SUPPORTING INFORMATION

Structure-activity relationships in DNA nanocarriers based on the amphipathic cell penetrating peptide transportan 10

Lucas Rodrigues de Mello[1], Tâmisa Seeko Bandeira Honda[1], Sang Won Han[1], Valeria Castelletto,[2] Ian William Hamley[2], Ly Porosk,[3] Ülo Langel,[3,4] Emerson Rodrigo da Silva[1,*]

[1] Departamento de Biofísica, Universidade Federal de São Paulo, São Paulo 04062-000, Brazil

[2] Department of Chemistry, University of Reading, Reading RG6 6AD, United Kingdom

[3] Institute of Technology, University of Tartu, Tartu 50411, Estonia

[4] Department of Biochemistry and Biophysics, Stockholm University, Stockholm SE-10691, Sweden

*Corresponding author: er.silva@unifesp.br

Summary

Supplementary methods:

Powder X-ray diffraction: Fiber X-ray diffraction (fXRD) was performed on stalks dried from concentrated TP10/DNA solutions. The sample was prepared from a solution containing 10 mM TP10 + 5 mM DNA (base pair concentration). At such a high concentration, the resulting solution exhibited a viscous aspect after a few days in the fridge. To prepare the stalks, about 20 µL of TP10/DNA solution were cast in-between wax-coated capillaries in a Petri dish. A suspended film was formed, and the Petri dish was then kept within a desiccator to dry. After three days, the solution evaporated, and a dry peptide/DNA stalk linked the capillaries. The arrangement was then separated, and one of the capillaries held an oriented stalk. The stalk was positioned on the holder of a Rigaku RAXIS IV diffractor equipped with a rotating anode. The X-rays were provided by a CuKα X-ray source (λ = 1.54 Å), and the diffraction pattern was measured using a Saturn 992 CCD camera at a sample-to-detector distance of 40 nm. The experiments were performed at room temperature.

Supplementary bioinformatics simulations: the secondary structure of the peptide was assessed by loading the TP10 amino acid composition on the I-TASSER webserver, a deep-learning platform dedicated to the prediction of the protein tridimensional structure from the primary amino acid sequence, develop by Prof. Yang Zhang at the University of Michigan.[1] The server was accessed on Feb 22, 2024, at the following address: https://zhanggroup.org/I-TASSER/. The top-scoring models were downloaded and subsequently visualized using the Pymol software. Helix wheel projections were obtained using the HeliQuest webserver (accessed on Feb 29, 2024, at https://heliquest.ipmc.cnrs.fr/).[2] The residues identified as belonging to a hydrophilic lysine cluster and to a hydrophobic face were then highlighted in the top-scoring model provided by I-TASSER in order to visualize the distribution of the different amphipathic domain in the tridimensional structure of the peptide.

Dynamic Light Scattering: DLS measurements were conducted using a homemade setup mounted on a Malvern goniometer model PQS-160. The photomultiplier was positioned at 90° relative to the laser path (λ = 633 nm), and data was collected for 90 seconds at temperature of 21 °C. Samples were prepared in filtered HEPES buffer, with a peptide concentration of 1 mg/ml and a charge ratio of 5:1. Volume vs. diameter plots were obtained from data fitting over the obtained correlograms using the sparse Bayesian learning (SBL) algorithm available in the control software (NanoKin, Cordouan Technologies). The different populations were characterized by using the mean and the standard deviation of Gaussian fits adjusted to their size distributions.

Electrophoresis: Electrophoretic runs were performed in 1% agarose gels using a 10 cm cube at 90 V for 100 minutes. Complexes were prepared in autoclaved PBS buffer (pH 7.2, 150 mM NaCl) by mixing appropriate amounts of DNA and peptide from previously prepared stocks. DNA

stocks were equilibrated in the fridge for a few days to ensure full hydration, while peptide solutions were freshly prepared prior to the experiments. The mixing procedure was performed in test tubes, and the mixtures were left to equilibrate for about 30 minutes before being loaded into the gels. The final amount of DNA per well was kept at 5 μg, with corresponding amounts of peptide added to achieve charge ratios (amine-to-phosphate) of 0:1, 0.15:1, 0.6:1, and 1.2:1. Gels were stained with SYBR Green, and gel documentation was carried out using a UV transilluminator.

Transportan 10 structural formula:

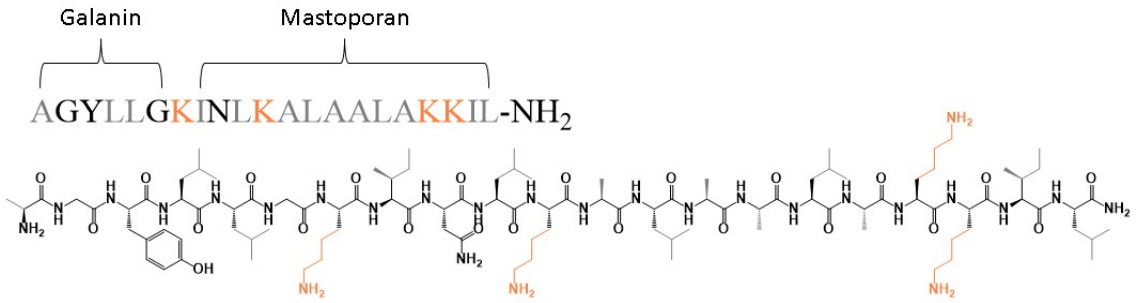

Figure S1: Amino acid sequence and structure of TP10. Hydrophobic residues are highlighted in gray and lysines are marked in orange. The N-terminal portion of the strand carries amino acids derived from the neuropeptide galanin, whereas the C-terminus segment is composed of residues derived from mastoporan. The chimera is linked together through a lysine residue.

Liquid chromatography and mass spectroscopy:

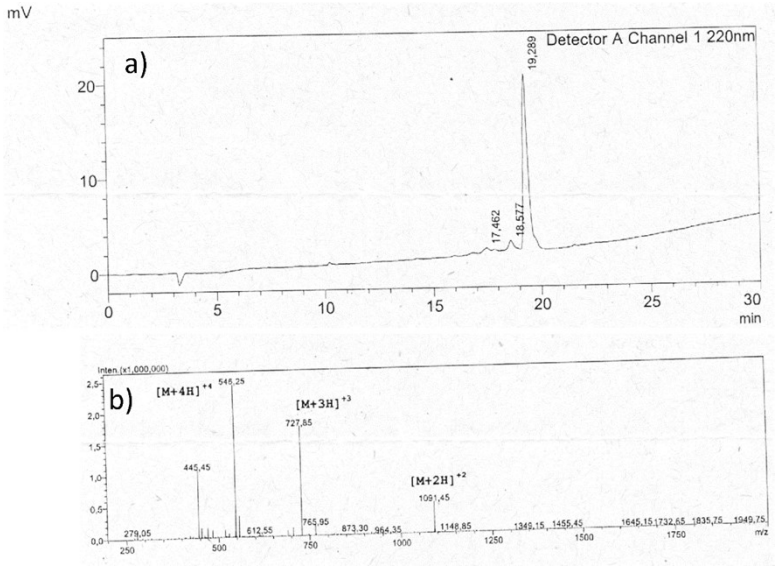

Figure S2: Liquid-chromatography and mass spectroscopy data from TP10 peptide. (a) chromatogram showing the TP10 fraction at elution time ~19.2 minutes (peak area calculated at 95.6%). b) mass spectrum displaying the fragmentation pattern of TP10, $[M+H]^+ = 2181.4$.

Electrophoretic runs:

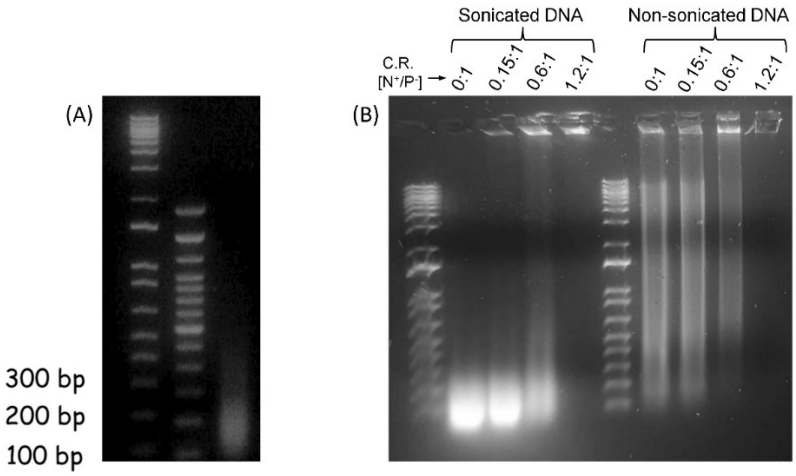

Figure S3: Electrophoretic runs from fragmented DNA samples (A) and from TP10/DNA complexes (B). In (B) the left lanes were loaded with complexes prepared with sonicated DNA with lengths averaging around 200 bp. On the right, complexes were prepared with non-sonicated DNA, resulting in lengths raging from a few hundred base pairs to > 20 kbp. When complexes have charges above 1:1, DNA is no longer observed in the lanes indicating full complexation. The behavior is the same irrespective of DNA length.

Supplementary bioinformatics simulations:

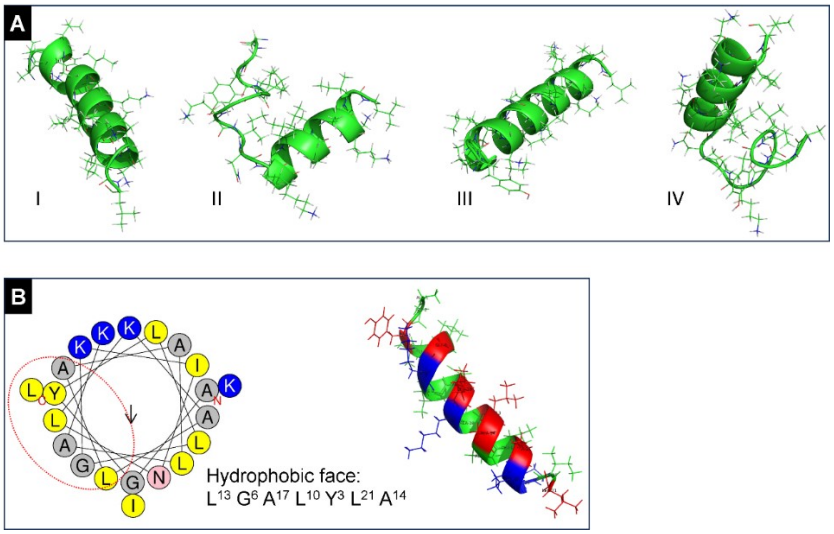

Figure S4: Propensity for α-helix structuration assessed through computational tools. (A) 3-D representations of the highest-scoring secondary structure results obtained from I-TASSER runs (Zhang et al., Nat Protoc 2022, 17, 2326–2353).[1] The C-terminus consistently adopts helical conformation, whereas the galanin-derived N-terminal portion may also exhibit disordered folding (outputs II and IV). (B) Helix wheel projection of TP10 reveals the formation of a hydrophilic lysine cluster (depicted in blue). A hydrophobic face is formed by residues $L^{13}G^{6}A^{17}L^{10}Y^{3}L^{21}A^{14}$ (dotted red line). In the 3-D representation on the right, lysine residues are highlighted in blue, and residues belonging to the hydrophobic face are colored in red, emphasizing the amphipathic nature of the TP10 helix.

Induced CD spectra:

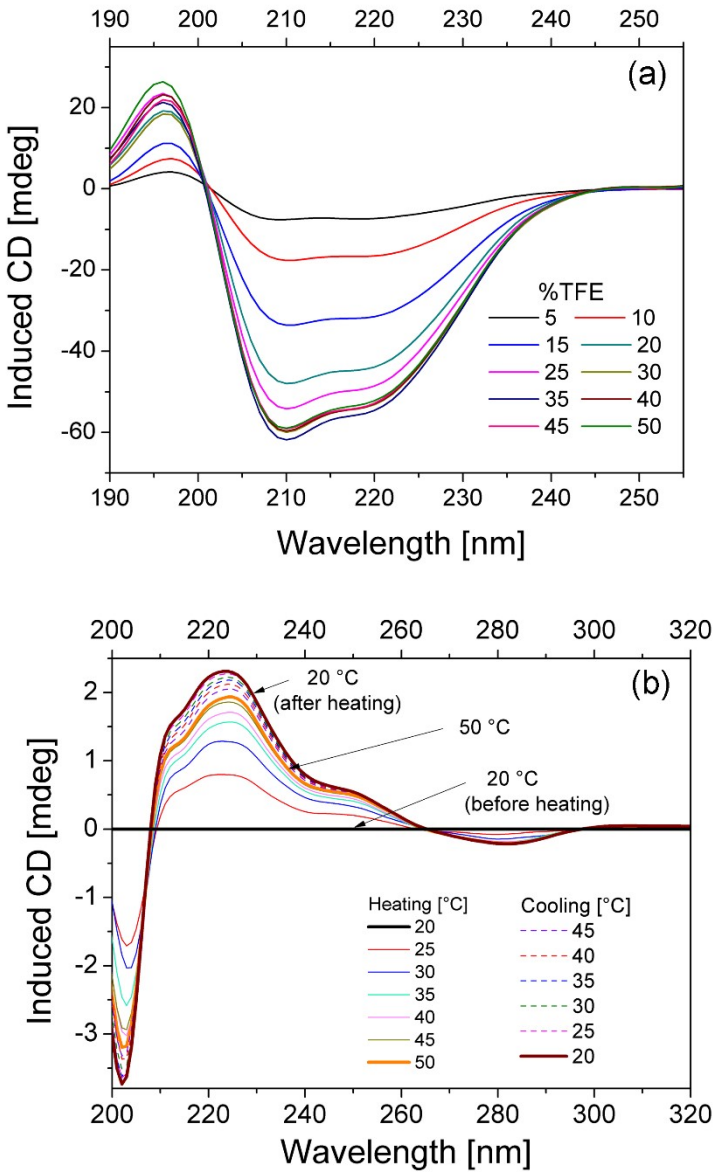

Figure S5: Induced circular dichroism (ICD) obtained from spectra presented in Figure 1 (main text). The spectra shown here were generated by subtracting a reference spectrum (i.e., the initial measurement) from the samples in which a physicochemical variable (%TFE or temperature) was altered. (a) ICD observed in solutions containing different %TFE. (b) ICD resulting from a 1:1 TP10:DNA sample subjected to an up-down temperature cycle.

Best fitting parameters and model equations for SAXS data:

Data fitting was performed using mathematical models widely described in the literature.[3–5] In the case of solutions containing only DNA fragments, the fit was performed using the Porod cylinder form factor. For scattering data from TP10 or TP10/DNA samples, were fitted using a combination of power laws plus the generalized Gaussian coil (CGC) form factor, or power laws along with functional formulas for broad peaks was applied.

The Porod cylinder form factor used here to fit data from DNA solutions is described as it follows:[3]

$$P_{cyl}(q,R,L) = \frac{2}{qL}\left(\Delta\eta R^2 L\right)^2 \times$$

$$\times\left\{Si_{\frac{\pi}{2}}(qL)\Lambda_1^2(qR) - \frac{2\Lambda_2(2qR) - \Phi(2qR)}{qL} - \frac{\sin(qL)}{(qL)^2}\right\} \quad (Eq.S1)$$

R and *L* are the radius and length of the cylinder, respectively. *Si* is the sine integral function given by:

$$Si_{\frac{\pi}{2}}(x) = \left(Si(x) + \frac{cos(x)}{x} + \frac{sin(x)}{x^2}\right) \overset{x\to\infty}{\to} \frac{\pi}{2}$$

Λ is related to special Bessel functions of first and second kinds:

$$\Lambda_1(x) = \frac{2}{x}J_1(x)$$

$$\Lambda_2(x) = \frac{8}{x^2}J_2(x)$$

Φ is given by:

$$\Phi(x) = \frac{2}{x^2}[1 - \Lambda_1(x)]$$

In the case of DNA data shown in Fig. 3A, it is evident that the Guinier region is not achieved within the q-range investigated, and therefore the length of the cylinders is bigger than the resolution of the experiment. In this case, we have thus arbitrarily kept the parameter L fixed at 1000, and the cylinder radius (R) was kept free to fit.

Data fitting for TP10 solutions was performed using the combination of a power law plus the GGC model given by:

$$I(q) \cong Bkg + \frac{A}{q^{\alpha}} + \frac{I_0 \cdot U^{\frac{1}{2\upsilon}} \cdot \Gamma\left(\frac{1}{2\upsilon}\right) - \Gamma\left(\frac{1}{\upsilon}\right) - U^{\frac{1}{2\upsilon}} \cdot \Gamma\left(\frac{1}{2\upsilon}, U\right) + \Gamma\left(\frac{1}{\upsilon}, U\right)}{\upsilon \cdot U^{\frac{1}{\upsilon}}} \qquad (S2)$$

Where:

$$U = (2\nu + 1)(2\nu + 2)\frac{q^2 \cdot {r_g}^2}{6}$$

In Eq. S2, Bkg is just an additive constant accounting for incoherent background. The second term is a scaling law that describe the scattering at low-q region. The scaling exponent α carries information on the fractal dimensionality of the scattering particles.[4,6] This exponent can be used to derive the dimensionality (D) of mass fractals through the relationship:[7]

$$\alpha = D, \; for \; 1 < \alpha < 3 \qquad (S3)$$

Or of surface fractals, through:

$$\alpha = 6 - D, \; for \; 3 < \alpha < 4 \qquad (S4)$$

The third term is the GGC model[3] is used to describe the high-q region that carries scattering information from free peptide chains.

The GGC model carries two structural parameters, the radius of gyration r_g and the Flory exponent ν. r_g is the coil radius assumed by a free peptide chain in solution whereas ν carries information on chain/solvent interactions. A value of $\nu \approx 0.5$ indicates that the chains are in theta solvents. A value of $\nu \approx 0.33$ is related to collapsed chains and $\nu \approx 0.6$ is associated to self-avoiding swollen chains.

Finally, scattering from TP10/DNA samples in which interference peaks appeared were described by using the empirical formula:

$$I(q) = Bkg + A \cdot q^{-\alpha} + B \cdot q^{-\beta} + \frac{C}{1 + \left(|q - q_0|\xi_0\right)^2} + \frac{D}{1 + \left(|q - q_1|\xi_1\right)^2} \qquad (S5)$$

Equation S5 presents a combination of two power laws and two broad peak functions. *A*, *B*, *C* and *D* are scaling factors adjusting the weight of each term, alongside the background *BKg*. In contrast, the parameters α, β, q_0, q_1,and ξ carry structural information. The terms α and β are Porod exponents related to the shape of scattering particles. They can describe a fractal dimensionality as described above. Also, when they assume values closer to 0, this is usually associated to globular shapes, while values near to 1 indicate elongated morphologies.[6,8]

The last terms in S5 are broad peak functions that describe interference peak in the data. q_0 and q_1 are peak positions that are used to determine the average distances between scattering inhomogeneities using Bragg's law, $d = 2\pi/q$. ξ_1 and ξ_2 are correlation length associated with the size of the ordered domains that give rise to the corresponding interference peaks.

Data shown in Figure 3A and 3B (main text) were fitted using Equations S1-S5, and the best fitting parameters are shown in Table S1 below:

Table S1: Best fitting parameters for the SAXS data displayed in Figure 3 (main text). GGC = generalized Gaussian coil model, BP = broad peak model. Model equations are detailed below.

Sample	Fitting model	Parameters								
		Radius [nm]	Gyration radius, Rg [nm]	Flory exponent, ν	Scaling exponent, α	Scaling exponent, β	Peak position,q_0 [nm^{-1}]	Correlation length, ξ_0 [nm]	Peak position, q_1 [nm^{-1}]	Correlation length, ξ_1 [nm]
2.3 mM DNA	Porod	0.91±0.28	####	####	####	####	####	####	####	####
0.5 mM TP10	Power law + GGC	####	0.7	0.39	1.95	####	####	####	####	####
1.4 mM TP10	Power law + GGC	####	1.5	0.54	1.86	####	####	####	####	####
2.3 mM TP10	Power law + GGC	####	1.6	0.31	1.92	####	####	####	####	####
2:1 TP10:DNA	Power law + GGC	####	1.2	0.33	3.96	####	####	####	####	####
10:1 TP10:DNA	Double power law+BP	####	####	####	3.80	0.39	1.62	4.7	####	####
50:1 TP10:DNA	Power law+BP	####	####	####	3.48	####	1.63	6.7	2.11	1.9

DLS measurements on TP10/DNA peptiplexes:

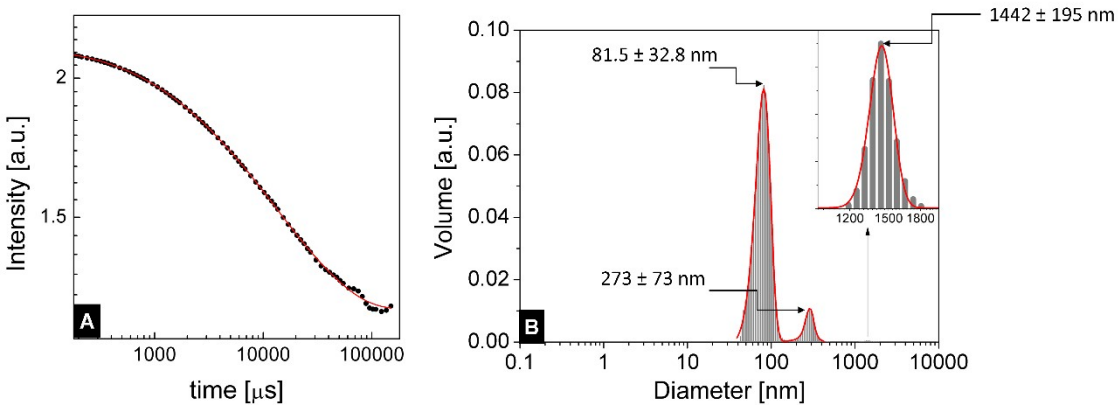

Figure S6: DLS measurements on TP10/DNA mixtures (peptides at 1 mg/mL, charge ratio 5:1). (A) Correlogram obtained over 90 seconds. Data fitting, red curve, was performed using the SBL (sparse Bayesian learning) algorithm. (B) Plot of volume vs. size distribution of the particles found in the sample. The values indicated are the means ± standard deviations of Gaussian functions fitted to size distributions. In the inset, the region between 1,000 and 2,000 nm is shown in detail, revealing the presence of a small fraction of large aggregates in the sample.

Fibre X-ray diffraction data:

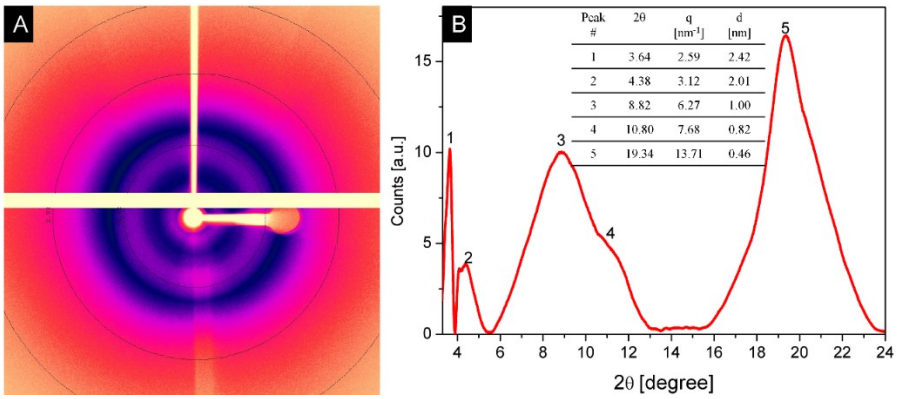

Figure S7: Fiber X-ray diffraction from a TP10 sample. An oriented stalk was dried from a 2.3 mM peptide solution. (A) 2D image showing the diffraction pattern dominated by two major rings corresponding to the most intense reflections. (B) a radially averaged 1D diffractogram reveals the peak positions of five identifiable bands.

Supplementary AFM-IR images:

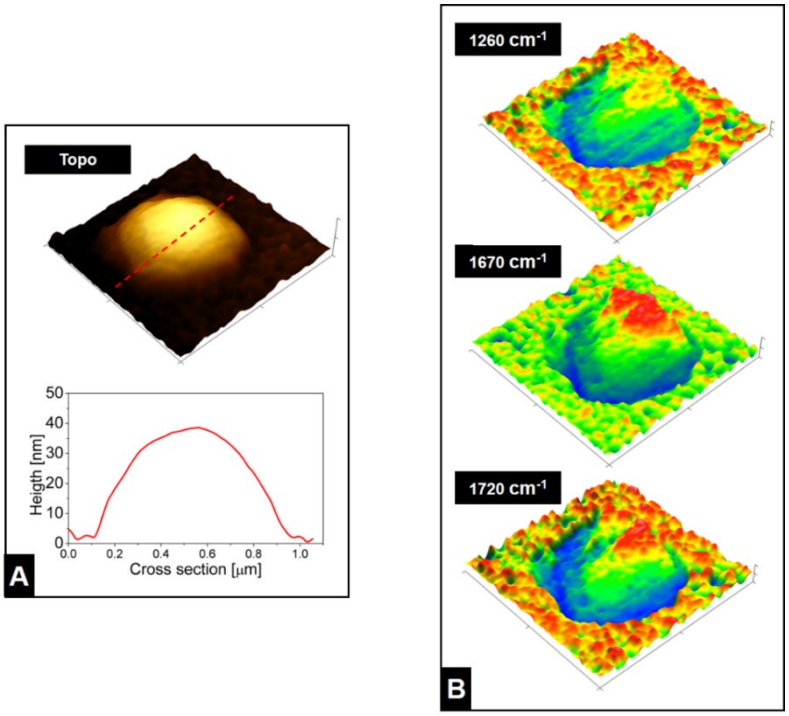

Figure S8: AFM-IR data showing (A) the 3D topographic image along with the cross-sectional profile of a large particle found in TP10/DNA mixtures, and (B) the infrared absorbance maps at wavenumbers associated with the DNA fingerprint region (1260 cm^{-1}), the amide I band (1670 cm^{-1}), and the C=O vibrations (1720 cm^{-1}).

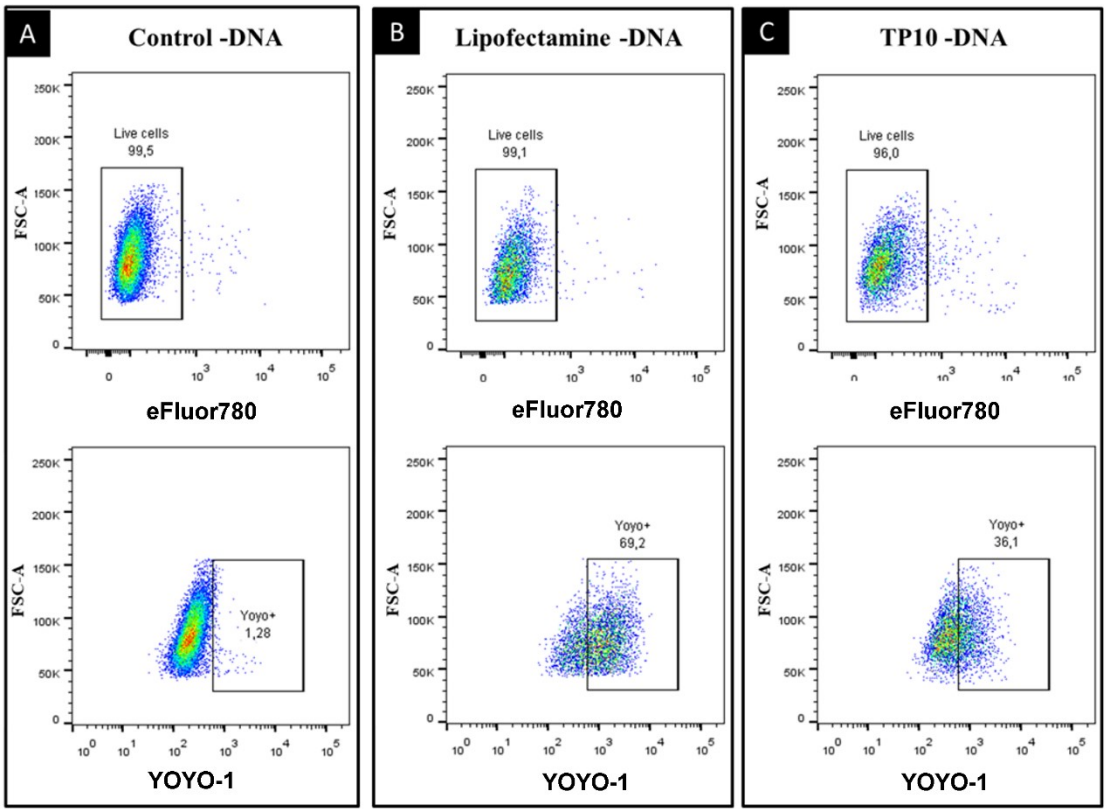

Figure S9: Flow cytometry replicates used for quantitative determinations shown in Figure 7 (main text).

References:

1. Zhou, X. *et al.* I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat. Protoc.* **17**, 2326–2353 (2022).
2. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* **24**, 2101–2102 (2008).
3. Pedersen, J. S. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv. Colloid Interface Sci.* **70**, 171–210 (1997).
4. Zemb, T. & Lindner, P. *Neutrons, X-rays and light : scattering methods applied to soft condensed matter*. (Elsevier, 2002).
5. Hamley, I. W. *Small-Angle Scattering: Theory, Instrumentation, Data and Applications*. (Wiley, 2021).

6. Hammouda, B. Clustering in polar media. *J. Chem. Phys.* **133**, 84901 (2010).

7. Teixeira, J. Small-Angle Scattering by Fractal Systems. *J. Appl. Crystallogr.* **21**, 781–785 (1988).

8. Kogikoski, S. *et al.* Polycaprolactone–Polyaniline Blend: Effects of the Addition of Cysteine on the Structural and Molecular Properties. *J. Phys. Chem. C* **121**, 863–877 (2017).