Supporting Information for:

Self-assembly properties and dynamic of synthetic proteo-nucleic building blocks in solution and on surfaces

Aude LAISNE[†], Maxime EWALD[‡], Toshio ANDO[§], Eric LESNIEWSKA[‡], and Denis POMPON[†*]

[†] Centre de Génétique Moléculaire. CNRS.UPR3404, Avenue de la Terrasse, F91190 Gif-sur-Yvette, France.

[‡] Institut Carnot Bourgogne UMR CNRS 5209, University of Bourgogne, F-21078 Dijon, France.

[§] Department of Physics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan.

***Corresponding Author**: Denis POMPON

*To whom correspondence should be addressed. Tel: +33 567048806; Fax: +33 561559400; Email: denis.pompon@insa-toulouse.fr

Present address:

LISBP/INSA, UMR INSA/CNRS 5504, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France.

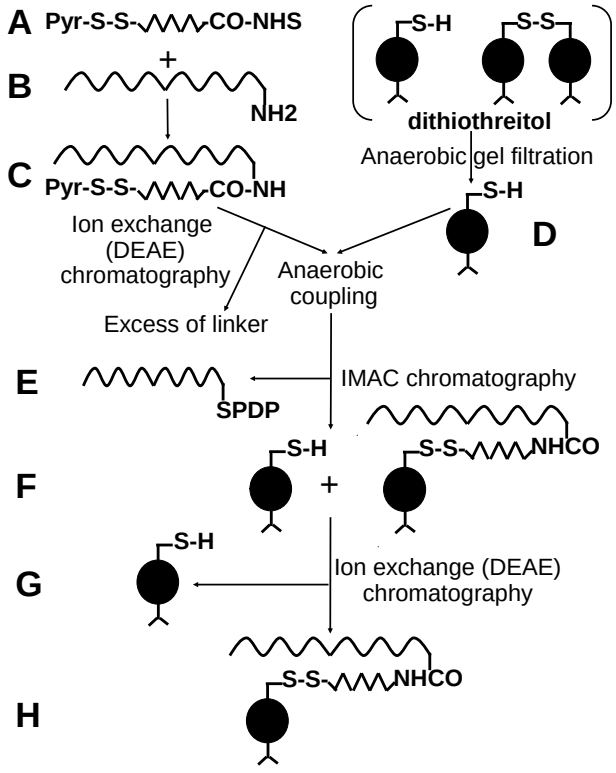
Section 1. DNA tile sequences and matches.

Tiles	Sequence (5'- - 3')	Bases
A---A'	CCGTAGGGTCTCCATTTTGA---AGCGAAGTCACACTGTACA	39
B---B'	GTCCATATGCAAGTCAGCGT---CAGTGTACAGTCACGTCAG	39
a---b'	TCAAAATGGAGACCCTACGG---CTGACGTGACTGTACACTG	39
b---a'	ACGCTGACTTGCATATGGAC---TGTACAGTGTGACTTCGCT	39
a---m	TCAAAATGGAGACCCTACGG---GTAAAACGACGGCCAG	36
b---m	ACGCTGACTTGCATATGGAC---GTAAAACGACGGCCAG	36
M---a'	CAGGAAACAGCTATGAC---TGTACAGTGTGACTTCGCT	36
M---b'	CAGGAAACAGCTATGAC---CTGACGTGACTGTACACTG	36

Matching tile moieties are A ⇌ a, A' ⇌ a', B ⇌ b, B' ⇌ b'

BB' sequence was synthetized in three versions: without, with a 3' amino-linker and with an internally amino-linker coupled thymidine base (underlined). In the case of the AA' sequence, only the unmodified and 3' amino-linker versions were used. The amino-linker modifications were used for protein derivatization in PDNAs.

Section 2. Schematic for the synthesis of protein-DNA (PDNA) structures.



Section 3. Synthesis of protein-DNA (PDNA) structures.

Genetic Engineering of Cytochrome b5: A human microsomal cytochrome b5 cDNA (GenBank accession number L39792) has been previously engineered to substitute the 26 C-terminal residues constituting the hydrophobic membrane anchor tail by the -NGHHHH-COOH sequence (16) encoding an His-tag. Engineered cDNA was subcloned into pCR®2.1-TOPO® cloning (Invitrogen) and the *Sph*I-*Bam*H1 fragment was ligated into the corresponding sites o*f* digested pUHE25-2 expression vector giving pUHE-[Hb5-(HIS_4)C-term]. This construction includes an additional Leu codon inserted immediately after the normal Met initiation codon. Site directed mutagenesis of plasmid was performed with QuickChange II XL Site-Directed Mutagenesis kit (Stratagene) to introduce a Ser-24-Cys aa-sequence change. This brought a unique solvent accessible thiol-containing residue at the surface of the b5 globular domain. Mutation has been checked by sequencing resulting in the pUHE-[Hb5-24C-(HIS_4) C-term] expression plasmid.

Protein expression and purification: XL1-Blue E. *coli* cells transformed by pUHE-[Hb5-24C-(HIS_4)C-term] were transferred in 50 mL liquid Luria-Bertani containing 100 µg/mL ampicillin and grown over-night at 37°C, 160 rpm. 45 mL of this culture were transferred into 1 L of Terrific Broth (TB) containing 100 µg/mL ampicillin and bacteria were grown at 28°C for 24 hrs. 1 mL aliquotes of the culture were stored at -80°C in 1X Hogness. Protein expression was induced with 0.5 mM IPTG and 0.5 mM δ-ALA added with 500 mL of TB and occured at 28°C for 24 hrs with shaking at 160 rpm. Cells were harvested after 10 minutes 8500 x g centrifugation at 4°C, washed 2 times with 50 mL 20 mM Na/K phosphate, 0.5 M NaCl, pH 6.5 in the presence of 1 mM PMSF, and suspended into the same buffer with anti-protease cocktail. Cells lysis was performed by repeated sonication following treatment with 1.25 g/L lysozyme , 50 mg/L ribonuclease A and 50 mg/L DNAse I. After removal of cell debris by centrifugation at 18500 g for 10 min, the supernatant was applied onto a 1.5 x 6 cm column of imminodiacetate immobilised on agarose (Sigma) equilibrated successively with 0.05 M $NiCl_2$ in 0.1 M KAc, 0.4 M KCl adjusted to pH4.5, with HCl, washed with the same buffer without $NiCl_2$ and finally equilibrated with 50 mM Na/K phosphate buffer pH 7.4 (buffer A). For elution, column was washed with buffer A containing increasing NaCl concentrations (0.1 M to 0.5 M) until OD_{280} of the eluate was lower than 0.3. The protein was eluted with 30 mM L-histidine solution in buffer A containing 1 mM PMSF. 5 mM EDTA was added to fractions containing the protein before loading onto a DEAE-Sephacel column (1.5 x 9 cm) equilibrated with buffer A. The column was washed with buffer A supplemented

with 0.1 M NaCl. The protein was eluted with buffer A containing 0.35 M NaCl. Fractions were pooled and dialysed against 1 L of 20 mM Na/K phosphate, 1 mM DTT pH 6.5. Eluate was concentrated using a Vivaspin 10000 MWCO device (Vivascience) before loading on a gel filtration column (Superdex 75 prep grade Amersham Biosciences) 1.6 x 50 cm equilibrated with 20 mM Na/K phosphate pH 6.5 and eluted at 2 ml/min. Concentration of the purified protein was determined by spectrophotometry using ε_{412nm}=118 $mM^{-1}.cm^{-1}$.

Linker coupling to ss-DNA: the SPDP reagents (Pierce) are heterobifunctionnal cross-linkers (SI Section 2A). Succinimidyl 6-[3'-(2-pyridyldithio) propionamido]hexanoate (LC-SPDP) is a 15.6 Å long spacer arm which can form, on one side an amid bond with the 3'-amino-modified oligonucleotide and, on the other side a disulfide bond with the engineered protein : Hb5-24C-(His_4)Cterm. Poorly soluble in aqueous buffers, LC-SPDP was first dissolved in DMSO (90 mM) before all dilution. The 39 mers 3'-amino-modified ss-DNA (SI Section 2B) were coupled to the bifunctional linker in 50 mM Na/K phosphate buffer pH 8, using 15 mM LC-SPDP for a DNA concentration ranging between 50 µM to 100 µM. The reaction occurred at room temperature. The coupling efficiency was checked by ion-pairing reverse phase chromatography on a Waters XTerra MS C18 column (3.5 µM, 4.6 x 100 mm) equilibrated with 0.1 M triethylamonium acetate pH 7.4 in water and eluted at 1 mL/min using a linear gradient between the aqueous buffer and acetonitrile (0 to 21.4% by vol. in 10 min). Coupled ssDNA (SI Section 2C) were purified from excess of reagent and of hydrolysis subproducts by chloroform extractions prior loading onto an 0.6 x 2 cm anion exchange column (DEAE Sephacel) equilibrated with 50 mM Na/K phosphate pH 7.4 (buffer A). After washing with 0.25 M NaCl in buffer A, DNA was eluted with 1 M NaCl in buffer A. DNAs were desalted against water using Float-A-Lyzer (Fisher) ∅5 mm, MWCO 3,5 K and concentrated by speed vacuum without heating (concentrator 5301, Eppendorf). Concentrations of coupled oligonucleotides were calculated from their absorption spectra and were stored at -20 °C.

Hb5-24C($HIS)_4$ coupling with LC-SPDP-DNA: To prevent a high level of spontaneous dimerisation of Hb5-24C($HIS)_4$, the protein was reduced and the monomer separated from DTT excess by gel filtration on G25 sephadex column equilibrated in 10 mM Tris-HCl pH 8 1mM EDTA under anaerobic conditions (SI Section 2D). LC-SPDP-oligonucleotides were coupled for 24 h at room temperature with 1.5 molar excess of reduced engineered b5 in a hermetic flask filled with argon. Coupling products were analyzed by ion pairing reverse phase HPLC.

Protein-DNA complex purification: Combined chromatographies were used to separate the complex from non-coupled DNA and non-coupled engineered b5. The reaction mixtures were loaded separately onto a 0.5 x 1.5 cm nickel chelate column equilibrated with buffer A. Column was washed with buffer A to remove uncoupled oligonucleotides (SI Section 2E) and pyridine-thiol subproducts until A_{260} absorbance was no more detectable. The his-tagged proteins were eluted by a solution of 30 mM histidine in buffer A. DNA-coupled and non-coupled protein fractions (SI Section 2F) were pooled based on the 413 nm absorption spectra and were treated with 5 mM EDTA in order to disrupt potential nickel-proteine complexes. Fractions were applied onto a 0.5 x 3 cm DE52 anion exchange column equilibrated with buffer A. Following a wash step with 0.35 M NaCl in buffer A to elute protein which is not coupled to oligonucleotide (SI Section 2G), the PDNA complex (H) was finally eluted with 1 M NaCl in Buffer A. The concentrations of PDNA complexes were determined by spectrophotometry at 260 and 413 nm and the purity checked by reverse phase HPLC for chemical contaminants and by reducing and non-reducing SDS-PAGE gels for protein contaminants. Both controls demonstrated the absence of any detectable DNA, protein or chemical contamination.

Section 4. Contribution of building blocks to apparent size in gel filtration and electrophoresis.

Methods	Unit	ss-DNA[1] (base)	ss-DNA[2] (base)	ds-DNA (bp)	Prot (unit)
Gel filtration[a]	kDa (bp)	0.30 (0.50)	0.30 (0.50)	0.70 (1.00)	6.6 (10)
SDS-PAGE[b]	kDa (bp)	0.28 (0.46)	0.28 (0.46)	0.36 (0.51)	15.0 (25)
Agarose gel[a,c]	kDa (bp)	0.36 (0.60)	0.94 (1.55)	0.70 (1.00)	16.5 (25)
Agarose gel[a,d]	kDa (bp)	0.42 (0.69)	1.08 (1.78)	0.73 (1.04)	17.5 (26.5)

Coefficients were calculated from the least square adjustment of calculated to experimental values.

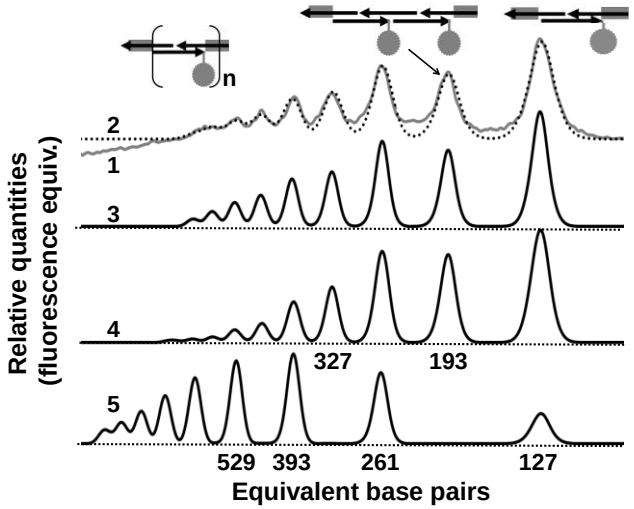
[1]Contribution per base for m and M ss-DNA extensions in protein free constructs

[2]Contribution per base for m and M ss-DNA extensions in protein containing constructs

[a]Calibrated with ds-DNA standards standards assuming an average of 330 Daltons per base.[b]Calibrated with protein standards. [c]Simple assemblies of PDNA and DNA from Figure 6. [d]Data from Figure 8.

Section 5. Experimental and simulated profiles for PDNA copolymers.

Scan (line 1 on the next Figure) of the experimental PDNA gel profile from Figure 8A and its deconvolution using an optimized expansion factor of 1.5 (line 2) and regular resolution for ds-DNA (line 3). Simulations using parameters defined in SI Section 6 with (line 4) or without (line 5) taking into account the SIC value. Regular ds-DNA gel resolution was used for reconstruction.



Section 6. Parameter sets for simulations.

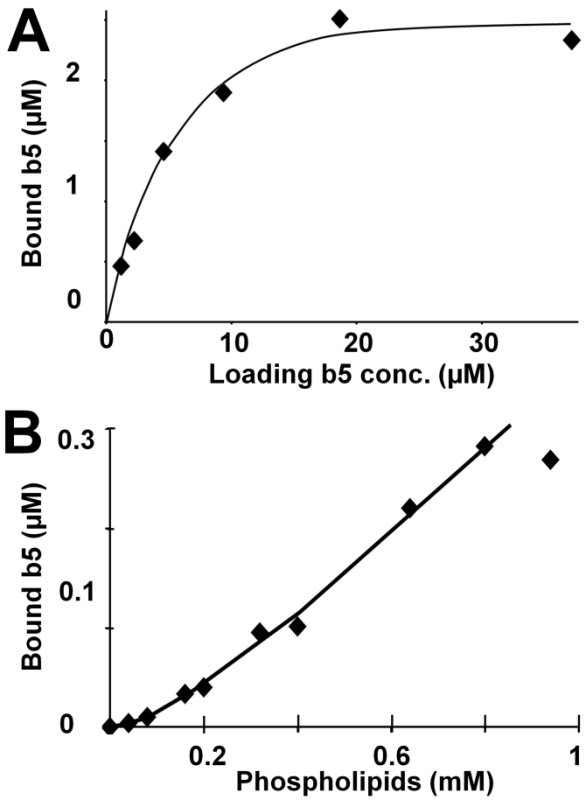
All simulations were performed and parameters defined as described in (21) and take into account the environment and length modulation of rate constants (with RL =19 bases), and the J-Factor table PT2 of the same paper. For rapid–mixing experiments (Figure 7) final concentrations used were 0.28 µM of [AA'PDNA-ba'DNA] and [BB'PDNA-ab'DNA] hybrids (panel A) and 0.35µM of [AA'-ba'DNA] and [BB'-ab'DNA] hybrids (panel B). For simulation of agarose gel profile in SI Section 5, concentrations were AA'b5 (0.7), BB'b5 (1.1), ab' (0.95), ba' (0.85), Ma' (1.54), Mb' (1.5), am (1.74), bm (1.38) µM. Rate constants used for simulation are given in the following table. Others parameters, for rapid-mixing experiments simulation, were hypochromic factors (Hf) of 11.3 and 10.4 % for PDNA and DNA, respectively. For the simulation of agarose gel profiles visualization parameters were SSEH = 0.5, DSEQ: yes, DSSS =1, DSDS =1.05 and DSCS =1. A value of 0.09 µM for SIC parameter was only used for curve 4 in SI Section 5.

Rate constants ($\mu M^{-1}. min^{-1}$)	PDNA	DNA	PDNA	DNA
	Type 1		Type 2	
A'/a'	5.6	7.0	8.8	11.1
A/a	3.7	3.7	5.8	5.8
B'/b'	4	3.4	6.3	5.4
B/b	5.5	4.1	8.7	6.5

Type 1 corresponds to the hybridization of ss-regions flanked by ss-regions. Type 2 refers to the hybridization of the ss-regions flanked by ds-regions.

Section 7. Cytochrome b5 and phospholipid concentration dependence for the formation of protein-lipid complexes.

A, saturation curve for the association of cytochrome b5 to liposomes using a total phospholipid concentration of 700 μM and variable b5 concentrations. B, same experiment using a constant b5 concentration of 10 μM and variable phospholipid concentrations. Freshly extruded (DOPC, DMPC, DOGS, 45: 45: 10, by mol.) phospholipids (0.7 mM) were incubated for 10 min at 24°C with different concentrations of cysteine-containing protein (1.2, 2.3, 4.6, 9.2, 18.5 and 37 μM). The mix (70 μL) was diluted to 400 μL with 1 x PBS buffer, before loading onto a nickel-charged IMAC column (0.5 x 1 cm). The column was washed with 2 x 400 μL buffer. Spectra were recorded on each eluted fraction and the amount of unbound material calculated by deconvolution of the cytochrome b5 spectra. The procedure was repeated using variable concentrations of phospholipids and a fixed concentration of cytochrome b5.



Section 8. Purification of the DNA and PDNA assemblies described in the paper

Bi-block construction (AA'/ba', BB'/ab', AA'-b5/ba', BB'-b5/ab'). Corresponding tiles (PDNA or/and DNA) 1-20 µM (depending on experiment) were hybridized in 1x PBS buffer for 1h at 42°C, then centrifugated prior purification by HPLC gel filtration performed at 0.3 ml/ min in 0.25 M NaCl, 0.15 M NaKPO4 buffer at pH 6.8 using a 4µm UHR SEC, 4.6 x 300 mm column (Waters). When PDNA and DNA were hybridized together a slight molar excess (1.2 x) of the DNA tile was used to improve yield.

Tri-block construction (AA'/ba'/ab', BB'/ba'/ab', AA'/BB'/ab', AA'-b5/ba'/ab', BB'-b5/ba'/ab', AA'-b5/BB'-b5/ab'). The procedure was formally the same than for diblock synthesis except that three tiles were used.

Penta-block construction (AA'/BB'/ba'/ab'/ba', AA'-b5/BB'-b5/ba'/ab'/ba'): suitable HPLC purified di-block and tri-block (see previously) were mixed and let to hybridize at 42°C for 60 minutes prior purification by HPLC gel filtration.

Framework bound PDNA (Figure 11): The [Mb'-ba'-ab'-ba'-ab'-reverse of m] sequence block was cloned on the M13 minus strand orientation (see SI Section 1 for letter codes) in the PCR cloning site of pCR2.1 (Invitrogen) using standard molecular biology techniques. A 513 bp fragment of the plasmid was amplified by 30 cycles of unsymmetrical PCR using a 10:1 molar ratio of primers 5'-CTGTTGGGAAGGGCGATCG-3' and 5'-ACAGCTATGACCATGATTACGCC-3', respectively. The PCR product was purified from protein by phenol-chloroform extraction and ethanol precipitation. Following quantification, the PCR product was heated for 3 min at 100°C and frozen in liquid nitrogen. 1.5-stoichiometry of 3'-end grafted AA'-b5 and internally grafted BB'-b5 PDNAs (taking into account the presence of 2 binding sites for AA' and 3 for BB' on the matrix) were added to the frozen tube and let to warm to room temperature. Following incubation for 30 min at 42°C, resulting hybrids were purified by HPLC size exclusion chromatography. While asymmetric PCR mostly generates ss-matrix, the construct might contain small amount of re-hybridized double stranded matrix contaminating the expected PDNA-framework hybrid due to the impossibility to discriminate large DNA structures by gel filtration.

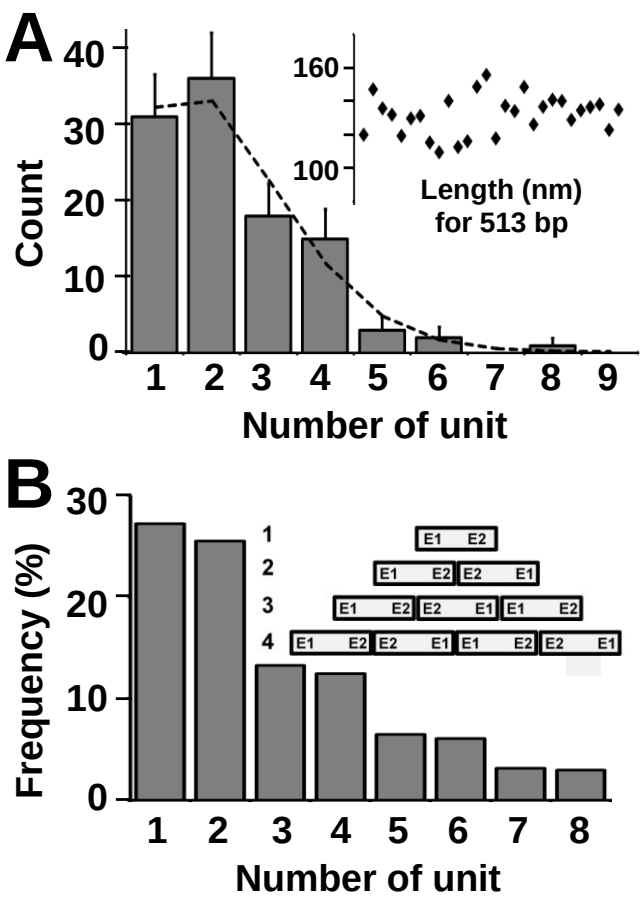
Section 9. HS-AFM image sequence.

Section 10. Thermodynamic parameters calculation

ΔG (kcal/mol) for formation of partial base pairing between ss-extension of the PDNA construct analyzed on Figure 10 was found to be -7.8 for (b/b), -7.5 for (a'/a') and -3.6 for (b/a') extension couples. Calculation was performed assuming a solution including 150 mM NaCl and 1 µM DNA using the online calculator (http://eu.idtdna.com/analyzer/ Applications/ OligoAnalyzer). The energy of the stronger possible interaction was taken in each case.

For calculation of interaction energies of ss-extension in the dynamic AFM conditions, the experimental distribution of concatemer lengths determined from AFM images was first simulated using as model a monomer block bordered by two sticky extensions, considering only the two stronger classes (b/b and a'/a') of interactions. The model was simulated with different affinities of sticky ends until the global shape of simulated length distribution (SD figure 11 B) optimally fit experimental distribution. Finally, affinities were converted to free energies using standard thermodynamic formula.

Section 11. Size distribution of concatemers observed in AFM imaging (main text Figure 10)



Panel A, the contour length distribution of concatemers was characterized using several sets of images (time range of ~3 min). Figure insert, contour length distribution for a 513 bp DNA.

Panel B, simulation of the concatemer formation assuming end-to-end interactions and ratios between rates of dissociation and formation of 0.05 and 0.25 for E1-E1 and E2-E2 interactions, respectively. No E1-E2 interaction was considered.

Method: Statistical analysis of the formation of concatemers was performed on a large set (several tenths of images, 106 objects) resulting in determining the size distribution of species. Calibration was performed using a 513 bp DNA, leading in high speed imaging conditions to apparent contour length of 0.26 nm/bp. The observed PDNA assemblies indeed correspond to structures including one to eight repeats (2-16 PDNAs) of the 36 nm long construct.