Supporting Information

A Nucleic Acid Mimetic Hexaplex Assembly Stabilized by the Synergistic Association of Non-Canonical Recognition Units and Peptide Backbones

Sajena Kanangat Saraswathy, Suneesh C. Karunakaran, Shibna Balakrishnan, Gary Newnam, Elizabeth Duselis, Luis Andres Casavilca Ramirez, Howard Page, Asem Alenaizan, Vicki Wysocki, Luke J. Leman, Jonathan B. Chaires, Gary B. Schuster, and Nicholas V. Hud*

School of Chemistry and Biochemistry, Georgia Institute of Technology,
Atlanta, GA, 30332, USA.
Chemistry Department and Interdisciplinary Research Center for Biosystems and Machines,
King Fahd University of Petroleum and Minerals,
Dhahran 31261, Saudi Arabia.
Department of Chemistry, Scripps Research Institute,
La Jolla, California, USA.

TABLE OF CONTENTS

SYNTHESIS AND CHARACTERIZATION **1**

SAMPLE PREPARATION FOR PHYSICAL STUDIES **8**

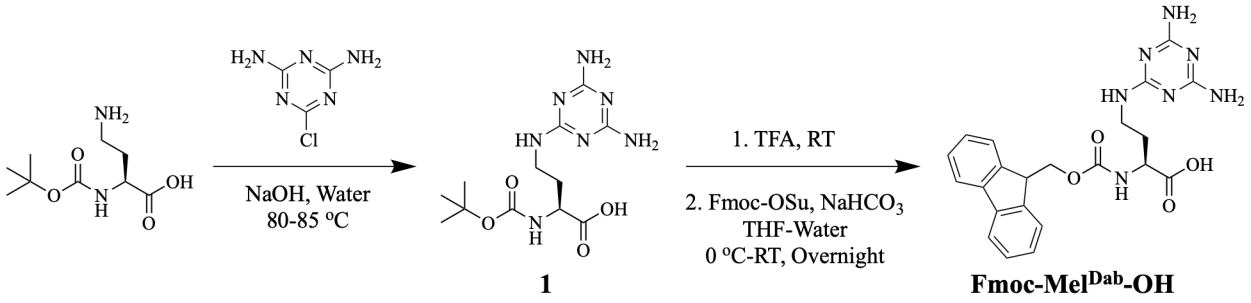
EXPERIMENTAL TECHNIQUES **9**

SUPPLEMENTARY FIGURES **11**

REFERENCES **31**

SYNTHESIS AND CHARACTERIZATION

Methods: 100-200 mesh silica gel was used for preparatory column chromatography. An Agilent 1200 analytical High Performance Liquid Chromatography (HPLC) integrated with a DAD SL UV detector and a 6130 Quadrupole LC/MS mass spectrometer was used to obtain HPLC traces and High-Resolution Mass Spectrometry-Electron Spray Ionization (HRMS-ESI) spectra of purified small molecules and peptides. ^{1}H NMR and ^{13}C NMR analyses were performed on Bruker Avance III HD 500 and Bruker Avance III HD 700 NMR spectrometers. Deuterated solvents: dimethyl sulfoxide-d_6 (DMSO-d_6), chloroform-d ($CDCl_3$), and deuterium oxide (D_2O) were used for NMR spectroscopic analysis. A LABCONCO FreeZone lyophilizer was used to dry the purified peptides and partially water-soluble compounds.



Scheme 1. Synthetic Scheme of **Fmoc-Mel^{Dab}-OH**

Synthesis of 2-((tert-butoxycarbonyl)amino)-4-((4,6-diamino-1,3,5-triazin-2-yl)amino) butanoic acid, (1)[1]: To a suspension of 6-chloro-1,3,5-triazine-2,4-diamine (5.6 g, 38.5 mmol) in 20 mL of water, Boc-L-2,4-diaminobutyric acid (7 g, 32.1 mmol) dissolved in water (80 mL) was added with continuous stirring. To this mixture, NaOH (2.57 g, 64.1 mmol, 60 mL of water) was added slowly and stirred overnight at 80-85 °C. The reaction mixture was then cooled down to room temperature (RT) and centrifuged to remove the excess 6-chloro-1,3,5-triazine-2,4-diamine. The separated clear liquid was decanted into a beaker and acidified to pH 5 with 1 M HCl at 0 °C. The contents of the beaker (both liquid and precipitate) showed HRMS mass spectra indicating that product was present. The contents of the beaker was transferred to 50 mL centrifuge tubes, frozen at -80 °C, and lyophilized to yield compound **1** as a white powder (10.2 g, Yield = 97%). NMR and HRMS spectra were recorded on a small portion of the powder purified by means of an Agilent 1260 infinity II preparative LC system fitted with an XBridge BEH C18 OBD Prep column

monitoring the UV absorbance at 215 nm. Solvent system: 0.1% formic acid (A) – acetonitrile (ACN, B); Gradient: 0-50% B for 30 min; 50-100% B for 30-40 min; 100-0% B for 30-40 min: remaining sample was directly used for the next step.

^{1}H NMR (500 MHz, DMSO-d_6), δ (ppm): 7.17 (d, 1H, J = 7.9 Hz), 6.46 (broad s, 1H), 6.08 (broad d, 4H), 3.92-3.88 (m, 1H), 3.17-3.12 (m, 2H), 1.93-1.87 (m, 1H), 1.74-1.67 (m, 1H), 1.38 (s, 9H); ^{13}C NMR (700 MHz, DMSO-d_6), δ (ppm): 174.59, 156.06, 78.55, 52.04, 40.46, 37.51, 31.39, 28.67, 28.41; HRMS (ESI) *m/z* calculated for $C_{12}H_{21}N_7O_4$: 328.17 $[M+H]^+$; found, 328.2 $[M+H]^+$.

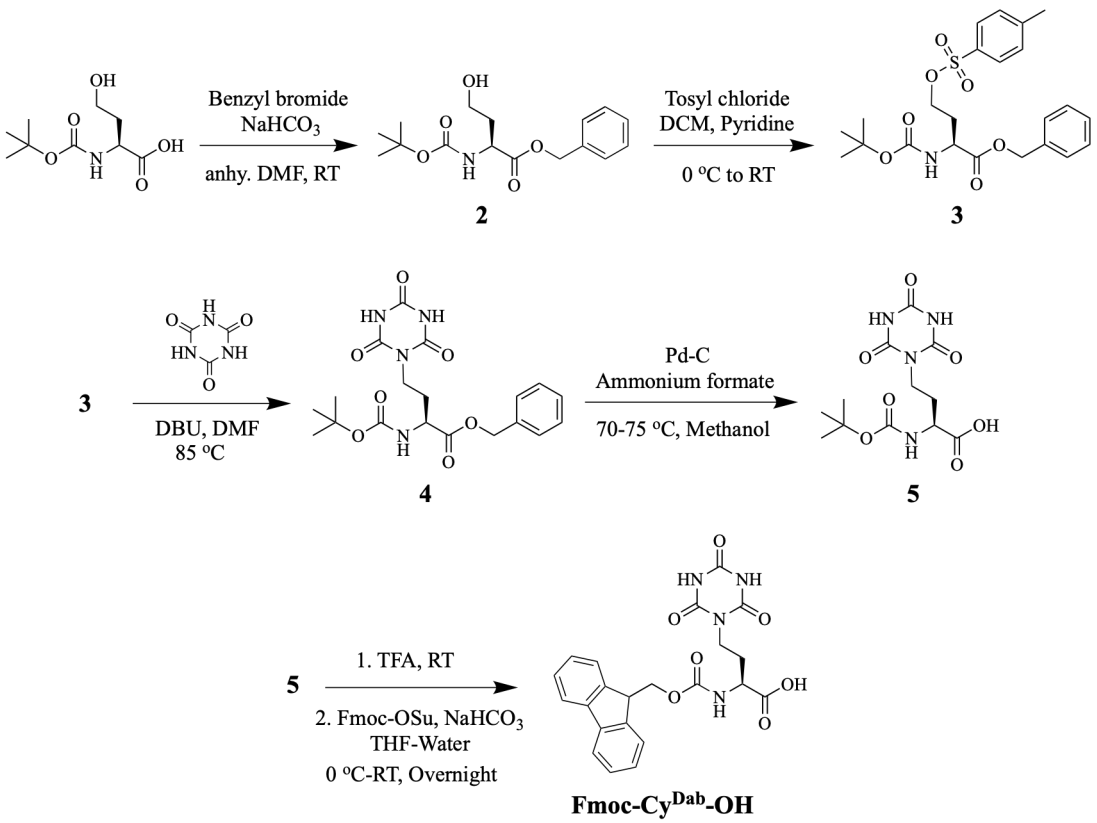
Synthesis of 2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-((4,6-diamino-1,3,5-triazin-2-yl)amino)butanoic acid, (Fmoc-Mel^{Dab}-OH): Fmoc-Mel^{Dab}-OH was synthesized by modification of a previously reported procedure.[1] Compound **1** (10 g, 44 mmol) was dissolved in trifluoroacetic acid (TFA, 80 mL) and water (1 mL). The resulting suspension was stirred at RT for three hours, at which time complete consumption of **1** was confirmed by HRMS-ESI. Most of the TFA was removed by purging the sample with N_2 gas and the residue was suspended in 200 mL of water. The resulting solution was neutralized to pH 7 using $NaHCO_3$ and an additional 1.5 equiv. (5.5 g, 66 mmol) of the $NaHCO_3$ solution was added to keep the solution basic for next step. The mixture was stirred at 0 °C for 20 min followed by the slow addition of 1.5 eq. Of Fmoc-OSu (22 g, 66 mmol) dissolved in 200 mL of ice-cold tetrahydrofuran. The resulting mixture was stirred at 0 °C for 2 hours and then overnight at RT after which 150 mL of water was added to the reaction and the THF was evaporated at reduced pressure. The aqueous layer was washed with ethyl acetate (EtOAc, 2 × 100 mL) and acidified to pH 1 using 1 M HCl at 0 °C. The acidified aqueous layer was washed three times with 50 mL portions of EtOAc and the organic layer was dried over anhydrous Na_2SO_4. The crude product was obtained by removing EtOAc under reduced pressure and was purified by Teledyne Isco CombiFlash Rf+ purification system (monitored by the UV absorbance at 215 nm) equipped with RediSep C18Aq 150 g Gold column (Solvent system: water-ACN; Gradient: 100% water for 0-5 min, 0-65% ACN for 5-30 min, 100% ACN for 30-40 min; Flow rate: 75 mL/min), and lyophilized to provide **Fmoc Mel^{Dab}-OH** as a white powder (4.2 g, yield = 30%).

Characterization data of **Fmoc-Mel^{Dab}-OH** are provided below in Figures 16, 17, 18 and 19.

^{1}H NMR (700 MHz, DMSO-d_6), δ (ppm): 7.89 (d, 2H, J = 7.28 Hz), 7.74 (d, 2H, J =7.21 Hz), 7.67 (d, 1H, J =7.77 Hz), 7.43-7.32 (m, 4H), 6.54 (broad s, 1H), 6.20-6.08 (broad d, 4H), 4.29-4.23 (m, 3H), 4.00 (broad d, 1H), 3.30-3.28 (m, 2H), 1.99-198 (m, 1H), 1.77-1.76 (m, 1H); ^{13}C NMR (700 MHz, DMSO-d_6), δ (ppm): 174.54, 166.49, 156.62, 144.31, 141.18, 128.12, 127.58, 125.78, 120.59, 66.14, 52.46, 47.10, 37.59, 31.41; HRMS (ESI) *m/z* calculated for $C_{22}H_{23}N_7O_4$: 450.19 $[M+H]^+$; found, 450.2 $[M+H]^+$.

Synthesis of benzyl (tert-butoxycarbonyl)homoserinate, (2)[2]: Benzyl bromide (5.4 mL, 45 mmol) was added to a stirred mixture of N-tert-butoxycarbonyl-L-homoserine (5 g, 22 mmol) and $NaHCO_3$ (2.8 g, 33 mmol) in DMF (40 mL) at room temperature. TLC was used to follow product formation. After 3 days of continuous stirring, the reaction mixture was partitioned between water (150 mL) and EtOAc (200 mL). The resulting organic layer was washed with 10% sodium thiosulfate (30 mL) and then brine (50 mL), dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to yield an oily liquid. The crude product was purified using silica gel column chromatography with 55% ethyl acetate-hexane as eluent to give **2** as a colourless viscous oil (4.2 g, yield = 62%).

^{1}H NMR (700 MHz, DMSO-d_6), δ (ppm): 7.38-7.31 (m, 5H), 7.23 (d, 1H, J = 7.7 Hz), 5.16-5.07 (m, 2H), 4.61 (t, 1H, J = 5 Hz), 4.16-4.13 (m, 1H), 3.46-3.44 (m, 1H), 1.85-1.80 (m, 1H), 1.75-1.70 (m, 1H), 1.39 (s, 9H); ^{13}C NMR (700 MHz, DMSO-d_6), δ (ppm): 173.33, 156.10, 136.51, 128.77, 128.38, 128.14, 78.61, 66.20, 57.63, 51.43, 34.15, 28.57; HRMS (ESI) *m/z* calculated for $C_{16}H_{23}NO_5$: 332.15 $[M+Na]^+$; found, 332.2 $[M+Na]^+$.



Scheme 2. Synthetic Scheme of **Fmoc-Cy^{Dab}-OH**

Synthesis of benzyl N-(tert-butoxycarbonyl)-O-tosylhomoserinate, (3)[3]: p-Toluenesulfonyl chloride (TsCl, 3.9 g, 20.4 mmol, 1.5 equiv.) was added to an anhydrous DCM (30 mL) solution of **2** (4.2 g, 13.6 mmol). After purging the solution with nitrogen for 10 minutes, anhydrous pyridine (4.46 mL, 55.7 mmol) was added, and the reaction mixture was stirred overnight at 0 °C. Product formation was monitored by TLC. The reaction mixture was washed with a 10% citric acid solution, dried over anhydrous Na_2SO_4 and evaporated under reduced pressure to yield a yellow oil. Silica gel column chromatography of the crude mixture eluting with 50% hexane-ethyl acetate gave compound **3** as a pale-yellow oil (3.8 g, yield = 60%).

^{1}H NMR (700 MHz, $CDCl_3$), δ (ppm): 7.78 (d, 2H, J = 8.2 Hz), 7.40-7.34 (m, 7H), 5.17-5.15 (m, 2H), 4.37 (m, 1H), 4.14-4.09 (m, 2H), 2.46 (s, 3H), 2.28-2.25 (m, 1H), 2.15-2.12 (m, 1H), 1.42 (s, 9H).

Synthesis of benzyl 2-((tert-butoxycarbonyl)amino)-4-(2,4,6-trioxo-1,3,5-triazinan-1-yl)butanoate, (4)[4]: Cyanuric acid (5.3 g, 41 mmol) was added to a 250 mL round bottom flask

that was then purged with nitrogen for 10 minutes. Dry DMF (50 mL) was added followed by the dropwise addition of DBU (1.47 mL, 9.7 mmol). The mixture was heated to 85 °C until the cyanuric acid dissolved completely, then compound **3** (3.8 g, 8.2 mmol) was added to the clear reaction mixture and stirred at 85 °C for 3 days. After cooling to room temperature, the reaction mixture was poured into water and extracted with EtOAc. The aqueous layer was washed multiple times with EtOAc and any solid material was removed by filtration. The combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. The crude product was purified by Teledyne Isco CombiFlash Rf+ purification system (monitored by the UV absorbance at 215 nm) connected to a RediSep C18Aq 150 g Gold column (Solvent Gradient: 100% water for 0-5 min, 0-50% ACN for 5-24 min, 50-100% ACN for 24-30 min, 100% ACN for 30-40 min; Flow rate 75 mL/min) to provide compound **4** as a white solid (2.3 g, yield = 67%).

^{1}H NMR (700 MHz, DMSO-d_6), δ (ppm): 11.40 (s, 2H), 7.37-7.32 (m, 6H), 5.11-5.10 (d, 2H, J = 3.3 Hz), 4.07-4.04 (m, 1H), 3.75-3.68 (m, 2H), 2.05-2.01 (m, 1H), 1.89-1.83 (m, 1H), 1.37 (s, 8H), 1.28 (s, 1H); ^{13}C NMR (700 MHz, DMSO-d_6), δ (ppm): 172.31, 155.91, 150.27, 149.10, 136.36, 128.82, 128.44, 128.14, 78.78, 66.45, 52.19, 38.26, 29.27, 28.61; HRMS (ESI) *m/z* calculated for $C_{19}H_{24}N_4O_7$: 419.16 $[M-H]^-$; found, 419.10 $[M-H]^-$.

2-((tert-Butoxycarbonyl)amino)-4-(2,4,6-trioxo-1,3,5-triazinan-1-yl)butanoic acid (5): Compound **4** (2.3 g, 5.4 mmol) and 10% Pd/C (10% of the weight of **4**) were dissolved in 35 mL of dry methanol. Ammonium formate (1.7 g, 27.3 mmol) was added to this solution under a N_2 atmosphere and heated at reflux for 3 hours. The extend of reaction was determined by HRMS-ESI. Upon completion, the catalyst was removed through centrifugation, and the resulting supernatant was filtered. The filtrate was concentrated under reduced pressure to give crude **5** as a white solid which was used without further purification.

HRMS (ESI) *m/z* calculated for $C_{12}H_{18}N_4O_7$: 329.11 $[M-H]^-$; found, 329.10 $[M-H]^-$.

Synthesis of 2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-(2,4,6-trioxo-1,3,5-triazinan-1-yl)butanoic acid, (Fmoc-Cy^{Dab}-OH): Compound **5** (1.78 g) was dissolved in TFA (20 mL) and water (0.5 mL). The reaction mixture was stirred at RT for ca. 2 hours at which time completion of the reaction was confirmed by HRMS-ESI. The bulk of the TFA was removed by purging with N_2 gas. The remaining residue was resuspended in 100 mL of water, and the solution was

neutralized to pH 7 using $NaHCO_3$. An additional, 1.5 equivalent of $NaHCO_3$ (0.68 g, 8.1 mmol) was added and the resulting solution was stirred at 0 °C for 20 minutes. Fmoc-OSu (2.7 g, 8.1 mmol) in 100 mL of THF (cooled in ice bath) was added slowly to the reaction mixture and stirred at 0 °C for 2 hours and then overnight at RT. The THF was removed under reduced pressure and the aqueous layer obtained was first washed with EtOAc (2×100 mL), then acidified to pH 1 with 1 M HCl at 0 °C and extracted again with EtOAc (3×120 mL). The combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. The crude product was purified by Teledyne Isco CombiFlash Rf+ purification system (monitoring the UV absorbance at 215 nm) equipped with RediSep C18Aq 150 g Gold column (Solvent Gradient: 100% water for 0-10 min, 0-100% ACN for 10-30 min, 100% ACN for 30-40 min; Flow rate 75 mL/min) to provide **Fmoc-Cy^{Dab}-OH** as a white powder (1.4 g, yield = 56%).

Characterization data for **Fmoc-Cy^{Dab}-OH** are provided below as Figures 20, 21, 22 and 23.

^{1}H NMR (500 MHz, DMSO-d_6), δ (ppm): 12.71 (broad s, 1H), 11.41 (s, 2H), 7.89 (d, 2H, J = 7.5 Hz), 7.74 (m, 3H), 7.42 (t, 2H, J = 7.4 Hz), 7.36-7.32 (m, 2H), 4.30-4.22 (m, 3H), 4.02-3.97 (m, 1H), 3.80-3.70 (m, 2H), 2.05-2.03 (m, 1H), 1.89-1.85 (m, 1H); ^{13}C NMR (700 MHz, DMSO-d_6), δ (ppm): 173.75, 156.53, 150.30, 149.13, 144.28, 141.16, 128.11, 127.59, 125.83, 120.57, 66.23, 52.39, 47.10, 38.39, 29.51; HRMS (ESI) *m/z* calculated for $C_{22}H_{20}N_4O_7$: 451.13 $[M-H]^-$; found, 451.10 $[M-H]^-$.

Solid phase peptide synthesis: Peptide synthesis was performed manually using a Wang resin (100-200 mesh, 0.87 mmol/g loading) following the modified Fmoc protection-deprotection procedure.[1] The Fmoc-protected monomers used in this work for the synthesis of proto-nucleic acids are **Fmoc-Mel^{Dab}-OH**, **Fmoc-Cy^{Dab}-OH** and Fmoc-Glu(OtBu)-OH. The synthesis began by swelling the resin (0.05 M) in 2 mL of dry DMF (TORVIQ 10 mL luer lock fritted syringe) for 1 hour and then washing it twice with 2 mL portions of DMF. To achieve maximum loading of the resin, the first coupling to the C-terminus monomer was carried out with 5 equiv. of the Fmoc-Glu(OtBu)-OH, 5 equiv. of HBTU and 12.5 equiv. of DIPEA for 3 hours. The resin was washed twice with 2 mL of DMF followed by 2 mL of DCM between each coupling and deprotection step. After the first coupling, the unreacted hydroxyl groups were capped by adding 5 equiv. of Ac_2O and 10 equiv. of DIPEA to the resin. Deprotection of the Fmoc group was accomplished with 20%

piperidine in DMF. From the second coupling step onward, the Fmoc protected monomer (corresponding to the respective proto-nucleic acid), the coupling reagent and DIPEA were used in a 1:1:2.5 ratio and 10 % piperidine in DMF was used for deprotection to reduce any side reaction or degradation of Cy and Mel. HBTU was used for **Fmoc-Cy^{Dab}-OH** and Fmoc-Glu(OtBu)-OH coupling, whereas DEPBT was used for **Fmoc-Mel^{Dab}-OH** coupling to minimize the competing lactamization of Dab under these coupling conditions.[5] These coupling steps were repeated until the Kaiser test (detection of free primary amines) is negative on a small amount of resin before the complete deprotection of Fmoc group for the next coupling. After the completion of all coupling and deprotection steps, deprotection of the Boc groups on the peptide and cleavage of the C-terminal of peptide from resin was carried out using a mixture of 95% TFA, 2.5% TIS and 2.5% water (2 mL). After 3 hours, the TFA solution of the peptide was reduced in volume by purging with N_2 gas. The remaining TFA solution was cooled in an ice bath and chilled diethyl ether was added to precipitate the peptides. The precipitated peptide was separated by centrifugation and washed again with chilled diethyl ether and dried to provide the peptide in powder form.

The crude peptide was dissolved in water and purified using an Agilent 1260 infinity II preparative LC system fitted with XBridge BEH C18 OBD Prep column (Solvent system: 0.2% TFA in water (A) - 0.2% TFA in ACN (B); Gradient: 0% B for 0-8 mins, 0 to 30% B for 8-30 mins, 30 to 100% B for 30-40 mins and 100 to 0% B for 40-50 mins) monitored at 215 nm and 255 nm. The purification step was repeated to remove "shoulder peaks" and small amounts of similarly retained compounds from the peptide. The purified peptides were lyophilized to dryness, and the purity was checked by means of an Agilent 1200 analytical HPLC (with DAD SL) connected to 6130 Quadrupole LC/MS and equipped with Kinetex 2.6 µm XB-C18 column. The percentage yield ((actual yield/theoretical yield) × 100) of peptides varied between 50 to 60 % depending on the peptide sequences. ^{1}H NMR spectral characterization of all the peptides was carried out in D_2O with an internal proton intensity standard isopropyl alcohol (IPA) to obtain an accurate measure of peptide concentrations. Following the ^{1}H NMR spectral collection with IPA, the IPA was removed from the samples by lyophilization.

Characterization data of α-(PNA)s are provided in Supplementary Figures: α-$(M)_4$, Figures 24, 25 and 26; α-$(C)_4$, Figures 27, 28 and 29.

SAMPLE PREPARATION FOR PHYSICAL STUDIES

Stock solutions of peptides were prepared in water containing a very small amounts of NH_4OH (0.74 M) to adjust sample pH to keep the peptides completely soluble and stable over long periods at either 4 °C or RT. pH measurements were performed with VWR Scientific Model 8100 pH/Temp/mV meter. Deionized water (18.2 MΩ/cm) from Barnstead NANO pure Diamond water purification system was used for all the sample preparations and experimental studies.

Two methods were used to prepare samples for analysis: Method 1: A 1 μL volume of a 10 mM sample of α-$(M)_4$ and 1 μL volume of a 10 mM sample of α-$(C)_4$ were added to a solution of 40 μL of 1 M sodium phosphate buffer (pH 7), vortexed and diluted to a final volume (200 μL) with water and then vortexed again to provide a final mixture containing 1:1 α-$(M)_4$/α-$(C)_4$ (50 μM each) in 200 mM sodium phosphate buffer (final pH = 7.2). This method of preparation is denoted as the "High Concentration" (HC) assembly method. In this case, the assembly forms at a higher concentration (0.24 mM) and higher ionic strength (0.952 M sodium phosphate buffer) and then is diluted to the final volume. This method was used to prepare the samples for most of the experiments, unless otherwise stated. In a second method of sample preparation, "Low Concentration" (LC), separate solutions containing 100 μL of 100 μM α-$(M)_4$ and 100 μL of 100 μM α-$(C)_4$ in 400 mM of sodium phosphate buffer solution were prepared and then mixed to provide 200 μL of a solution containing a 1:1 mixture of α-$(M)_4$/α-$(C)_4$ (50 μM each) in 200 mM sodium phosphate buffer. In this case, the peptides are together only at lower concentration (50 μM) compared to the HC (0.24 mM) method. The final concentration of sodium phosphate buffer was 200 mM in all samples, except for the samples prepared for pH dependent studies (which used a 450 mM sodium phosphate buffer/450 mM cacodylic acid buffer) and native mass spectrometry (which used 200 mM ammonium acetate buffer). Samples for pH dependent experiments were prepared by the HC method and the final buffer and sodium ion concentrations in all those samples were adjusted to 450 mM and 1M, respectively by the addition of a different pH buffer and 5 M NaCl. For samples with final pH of 6 to 8, sodium phosphate buffer was used and for samples with final pH less than 6, cacodylic acid buffer was utilized.

EXPERIMENTAL TECHNIQUES

Circular dichroism (CD) spectral measurements were performed on Jasco J-810 Spectro-polarimeter equipped with a Peltier temperature controller cell holder. To conserve sample, a 1 mm quartz cuvette was used to record the CD spectra. For all the normal CD spectral measurements, the wavelength range was set at 300-200 nm with a 100 nm/min scanning speed; 1 second response time; 1 nm data pitch; and 4 nm band width. Four accumulations were collected and corrected for blanks. CD spectra were converted to units of molar circular dichroism ($\Delta\varepsilon$), using $\Delta\varepsilon = \theta/(32980 \cdot c \cdot l)$, where θ is CD signal in units of mdeg, c is molar concentration in amino acid residues and l is cuvette pathlength in cm.

Atomic Force Microscopic (AFM) imaging of α-$(M)_4$/α-$(C)_4$ assembly was performed with Nanoscope IIIa (Digital Instruments) in tapping mode using silicon tips (Vistaprobes, 48 Nm^{-1}). A 1:1 mixture of α-$(M)_4$/α-$(C)_4$ (400 μM each) in 200 mM sodium phosphate buffer, pH 7.2 was used to image the morphology of the assemblies. A 3 μL sample was drop-cast onto a freshly cleaved mica surface, incubated at 4 °C for one hour and blown dry with N_2. The larger salt particles that crystalized over the sample were washed off with aliquots of water, and the sample was again blown dry with N_2 before imaging.

Native Mass spectrometric (nMS) analyses were performed on Orbitrap platforms (Q-Exactive Ultra High Mass Range Orbitrap MS [UHMR] and Orbitrap Exploris 480, Thermo Fisher Scientific, Bremen, Germany) using nano-ESI from in-house prepared borosilicate capillaries. Samples were diluted to 1–50 μM in 200 mM ammonium acetate. Data were acquired with resolution 200,000 for UHMR and 240,000 for Exploris 480 at 200 *m/z*. UHMR instrument settings were optimized for high- and low-*m/z* ion transmission, in-source trapping enabled (−50 V), ion transfer optics set to 5/4/2 V (injection flatapole/interflatapole lens/bent flatapole), and N_2 was used as collision gas (UHV 2×10^{-10} to 5×10^{-10} mbar).

Polyacrylamide gel electrophoresis (PAGE) of the assemblies was carried out by means of a 10% PA denaturing gel containing 12.5 mM KCl and 1X TBE (1 mm thickness glass plates;14 cm × 14 cm). The gels were polymerized for 2 hours and incubated at 4 °C for one hour. A 2 μL (50 μM in each strand) sample along with 7 μL of 12.5 mM KCl and 1 μL of 10X loading dye (50% Glycerol, 1% SDS, 10 mM EDTA pH 8, 0.25% Bromophenol blue and 0.25% xylene cyanole) were loaded

on the gel and run at 250 V and 4 ºC. After one hour, the gel was stained with SYBR Green II for 15 min. with gentle shaking and then destained in water for 5 mins. The destained gels were then imaged using Typhoon FLA 9500 (GE Health care) laser scanner and processed with an Image J 1.51j8 version from NIH. The oligonucleotide sequence d(TGGGGT), well known to form a four-stranded G-quadruplex structure, was used as a mobility standard (Figure 2D, Lane 1).

Determination of degree of strand assembly. Assembly was monitored by circular dichroism at 218 nm over a range of 20-100 μM total strand concentration (S_0), allowing 240 minutes for equilibration (Figure S11). The concentration of free strands (S) was calculated by the relation

$$S = S_0\left(\frac{\Delta\varepsilon_S - \Delta\varepsilon}{\Delta\varepsilon_S - \Delta\varepsilon_H}\right)$$

where $\Delta\varepsilon_S$ = 10.9 and $\Delta\varepsilon_H$ = 30.4 $M^{-1}cm^{-1}$ are the molar circular dichroism of the unassembled strands and assembled hexad, respectively, expressed in strand concentration. The concentration of assembled hexaplex (H) was $H = (S_0 - S)/6$.

Computational models. The proto-Nucleic Acid Builder[6] was used to generate preliminary model structures for these assemblies. Because Cy and Mel nucleobases are connected to the backbone in different patterns (endocyclic for the former and exocyclic for the latter), the first three carbon atoms were frozen in a manner that allowed the formation of a periodic backbone. A search was performed on the remaining torsional angles in the backbone to find plausible structures with no steric clashes. The initial models were further refined with energy minimization using the Open Babel packate.[7]

Parallel and antiparallel hexameric model structures were constructed with the proto-Nucleic Acid Builder (pNAB) using a development version that adds an inter-residue peptide planarity filter. Each model comprised six 16-residue strands assembled as a hexad from the peptide backbone with the cyanuric acid/melamine base pair. The helical rise was fixed at 3.40 Å and all remaining rigid-body parameters (x- and y-displacement, inclination, tip, shear, stretch, stagger, buckle, propeller and opening) were set to zero, while the helical twist was scanned from 0° to 60° in 1° increments, giving 61 independent helical configurations. For each configuration, four rotatable backbone dihedrals were enumerated exhaustively on a 2° grid, corresponding to $(360/2)^4 \approx 1.05 \times 10^9$ conformers per configuration. The four rotatable bonds correspond to the psi and phi angles

in each peptide unit. The intra-residue omega angle was fixed at 180°. Additionally, the two exocyclic carbons that connect the bases to the backbone were manually fixed in order to maintain the same attachment vector for cyanuric acid and melamine bases to the peptide backbone despite using an exo-cyclic amine for melamine to connect the backbone. A conformer was accepted as a candidate when the distance between the head and tail backbone atoms joining successive periodic units was below 0.8 Å, when its MMFF94 energies fell below thresholds of 2 and 4 kcal/mol for the bond and angle terms and 8, 150 and 1000 kcal/mol per nucleotide for the torsional, van der Waals and total energies, and when the peptide inter-residue linkage dihedral deviated from planarity by no more than 8°. To simplify the pNAB search, the glutamic acid side chains were not included during the dihedral scans. Instead, fully extended chains were attached to the generated conformers in a post-processing step.

SUPPLEMENTARY FIGURES

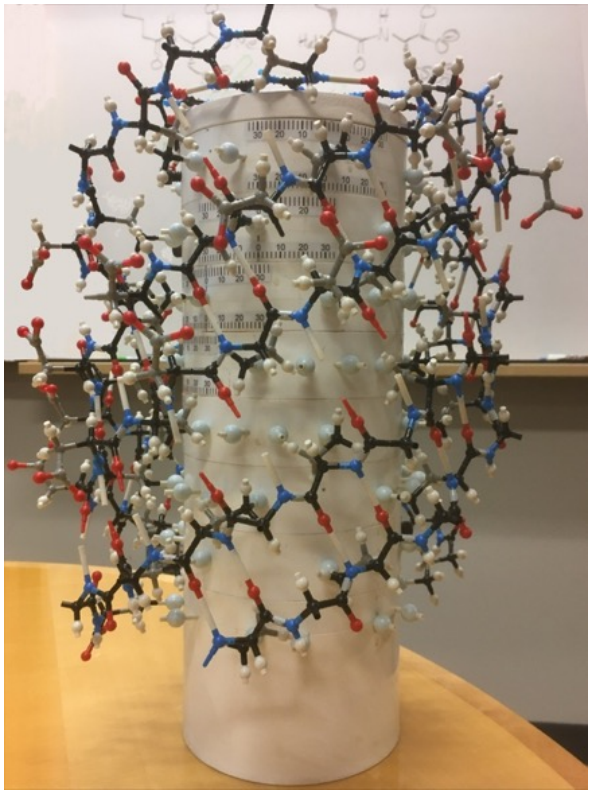


Figure 1. Physical model of α-peptide backbones attached to stacked hexads. HGS Molecular Models were used for the backbones, with a scale of 1 cm = 1 Å (Maruzen International Co., Ltd., Japan). Central tower representing connection points to a stack of hexads, with variable twist, was homebuilt from PVC tubing with a scale of 1 cm = 1 Å (tolerance ± 0.1 Å). The helical twist for the model shown was set to 30°.

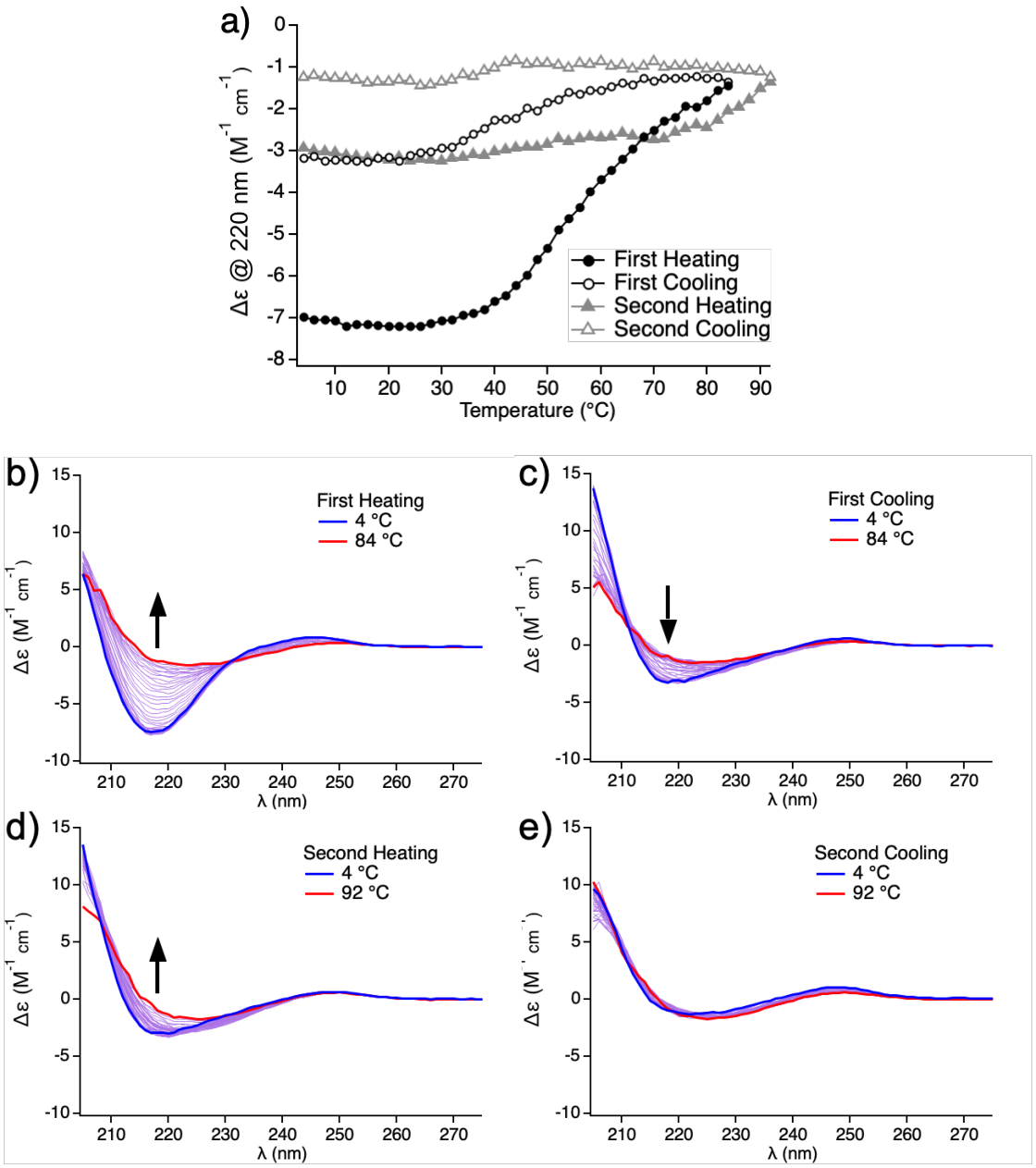


Figure 2. Repeated thermal denaturation of the $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assembly as monitored by CD. a) Heating and cooling curves generated by monitoring CD signal at 220 nm of a sample containing a 1:1 mixture of $\alpha\text{-}(M)_4$ and $\alpha\text{-}(C)_4$ (100 μM each). The first heating trace (solid black circles) was followed immediately by a first cooling trace (open black circles). A second heating trace (solid grey triangles) of the same sample was started 11 hr after the first cooling trace was completed. The sample remained at 4 °C for the intervening 11 hr. The second cooling trace (open grey triangles) was initiated immediately after the second heating trace finished. Note that the similar CD signal at the end of the first cooling trace and at the start of the second heating trace (both at 4 °C) indicates that additional $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assembly did not reform during the 11 hr incubation at 4 °C. b-e) CD spectra collected during the first and second heating-cooling cycles, from which curves shown in (a) were derived. The sample was prepared by the HC assembly method, and the final assembly was in 200 mM phosphate buffer at pH 7.2.

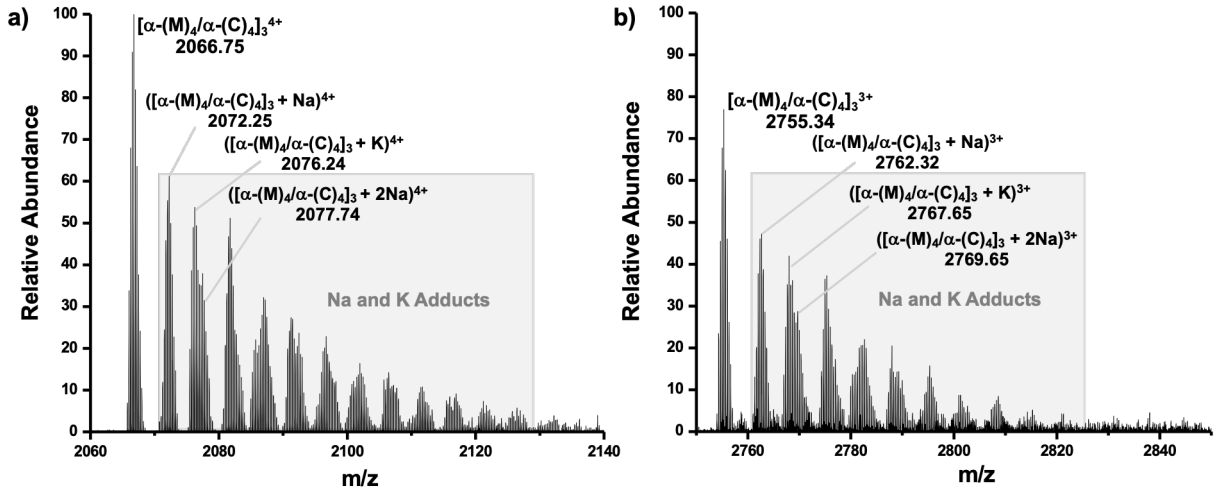


Figure 3. Expanded Positive mode native mass spectra (nMS) showing Na and K adducts of 1:1 α-$(M)_4$/α-$(C)_4$ (1 μM each) assemblies with 4 μM EtBr in the quadruply (a) and triply (b) charged hexaplex region of Figure 2E. Sample was prepared in 200 mM ammonium acetate at pH 7.

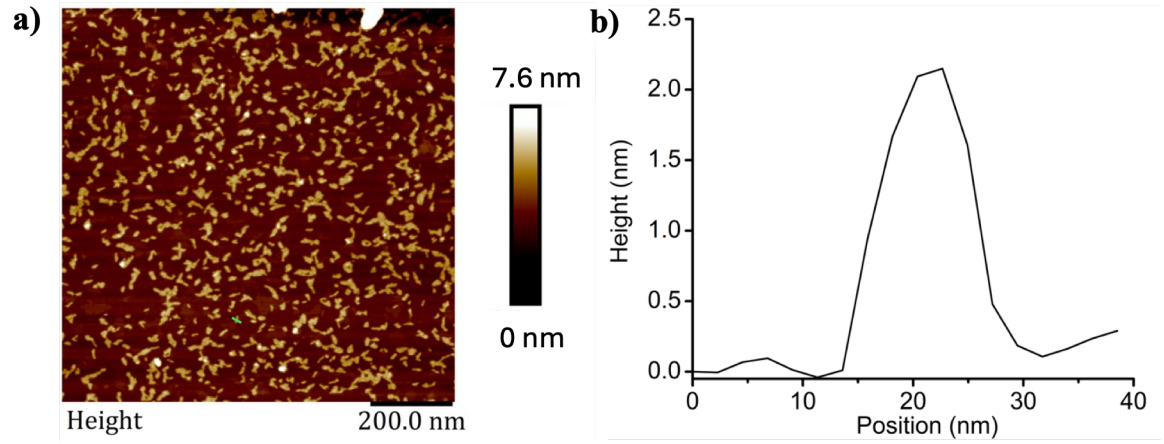


Figure 4. a) AFM image of assemblies formed by a 1:1 mixture of α-$(M)_4$ and α-$(C)_4$ (400 μM each strand) in presence of 1.6 mM of EtBr when dried on a freshly cleaved mica substrate. b) Height measurement of an individual assembly at the location indicated by the green horizontal line in AFM image. The 400 μM strand concentration was used for this experiment because samples containing lower strand concentrations lost too much material when buffer salt deposits were washed off with water after initial sample drying.

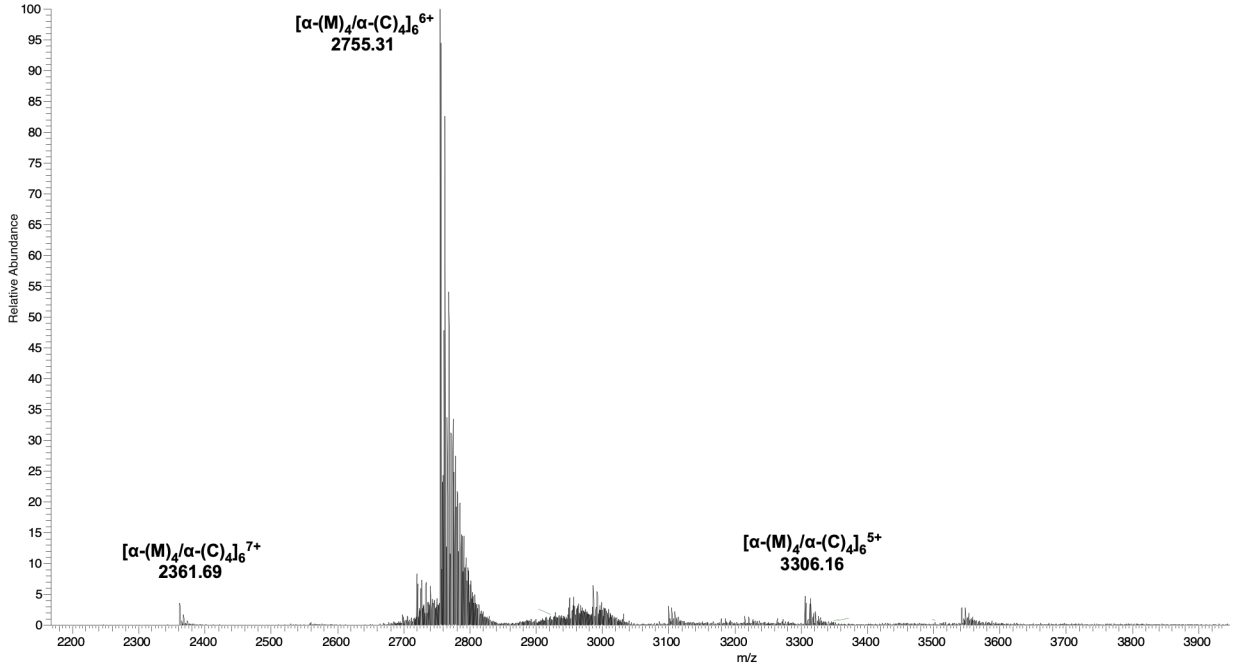


Figure 5. Positive mode high *m/z* native mass spectra (nMS) of 1:1 α-$(M)_4$/α-$(C)_4$ (50 μM each) assemblies without the addition of EtBr. Sample was prepared in 200 mM ammonium acetate at pH 7.

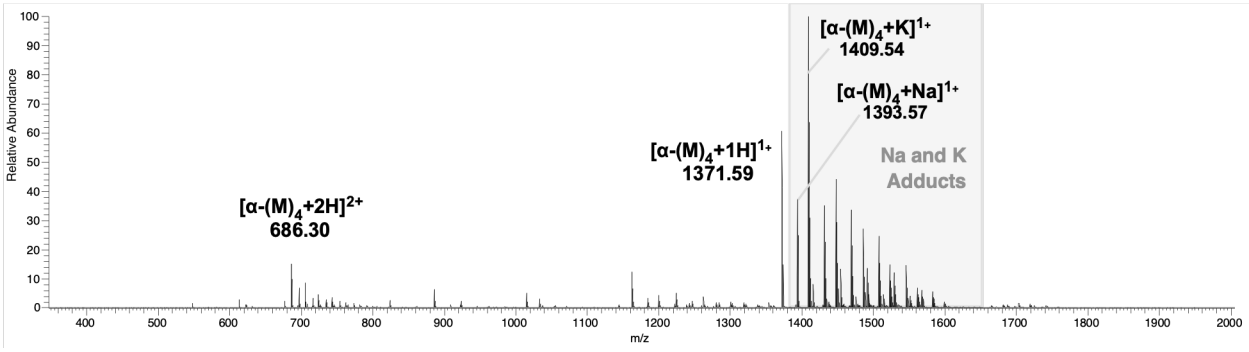


Figure 6. Positive mode low *m/z* native mass spectra (nMS) of 50 μM α-$(M)_4$ in 200 mM ammonium acetate at pH 7. Na and K adducts of α-$(M)_4$ were observed predominantly.

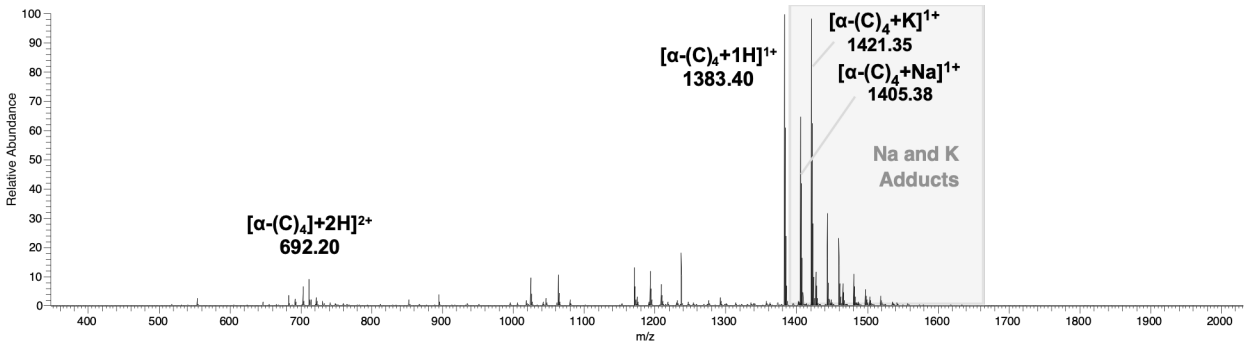


Figure 7. Positive mode low *m/z* native mass spectra (nMS) of 50 μM α-$(C)_4$ in 200 mM ammonium acetate at pH 7. Na and K adducts of α-$(C)_4$ were observed predominantly.

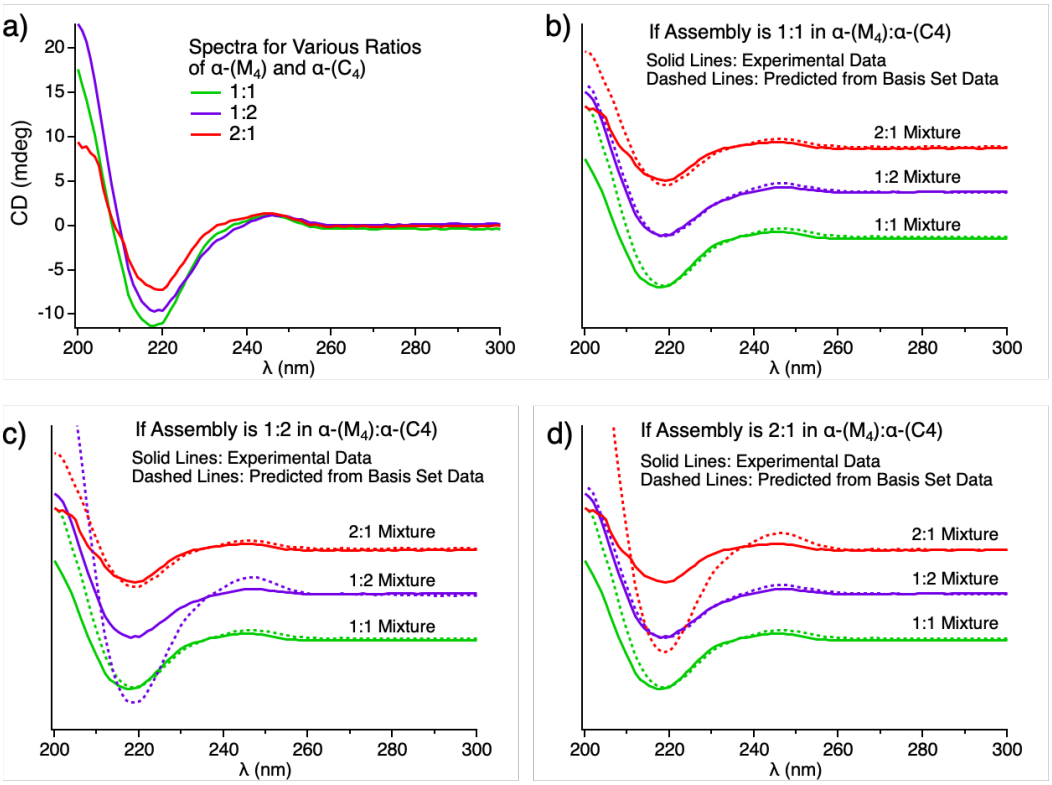


Figure 8. Experimental and predicted CD spectra which confirm that a 1:1 molar stoichiometry of α-(M_4) and α-(C_4) is more favored for assembly than 1:2 or 2:1 molar strand ratios. a) CD spectra of samples containing various molar ratios of α-(M_4) and α-(C_4). In all samples, 1 molar ratio corresponds to 50 μM strand concentration and 2 molar ratio corresponds to 100 μM strand concentration. The reduction of the CD band at 218 nm for samples containing 1:2 and 2:1 molar strand ratios is consistent with the signal of the 1:1 assembly being reduced by the presence of either octapeptides in the single-stranded state, both of which have positive CD band in this wavelength region. b) Comparison of experimental data (solid lines, data shown in (a)) with spectra predicted if the most favored assembly has a 1:1 stoichiometry of α-(M_4) and α-(C_4). Specifically, the dotted line for the spectrum labeled 2:1 mixture is the sum of the spectrum of a 1:1 sample and the spectrum of single stranded α-(M_4). The dotted line for the spectrum labeled 1:2 is the sum of the spectrum of a 1:1 sample and the spectrum of single stranded α-(C_4). The dotted line for the spectrum labeled 1:1 is simply that of a 1:1 sample. c) Comparison of experimental data (solid lines, data shown in (a)) with spectra predicted if the most favored assembly has a 1:2 stoichiometry of α-(M_4) and α-(C_4). Specifically, the dotted line for the spectrum labeled 2:1 mixture is the sum of the spectrum of a 1:1 sample and the spectrum of singled stranded α-(M_4). The dotted line for the spectrum labeled 1:2 is the sum of 2x the spectrum of a 1:1 sample minus 0.5x the spectrum of single stranded α-(M_4). The dotted line for the spectrum labeled 1:1 is simply that of a 1:1 sample. d) Comparison of experimental data (solid lines, data shown in (a)) with spectra predicted if the most favored assembly has a 2:1 stoichiometry of α-(M_4) and α-(C_4). Specifically, the dotted line for the spectrum labeled 2:1 mixture is the sum of 2x the spectrum of a 1:1 sample minus 0.5x the spectrum of singled stranded α-(C_4). The dotted line for the spectrum labeled 1:2 is the sum of the spectrum of a 1:1 sample and the spectrum of single stranded α-(C_4). The dotted line for the spectrum labeled 1:1 is simply that of a 1:1 sample. The spectra used for all predicted spectra are those shown in Figure 2 of the main text. The difference between the experimental data shown in this figure for a 1:1 sample and the predicted 1:1 spectrum illustrates the experimental variation between the samples of Figure 1 and those of this experiment. The best match of the 2:1 and 1:2 mixture experimental and predicted spectra in panel (b) confirm that assembly formed has a 1:1 molar strand stoichiometry.

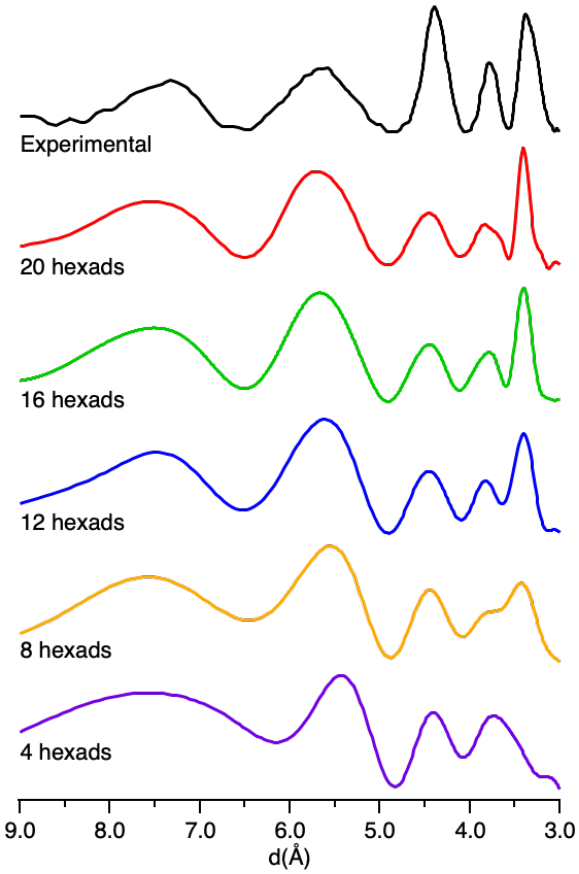


Figure 9. A comparison of the experimental powder diffraction pattern with calculated powder diffraction patterns based on different numbers of co-axially stacked $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assemblies. Based on the comparison of these curves with the experimental data, models with four stacked assemblies (16 hexads) were used for the calculation of other simulated patterns presented in the current study.

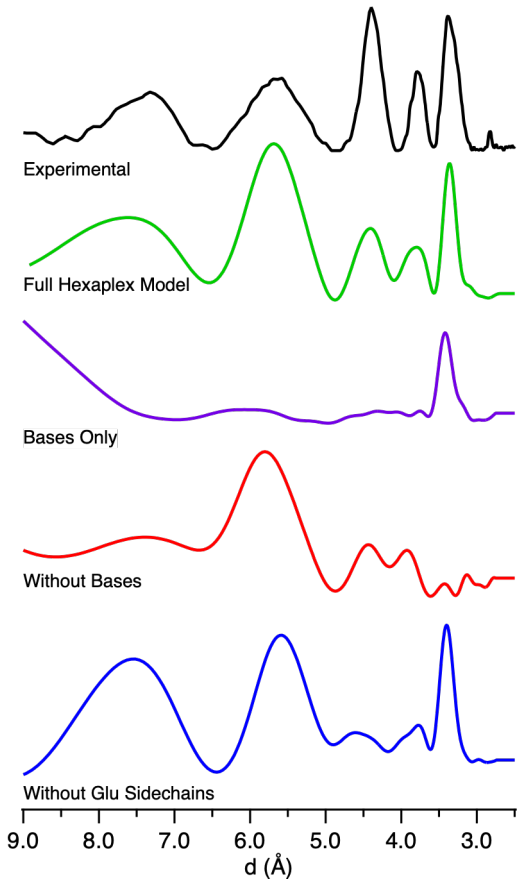


Figure 10. A comparison of the experimental powder diffraction pattern with the calculated powder diffraction pattern for the hexaplex model structure of $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ and components of this structure. The four lower curves are calculated powder diffraction patterns (from top to bottom) of the full hexaplex model of the $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assembly, of the assembly without the Glu sidechains, of the assembly without the bases, and of just the bases of the assembly.

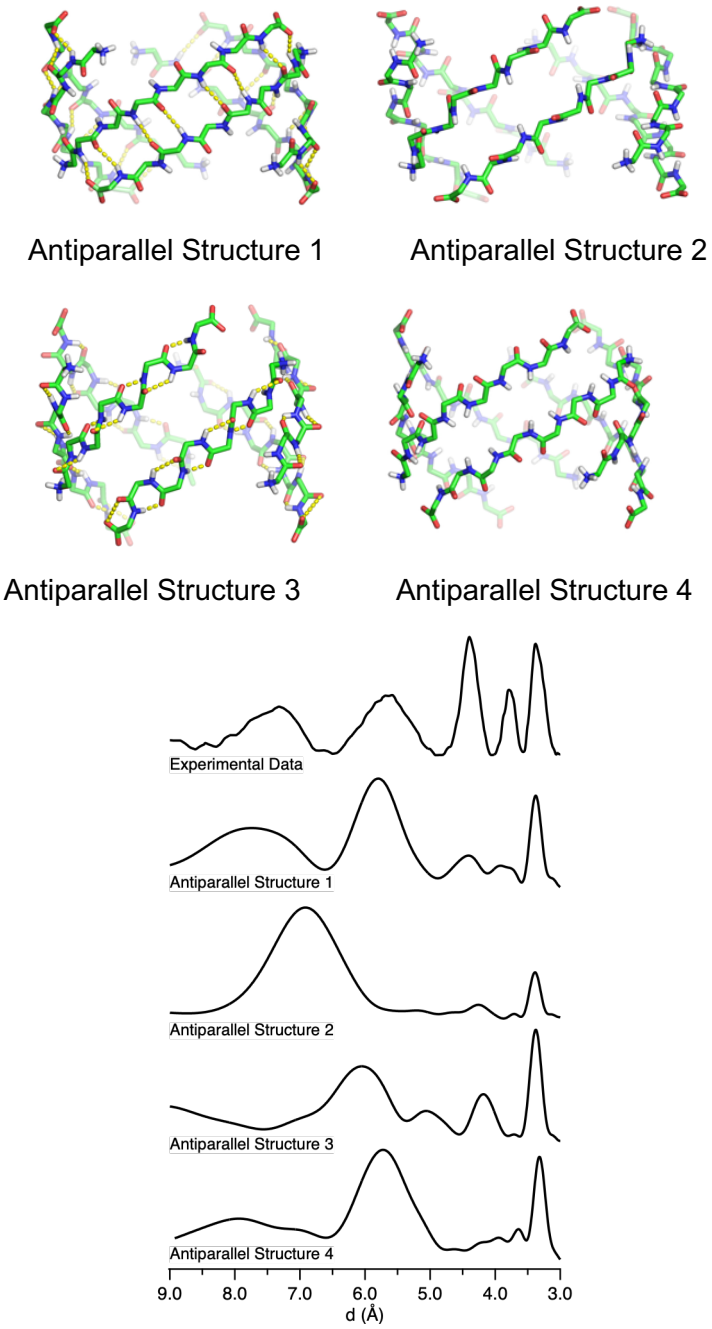


Figure 11. Representative structures of the four families of possible backbone solutions with antiparallel strands identified for the $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ hexaplex along with their predicted powder diffraction patterns. Top) Structure 1 (also shown in Figure 3 of main text) has the β-sheet secondary structure with inter-strand H-bonds between every-other adjacent strand. Structure 2 is an extended conformation that lack any form of intra- or inter-strand H-bonds. Structure 3 is a 2_7 helix, a conformation which was originally proposed keratin[8] but which has not been observed in proteins. Structure 4 is expected to contain significant inter-strand electrostatic repulsions between NH group and carbonyl oxygens. (Hydrogen bonds are shown as yellow dashes). Bottom) A comparison of experimental diffraction data with the four calculated patterns based on the four models shown in (A) for the $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assembly with antiparallel strands.

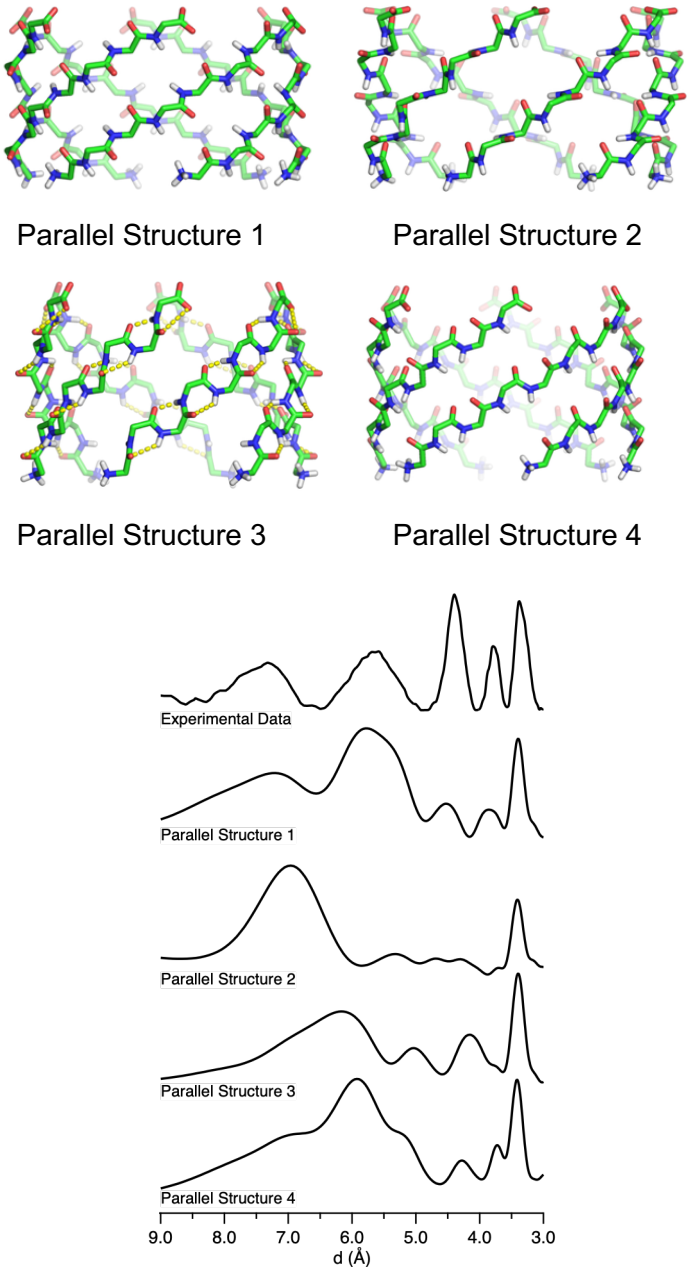


Figure 12. Representative structures of the four families of possible backbone solutions with parallel strands identified for the $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ hexaplex along with their predicted powder diffraction patterns. Top) Structure 1 has the β-sheet secondary structure with inter-strand H-bonds between every-other adjacent strand. Structure 2 is an extended conformation that lack any form of intra- or inter-strand H-bonds. Structure 3 is a 2_7 helix, a conformation which was originally proposed keratin[8] but which has not been observed in proteins. Structure 4 is expected to contain significant inter-strand electrostatic repulsions between NH group and carbonyl oxygens. (Hydrogen bonds are shown as yellow dashes). Bottom) A comparison of experimental data with the four calculated patterns based on the four models shown in (A) for the $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assembly with parallel strands.

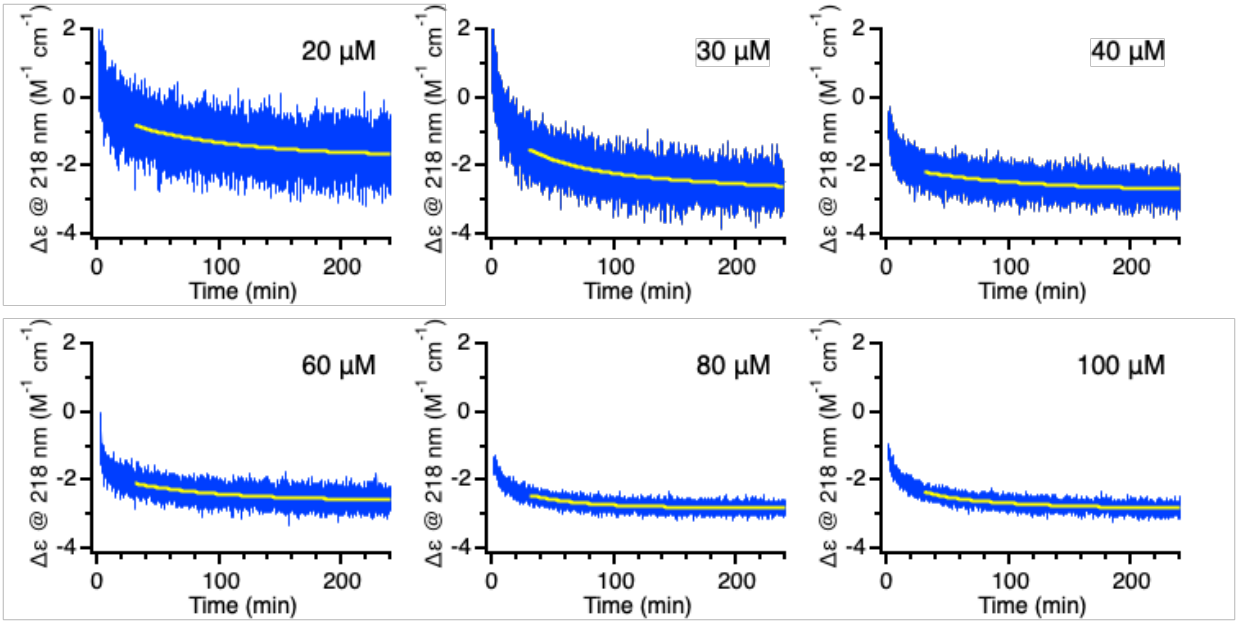


Figure 13. Data used to determine the degree of $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ assembly as a function of mixing concentration and time. Time-dependent measurements of the CD signal at 218 nm for samples with total strand concentrations indicated for 1:1 mixtures of $\alpha\text{-}(M)_4$ and $\alpha\text{-}(C)_4$ generated by mixing of stock solutions with 2x concentrations of $\alpha\text{-}(M)_4$ and $\alpha\text{-}(C)_4$ (LC method of sample preparation). Yellow lines are double exponential fits of experimental data, which provide approximate values for the final signal values of each sample when equilibrium would ultimately be reached. Values obtained by this fitting and extrapolation are between 1 and 5% of the last experimental value collected at 240 min.

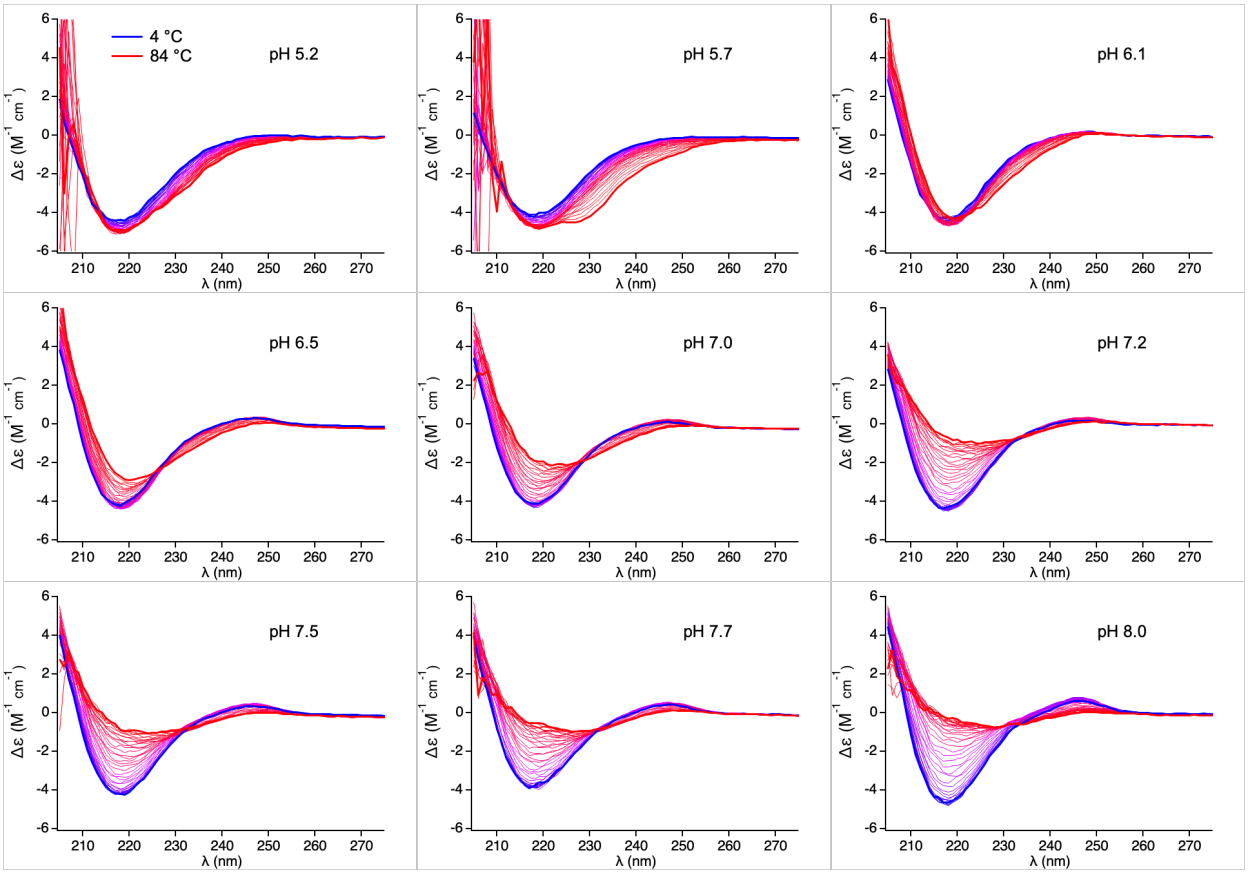


Figure 14. CD Spectra of $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ during thermal denaturation experiments at various sample pH values. CD spectra were recorded as a function of temperature from 4 °C to 84 °C. All samples contained a 1:1 mixture of $\alpha\text{-}(M)_4$ and $\alpha\text{-}(C)_4$, 50 μM in each strand, 450 mM sodium phosphate buffer/450 mM cacodylic acid buffer and 1M NaCl, except pH 7.2 which contained 200 mM sodium phosphate buffer.

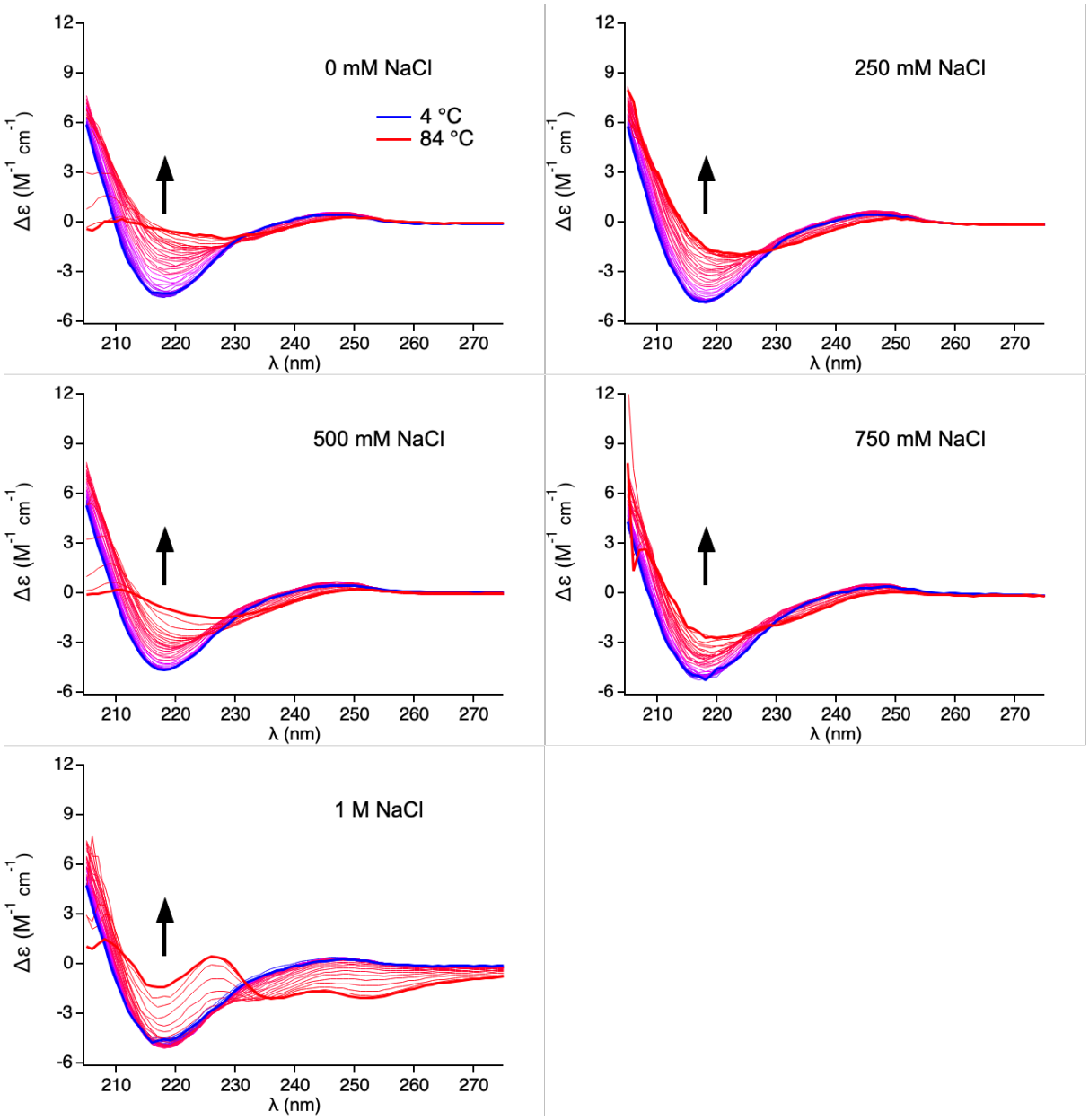


Figure 15. CD Spectra of $[\alpha\text{-}(M)_4/\alpha\text{-}(C)_4]_3$ during thermal denaturation experiments with various NaCl concentrations. CD spectra were recorded as a function of temperature from 4 °C to 84 °C. All samples contained a 1:1 mixture of $\alpha\text{-}(M)_4$ and $\alpha\text{-}(C)_4$, 50 μM in each strand and 200 mM sodium phosphate buffer (the pH was maintained in the 6.8 to 7.2 range).

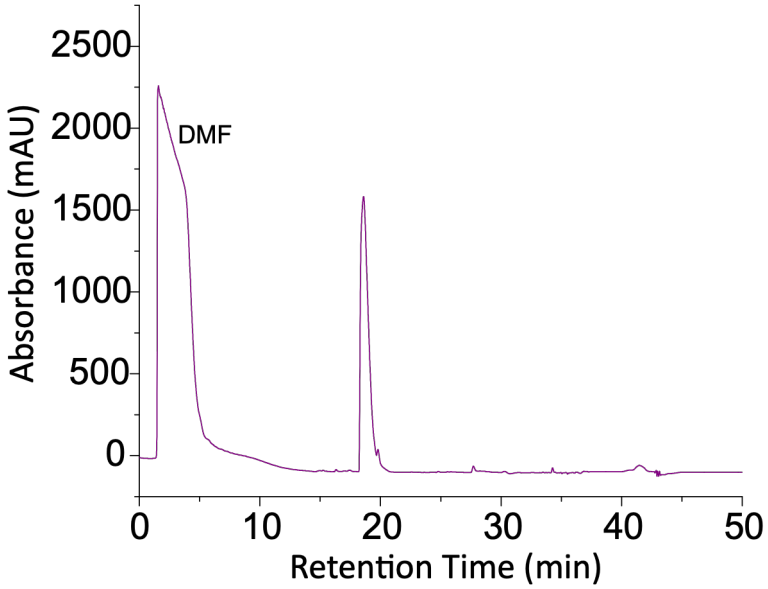


Figure 16. HPLC trace of **Fmoc-Mel^{Dab}-OH** (DMF) on a C18 analytical column monitored at 215 nm with a 0.1% formic acid (A) - acetonitrile (B) solvent system: Gradient; 0-50% B for 0-30 min, 50-100% B for 30-40 min and 100-0% B for 40-50 mins.

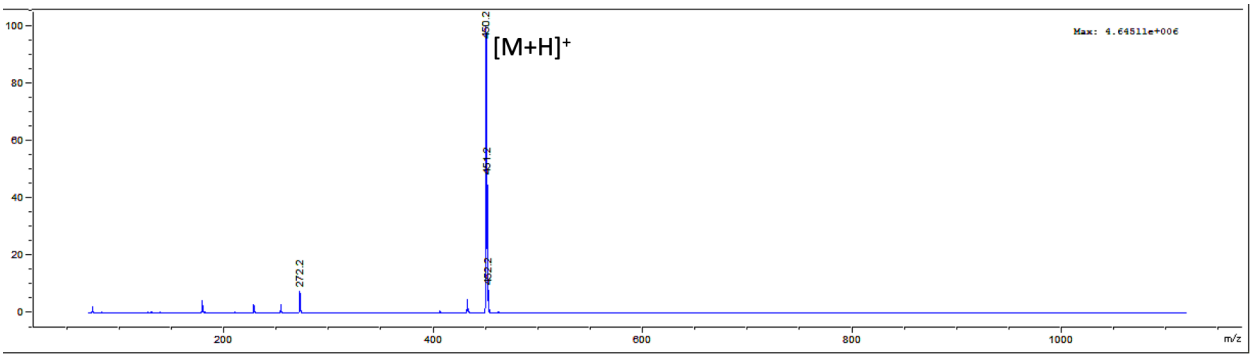


Figure 17. HRMS-ESI positive mode spectrum of **Fmoc-Mel^{Dab}-OH** extracted by averaging the corresponding peak area in the LC/MS: m/z calculated for $C_{22}H_{23}N_7O_4$: 450.19 $[M+H]^+$; found, 450.2.

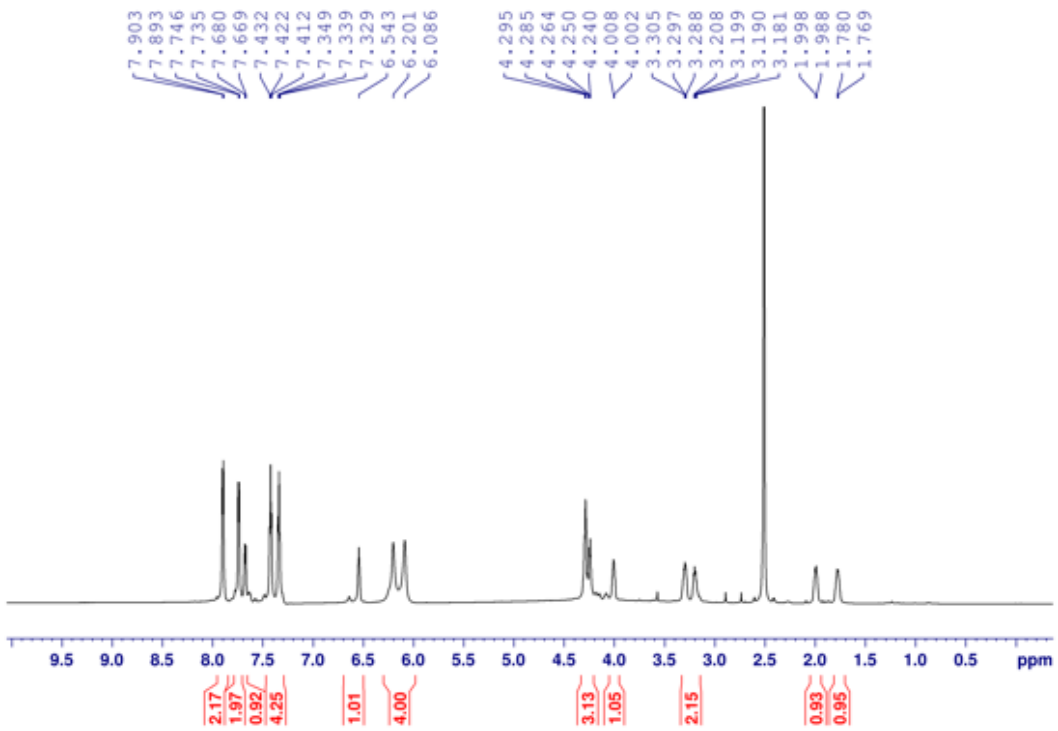


Figure 18. [1]HNMR (700 MHz) spectrum of **Fmoc-Mel^{Dab}-OH** in DMSO-d_6.

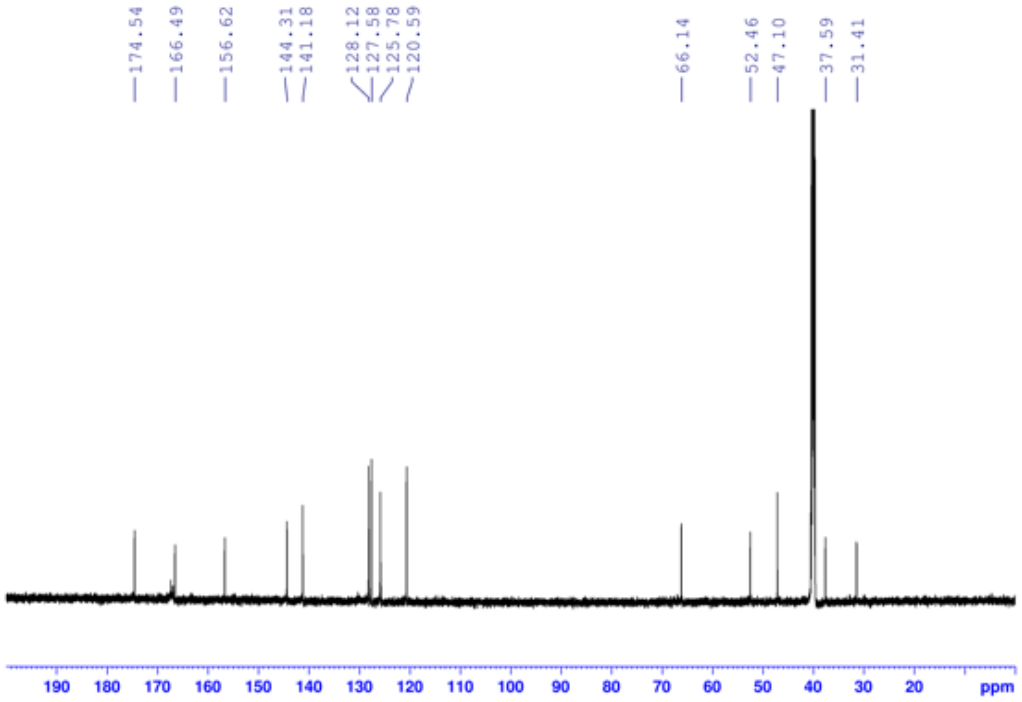


Figure 19. [13]C NMR (700 MHz) spectrum of **Fmoc-Mel^{Dab}-OH** in DMSO-d_6.

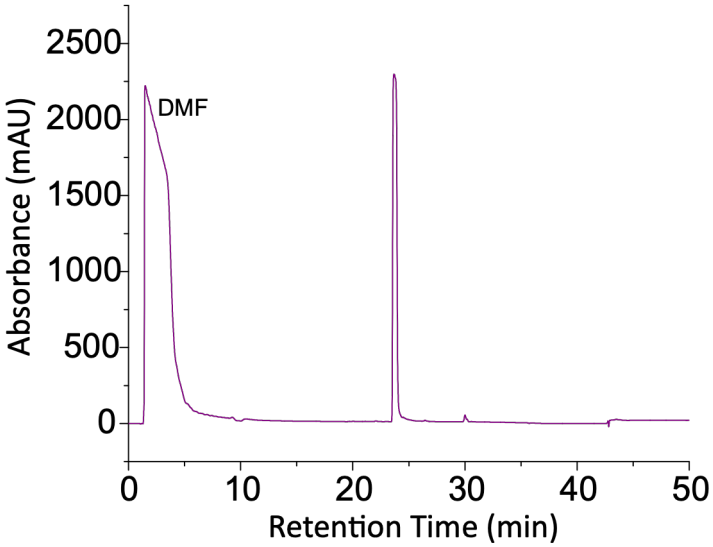


Figure 20. HPLC trace of **Fmoc-Cy^{Dab}-OH** (DMF) on a C18 analytical column monitored at 215 nm with a 0.1% formic acid (A) - acetonitrile (B) solvent system: Gradient; 0-50% B for 0-30 min, 50-100% B for 30-40 min and 100-0% B for 40-50 mins.

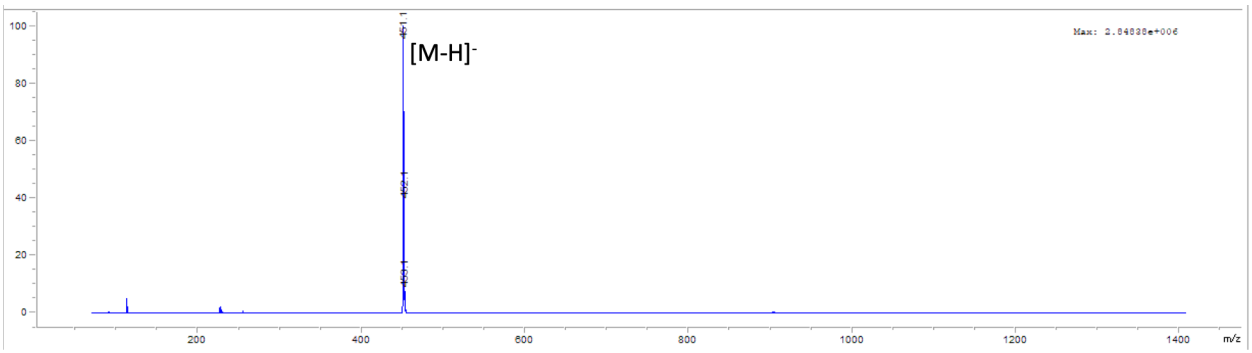


Figure 21. HRMS-ESI negative mode spectrum of **Fmoc-Cy^{Dab}-OH** extracted by averaging the corresponding peak area in the LC/MS: m/z calculated for $C_{22}H_{20}N_4O_7$: 451.13 $[M-H]^-$; found, 451.1.

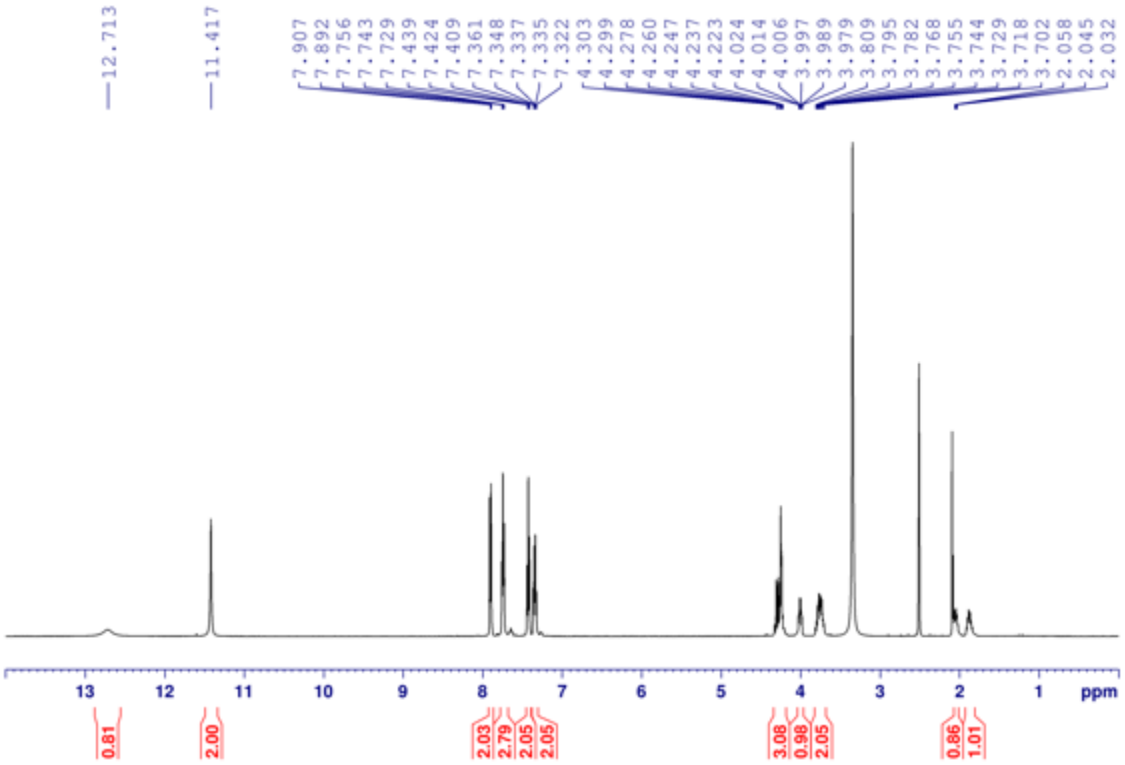


Figure 22. ^{1}H NMR (500 MHz) spectrum of **Fmoc-Cy^{Dab}-OH** in DMSO-d$_6$.

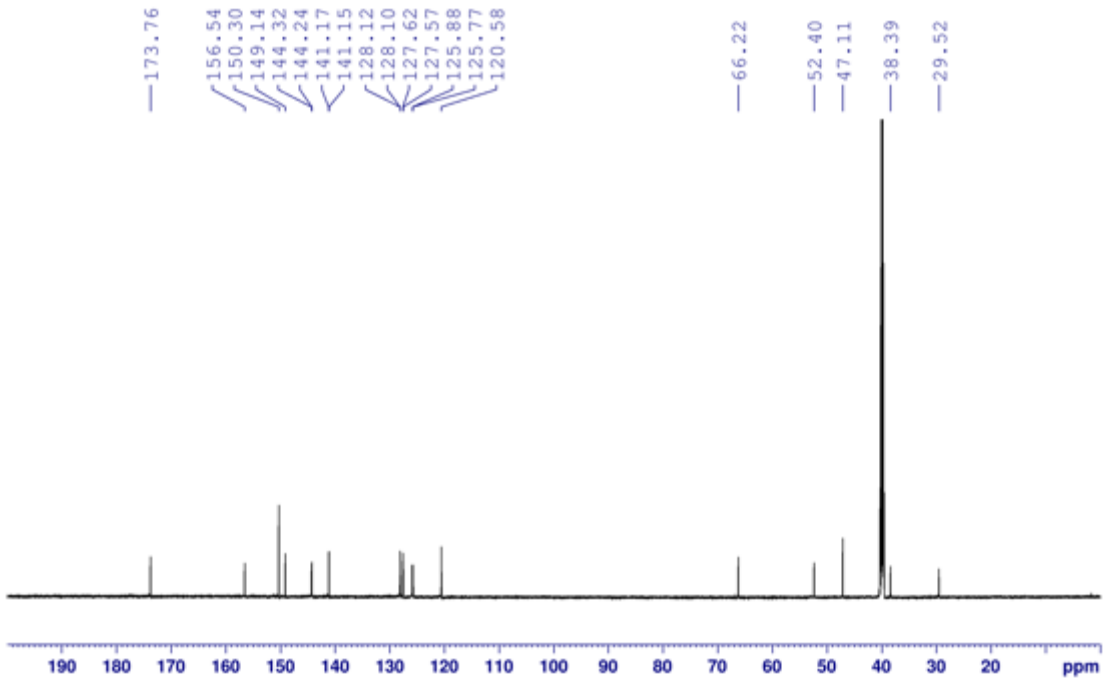


Figure 23. ^{13}C NMR (700 MHz) spectrum of **Fmoc-Cy^{Dab}-OH** in DMSO-d$_6$.

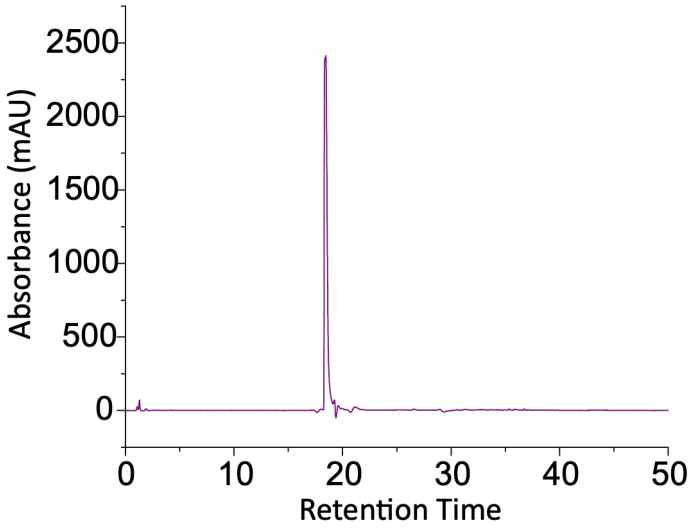


Figure 24. HPLC trace of α-$(M)_4$ on a C18 analytical column monitored at 215 nm with a 0.1% formic acid (A) - acetonitrile (B) solvent system: Gradient; 100% A for 0-14 min, 0-50% B for 14-35 min, 50-100% B for 35-45 min and 100-0% B for 45-50 mins.

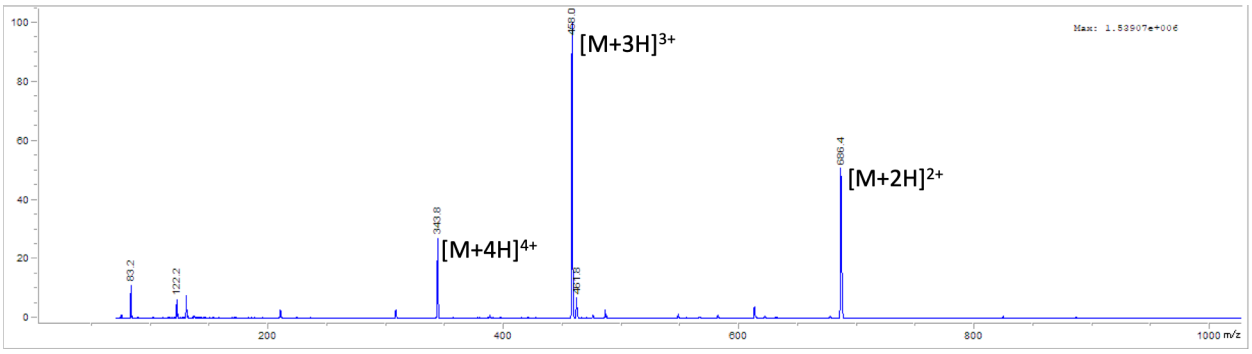


Figure 25. HRMS-ESI positive mode spectrum of α-$(M)_4$ extracted by averaging the corresponding peak area in the LC/MS: m/z calculated for $C_{48}H_{74}N_{32}O_{17}$: 343.65 $[M+4H]^{4+}$; found 343.8, 457.87 $[M+3H]^{3+}$; found 458, 686.3 $[M+2H]^{2+}$; found 686.4.

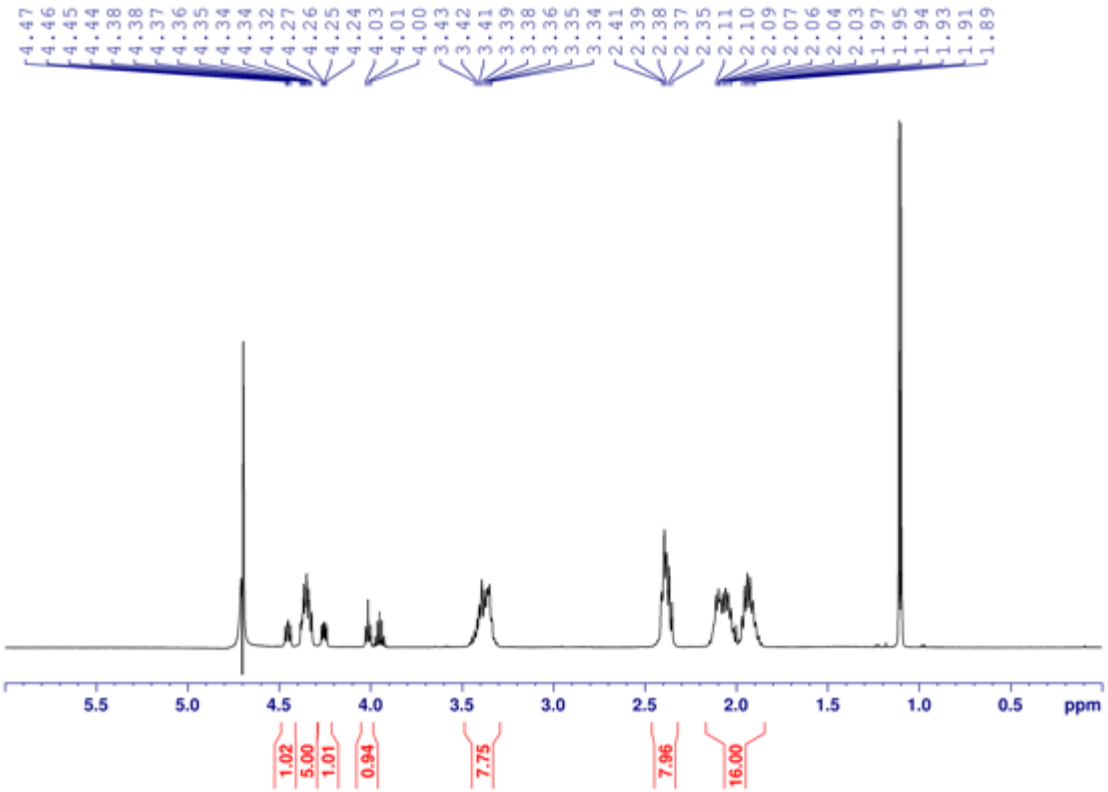


Figure 26. ^{1}H NMR (500 MHz) spectrum of α-$(M)_4$ in D_2O after water suppression with IPA as internal standard.

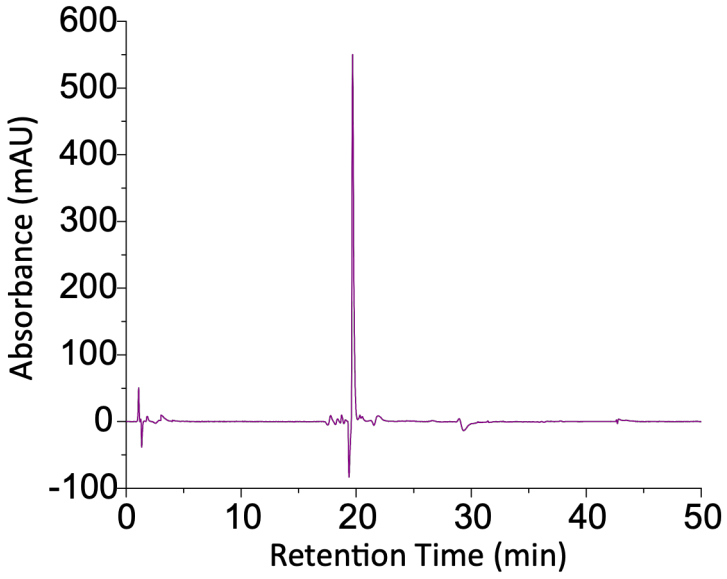


Figure 27. HPLC trace of α-$(C)_4$ on a C18 analytical column monitored at 215 nm with a 0.1% formic acid (A) - acetonitrile (B) solvent system: Gradient; 100% A for 0-14 min, 0-50% B for 14-35 min, 50-100% B for 35-45 min and 100-0% B for 45-50 mins.

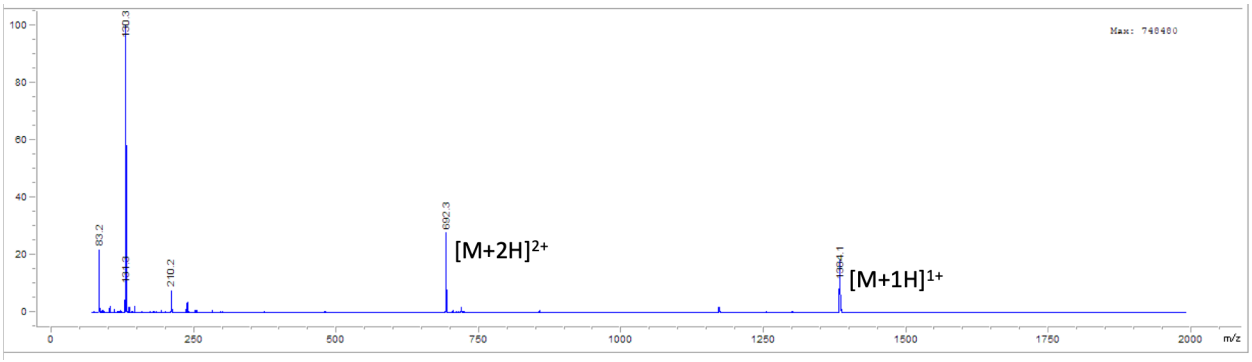


Figure28. HRMS-ESI positive mode spectrum of α-$(C)_4$ extracted by averaging the corresponding peak area in the LC/MS: m/z calculated for $C_{48}H_{62}N_{20}O_{29}$: 692.2 $[M+2H]^{2+}$; found 692.3, 1383.41 $[M+H]^{+}$; found 1383.1.

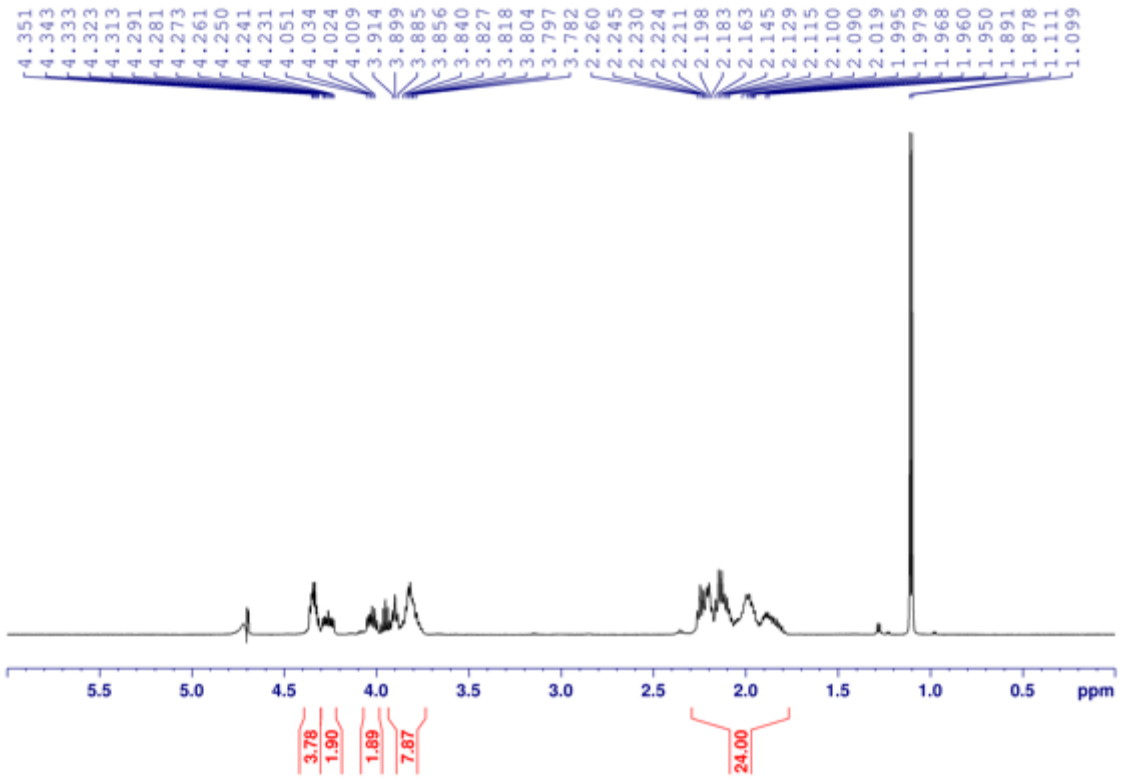


Figure 29. ^{1}H NMR (500 MHz) spectrum of α-$(C)_4$ in D_2O after water suppression with IPA as internal standard.

REFERENCES

1 Zeng, Y., Pratumyot, Y., Piao, X. & Bong, D. Discrete assembly of synthetic peptide-DNA triplex structures from polyvalent melamine-thymine bifacial recognition. *J. Am. Chem. Soc.* **134**, 832-835 (2012). https://doi.org/10.1021/ja2099326

2 Nair, S. A., Lee, B. & Hangauer, D. G. Synthesis of orthogonally protected L-homocysteine and L-2-amino-4-phosphonobutanoic acid from L-homoserine. *Synthesis* **1995**, 810-814 (1995). DOI: 10.1055/s-1995-4015

3 Chen, S.-Y. *et al.* Synthesis of a cyclen-functionalized α-amino acid and its incorporation into peptide sequence. *Synthesis* **2005**, 888-892 (2005). DOI: 10.1055/s-2005-861854

4 Wang, D.-H., Zhu, D.-J., Ding, W., Xue, M. & Yang, Y. Tuning the fluorescence of tetraphenylethylene in dilute solutions via modulating multiple-hydrogen-bonding interactions between a Hamilton receptor and cyanuric acid. *Org. Biomol. Chem.* **16**, 4429-4432 (2018). https://doi.org/10.1039/c8ob00814k

5 Lam, P.-L., Wu, Y. & Wong, K.-L. Incorporation of Fmoc-Dab (Mtt)-OH during solid-phase peptide synthesis: a word of caution. *Org. Biomol. Chem.* **20**, 2601-2604 (2022). https://doi.org/10.1039/d2ob00070a

6 Alenaizan, A., Barnett, J. L., Hud, N. V., Sherrill, C. D. & Petrov, A. S. The proto-Nucleic Acid Builder: a software tool for constructing nucleic acid analogs. *Nucleic Acids Res.* **49**, 79-89 (2021). https://doi.org/10.1093/nar/gkaa1159

7 O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011). https://doi.org/10.1186/1758-2946-3-33

8 Ambrose, E. J. & Hanby, W. E. Evidence of chain folding in a synthetic polypeptide and in keratin. *Nature* **163**, 483-484 (1949). https://doi.org/10.1038/163483a0