Nucleopolypeptide with DNA-triggered α helix-to-β sheet transition

Michel Nguyen,[ab] Jean-Luc Stigliani,[ab] Genevieve Pratviel,[ab] Colin Bonduelle,[ab*]

[a] *CNRS, LCC (Laboratoire de Chimie de Coordination (UPR8241)), 205 route de Narbonne, F-31077 Toulouse, France.*
[b] *Université de Toulouse; Université Paul Sabatier, Toulouse. France.*

E-mail : colin.bonduelle@lcc-toulouse.fr. Tel : +335 61 33 3236

13th March 2018

Note added after first publication: This Supplementary Information file replaces that originally published on 7th June 2017, in which there was a minor labelling error in Fig. S3.

Summary

1. Materials and Methods

L-glutamic acid, propargyl alcohol, chlorotrimethylsilane, diethyl ether, ethanol, diphosgene, sodium ascorbate, copper sulfate, dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich and used as received unless otherwise noted. **Diphosgene requires careful handling (syringe) and may form *in situ* phosgene that is extremely toxic and volatile. Its fume hood's use includes careful wash of the shlencks, syringes, and personal gloves with NH_4OH once reactions achieved**. 3'-azido-2', 3'-dideoxythymidine (AZT) was purchased from Jena Bioscience. DNA was purchased from Genscript. The following sequences were used for CD measurements (d stands for deoxy): dA_{30}, the DNA sequence AAAAAAAAAA-AAAAAAAAAAAAAAAAAA; dT30, the DNA sequence TTTTTTTTTT-TTTTTTTTTTTTTTTTTT; $d[(ACCC)_7AC]$, the DNA sequence ACCCACCC-ACCCACCCACCCACCCACCCAC. γ-Propargyl-L-glutamate *N*-carboxyanhydride (PLG-NCA) was synthesized according to the procedure given below. DMSO was deoxygenated before use by several freeze-thawing cycles. Anhydrous tetrahydrofuran (THF), and dichloromethane (CH_2Cl_2) were used directly from a solvent purification system under an inert and dry atmosphere. Hexylamine (98%) was purchased from Sigma-Aldrich and was double distilled before use. Dimethyl formamide (DMF) was obtained from a Solvent Purification System (SPS) and freshly used for the polymerization. NMR spectra were recorded on a Bruker AC 400 spectrometer and chemical shifts are reported relative to the deuterated solvents used ($CDCl_3$, DMF-d^7). Infrared measurements were performed on a Bruker Tensor 27 spectrometer using the attenuated total reflection (ATR) method. Molar masses were determined by SEC in DMF/LiBr (1%) using an Acquity Advanced Polymer Chromatography System (Waters) equipped with an Acquity APC XT column for extended temperature organic-based separations (4.6 × 150 mm) and an Acquity Refractive Index detector. Calibration was performed by using polystyrene standards. Samples (5 $mg.mL^{-1}$) were dissolved in DMF and were run at a flow rate of 0.5 mL min^{-1} at

55 °C. The circular dichroism measurements were performed on a JASCO J-815 spectropolarimeter between 200 nm and 260 nm (far-UV), by using a quartz cell of 1 cm path length, at 20 °C. The measure parameters were optimized as follows: sensitivity mode high, 0.1 nm data pitch, 0.01 mdeg resolution, 8 seconds response time (Digital Integration Time), 1 nm bandwidth and 10 nm/ min scanning rate. The polypeptide solutions were diluted with Milli-Q water. Important: the blanks of the CD analysis were prepared with nucleic acids that significantly responded to circular dichroism (cf figure S1 for dA_{30}). Parallel CD measurements performed with Poly(*L*-glutamic acids) with similar molecular weights in aqueous Milli-Q solutions did not evidence changes in shape (coil, alpha helix) upon DNA addition (if DNA signal is included for the blank)**.** Data from circular dichroism were analyzed using the CAPITO software, a web server based analysis and plotting tool for circular dichroism data.[1]

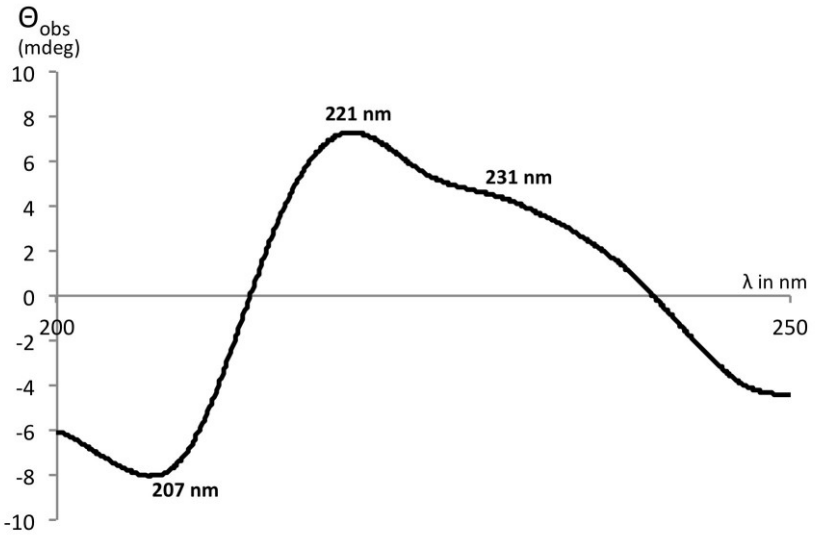
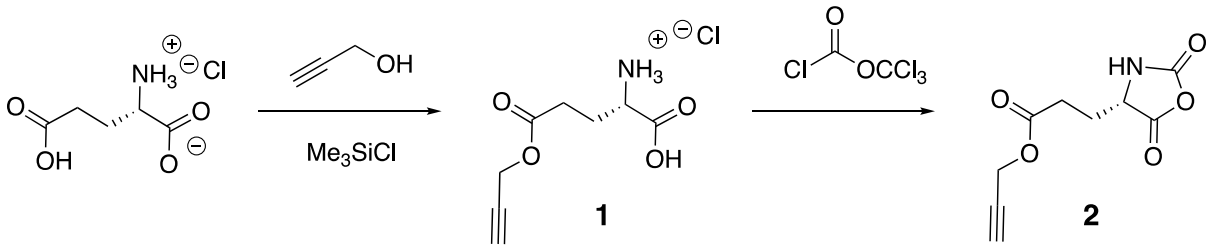


Figure S1. CD spectra of an aqueous Milli-Q solution of dA30 at 23 μM in nucleotide units.

2. Synthesis of the nucleopolypeptide.

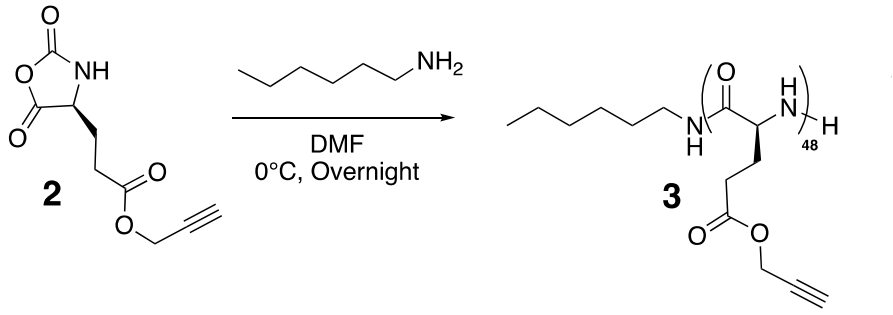


Scheme 1. PLG-NCA **2** synthesis from PLG **1**

*a) γ-Propargyl-L-glutamate NCA **2** (scheme S1)*

*1) Synthesis of γ-propargyl L-glutamate hydrochloride **1*** (It was synthesized following the P. Hammond procedure Angew Chem, 2009, 121, 9498-9502 and Soft Matter, 2012, 8, 10887-10895). *L*-glutamic acid (10 g, 68 mmol) was suspended in propargyl alcohol (370 mL) under argon. Chlorotrimethylsilane (19 mL, 150 mmol) was added dropwise to the suspension at 0°C. The resulting solution was stirred at 0°C for 1 hour and then at room temperature for two days until there was no undissolved *L*-glutamic acid. The reaction solution was filtered through a Chromafil GF/PET-45/25 filter to remove undissolved material, and then precipitated into diethyl ether giving a white solid. The precipitate was filtered, washed with diethyl ether, and re-dissolved in ethanol at 50°C. After filtration and precipitation of filtrate into diethyl ether, the white solid was dried under vacuum to yield 9.98 g (66%). ^{1}H NMR (400MHz, D_2O, δ in ppm): 2.19 (m, 2H, CO-CH_2-**CH_2**-CH), 2.61 (dt, 2H, CO-**CH_2**-CH_2-CH), 2.85 (t, 1H, C≡CH), 4.02 (t, 1H, CH), 4.68 (d, 2H, CH_2O) ^{13}C NMR (100.6 MHz, D_2O, δ in ppm): 24.74 (CO-**C**H_2-CH_2-CH), 29.42 (CO-CH_2-**C**H_2-CH), 52.14 (CH), 52.77 (CH_2O), 76.07 (C≡**C**H), 77.60 (**C**≡CH), 171.67 (**C**O-CH_2), 173.44 (**C**(O)OH). Significant CO stretch by FTIR : 1747 cm^{-1}, 1725 cm^{-1}. HRMS (ESI positif, m/z) calculated for $C_8H_{12}NO_4$, 186.0766, found 186.0766. *2) Synthesis of N-carboxyanhydride of γ-propargyl L-glutamate **2***. γ-propargyl *L*-glutamate hydrochloride **1** (5 g, 22.56 mmol) was suspended in

dry THF (65 mL) and the suspension was stirred at room temperature. Diphosgene (3.11 g, 15.74 mmol) was added and the reaction was left stirring for 2 hours until complete solubilization of the powder. The reaction solution was then dried, redissolved in dry THF and the solution was degassed during 20 min. Then the THF was evaporated and dry pentane was added. After 15 min of stirring, the solvent was removed with a syringe. NCA was dissolved in a minimal amount of dry CH_2Cl_2 and passed through a column of dry silica gel (L=10 cm X θ 3 cm) (Silica gel 40-63 µm was previously dry 2 days at 150°C).[2] After elution with dry THF and elimination of the solvent, a yellowish solid was obtained (4.7 g, 99%). ^{1}H NMR (400MHz, $CDCl_3$, δ in ppm): 2.25 (m, 2H, $COCH_2\mathbf{CH_2}CH$), 2.54 (t, 1H, C≡CH), 2.63 (t, 2H, $CO\mathbf{CH_2}CH_2CH$), 4.45 (t, 1H, CH), 4.73 (d, 2H, CH_2O), 6.77 (s, 1H, NH). ^{13}C NMR (100.6 MHz, D_2O, δ in ppm): 26.78 (CO-CH_2-**C**H_2-CH), 29.46 (CO-**C**H_2-CH_2-CH), 52.63 (CH_2O), 56.79 (CH), 75.52 (C≡**CH**), 77.05 (**C**≡CH), 151.95 (**C**O-NH), 169.31 (**C**(O)O), 171.71 (**C**(O)CH_2). Significant CO stretch by FTIR 1854 cm^{-1}, 1787 cm^{-1}, 1655 cm^{-1}.



Scheme 2. Synthesis of poly(γ-propargyl-*L*-glutamate) **3** from PLG-NCA **2**.

*b) Synthesis of Poly(γ-propargyl-L-glutamate) **3** (scheme S2).*

The NCA monomer of γ-propargyl-*L*-glutamate (0.52 g, 2.46 mmol) was dissolved in 5 ml of dry DMF in a Schlenk tube. A solution of hexylamine (7 µL, 0.053 mmol) in 2 mL dry DMF was added after NCA was dissolved. The reaction was left to stir in a cold bath at 10°C for 1 day under an inert atmosphere. After PLG-NCA had been completely consumed as

monitored by FTIR (disappearance of the peaks at 1787 cm^{-1} and 1854 cm^{-1}, figure S2), the polymer was recovered by precipitation in Et_2O and dried under vacuum. Yield (93%).

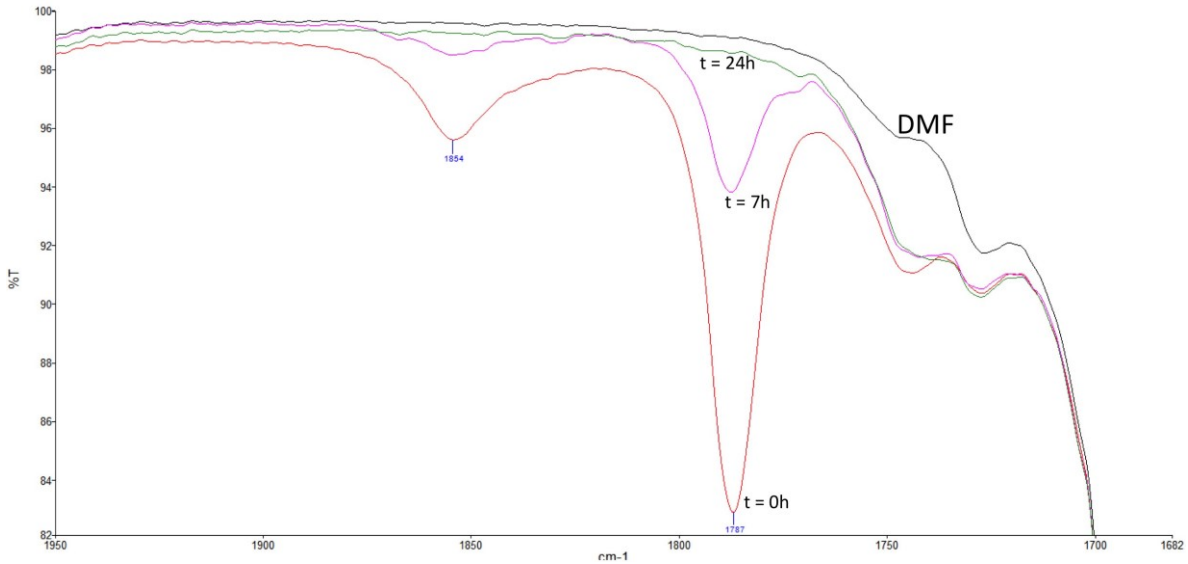
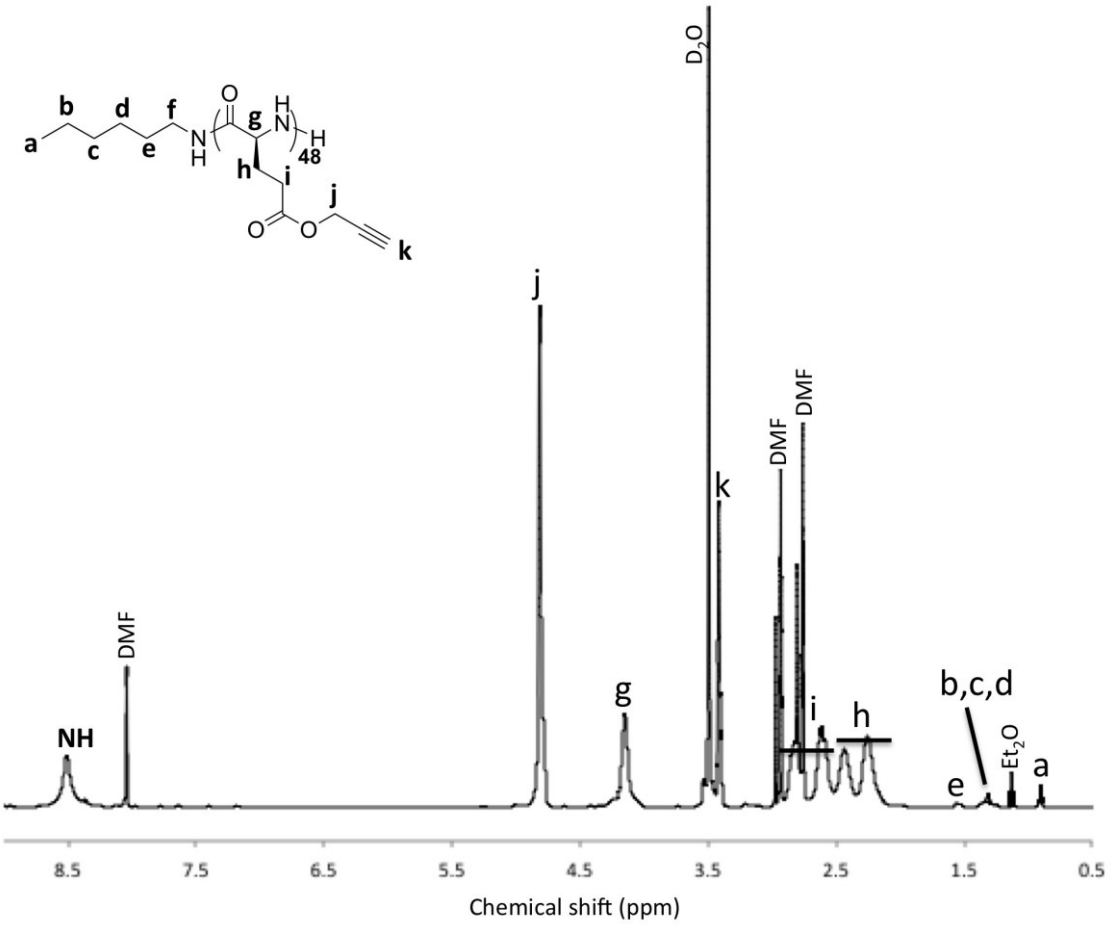


Figure S2. Ring-opening polymerization of **2**: kinetic of the polymerization in DMF.

Molar mass (Mn) was first determined by ^{1}H NMR using the following equation where Ia, Ig, M(PLGNCA), and M(hexylamine) are, respectively, the intensity of methylene protons a (hexylamine, figure S3), the intensity of methylene protons g (polypeptide backbone, figure S3), the molar mass of the PLG monomer unit and the molar mass of hexylamine initiator.

$$Mn = \frac{Ig * M\,(PLGNCA)}{Ia} + M\,(hexylamine)$$

A Mn value of 8200 g/mol and a number-average degree of polymerization DP of 48 was therefore derived from ^{1}H NMR analysis. SEC then provided a number-average molecular weight M_n of 9900 g/mol (calibration with polystyrene standards) and a polymolecularity index (M_w/M_n) of 1.10.



Corrected Figure S3. RMN ^{1}H of poly(γ-propargyl-*L*-glutamate) **3** (DMF-d^{7})

^{1}H NMR (400MHz, DMF-d^{7}, δ, ppm): 0.90 (t, 3H, CH_3 hexylamine), 1.22-1.42 (m, 3 x 2H, $CH_3C\mathbf{H_2}C\mathbf{H_2}C\mathbf{H_2}CH_2CH_2$ hexylamine), 1.54 (m, 2H, $CH_3CH_2CH_2CH_2C\mathbf{H_2}CH_2$ hexylamine), 2.26 (m, 47H, $COCH_2C\mathbf{H}HCH$), 2.45 (m, 48H, $COCH_2CH\mathbf{H}CH$), 2.59 (m, 47H, $COC\mathbf{H}HCH_2CH$), 2.83 (m, 49H, $COCH\mathbf{H}CH_2CH$), 3.45 (m, 43H, $COOCH_2CC\mathbf{H}$), 4.15 (m, 48H, $COC\mathbf{H}NH$), 4.82 (m, 95H, $COOC\mathbf{H_2}CCH$). ^{13}C NMR (100.6 MHz, DMF, δ in ppm): 13.7 ($\mathbf{C}H_3$ Hex), 22.5 (Hex), 25.6 ($CH\mathbf{C}H_2CH_2$), 26.4 (Hex), 30.5 ($CHCH_2\mathbf{C}H_2$), 31.5 ($\mathbf{C}H_2NH$ Hex), 51.9 ($COO\mathbf{C}H_2C{\equiv}C$), 56.7 ($NH\mathbf{C}HCO$), 76.5 (C≡C), 78.4 (C≡C), 171.7 (**C**O-NH), 175.7 (**C**$OOCH_2C{\equiv}C$).

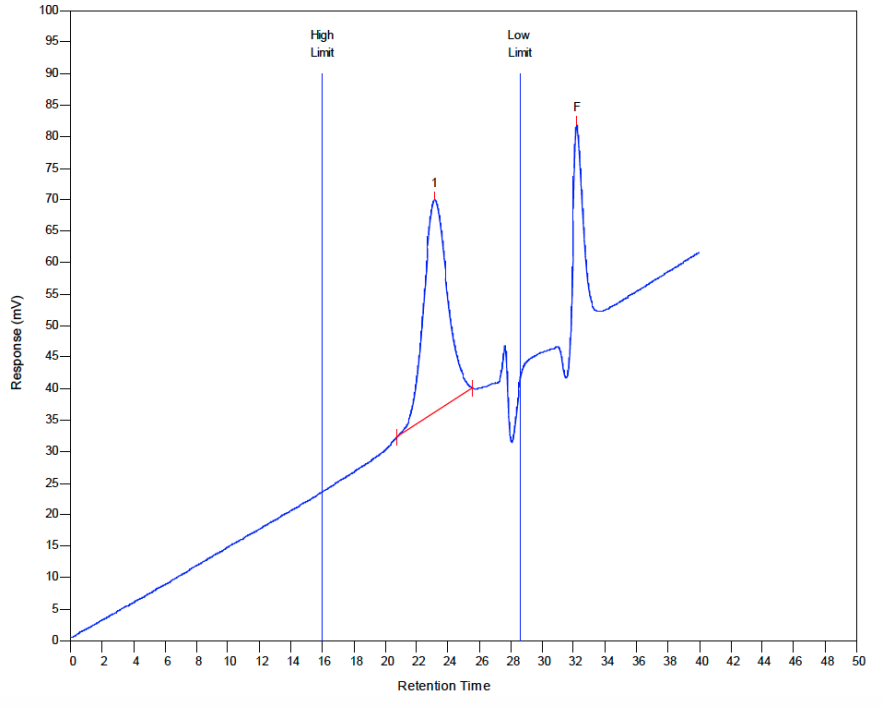
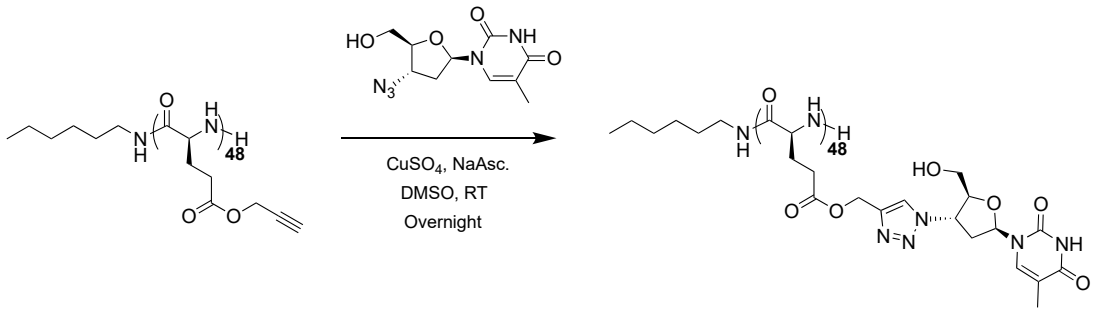


Figure S4. Size exclusion chromatogram of poly(γ-propargyl-*L*-glutamate) **3** in DMF (1% LiBr)



Scheme 3. Synthesis of nucleopolypeptide **4** from poly(γ-propargyl-*L*-glutamate) **3.**

c) Synthesis of the nucleopolypeptide 4 (scheme S3).

Poly(γ-propargyl-*L*-glutamate) **3** (55 mg, ca. 0.36 mmol of alkyne units), AZT (100 mg, 0.39 mmol, 1.1 equiv. to alkyne groups) and sodium ascorbate (230 mg, 1.27 mmol, 3.5 equiv. to alkyne groups) were dissolved in 5 mL of anhydrous and deoxygenated DMSO in a Schlenk tube. $CuSO_4$ (160 mg, 0.73 mmol, 2 equiv. to alkynes groups) was then added and the Schlenk tube was placed in an oil bath at 25°C for 16 hours. The reaction medium was then dialyzed for 4–5 days against milliQ water (Spectra/Por MWCO 2 kDa membrane),

containing EDTA the first 2 days, and then the polymer **4** was recovered upon centrifugation. Yield (92%).

^{1}H NMR (400MHz, DMF-d^{7}, δ in ppm) 0.82 (t, 3H, C**H$_3$** hexylamine), δ= 1.22-1.33 (m, 3X2H, $CH_3C\mathbf{H_2}C\mathbf{H_2}C\mathbf{H_2}CH_2CH_2$ hexylamine), 1.48 (m, 2H, $CH_3CH_2CH_2CH_2C\mathbf{H_2}CH_2$ hexylamine), 1.82 (s, 135H, C**H$_3$** AZT), 2.05-2.66 (m, 148H, $COCH_2C\mathbf{H}HCH$ and C**H$_2$** ribose AZT), 2.78-292 (m, 137H, $COCH_2CH\mathbf{H}CH$ and $COC\mathbf{H_2}CH_2CH$), 3.74-3.94 (d broad, 93H, C**H$_2$**OH AZT), 4.20 (s broad, 47H, COC**H**NH), 4.36 (s broad, 50H, C**H** ribose AZT), 5.07-5.38 (m, 91H, COOC**H$_2$**-triazole), 5.43 (m, 48H, C**H** ribose AZT), 6.55 (m, 49H, C**H** ribose AZT), 7.92 (s broad, 48H, C**H** thymidine AZT), 8.36 (s broad, 49H, C**H** triazole), 8.53 (s broad, N**H** polypeptide), 11.22 (s broad, N**H** thymidine).

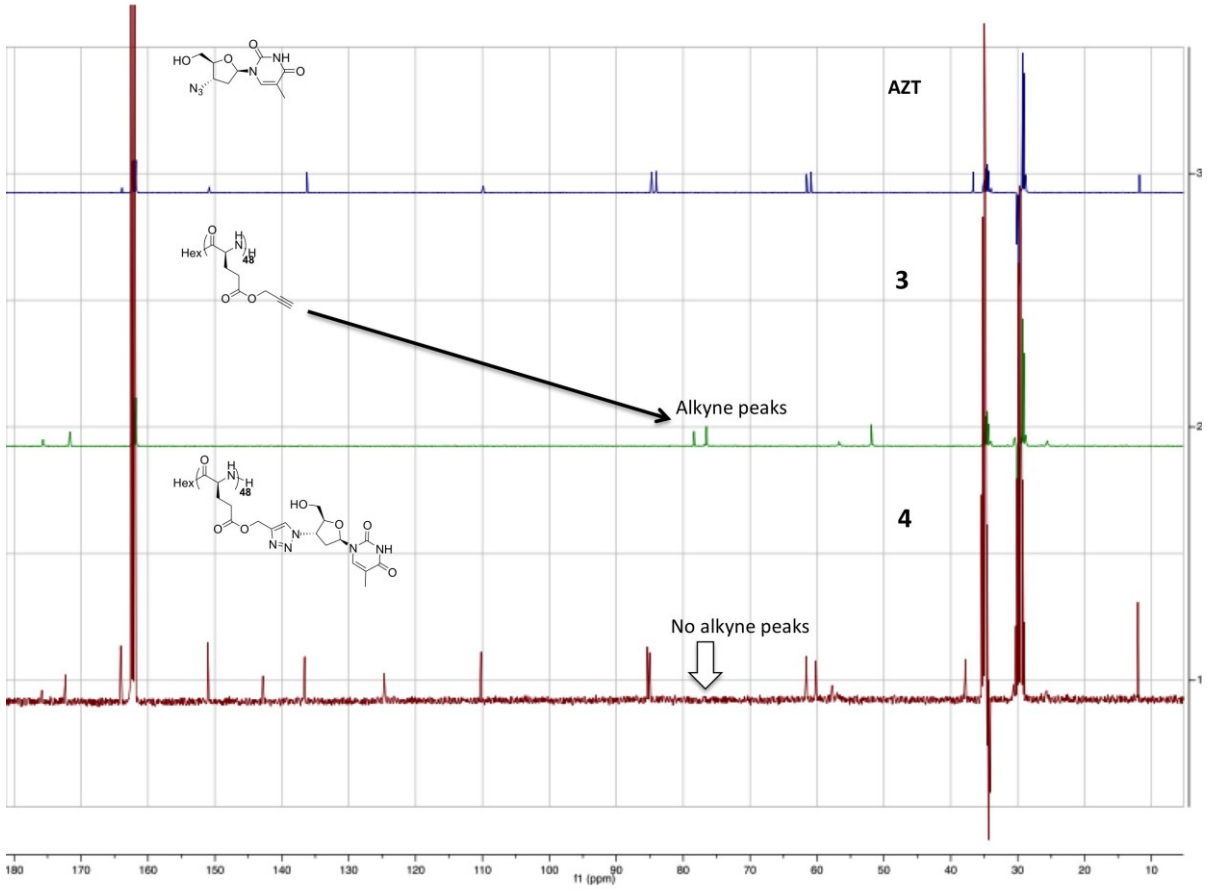


Figure S5. ^{13}C NMR (DMF-d^{7}) of AZT (top), PPLG **3** (middle) and nucleopolypeptide **4** (down).

^{13}C NMR (100.6 MHz, DMF-d^7, δ in ppm) 12.1(**C**H$_3$ thymidine AZT), 25.5 (CH**C**H$_2$CH$_2$), 30.7 (CHCH$_2$**C**H$_2$), 37.8 (**C**H$_2$ ribose AZT), 56.5 (NH**C**HCO), 57.7 (COO**C**H$_2$ triazole), 60.2 (**C**H ribose AZT), 61.6 (**C**H$_2$OH ribose AZT), 85.0 (**C**H-triazole ribose AZT), 85.4 (**C**H ribose AZT), 110.2 (**C**CH$_3$ thymidine AZT), 124.7 (**C**H triazole), 136.6 (**C**H thymidine AZT), 142.8 (**C** triazole), 151.0 (**C**O thymidine AZT), 164.1 (**C**O thymidine AZT), 172.4 (**C**OOCH$_2$), 175.9 (**C**OCHNH).

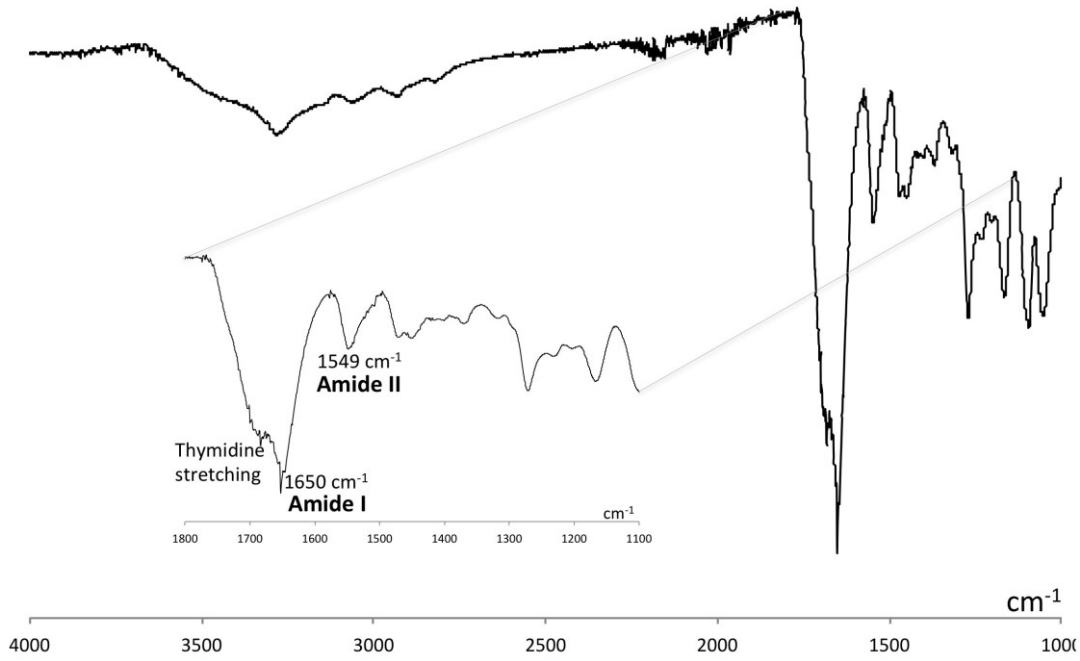


Figure S6. FTIR spectrum of **4** (powder): peptide bond stretching were found located at 1650 cm^{-1} and 1549 cm^{-1}.

3. Circular Dichroism Spectroscopy

a) Generalities:

The secondary structure of proteins/peptides (α-helix, β-sheet…) can be analyzed from the CD spectra in the range from 195 to 255 nm, corresponding to the peptide bond absorption.[3] The secondary structure of the polymers was studied by CD spectroscopy using the following procedure: the final concentration (the concentration in the cuvette used for the CD analyses) was always 23 μM in monomer units. The final volume in the cuvette was

always 1 mL (pathlength of 10 mm). The “molar ellipticity” also called the “mean residue ellipticity” has been calculated as follow:

$$[\theta] = (10 \times \theta_{obs}) / (l \times c)$$

$[\theta]$ is expressed in $deg.cm^2.dmol^{-1}$ and θ_{obs} was the observed ellipticity in degrees (deg), l is the path length in dm, and c is the polypeptide concentration in mol/L.

It is to note that specific work done previously with other polypeptides have shown that quantitative evaluation of the structuring may strongly depends on the molecular weight and on the solubility.[4]

b) Preparation of CD samples:

* Preparation of mother solutions used for the CD experiments: 1) Mother solution of **4**: the polymer solution was always prepared the day before CD analysis. The mother solution of the polymer was prepared by dissolving 4.5 mg of the polymer **4** in 4.5 mL of DMSO and 9.5 mg of 3'AMP was added. After extensive sonication, the reaction medium was dialyzed for 1 day against 1 L of milliQ water (Spectra/Por MWCO 2 kDa membrane, 4 water changes) and the resulting aqueous solutions was used as mother solution for CD analysis (polymer **4** concentration: 0.76 mM in monomer units). 2) Mother solution of DNA: DNA was diluted with milliQ water to afford an initial concentration of 3 mM in nucleic acid units.

* For each experience, an aliquot of the desired volume was taken off from polypeptide **4** solution and diluted in order to keep the final concentration at 23 μM in monomer units. Typical CD solution: 30 μL of polymer **4** mother solution was added in a tube, followed by

950 µL of Milli-Q water and if necessary eventually mixed with 7.5 uL of a mother solution of the desired DNA sequence. The volume was completed to 1 mL final volume with Milli-Q water. The samples were always prepared the day before analyses. Furthermore, for each experience, a blank sample was prepared containing 3'AMP and if necessary 7.5 uL of a mother solution of DNA that exhibited significant positive values upon CD analysis (cf for instance figure S1 for dA_{30}).

c) Raw data:

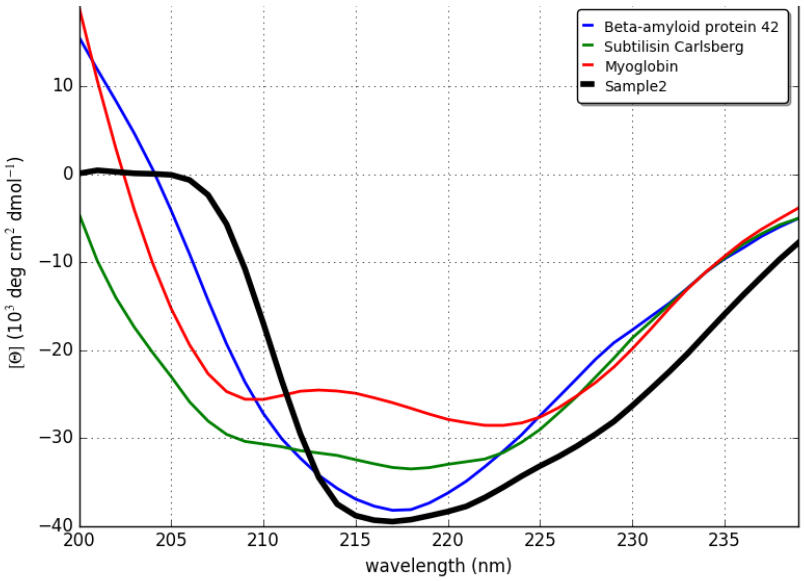


Figure S7. CD analysis of polymer **4** (sample 2, CD cuvette cut the signal below 205 nm) in interaction with dA_{30} and compared to CD analysis of various natural protein models (datas given by CAPITO software).[1]

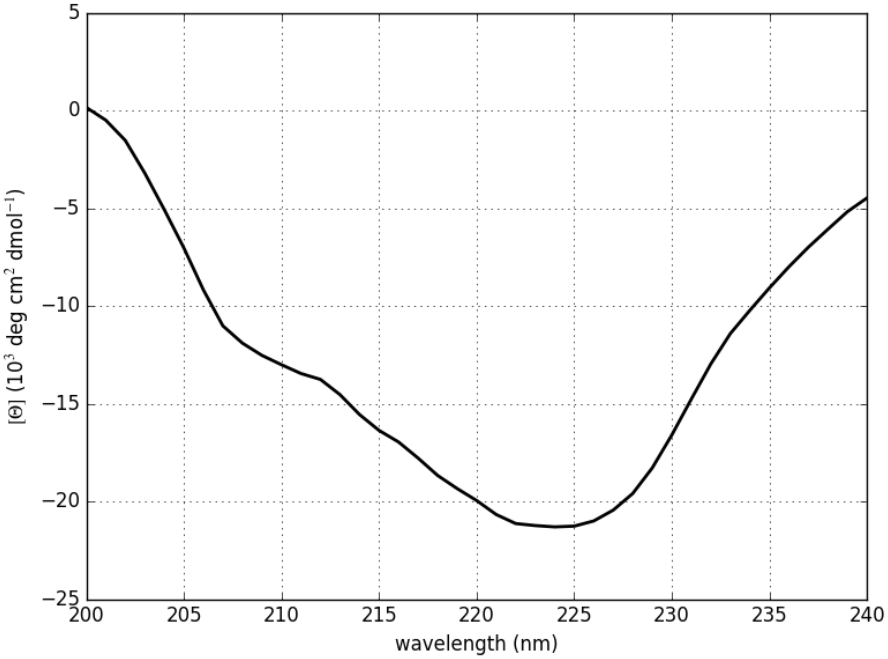


Figure S8. CD analysis of polymer **4** with d[(ACCC)$_7$AC].

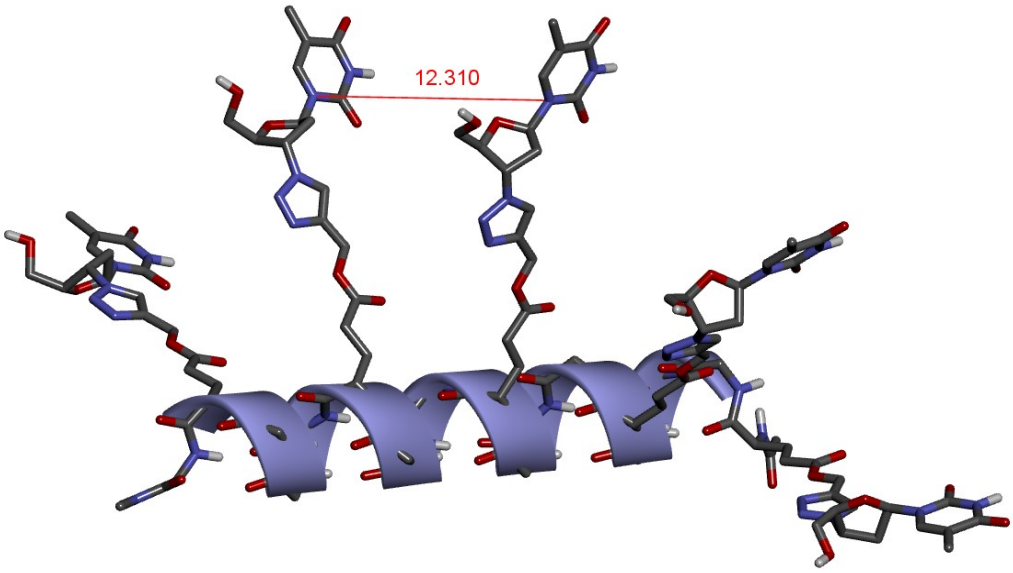
4. Molecular Modelling

a) Preparation of the nucleopeptide template and the nucleopolypeptide-dA$_5$ models:

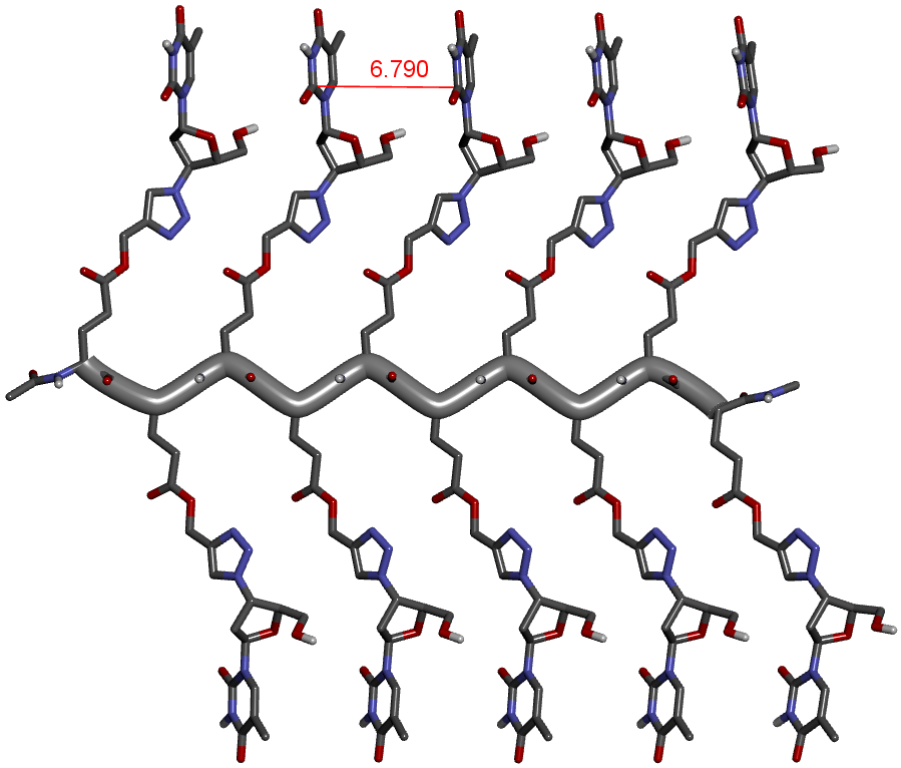
The monomer of the nucleopolypeptide **4** was built with the molecular modeling package Maestro 10.6.[5] The geometry was optimized using the PM7 method[6] implemented in the MOPAC 2016 quantum chemistry package.[7] Restrained electrostatic potential (RESP) charge derivation of the template was obtained from the R.E.D. server[8] and its topology was given from the Antechamber and Leap modules of the AmberTools16 suite.[9] The resulting library file was then used to set the atomic parameters of the studied nucleopolypeptides. Maestro was also used to prepare the α-helix or the β-sheet models of the nucleopolypeptide **4** in interaction with dA$_5$. The adenines of the oligonucleotide were manually positioned opposite to the thymine moieties of the nucleopolypeptides, so that the base pairs were in a same plan and positioned to generate two hydrogen bonds with distances of about two Å between the hydrogen atoms involved in the H-bonds and the corresponding nitrogen or oxygen atoms H-acceptors. These models were then submited to a molecular mechanics relaxation while keeping the H-bond freezed. N- and C-termini of the nucleopolypeptides were acetylated and

amidated, respectively, to mimic the continuation of the peptidic chain. This procedure was used to prepare an α-helix and a β-sheet model. For the α-helix, dA_5 was positionned on a 18-mers of **4** with a periodicity of n+4 ; indeed this nucleopeptide spacing appeared to be the more favourable to place the adenines. A 10-mers nucleopolypeptide was used to prepare the β-sheet model. Two dA_5 chains were positioned: one at each side of the peptidic backbone.

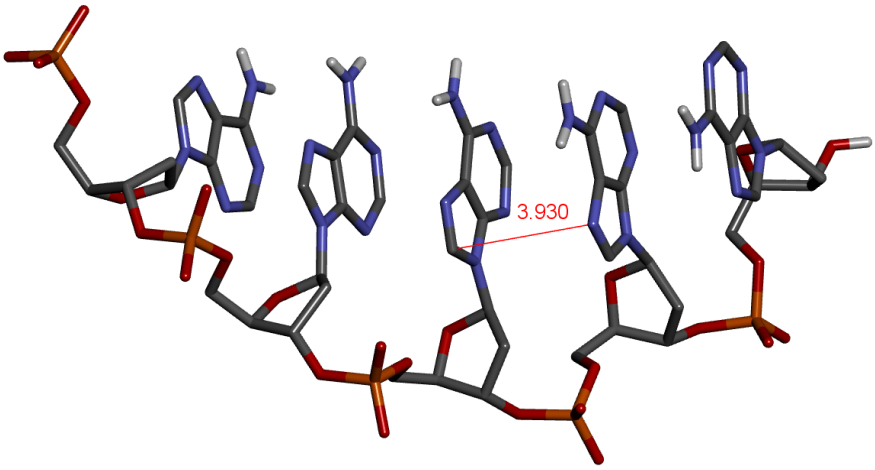
Figures S9 to S11 respectively represent the starting α-helix, the β-sheet and the dA_5 templates used to build the supramolecular models. The average distances (highlighted in red) between two consecutive bases show that the interaction between the dA_5 and the thymine moiety is more appropriate with the β-sheet model.



Figures S9. α-helix model of the nucleopolypeptide used to create the supramolecular model with dA_5. For clarity, only the n+4 residues of the nucleopolypeptide interacting with dA_5 and the polar hydrogens are represented. The distance between two consecutive bases (unit in Å) is highlighted in red.



Figures S10. β-sheet model of the nucleopolypeptide used to create the supramolecular model with dA_5. For clarity, only the polar hydrogens are represented. The distance between two consecutive bases (unit in Å) is highlighted in red.



Figures S11. dA_5 template used to create the supramolecular models. For clarity, only the polar hydrogens are represented. The distance between two consecutive bases (unit in Å) is highlighted in red.

b) Simulated annealing Simulations:

Simulations were performed using the SANDER module of Amber16. The protein.ff14SB[10] and DNA.OL15[11] force fields were employed. Solvation effects were incorporated using the

Generalized Born model[12] (igb = 1). The structures were subjected to 10000 steps of steepest descent minimization followed by 10000 steps of conjugate gradient. Simulated annealing was then used to optimize the positioning of both the oligonucleotides and the lateral chains of the nucleopolypeptides. In order to preserve the oligonucleotide-nucleopolypeptide interactions during the heating phase, distance restraints were introduced to maintain the H-bonds at an average distance of 1.8 Å with a force constants of 20 kcal/mol.Å. The system was heated from 0 to 600 K for 500 ps, with a time step of 1 fs, then gradually cooled to 300 K for 500 ps and lastly equilibrated at 300 K for 1 ns. The Langevin temperature equilibration scheme was used to equalize the system temperature.

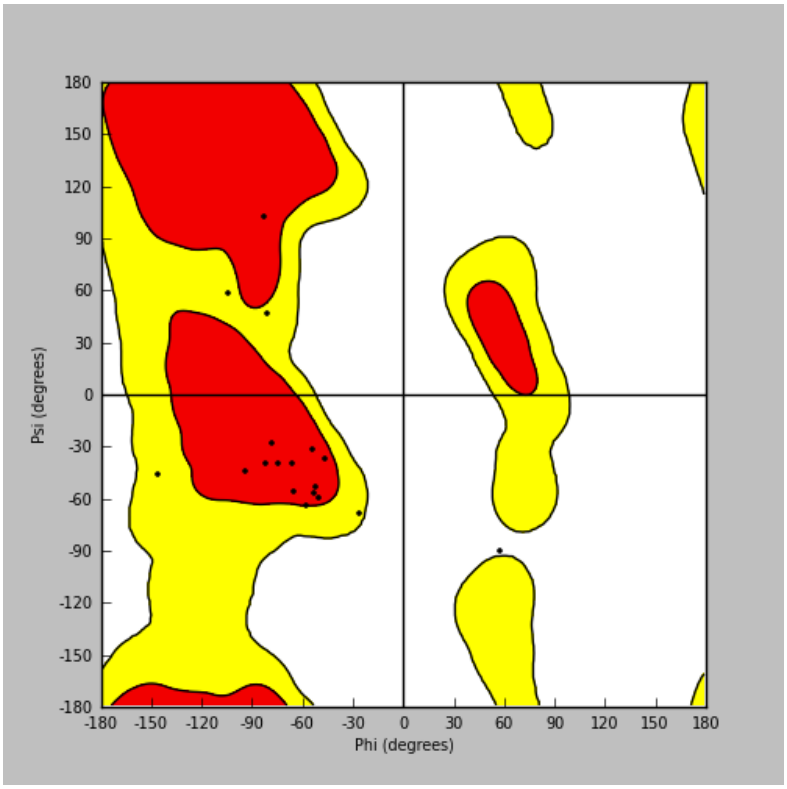


Figure S12. Ramachandran plot of the free α-helix nucleopolypeptide (Dp10), without the oligo A_5, after the Simulated annealing protocol. The plot shows that the helix shape is logicaly maintained after the heating phase, as most of the phi, psi values are located into the alpha zone. The outlying points correspond to the peripheral residues.

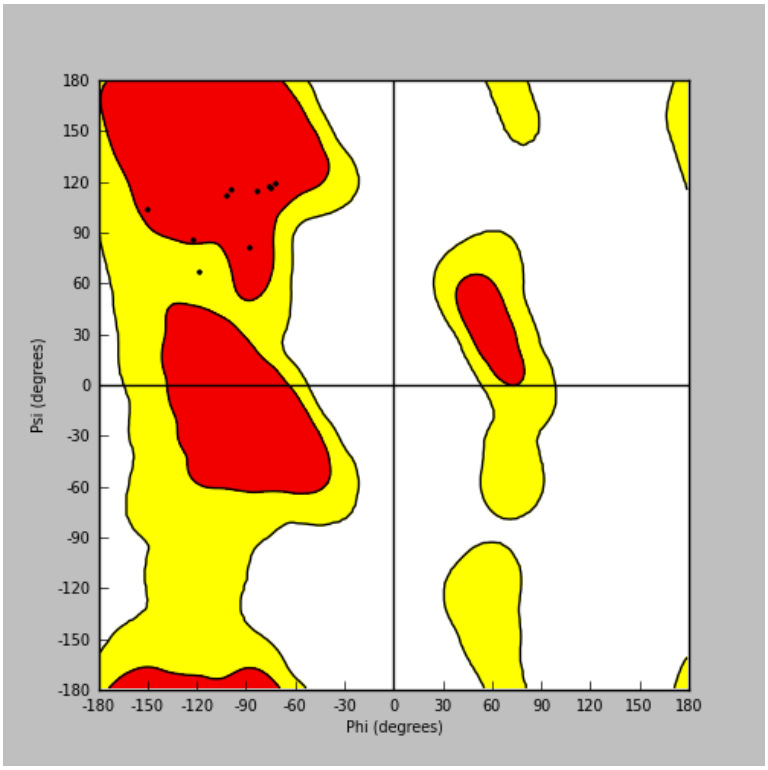


Figure S13. Ramachandran plot of the free β-sheet nucleopolypeptide (Dp10), without dA_5, before the simulated annealing protocol.

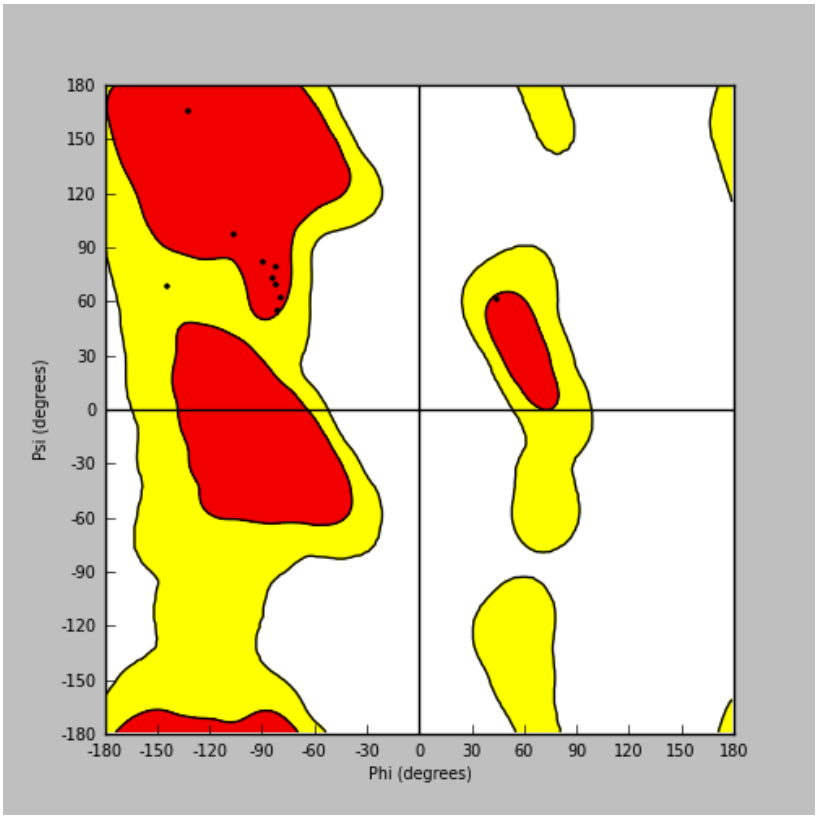


Figure S14. Ramachandran plot of the free β-sheet nucleopolypeptide (Dp10), without dA_5, after the Simulated annealing protocol. The plot shows that the beta shape of the nucleopolypeptide is maintained after the heating phase.

References

1 C. Wiedemann, P. Bellstedt and M. Görlach, *Bioinformatics*, 2013, **29**, 1750.
2 J. R. Kramer and T. J. Deming, *Biomacromolecules*, 2010, **11**, 3668.
3 B. M. Bulheller, A. Rodger and J. D. Hirst, *Physical Chemistry Chemical Physics*, 2007, **9**, 2020.
4 M. Rinaudo and A. Domard, *Journal of the American Chemical Society*, 1976, **98**, 6360.
5 Schrodinger release 2016-4 : Maestro, Schrodinger, LLC, New York, NY, 2016.
6 J. J. P. Stewart, *Journal of Molecular Modeling*, 2013, **19**, 1.
7 J. J. P. Stewart in *Colorado Springs*, CO, USA, p. http://OpenMOPACnet.
8 E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak and F.-Y. Dupradeau, *Nucleic Acids Research*, 2011, **39**, W511.
9 D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. Legrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, University of California, San Francisco, 2016.
10 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *Journal of Chemical Theory and Computation*, 2015, **11**, 3696.
11 M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *Journal of Chemical Theory and Computation*, 2015, **11**, 5723.
12 V. Tsui and D. A. Case, *Biopolymers*, 2000, **56**, 275.