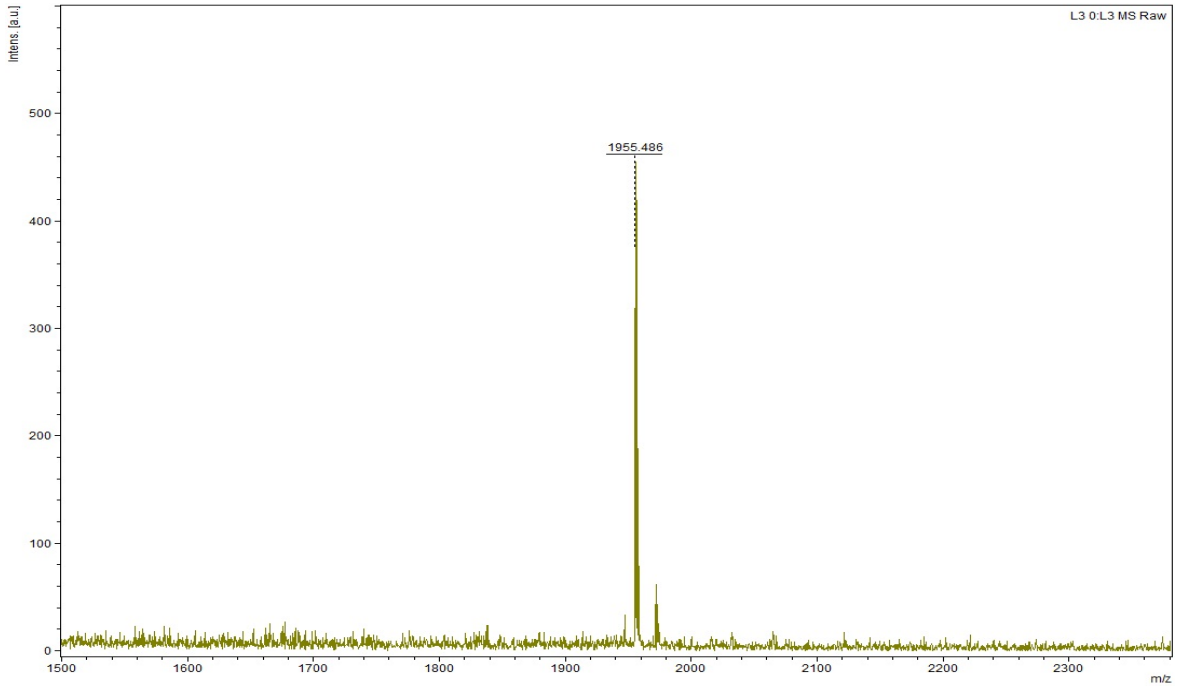
CMP **11** Ac-(Pro-Hyp-Gly)[(Pro-Hyp-azGly)(Pro-Hyp-Gly)]$_3$-NH_2



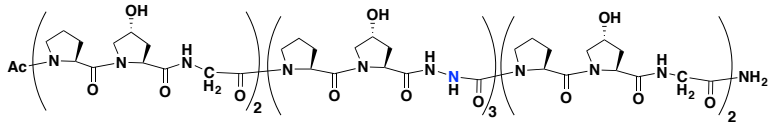
Synthesis: A + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + B''(FmocNH-NH_2) + B'''(O) + B'(P) + B(POG) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1954.87, found 1955.49





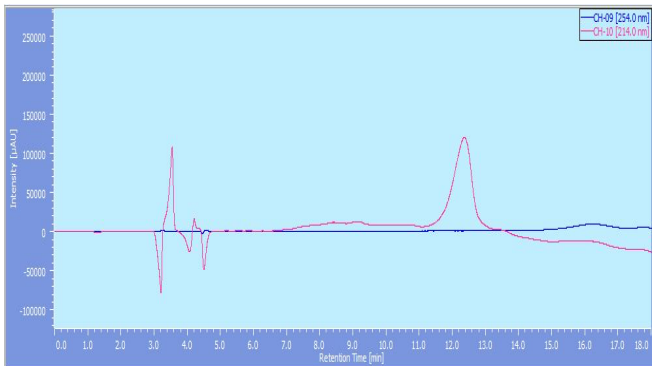
CMP **12** Ac-(Pro-Hyp-Gly)$_2$(Pro-Hyp-azGly)$_3$(Pro-Hyp-Gly)$_2$-NH_2

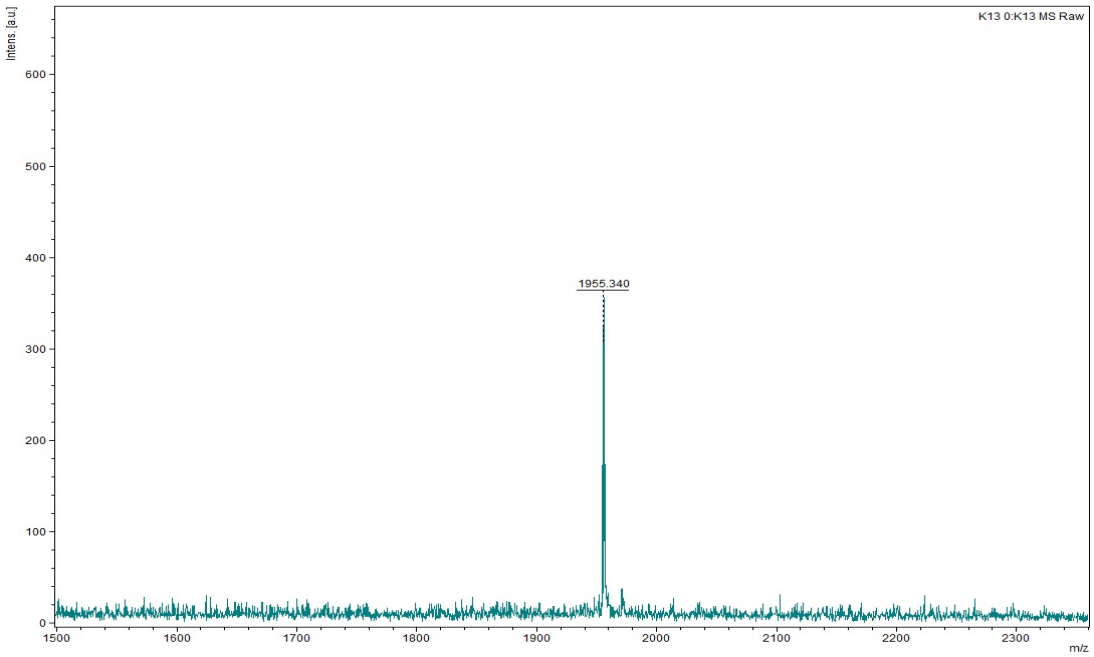


Synthesis: A + B(POG) + B'(G) + B'''(Fmoc-azGPO) + B'''(Fmoc-azGPO) + B'''(Fmoc-azGPO) + B'''(O) + B'(P) + B(POG) + C + D + E

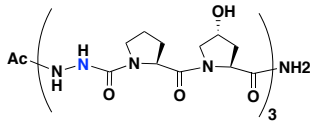
"Fmoc-azGPO" stands for Fmoc-azGly-Pro-Hyp(tBu)-OH, whose synthesis was described on page S3 and S4.

MALDI-MS calculated $[M+Na]^+$ 1954.87, found 1955.34



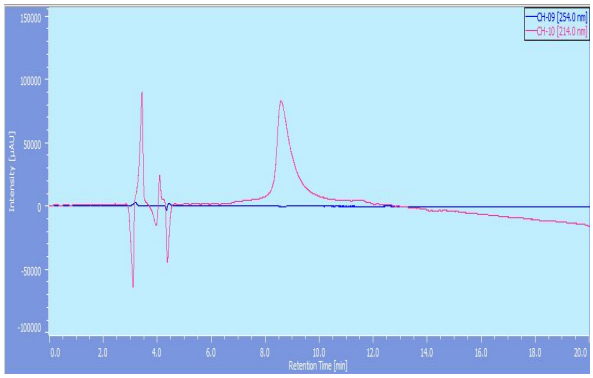


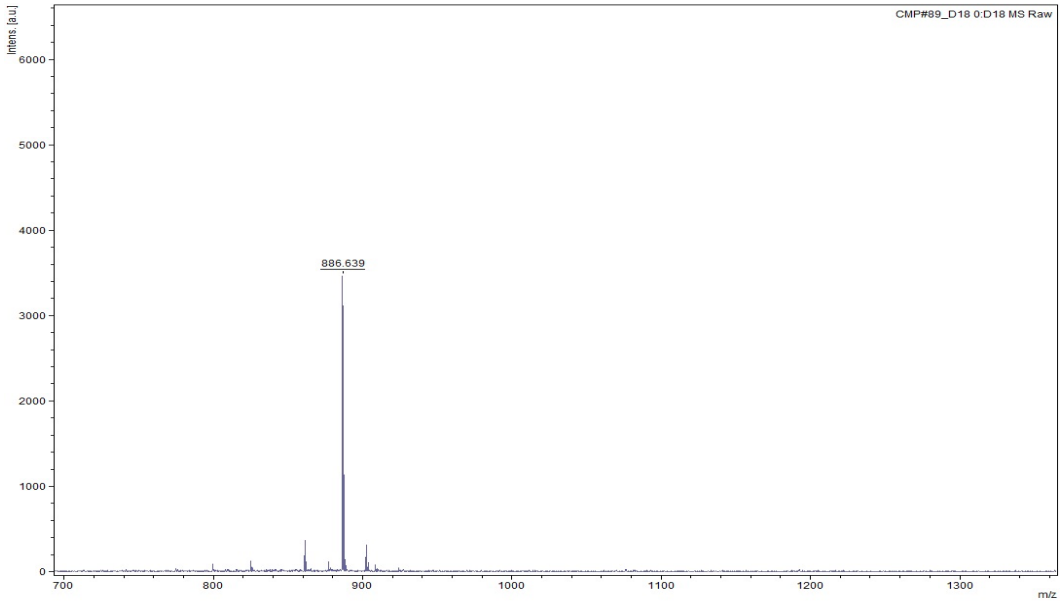
CMP **13** Ac-(azGly-Pro-Hyp)$_3$-NH$_2$



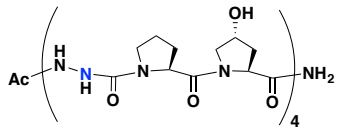
Synthesis: A + B(Fmoc-azGPO) + B'''(Fmoc-azGPO) + B'''(Fmoc-azGPO) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 886.38, found 886.64.



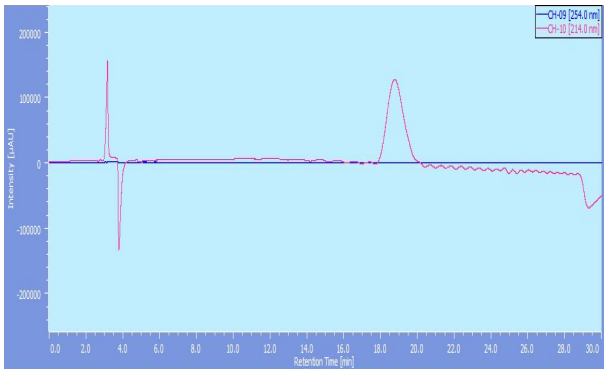


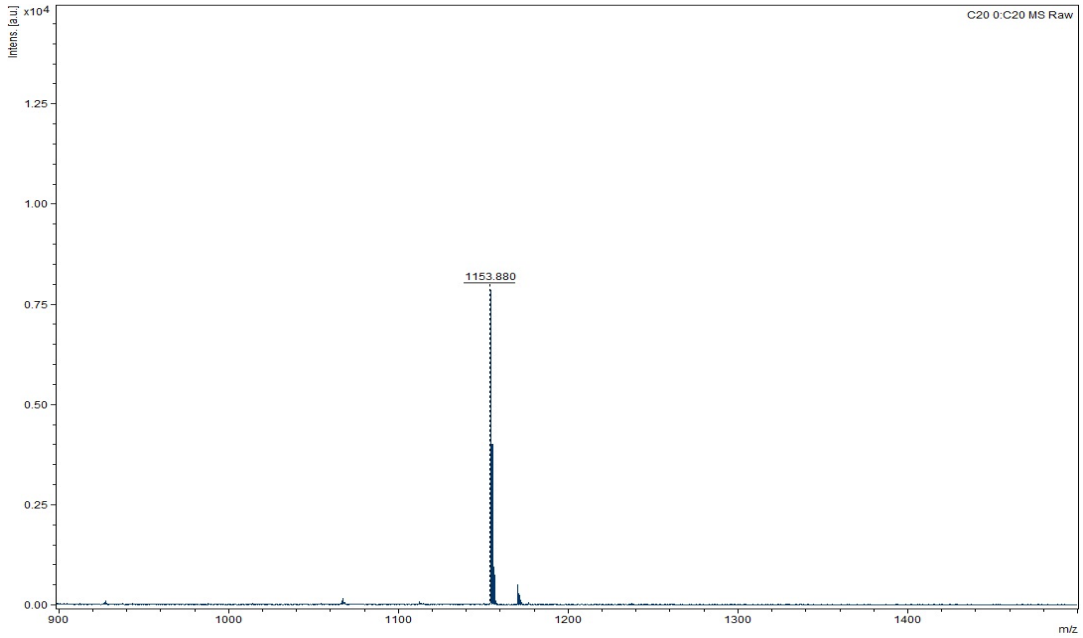
CMP **14** Ac-(azGly-Pro-Hyp)$_4$-NH$_2$



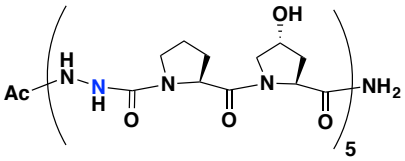
Synthesis: A + B(Fmoc-azGPO) + B’’’(Fmoc-azGPO) + B’’’(Fmoc-azGPO) + B’’’(Fmoc-azGPO) + C + D + E

MALDI-MS calculated [M+Na]$^+$ 1154.50, found 1153.88



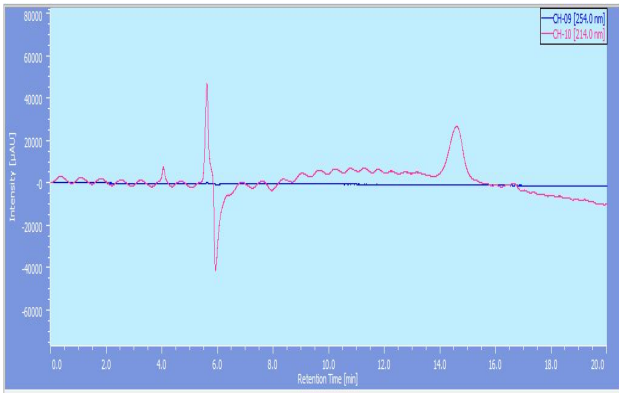


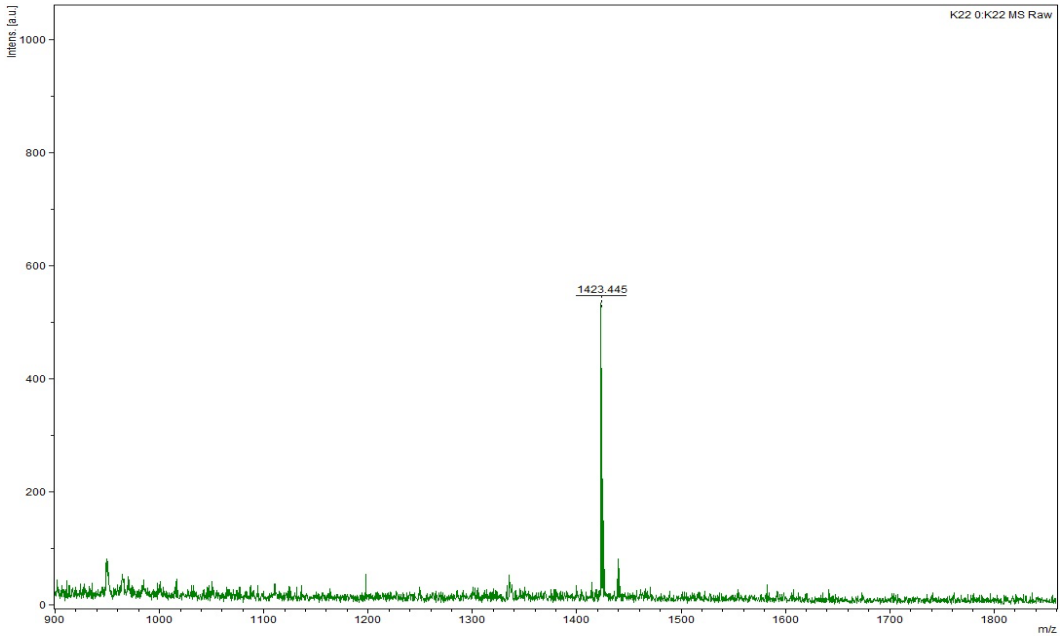
CMP **15** Ac-(azGly-Pro-Hyp)$_5$-NH$_2$



Synthesis: A + B(Fmoc-azGPO) + B'''(Fmoc-azGPO) + B'''(Fmoc-azGPO) + B'''(Fmoc-azGPO) + B'''(Fmoc-azGPO) + C + D + E

MALDI-MS calculated $[M+Na]^+$ 1422.61, found 1423.44





CD Experiment Protocols

Peptide sample preparation

Purified peptides used in this study were made into 0.20 mM solutions in PBS buffer (0.20 g KCl, 0.20 g KH_2PO_4, 8.0 g NaCl, 2.16 g $Na_2HPO_4 \bullet 7H_2O$ in 1.0 L H_2O). Samples were incubated at 4 ^{0}C for at least 24 h before CD experiments.

CD Wavelength scan

CD spectra were recorded at a step of 1.0 nm from 260 nm to 190 nm at 10 ^{0}C with a 1.0 s equilibration time.

CD Thermal denaturation experiment

The wavelength that gave the highest absorption in the range 224-227 nm was selected and monitored as a function of time in the thermal denaturation experiment. For complete aza-glycine CMPs **13 - 15**, the negative absorptions were monitored at five different wavelengths (210 nm, 215 nm, 220 nm, 225 nm, 230 nm). Averaging time was set to be 16 s with a heating rate of 12 ^{0}C/h and the ellipticity of every 1 ^{0}C was recorded. Data collected from these experiments were fitted to a two-state model according to Engel et al. to obtain the melting temperature (i.e. the temperature at which 50% of the triple helix unfolds) of each peptide.[4] We used the software Graphpad Prism 6 and followed the procedure described by Erdmann & Wennemers.[5]

CD kinetic refolding experiment

Peptide solution of 0.20 mM was heated at 95 ^{0}C for 15 min and transferred to a CD cuvette (pre-cooled at 4 ^{0}C.) After about 1 min, the ellipticity at 224 nm (for CMP **1**), or 215 nm (for CMP **14** and **15**), was monitored at 4 ^{0}C for 3 h, with a 10 s time constant and 4 s time interval. The fraction refolded percentage was calculated by the following equation:[6]

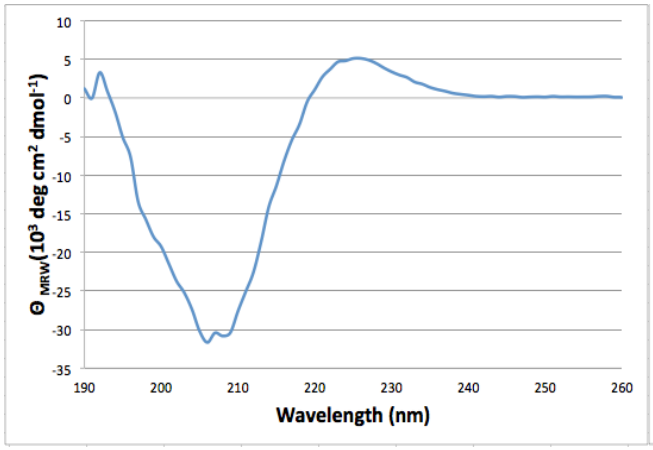
$$F = (\Theta_T - \Theta_I) / (\Theta_N - \Theta_I)$$

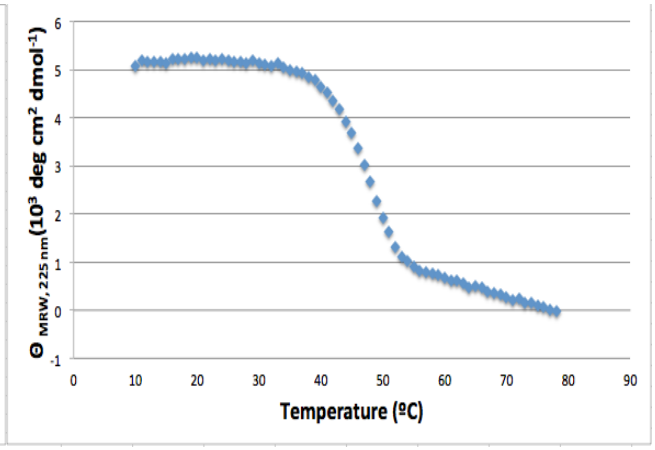
F = fraction refolded; Θ_T = ellipticity at time t; Θ_N = ellipticity before denaturation; Θ_I = first data (initial) recorded after ~65 s cooling (dead time). Plots of fraction refolded data versus time (minutes as units) were then fitted into a 3rd order kinetic equation as previously reported.[7] Half-refolding time ($t_{1/2}$) was then obtained as the time value at which 50% of each peptide recovered triple helicity. At least 3 kinetic refolding curves were collected and fitted for each peptide in this study.

CD Spectra

CD plots for CMP **3** Ac-(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_6$-NH$_2$

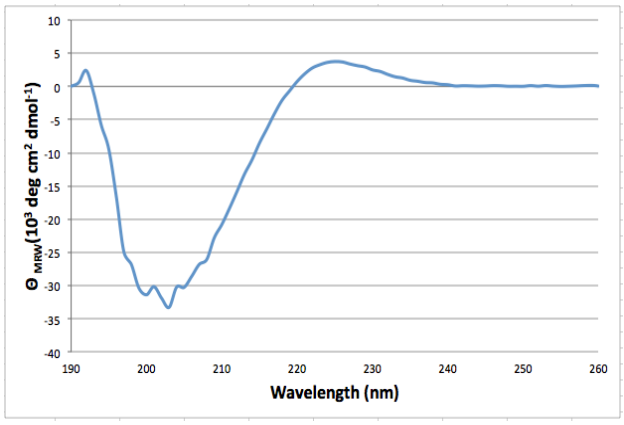
Left: CD wavelength scan; Right: CD thermal denaturation experiment;

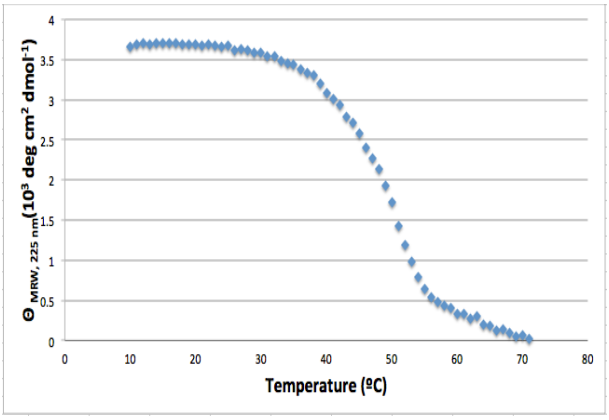




CD plots for CMP **4** Ac-(Pro-Hyp-Gly)(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_5$-NH$_2$

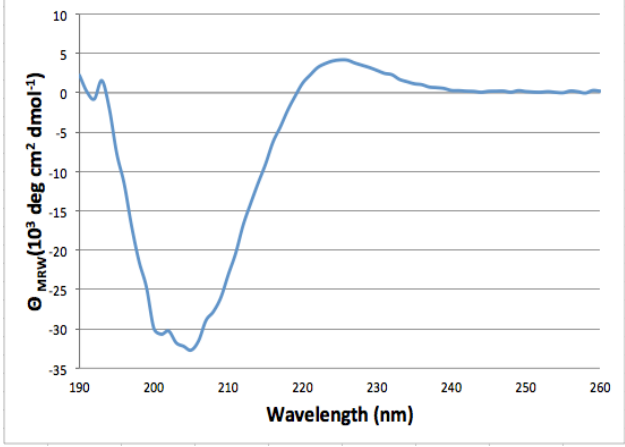
Left: CD wavelength scan; Right: CD thermal denaturation experiment;

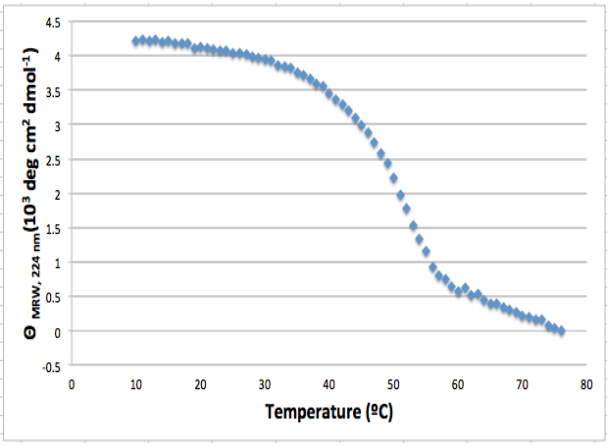




CD plots for CMP **5** Ac-(Pro-Hyp-Gly)$_2$(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_4$-NH$_2$

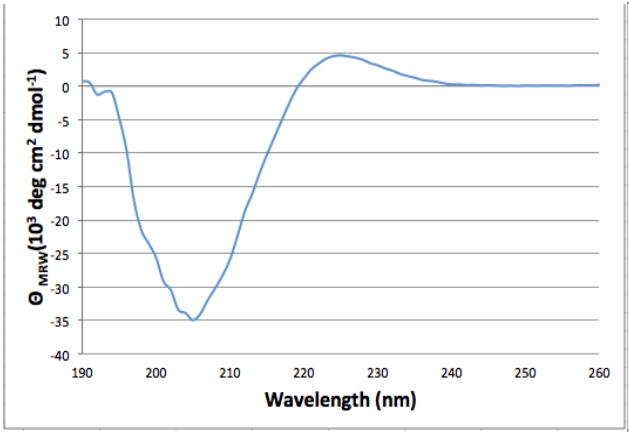
Left: CD wavelength scan; Right: CD thermal denaturation experiment;

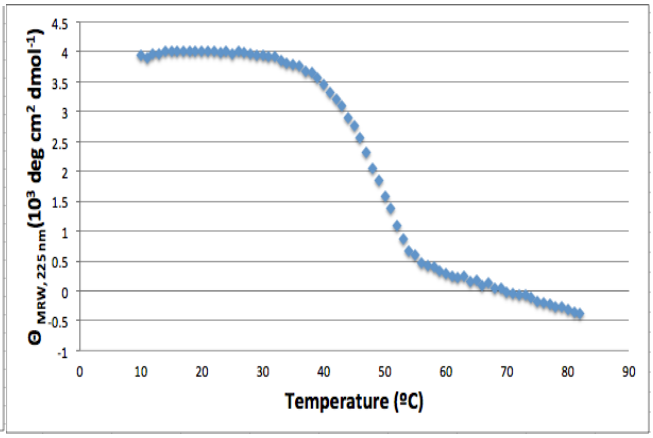




CD plots for CMP **6** Ac-(Pro-Hyp-Gly)$_4$(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_2$-NH$_2$

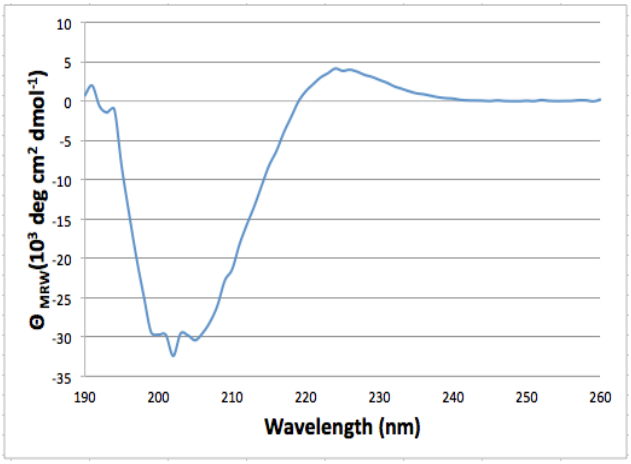
Left: CD wavelength scan; Right: CD thermal denaturation experiment;

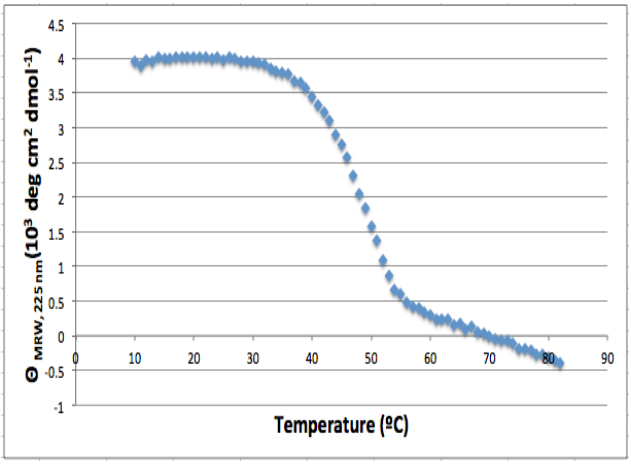




CD plots for CMP **7** Ac-(Pro-Hyp-Gly)$_5$(Pro-Hyp-azGly)(Pro-Hyp-Gly)-NH$_2$

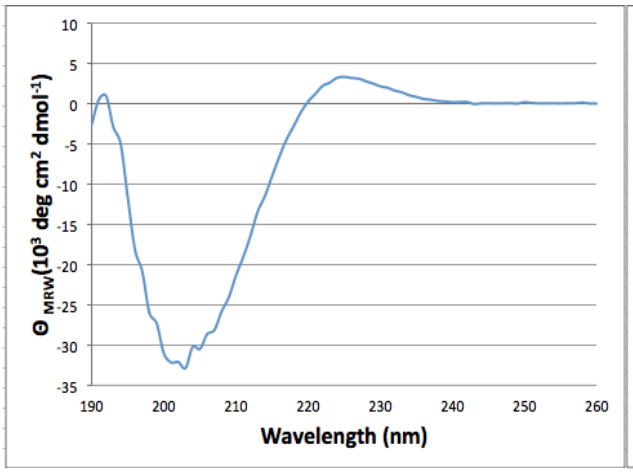
Left: CD wavelength scan; Right: CD thermal denaturation experiment;

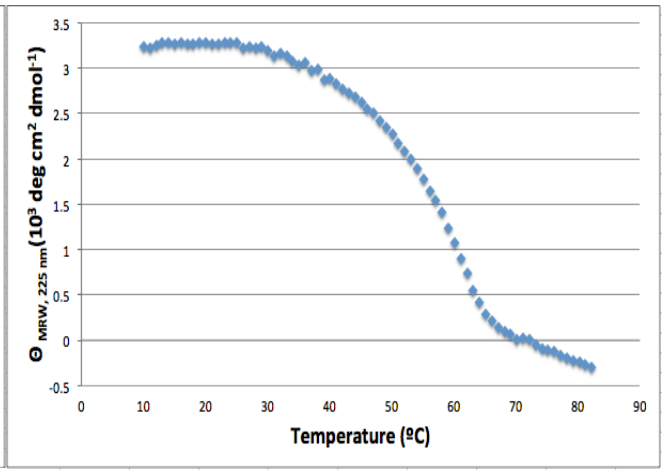




CD plots for CMP **8** Ac-(Pro-Hyp-Gly)(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_3$(Pro-Hyp-azGly)(Pro-Hyp-Gly)-NH$_2$

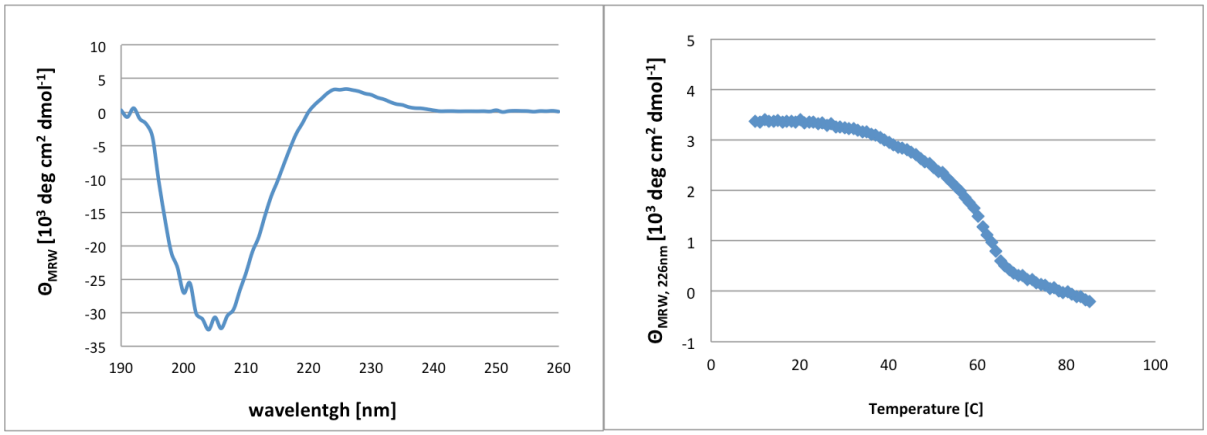
Left: CD wavelength scan; Right: CD thermal denaturation experiment;





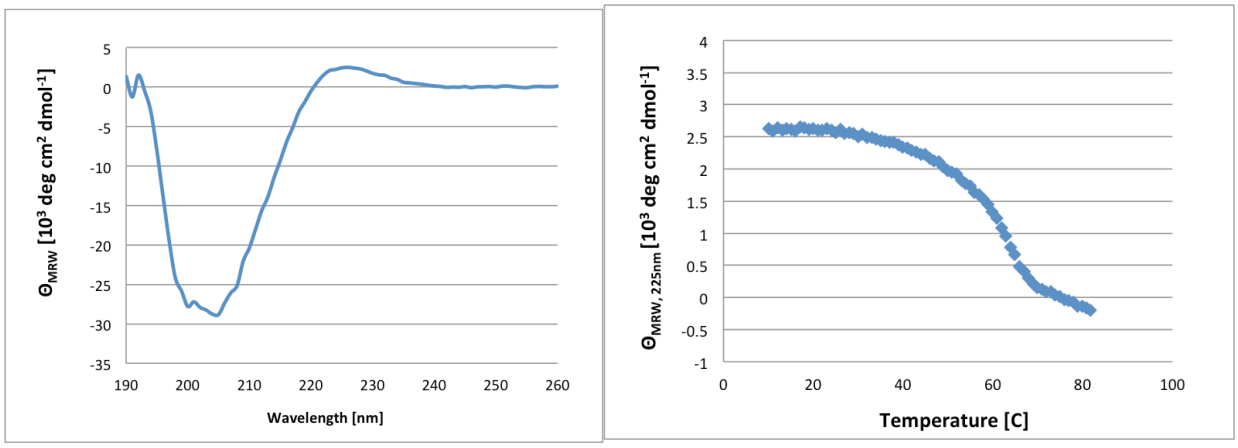
CD plots for CMP **9** Ac-(Pro-Hyp-Gly)$_2$(Pro-Hyp-azGly)(Pro-Hyp-Gly)(Pro-Hyp-azGly)(Pro-Hyp-Gly)$_2$-NH$_2$

Left: CD wavelength scan; Right: CD thermal denaturation experiment;



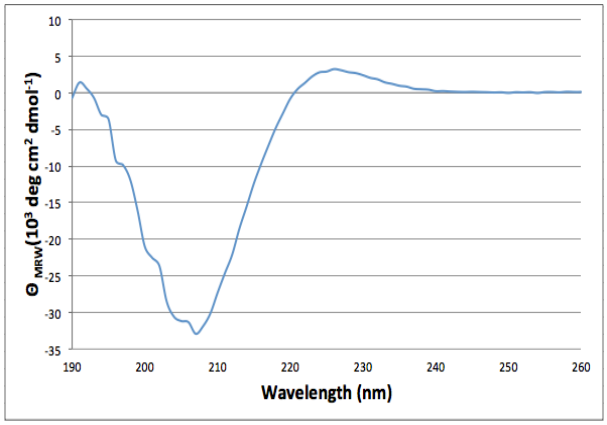
CD plots for CMP **10** Ac-(Pro-Hyp-Gly)$_3$(Pro-Hyp-azGly)$_2$(Pro-Hyp-Gly)$_2$-NH$_2$

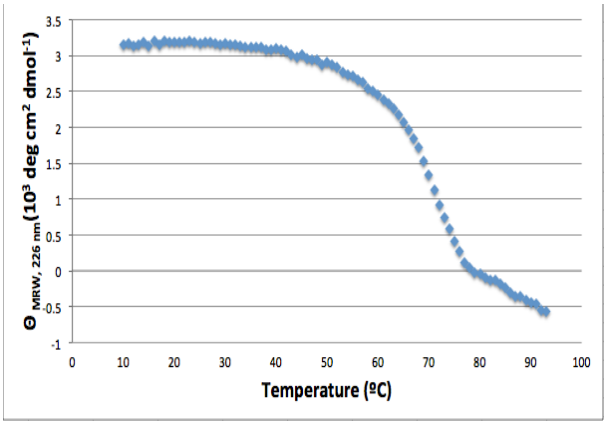
Left: CD wavelength scan; Right: CD thermal denaturation experiment;



CD plots for CMP **11** Ac-(Pro-Hyp-Gly)[(Pro-Hyp-azGly)(Pro-Hyp-Gly)]$_3$-NH$_2$

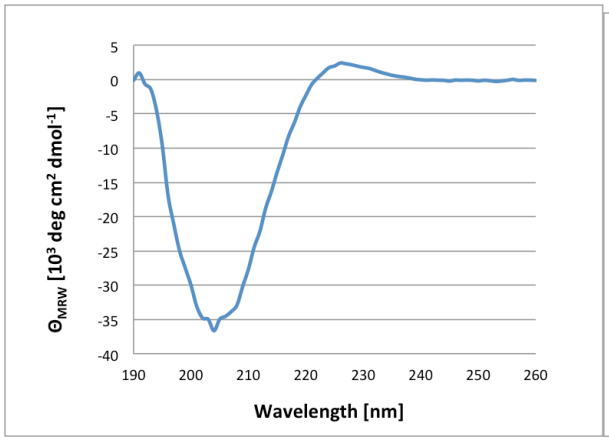
Left: CD wavelength scan; Right: CD thermal denaturation experiment;

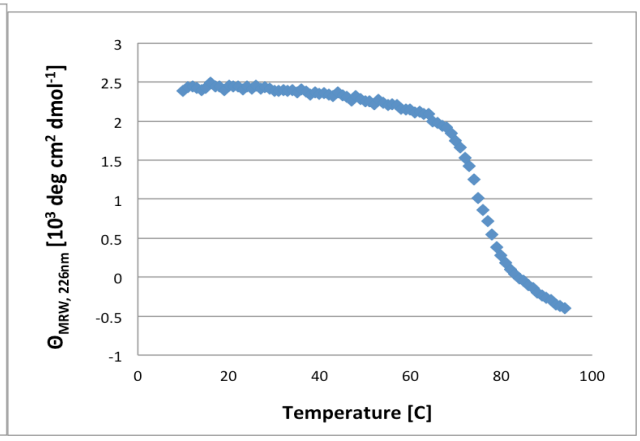




CD plots for CMP **12** Ac-(Pro-Hyp-Gly)$_2$(Pro-Hyp-azGly)$_3$(Pro-Hyp-Gly)$_2$-NH$_2$

Left: CD wavelength scan; Right: CD thermal denaturation experiment;





Summary of λ_{max} and CD Ellipticity at λ_{max} for CMP 1-12

CMP	λ_{max} / nm	CD Elipticity at λ_{max} / mdeg
1	224	18.9
2	225	15.7
3	225	21.4
4	225	15.7
5	225	17.5
6	225	19.1
7	225	16.7
8	225	13.9
9	226	14.3
10	225	11.1
11	226	13.5
12	227	8.5

UV-Vis spectrum of CMP 13-15

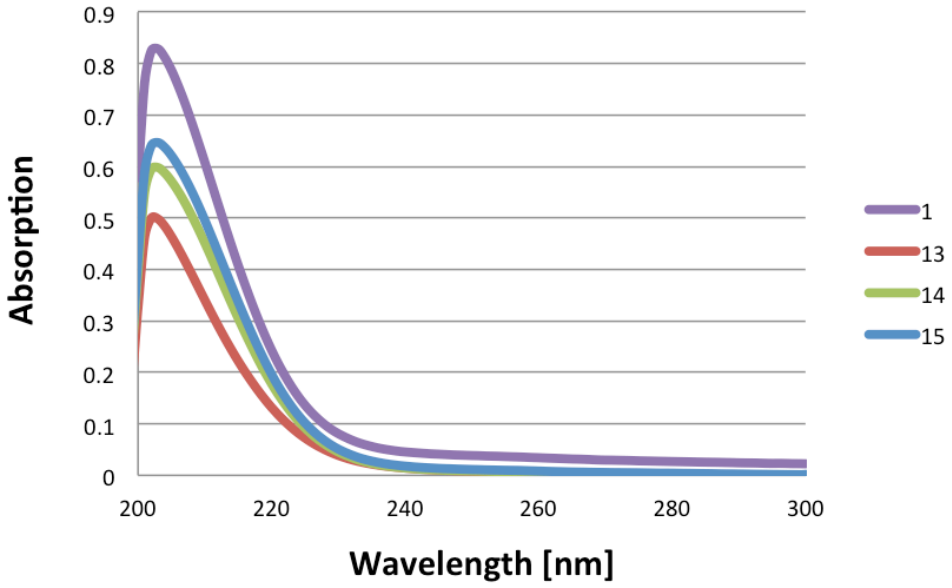
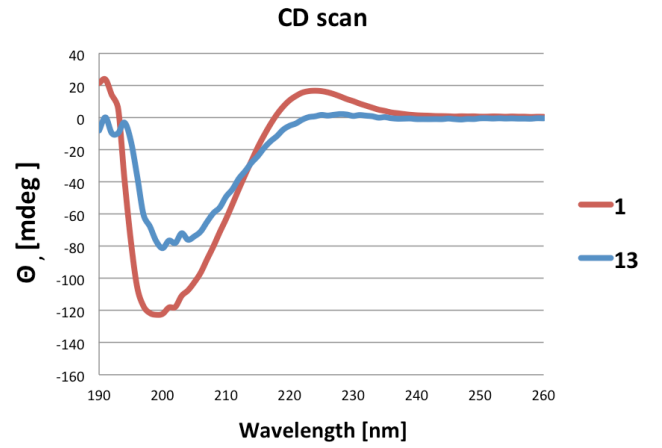
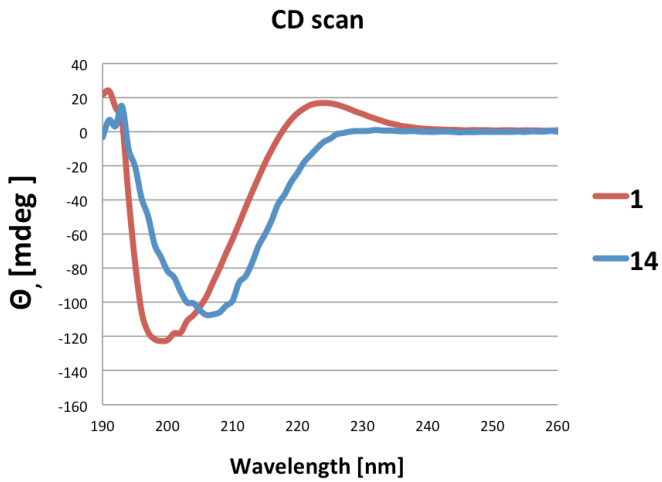


Figure S3. UV-Vis spectra of CMP 13-15 and control peptide 1, $Ac(POG)_7NH_2$.

CD scan of CMP 13-15





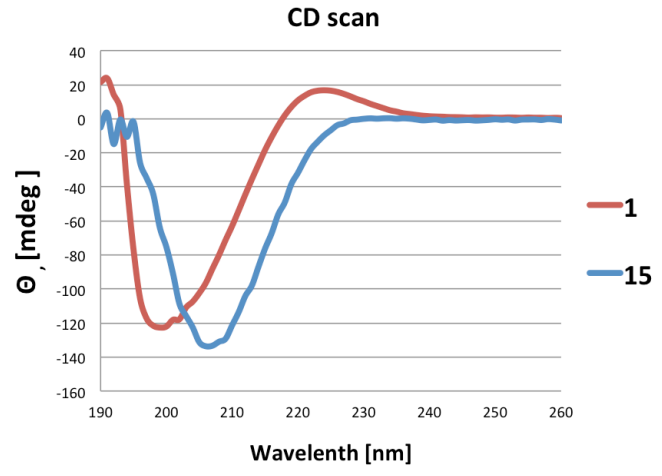


Figure S4. CD spectrum of CMP 13-15 with red curve from natural control peptide 1, $Ac(POG)_7NH_2$

Temperature-dependent CD experiments for CMP 13-15 and 1.

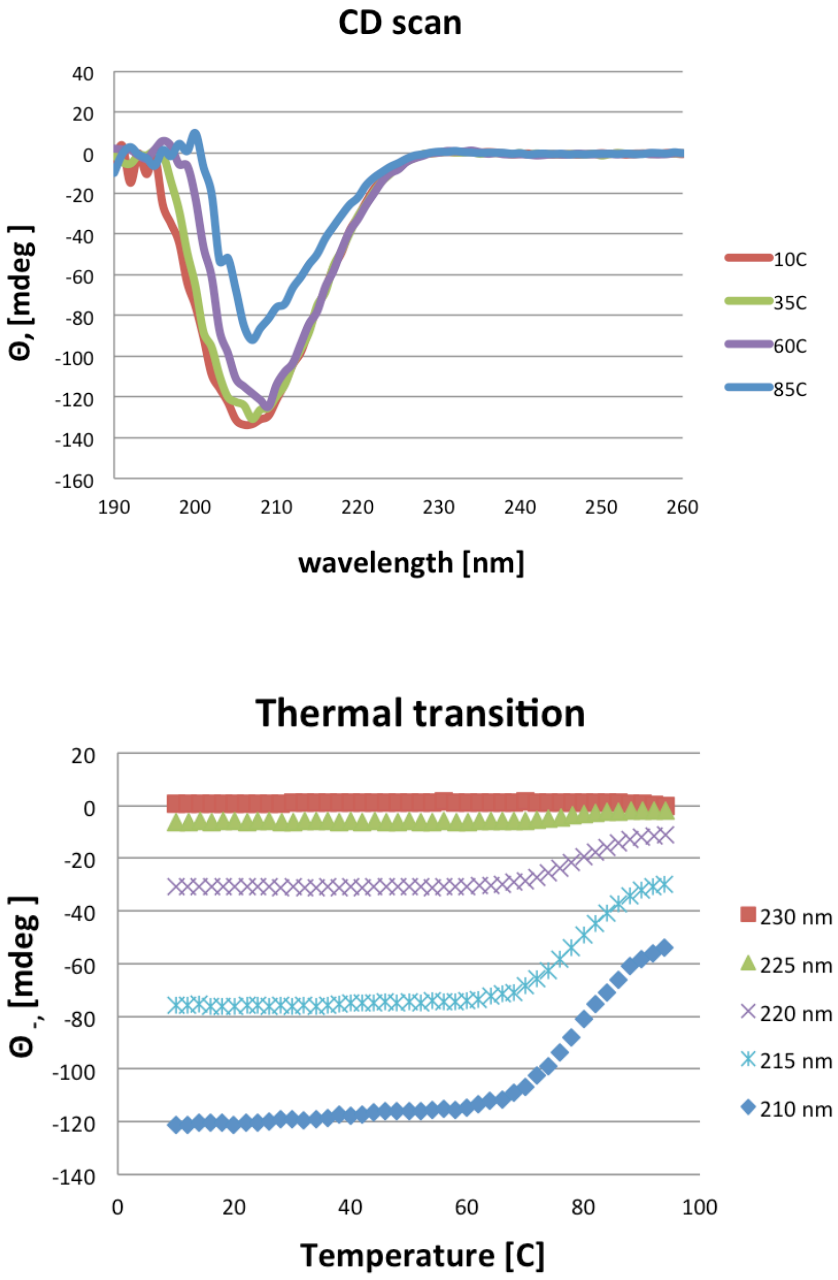


Figure S5. Variable temperature CD scan and thermal transition plots (multiple wavelength channels) for CMP 15 Ac-(azGly-Pro-Hyp)$_5$-NH$_2$.

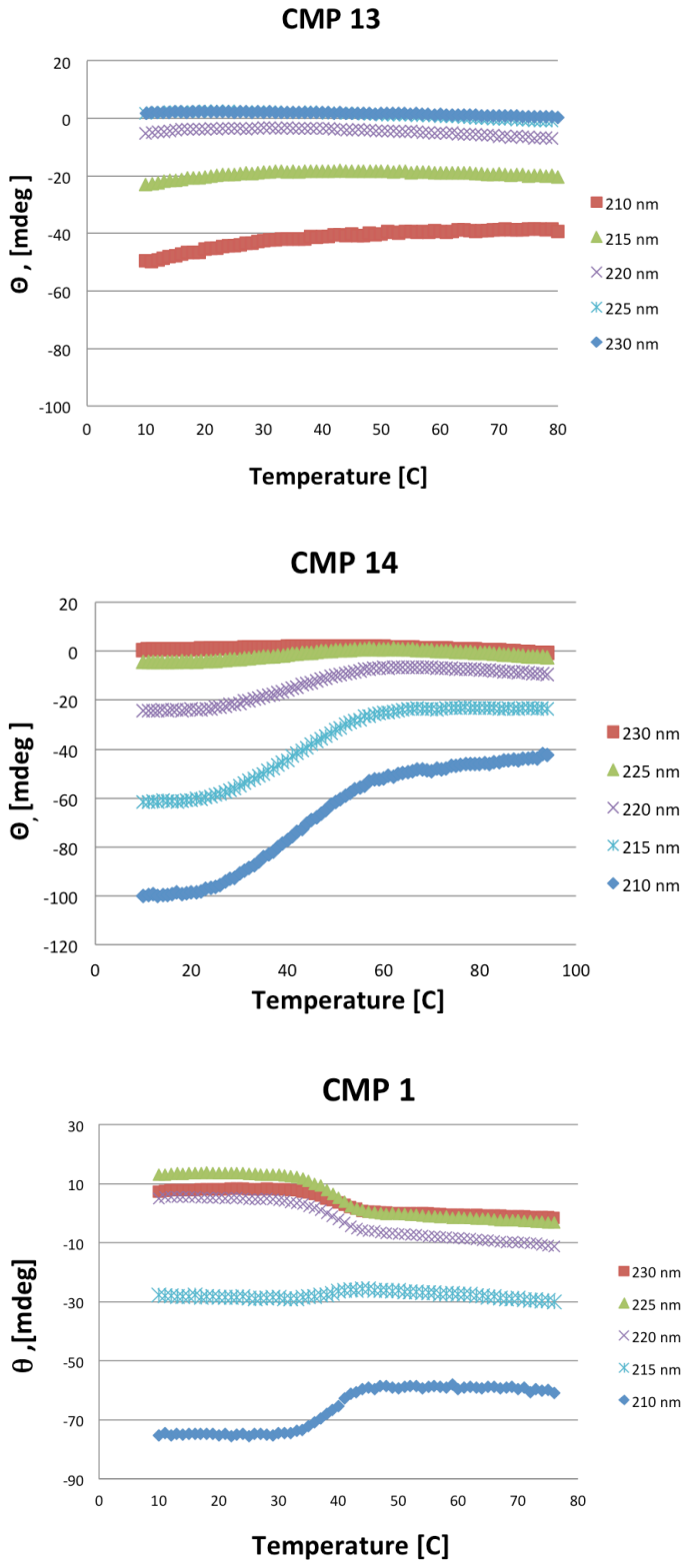


Figure S6. CD thermal transition plots (multiple wavelength channels) for CMP 13, 14, and control peptide 1.

References

1. Technical note: Determination of the Amino Acid Substitution Level via an Fmoc Assay, http://www3.appliedbiosystems.com/cms/groups/psm_marketing/documents/generaldocuments/cms_040640.pdf
2. Carpino, L. A.; Han, G. Y. J. *Org. Chem.* **1972**, 37, 3404-3409.
3. Sacca, B.; Moroder, L. *J. Peptide Sci.***2002**, *8*, 192-204.
4. Engel, J.; Chen, H. T.; Prockop, D. J.; Klump, H. *Biopolymers* **1977,** *16,* 601-622.
5. Erdmann, R. S.; Wennemers, H. *Angew. Chem., Int. Ed.* **2011**, *50*, 6835-6838.
6. Kotch, F. W.; Raines, R. T.*Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 3028-3033
7. Greenfield, N. J. *Nat. Protoc.***2006**, *1*, 2891-2899.

Appendix 1: SEC-MALS Report

HPLC Size Exclusion Chromatography/ Laser Light Scattering Determination of Native Protein Molecular Weights

From: Ewa Folta-Stogniew
HHMI Biopolymer Facility and
W.M. Keck Foundation Biotechnology Resource Laboratory
300 George Street, PO box 201
New Haven CT 06511
Telephone: (203) 737-4387 Fax: (203) 785-7023
Email: Ewa.Folta-Stogniew@yale.edu

File directory: ASTRA/Jun_16
Column: Superdex peptide #2, HR 10/300
Buffer: PBS+azide from 10x stock

Each sample was filtered through 20 nm Anotop filters before analysis. Samples were analyzed at 200ul and 50 ul injection from filtered samples.

During sample analysis, the MW is determined by solving the equation that relates the excess scattered light, measured at several angles, to the concentration of solute and the weight-average molar mass (ASTRA calculations, please refer to "http://info.med.yale.edu/wmkeck/6_16_98/Astra2a.htm#Calculation of MW by" _for details on the ASTRA calculations). During data analysis a dn/dc value of 0.188 mL/g was used as it proved satisfactory during analyses of protein standards in PBS.

The UV traces were recorded at 220 nm and 235 nm (to prevent saturation of the signal).

Sample Names:

CMP 14 in manuscript = CMP-87 in SEC-MALS report: Ac-$(azGPO)_4$-NH_2

CMP 15 in manuscript = CMP-88 in SEC-MALS report: Ac-$(azGPO)_5$-NH_2

Single strand peptide controls used for analysis:

CMP 90 represents a single strand control for CMP 14. The structure of CMP 90 is:

Ac-$(POG)_4$-NH_2

CMP 91 represents a single strand control for CMP 15. The structure of CMP 91 is:

Ac-$(POG)_5$-NH_2

Table 1. Results of SEC-LS/RI/UV analysis from ASTRA (reports for *major peaks* of selected analyses).

RUN	PEAK Elution at RI trace (mL)	MW Average of eluting peak (Da)	Molar mass range across eluting peak	Sequence Predicated MW for monomer (kDa)	Comments
#87_200ul	11.11	3271	3150-3300	1132	
#88_200ul	10.56	4149	4050-4150	1400	
#90_200ul	12.76	1098	1070-1100	1128	
#91_200ul	12.03	1388	1350-1400	1395	

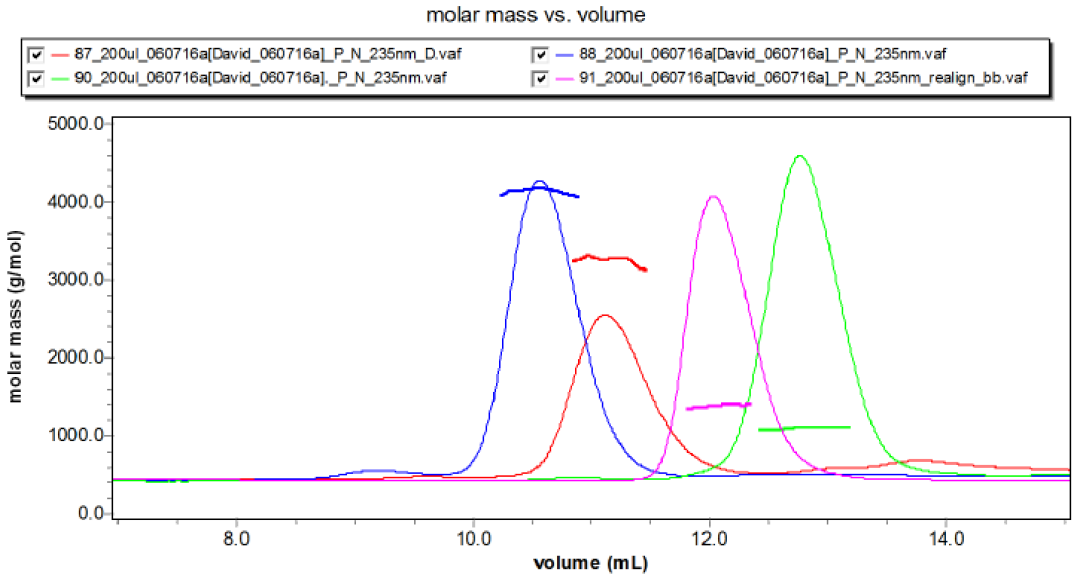
* A monodisperse peak is a peak that appears to contain only one type of molecule (*e.g.*, pure monomeric protein) and results in one apparent MW. Thus the average mass is independent of the averaging method and the polydispersity (see ASTRA Summary Report) will equal 1. If the peak contains a mixture of species of different molar masses (*i.e.*, polydisperse peak; for example, a mixture of dimer (or tetramer) and monomer) the average mass will depend on the averaging method and the polydispersity will be different then 1.

[a] as shown on Molar Mass Distribution Plot

**N/D not determined since the signal was weak; none of the eluting peaks contained 50ug of proteins precluding reliable determination of monodispersity

The distribution of molar masses across the eluting peak are shown below:

Solid lines: traces from UV recorded at 235 nm ; dots: MW measured every 2s across the eluting peak.



Summary Table:

SAMPLE	monomer	trimer	observed	oligomer	differential
CMP-87	1132	3250	3271	trimer	0.6%
CMP88	1400	4200	4149	trimer	1.2%
CMP90	1128	3384	1098	monomer	2.7%
CMP91	1395	4185	1388	monomer	0.5%

Appendix 2: Analytical Ultracentrifugation Report: Contained on following pages

University of Connecticut
James L. Cole & Jeffrey W. Lary
Analytical Ultracentrifugation Facility
Biotechnology • Bioservices Center

15June2016

Sample & Method

Contact Person

David M. Chenoweth, Ph.D.
Department of Chemistry
University of Pennsylvania
231 S. 34 Street, Philadelphia, PA 19104-6323

Sample

Two protein samples, CMP#87 and CMP#88, were submitted for analysis by sedimentation equilibrium (SE) ultracentrifugation. Each sample was supplied at 1 mg/ml. A 15 mL sample of the buffer was also supplied.

Buffer

PBS (0.20 g KCl, 0.20 g KH2PO4, 8.0 g NaCl, 2.16 g Na2HPO4●7H2O in 1.0 L H2O).

Physical Constants

The following physical constants were calculated from the amino acid composition for the proteins using the program Sednterp.[5] The calculated $\bar{v}_{20^o}$ are only a close approximation as there is no way to estimate the effect on the partial specific volume of the protein due to the presence of the "az" moieties.

CMP#87 - $Ac(azGPO)_4NH_2$: MW_{seq} = 1132 Da, $\bar{v}_{20^o}$ = 0.695 ml/g
CMP#88 - $Ac(azGPO)_5NH_2$: MW_{seq} = 1400 Da, $\bar{v}_{20^o}$ = 0.696 ml/g

The buffer density and viscosity were calculated to be 1.00533 g/mL and 0.01019 poise at 20°C, respectively using Sednterp.

Method

Sedimentation equilibrium analysis was conducted at 20°C using a single speed of 55,000 RPM and 2 loading concentrations, 0.3, and 1.0 mg/mL, for each sample. Interference optics were used with a Beckman-Coulter XL-I analytical ultracentrifuge. Standard, double-sector aluminum-filled Epon cells equipped with sapphire windows were used, filling the reference side with 360 µL of buffer and the sample side with 15 µL of FC-43 (an inert oil) plus 335 µL of the appropriate sample. This resulted in solution column heights of ~0.9 cm. Interference scans were acquired at 6 minute intervals and attainment of equilibrium was determined using the "Match" module of the program HeteroAnalysis[8]. Cell blanks were acquired at 55,000 RPM prior to the actual run by filling the cells with water in both the reference and samples sides. The rotor was accelerated to speed and after 30 minutes scans were recorded for each cell for use as cell blanks. The water was removed from the cells after the blank scans were recorded, without disassembly of the cells, and they were dried in a vacuum prior to their use in the actual run.

University of Connecticut
James L. Cole & Jeffrey W. Lary
Analytical Ultracentrifugation Facility
Biotechnology • Bioservices Center

15June2016

Analysis – 1

Program

HeteroAnalysis[8], version 1.1.0.60

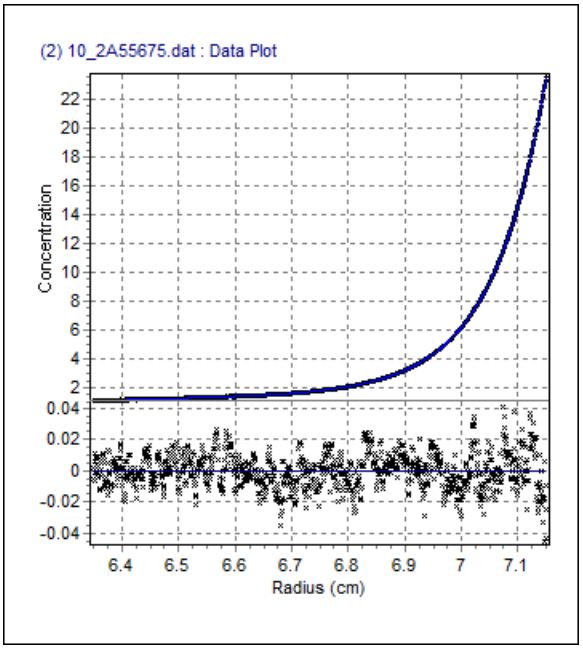
Method

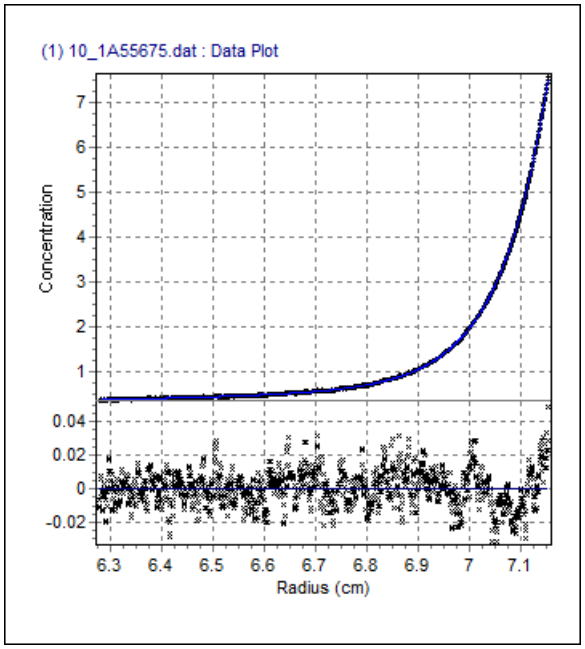
Model-based non-linear least squares analysis of sedimentation equilibrium data.

Results

CMP#87:

Blank-subtracted sedimentation equilibrium data for the two loading concentrations of CMP#87 was globally fit to a number of models using Heteroanalysis. The only model that gave good results was the model of a mixture of two, non-interacting species (denoted A and B). A screen capture of the best fits for the two cells are shown next and the results of the fit are given in the table below the plots. The residuals (shown at the bottom of each plot) are randomly distributed, indicating that this is a good model for the system.





(The concentrations shown in the plots are in units of fringes, where 1 mg/mL = 3.3 fringes)

Model	MW(A) (Da)	MW(B) (Da)	RMS (mg/ml)
Mixture of two, ideal species	250	3630 [3610, 3660]	0.0034

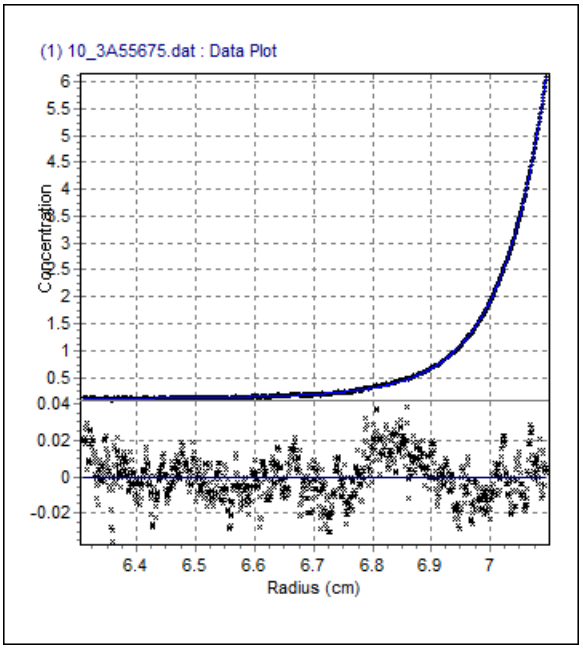
* Numbers in the brackets are the 95% confidence limits of the fitted values.

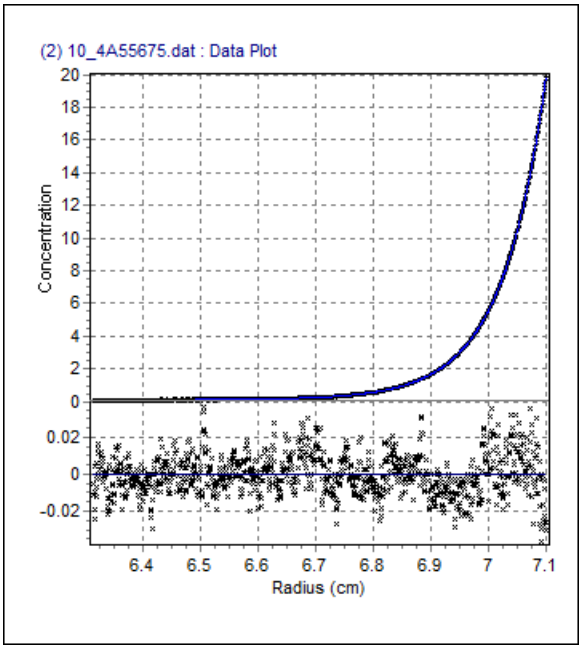
The fitted molecular weight is within 7% of the sequence value for the trimer of CMP#87. This is a quite good estimate keeping in mind that the partial specific volume ($\bar{v}$) that was calculated for the protein is, at best, a close approximation.

Results (cont'd)

CMP#88:

Blank-subtracted sedimentation equilibrium data for the two loading concentrations of CMP#88 was globally fit to a number of models using Heteroanalysis. As with the CMP#87 sample, the only model that gave good results was the model of a mixture of two, non-interacting species (denoted A and B). A screen capture of the best fit is shown next followed by a table containing the fitted values for the molecular weights of the two species. Here too, the random nature of the residuals is a good indication that this is an appropriate model for the data.





(The concentrations shown in the plots are in units of fringes, where 1 mg/mL = 3.3 fringes)

Model	MW(A) (Da)	MW(B) (Da)	RMS (mg/ml)
Mixture of two, ideal species	390	4480 [4460, 4510]	0.0035

* Numbers in the brackets are the 95% confidence limits of the fitted values.

Unfortunately, we cannot estimate the relative amounts of each species contained in the sample using this program.

University of Connecticut
James L. Cole & Jeffrey W. Lary
Analytical Ultracentrifugation Facility
Biotechnology • Bioservices Center

15June2016

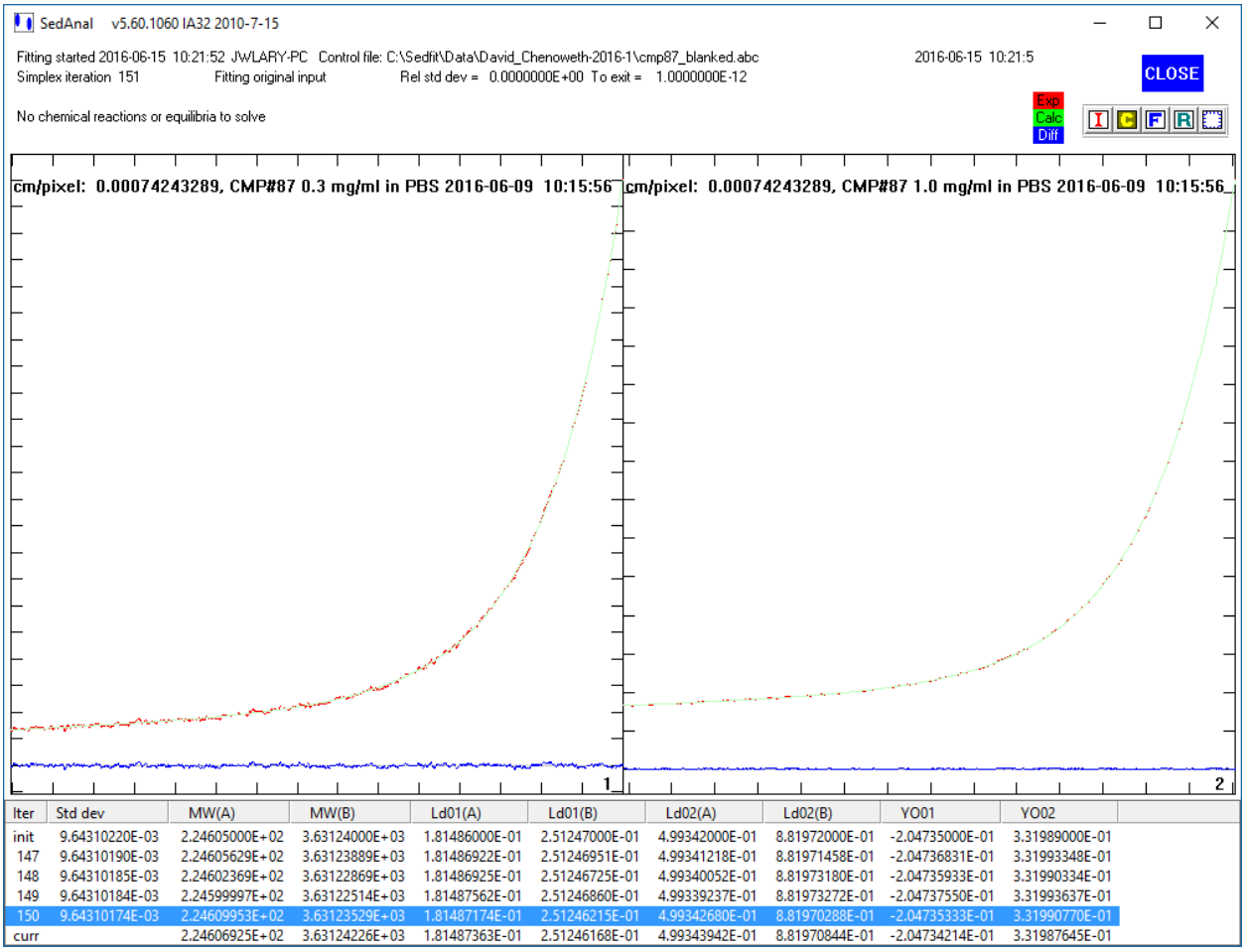
Analysis - 2

Program
Sedanal[6], version 5.60

Method
Direct boundary modeling program for global analysis (i.e. multiple data sets) of sedimentation velocity data using model based numerical solutions to the Lamm equation.

Model
Mixture of two ideal components.

Results
CMP#87:
The data set for CMP#87 was analyzed using Sedanal with the model of two, non-interacting ideal species – A and B. The fitting parameters were the molecular weights of the two species, the baseline (fringe offset), and the loading concentration of components A and B for each cell. A screen capture of the best fit is shown below. The random nature of the deviations (blue plotted points) and the low RMS deviation for the fit indicates that this is a good model for this system.



Iter	Std dev	MW(A)	MW(B)	Ld01(A)	Ld01(B)	Ld02(A)	Ld02(B)	YO01	YO02
init	9.64310220E-03	2.24605000E+02	3.63124000E+03	1.81486000E-01	2.51247000E-01	4.99342000E-01	8.81972000E-01	-2.04735000E-01	3.31989000E-01
147	9.64310190E-03	2.24605629E+02	3.63123889E+03	1.81486922E-01	2.51246951E-01	4.99341218E-01	8.81971458E-01	-2.04736831E-01	3.31993348E-01
148	9.64310185E-03	2.24602369E+02	3.63122869E+03	1.81486925E-01	2.51246725E-01	4.99340052E-01	8.81973180E-01	-2.04735933E-01	3.31990334E-01
149	9.64310184E-03	2.24599997E+02	3.63122514E+03	1.81487562E-01	2.51246860E-01	4.99339237E-01	8.81973272E-01	-2.04737550E-01	3.31993637E-01
150	9.64310174E-03	2.24609953E+02	3.63123529E+03	1.81487174E-01	2.51246215E-01	4.99342680E-01	8.81970288E-01	-2.04735333E-01	3.31990770E-01
curr		2.24606925E+02	3.63124226E+03	1.81487363E-01	2.51246168E-01	4.99343942E-01	8.81970844E-01	-2.04734214E-01	3.31987645E-01

Analysis-2

Results (cont'd)

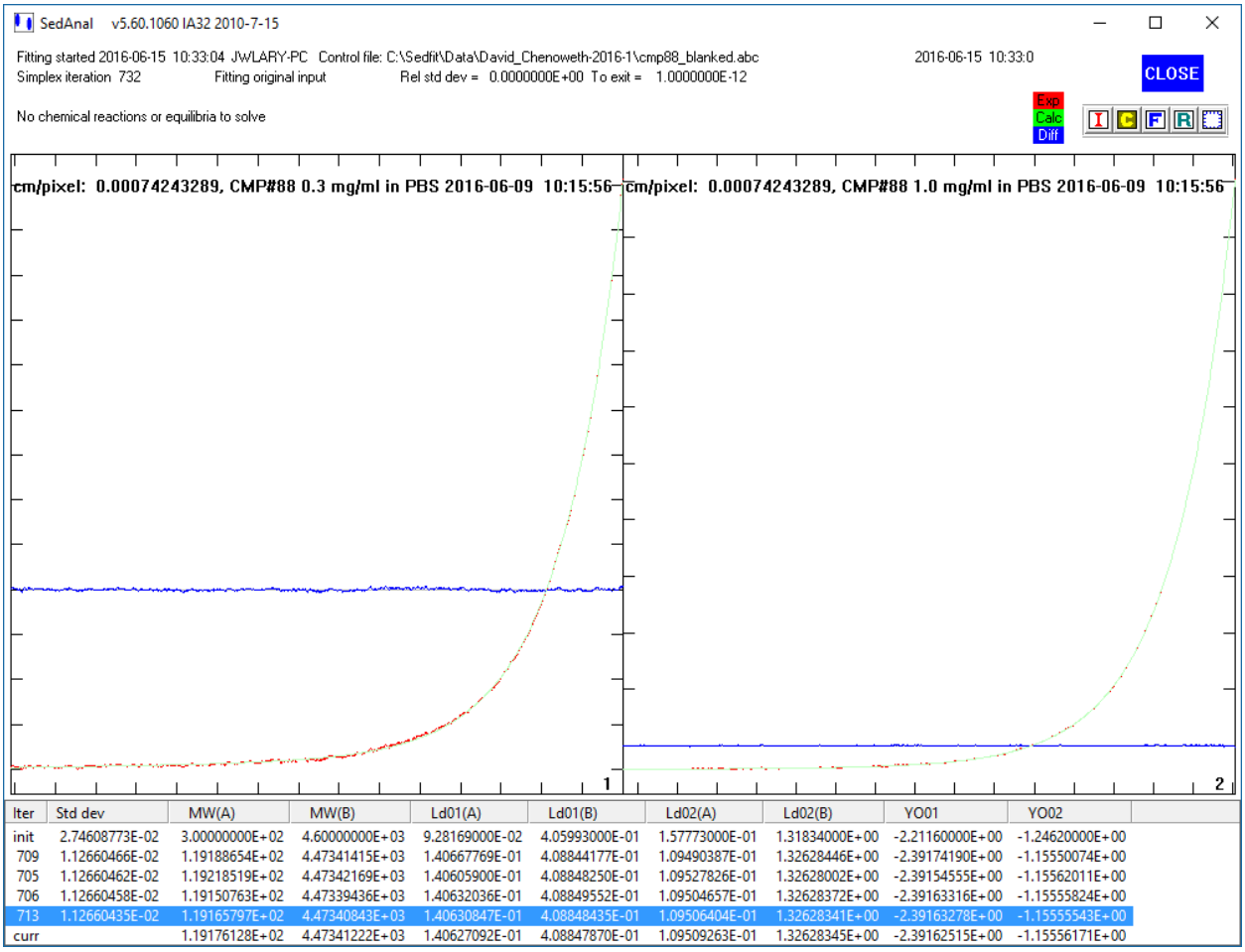
The results of the fit are given in the table below.

Model	MW(A) (Da)	MW(B) (Da)	Weight %		RMS (mg/ml)
			A	B	
Mixture of two, ideal species	225	3630 [3610, 3700]	0.3 mg/ml - 42% 1.0 mg/ml - 36%	0.3 mg/ml - 58% 1.0 mg/ml - 64%	0.0029

* Numbers in the brackets are the 95% confidence limits of the fitted values.

CMP#88:

The data set for CMP#88 was analyzed using Sedanal with the model of two, non-interacting ideal species – A and B. The fitting parameters were the molecular weights of the two species, the baseline (fringe offset), and the loading concentration of components A and B for each cell. A screen capture of the best fit is shown below. The random nature of the deviations (blue plotted points) and the low RMS deviation for the fit indicates that this is a good model for this system.



Iter	Std dev	MW(A)	MW(B)	Ld01(A)	Ld01(B)	Ld02(A)	Ld02(B)	YO01	YO02
init	2.74608773E-02	3.00000000E+02	4.60000000E+03	9.28169000E-02	4.05993000E-01	1.57773000E-01	1.31834000E+00	-2.21160000E+00	-1.24620000E+00
709	1.12660466E-02	1.19188654E+02	4.47341415E+03	1.40667769E-01	4.08844177E-01	1.09490387E-01	1.32628446E+00	-2.39174190E+00	-1.15550074E+00
705	1.12660462E-02	1.19218519E+02	4.47342169E+03	1.40605900E-01	4.08848250E-01	1.09527826E-01	1.32628002E+00	-2.39154555E+00	-1.15562011E+00
706	1.12660458E-02	1.19150763E+02	4.47339436E+03	1.40632036E-01	4.08849552E-01	1.09504657E-01	1.32628372E+00	-2.39163316E+00	-1.15555824E+00
713	1.12660435E-02	1.19165797E+02	4.47340843E+03	1.40630847E-01	4.08848435E-01	1.09506404E-01	1.32628341E+00	-2.39163278E+00	-1.15555543E+00
curr		1.19176128E+02	4.47341222E+03	1.40627092E-01	4.08847870E-01	1.09509263E-01	1.32628345E+00	-2.39162515E+00	-1.15556171E+00

Analysis-2

Results (cont'd)

The results of the fit are given in the table below.

Model	MW(A) (Da)	MW(B) (Da)	Weight %		RMS (mg/ml)
			A	B	
Mixture of two, ideal species	120	4470 [4460, 4490]	0.3 mg/ml - 26% 1.0 mg/ml - 8%	0.3 mg/ml - 74% 1.0 mg/ml - 92%	0.0034

* Numbers in the brackets are the 95% confidence limits of the fitted values.

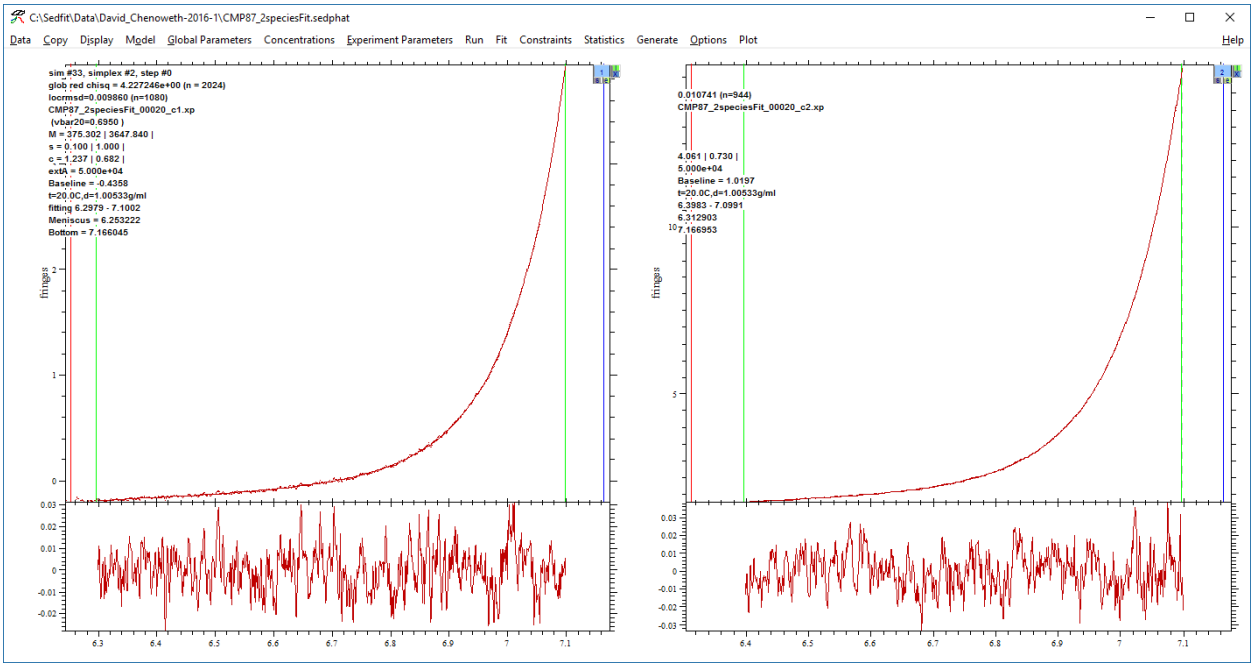
Analysis - 3

Program
Sedphat[9], version 12.1b

Method
Direct boundary modeling program for global analysis (i.e. multiple data sets) of sedimentation velocity and sedimentation equilibrium data using model based numerical solutions to the Lamm equation[9].

Model
Global discrete species model.

Results
CMP#87:
The blank-subtracted data from the two cells containing CMP#87 was globally fitted using Sedphat and the model of a two, non-interacting, ideal species – A and B. The fitting parameters were the molecular weights of the two species, the baseline (fringe offset), total loading concentration, and weight fraction of B for each cell. A screen capture of the best fit is shown below. The random nature of the deviations (bottom plot) and the low RMS deviation for the fit indicates that this is a good model for this system.



The results of the fit are given in the table below.

Model	MW(A) (Da)	MW(B) (Da)	Weight %		RMS (mg/ml)
			A	B	
Mixture of two, ideal species	375	3650 [3630, 3660]	0.3 mg/ml - 32% 1.0 mg/ml - 27%	0.3 mg/ml - 68% 1.0 mg/ml - 73%	0.0031

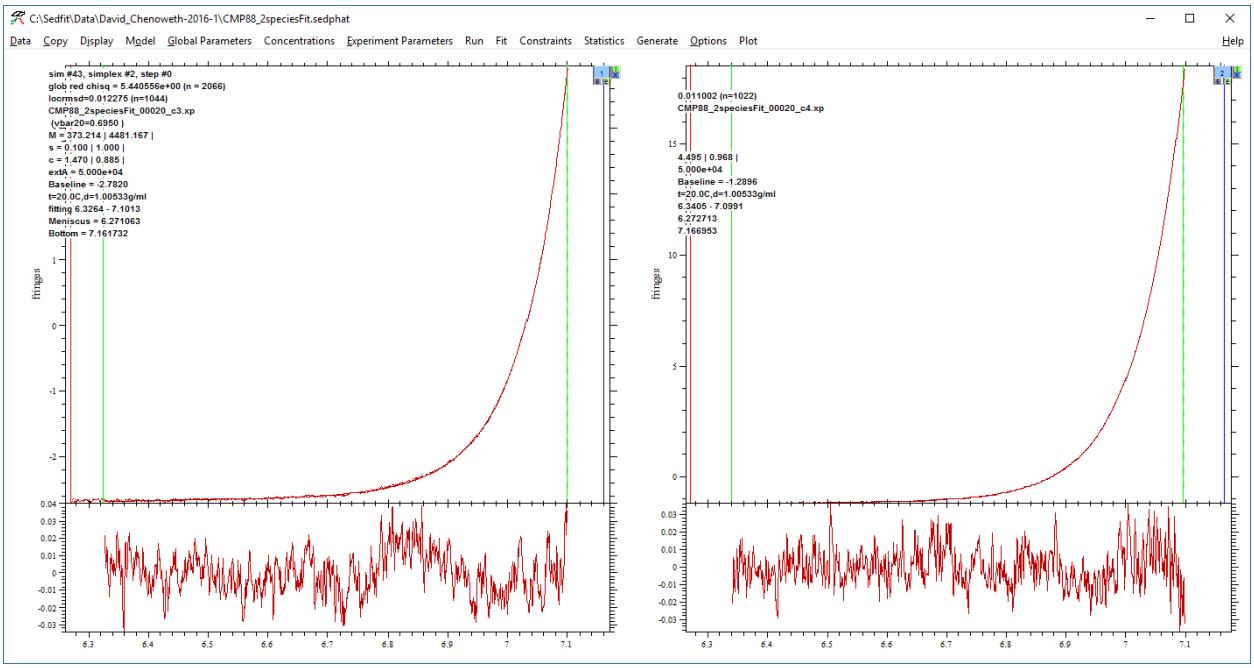
* Numbers in the brackets are the 95% confidence limits of the fitted values.

Analysis - 3

Results (cont'd)

CMP#88

The data set for CMP#88 was analyzed using Sedphat with the model of two, non-interacting, ideal species – A and B. The fitting parameters were the molecular weights of the two species, the baseline (fringe offset), total loading concentration, and weight fraction of B for each cell. A screen capture of the best fit is shown below. The random nature of the deviations (bottom plot) and the low RMS deviation for the fit indicates that this is a good model for this system.



The results of the fit are given in the table below.

Model	MW(A) (Da)	MW(B) (Da)	Weight %		RMS (mg/ml)
			A	B	
Mixture of two, ideal species	373	4480 [4470, 4490]	0.3 mg/ml - 11% 1.0 mg/ml - 3%	0.3 mg/ml - 89% 1.0 mg/ml - 97%	0.0035

* Numbers in the brackets are the 95% confidence limits of the fitted values.

Summary

The results from the SE run indicate that the CMP#87 sample can be best modeled by a non-interacting mixture of two species. One species of ~250 Da comprising ~35% of the material present and the second species with a molecular weight of ~3630 Da comprising ~65% of the material present. The molecular weight of the second and more abundant species is consistent with that of the trimer of CMP#87.

The CMP#88 sample was also best modeled by a non-interacting mixture of two species. One species of ~350 Da comprising ~15% of the material present, and the second species with a molecular weight of ~4480 Da comprising ~85% of the sample. The molecular weight of the second and more abundant species is consistent with that of the trimer of CMP#88. The following tables summarize the results from the various analyses for each sample.

CMP#87, Model: non-interacting mixture of two ideal species					
Program	MW (Da)		Weight %		RMS (mg/ml)
	A	B	A	B	
HeteroAnalysis	250	3630 [3610, 3660]	--	--	0.0034
Sedanal	225	3630 [3610, 3700]	0.3 mg/ml - 42% 1.0 mg/ml - 36%	0.3 mg/ml - 58% 1.0 mg/ml - 64%	0.0029
Sedphat	375	3650 [3630, 3660]	0.3 mg/ml - 32% 1.0 mg/ml - 27%	0.3 mg/ml - 68% 1.0 mg/ml - 73%	0.0031

* Numbers in the brackets are the 95% confidence limits of the fitted values.

CMP#88, Model: non-interacting mixture of two ideal species					
Program	MW (Da)		Weight %		RMS (mg/ml)
	A	B	A	B	
HeteroAnalysis	390	4480 [4460, 4510]	--	--	0.0035
Sedanal	120	4470 [4460, 4490]	0.3 mg/ml - 26% 1.0 mg/ml - 8%	0.3 mg/ml - 74% 1.0 mg/ml - 92%	0.0034
Sedphat	373	4480 [4470, 4490]	0.3 mg/ml - 11% 1.0 mg/ml - 3%	0.3 mg/ml - 89% 1.0 mg/ml - 97%	0.0035

* Numbers in the brackets are the 95% confidence limits of the fitted values.

References

1. Philo, J. S. (2000). A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Analytical Biochemistry* 279, 151-163.
2. Philo, J. S. (2006). Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Analytical Biochemistry* 354, 238-246.
3. Schuck, P. (2000). Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal* 78:1606-1619.
4. Stafford, W.F., III. (1992). Boundary analysis in sedimentation transport experiments: A procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Analytical Biochemistry* 203, 295-301.
5. T.M. Laue, B.D. Shah, T.M. Ridgeway and S.L. Pelletier, Analytical Ultracentrifugation in Biochemistry and Polymer Science, Royal Society of Chemistry (Edited by S. Harding and A. Rowe), 1992, pp. 90-125.
6. Walter F. Stafford and Peter J. Sherwood (2004). Analysis of heterologous interacting systems by sedimentation velocity: Curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. *Biophysical Chemistry* 108: 231-243.
7. J. Garcia de la Torre, M.L. Huertas and B. Carrasco (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal* 78, 719-730.
8. Cole, J.L. (2004). Analysis of heterogeneous interactions. *Methods Enzymol.* 384: 212-232.
9. Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Analytical Biochemistry* 320, 104-124.