

Electronic Supplementary Information (ESI)

DNA-protamine condensates under low salt conditions: Molecular dynamics simulation with a simple coarse-grained model focusing on electrostatic interactions

Yun Hee Jang,^{*abc} Eric Raspaud^c and Yves Lansac^{*abc}

^a GREMAN UMR 7347, Université de Tours, CNRS, INSA CVL, Tours, France (lansac@univ-tours.fr)

^b Department of Energy Science and Engineering, DGIST, Daegu, Korea (yhjang@dgist.ac.kr)

^c Laboratoire de Physique des Solides, CNRS UMR 8502, Université Paris-Saclay, Orsay, France

Table of contents

S1. DNA pairwise interaction	S1
S2. Diffusion coefficients	S3
S3. Persistence length of a single long DNA	S5
S4. 20 Independent simulations	S8
S5. Simulations from a pre-assembled bundle	S12
References	S13

S1. DNA pairwise interaction

Fig. S1 shows the effective interaction or potential of mean force (PMF; ΔF) between two dsDNA at each $R_{+/-}$ with respect to their non-interacting zero-free-energy limit ($\Delta F = 0$) reached at infinite r or practically when they are sufficiently far apart, i.e., $r = 140$ Å if they are kept parallel or 60 Å if they are free to rotate in parallel planes. Each curve in Fig. S1 shows a local maximum ΔF_{\max} , which represents a long-range repulsion or a free energy barrier to overcome in order to bring them together from the infinite r to a bound pair. At shorter r separated by this free energy barrier, each curve shows a local minimum ΔF_{\min} , which represents a short-range attraction or a binding free energy. The difference of these two values, $\Delta\Delta F = \Delta F_{\max} - \Delta F_{\min}$, would represent another free energy barrier required to break apart the bound pair. These three values at each $R_{+/-}$ are summarized in Table S1, along with those found in our previous simulations with a more-refined model^{S1} for comparison. Due to our simple and generic single-chain bead-spring model of dsDNA, the short-range r at the local minimum, i.e., the binding distance, is ~ 10 Å at all $R_{+/-}$, which is uniformly shorter than ~ 30 Å obtained previously with our more-refined model.^{S1} However, the strongly- $R_{+/-}$ -dependent binding energetics, e.g., the lower free energy barrier ΔF_{\max} and the lower (more negative) binding free energy ΔF_{\min} observed when $R_{+/-}$ is closer to the isoelectric point ($R_{+/-} = 1$), are well captured by the current model and qualitatively well agree with the qualitative trends shown previously with our more-refined model.^{S1} Quantitatively, we see in Table S1 some differences in the $R_{+/-}$ -dependent trends obtained with our two models. For example, the $R_{+/-}$ -dependence is stronger, i.e., the binding free energy ΔF_{\min} is lower at the isoelectric point and the free energy barrier ΔF_{\max} is higher when $R_{+/-} < 1$, with our current model than with our previous model,^{S1} indicating that the electrostatic interaction of DNA, both attraction toward protamine and repulsion between each other, is stronger with our current poly-bead DNA model than with our previous more-refined model.^{S1}

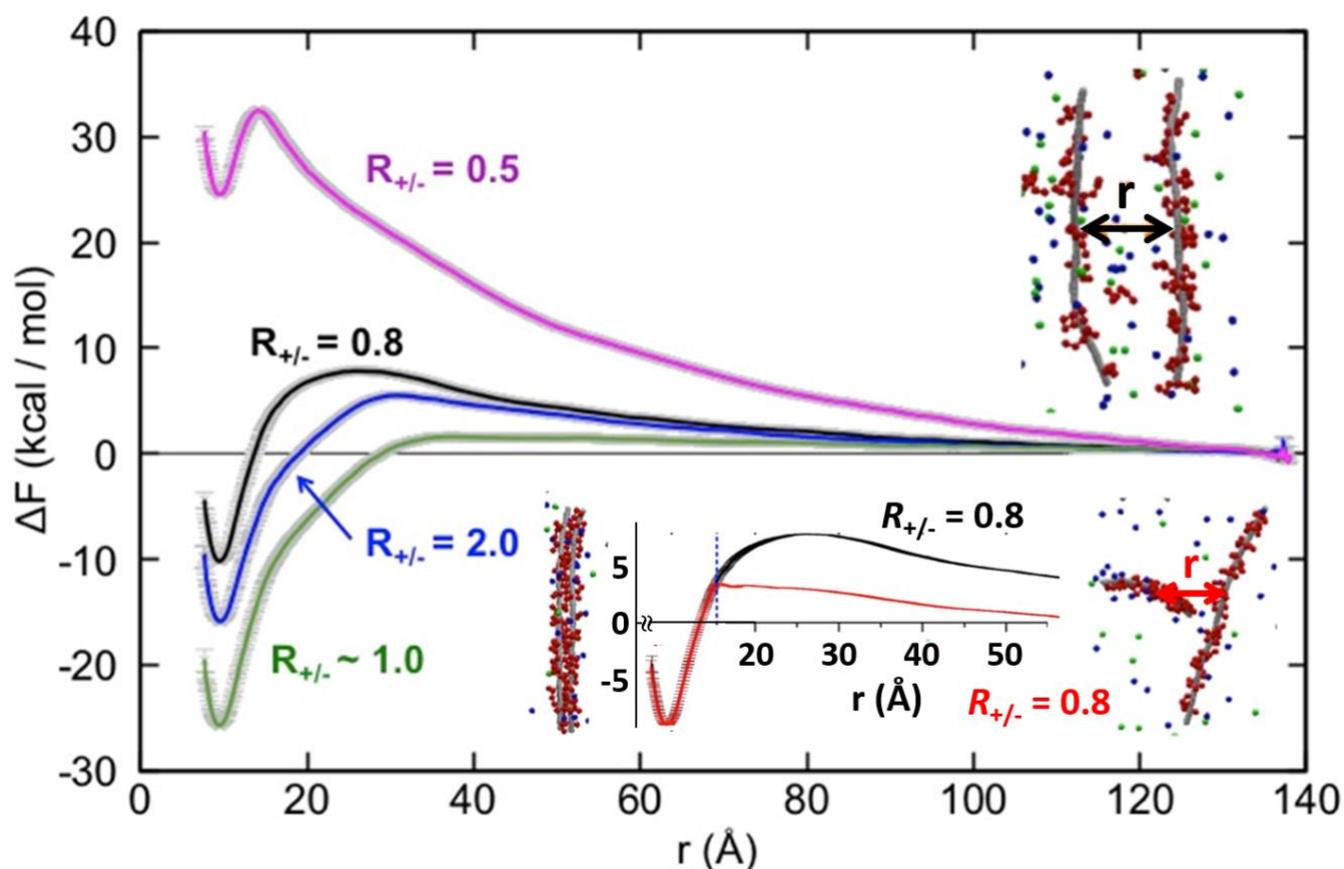


Figure S1. Effective interactions between a pair of dsDNA fragments held parallel to each other (upper inset; DNA in grey, protamine in red, and NaCl in blue and green) and their uncertainties, which are calculated by umbrella sampling along the reaction coordinate r (upper inset; black) with angular and dihedral restraints in excess of DNA ($R_{+/-} = 0.5$ and 0.8 ; magenta and black curves), in excess of protamine ($R_{+/-} = 2$; blue curve) and at the isoelectric point ($R_{+/-} = 1$; green curve). Together shown are the effective interaction calculated at $R_{+/-}$ of 0.8 after removing the dihedral restraints, i.e., when two dsDNA are free to rotate while held in two planes parallel to each other (lower inset; red curve). The effective interaction as well as the long-range free energy barrier at the local maximum (ΔF_{\max}) and the short-range binding free energy at the local minimum (ΔF_{\min}) are shown with respect to the non-interacting limit where two DNA are sufficiently far apart ($\Delta F = 0$), i.e., when r is 140 Å (parallel) or 60 Å (free-rotating).

Table S1. Pairwise DNA binding energetics (kcal mol^{-1})

$R_{+/-}$ ^a	ΔF_{\min} ^c	ΔF_{\max} ^d	$\Delta\Delta F_{\max}$ ^e	$R_{+/-}$ ^b	ΔF_{\min} ^c	ΔF_{\max} ^d	$\Delta\Delta F$ ^e
				0.25	1	5	5
0.5	24	32	6	0.375	-4	4	9
0.8	-10	9	19	0.75	-12	1	14
1.0	-25	1	26	1.0	-16	1	17
2.0	-15	6	21	1.25	-9	4	12
				2.5	0	7	7

^a This work, ^b Ref. S1

^c PMF at the short-range local minimum w.r.t. the non-interacting limit ($\Delta F = 0$), i.e., a binding free energy

^d PMF at the long-range local maximum w.r.t. the non-interacting limit ($\Delta F = 0$), i.e., a barrier to form a bound pair

^e PMF at the long-range local maximum w.r.t. the short-range local minimum, i.e., a barrier to break a bound pair

At the isoelectric point ($R_{+/-} = 1$, Fig. S1, green curve), the neutral DNA complexes experience essentially no repulsive barrier ($\Delta F_{\max} < 1 \text{ kcal mol}^{-1}$) to come close and assemble with the strongest effective attraction ΔF_{\min} of $-25 \text{ kcal mol}^{-1}$, which is ascribed to positional correlations including salt bridging of protamine adsorbed on the DNA pair.^{S2,S3,S4}

In excess of DNA but close to the isoelectric point ($R_{+/-} = 0.8$, Fig. S1, black curve), DNA complexes are negatively-charged with only 80% of charge neutralization and repel each other in a parallel configuration. A free energy barrier ΔF_{\max} of 9 kcal mol^{-1} needs to be overcome to achieve self-assembly stabilized by an attraction ΔF_{\min} of $-10 \text{ kcal mol}^{-1}$ with respect to the non-interacting limit ($r = 140 \text{ \AA}$). If the parallel-orientation restraint is removed but the position restraint is kept between their geometrical centres (Fig. S1, lower inset, red curve), the DNA complexes minimize the repulsion by adopting a perpendicular orientation to each other at large r , and then adopt the same parallel orientation at short r to maximize short-range attraction induced by position correlation of protamine. It lowers the free energy barrier ΔF_{\max} to 5 kcal mol^{-1} with respect to the non-interacting limit reached sooner ($r = 60 \text{ \AA}$). If even the position restraint between their geometrical centres is released, the DNA complexes would minimize their repulsion even further by approaching each other's tips, not the centres, in the perpendicular orientation. This would lower the free energy barrier even further and allow DNA assembly at this $R_{+/-}$ near the isoelectric point. Indeed, such transient configurations are often observed in unbiased MD simulations on DNA-protamine mixtures.

In larger excess of DNA far from the isoelectric point ($R_{+/-} = 0.5$, Fig. S1, magenta curve), negatively-charged DNA complexes with only 50% of charge neutralization repel each other even more strongly, creating a much higher free energy barrier ΔF_{\max} of 32 kcal mol^{-1} to overcome for self-assembly when they are restraint to align parallel to each other. This barrier is expected to stay still high even when the restraints are released, and the self-assembly would be still rare and transient.

In excess of protamine ($R_{+/-} = 2$, Fig. S1, blue curve), the positively-charged DNA complexes repel each other to create a free energy barrier ΔF_{\max} of 6 kcal mol^{-1} in a parallel orientation, but eventually the self-assembly with the attraction ΔF_{\min} of $-15 \text{ kcal mol}^{-1}$ can be achieved in the same manner as in a small excess of DNA ($R_{+/-} = 0.8$).

S2. Diffusion coefficients

We perform MD simulations in implicit solvent and no friction is present. We nevertheless choose to assign masses in order to roughly reproduce the ratio between the translational diffusion coefficient of protamine D_{PRO} and the translational diffusion coefficient of DNA D_{DNA} .

As stated in the main text, DNA and protamine length are decreased by a factor 3 with respect to the experimental values due to constraints imposed by the dilute regime condition and the need to be able to handle long-range electrostatic interactions. For DNA in saline water, D_{DNA} has been measured previously by a spot-photo-bleaching of fluorescein-labelled dsDNA fragments with a focused laser and then related to the DNA size by an empirical relationship,^{S5}

$$D_{\text{DNA}} = \frac{4.9 \times 10^{-6} \text{ cm}^2/\text{s}}{(\text{bp size})^{0.72}}$$

which leads to D_{DNA} of $1.35 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ for the 146-bp long fragment used in experiments and of $2.9 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ for the 50-bp long fragment used in the present simulation. For protamine in water, D_{PRO} has been

estimated to be $\sim 1.6 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ from its hydrodynamic radius at 22°C , which is 1.35 nm ,^{S6} using the Stokes-Einstein equation. D_{Na} and D_{Cl} for Na^+ and Cl^- counterions have been estimated as 1.33×10^{-9} and $2.03 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$, respectively.^{S7} The estimated relative ratio $(D_{\text{PRO}}/D_{\text{DNA}})_{\text{exp}}$ in experiments is therefore ~ 11.9 and it is kept fixed in our simulation, i.e. $(D_{\text{PRO}}/D_{\text{DNA}})_{\text{sim}} = 11.9$, leading to a diffusion coefficient for the short protamine (1/3 of the original one) of $\sim 3.5 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$.

The mass of 50-bp DNA model M_{DNA} is 33 kDa ($= 660 \text{ Da} \times \text{bp size}$), and each of its 100 beads has a mass m_{DNA} of 330 Da . The mass of salmon protamine is $\sim 4250 \text{ Da}$, and each of its 21 constituting beads has a mass m_{pro} of $\sim 202 \text{ Da}$. The mass of Na^+ and Cl^- , m_{Na} and m_{Cl} , are 23 and 35.5 Da , respectively. We thus assign to each bead the relative mass expressed in the unit of m_{Na} , i.e., 14.35 , 8.78 , 1.00 , and 1.53 as m_{DNA} , m_{pro} , m_{Na} , and m_{Cl} , respectively.

We then verify that these masses roughly reproduce the ratio between D_{PRO} and D_{DNA} by performing 200-ns long MD simulations in periodic cubic cells of size 48 nm , which contain a system constituted of 20 DNA and 2000 Na^+ counterions or a system constituted of 143 protamine and 1001 Cl^- counterions. The concentration of each species is the same as used in the unbiased MD simulations at the isoelectric point. For the DNA case, D_{DNA} is estimated with two bending coefficients K_b of 700 and $150 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. Based on the Einstein-Stokes equation, fluctuations in the mean-square displacement (MSD) $\Delta r^2(t)$ of a species are related to its translational diffusion coefficient D as

$$6tD = \lim_{t \rightarrow \infty} \langle |\mathbf{r}_{\text{CM}}(t_0 + t) - \mathbf{r}_{\text{CM}}(t_0)|^2 \rangle = \lim_{t \rightarrow \infty} \Delta r^2(t)$$

where $\mathbf{r}_{\text{CM}}(t)$ is the center-of-mass (CM) position of a given species at time t (see Fig. S2 for the DNA case).

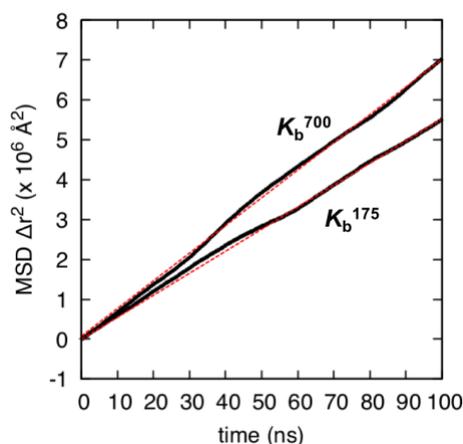


Figure S2. Mean-square displacement (MSD) of center of mass of 50-bp DNA fragments in a cubic box of side length 48 nm containing 20 DNA fragments and their Na^+ counterions. Results on the DNA fragments with a bending coefficient K_b of $700 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ (K_b^{700} , upper curve) and $175 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ (K_b^{175} , lower curve) are shown together. Fits performed on the long-time motion behaviour of MSD are shown as dashed red lines.

The fits of the MSD curves corresponding to the long-time motion behaviour, i.e., done when $\Delta r^2 \propto t$, give $\sim 1.2 \times 10^{-7}$ and $\sim 9.3 \times 10^{-8} \text{ m}^2 \text{ s}^{-1}$ as D_{DNA} when $K_b = 700$ and $175 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, respectively (Fig. S2, red lines), as well as $\sim 1.4 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$ as D_{PRO} for 7-bead short protamine model. The $D_{\text{PRO}}/D_{\text{DNA}}$ ratio is ~ 12 when $K_b = 700 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ and ~ 15 when $175 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ in rather good agreement with the estimated experimental ratio.

The dynamics in MD with implicit solvents is much faster than experimentally observed. Experimental time scale is anyway too long to allow an adequate sampling of the systems. Nevertheless, the relative diffusion coefficients of DNA and protamine are still rather well reproduced.

S3. Persistence length of a single long DNA

We perform a set of 10 independent simulations of 10×10^6 time steps each on a DNA chain contained in a cubic box of 48 nm on each side. The DNA chain is made of 240 positively-charged beads bonded at an equilibrium distance r_{eq} of 1.7 Å, which corresponds to 41 nm when fully extended. Compared to the system used for the simulations on the DNA-protamine mixture, the DNA chain is 2.4 times longer and the size of the simulation box is the same. We repeat this set of simulations at three different conditions, increasing the strength of the electrostatic interactions,

- (1) with no consideration of electrostatic interaction, i.e., considering only steric non-bonding interaction of a neutral polymer (denoted as **neutral**),
- (2) with consideration of both steric and electrostatic interactions at the physiological condition, i.e., with 150 mM of added monovalent salts (denoted as **150 mM**), and
- (3) with consideration of both steric and electrostatic interactions, but with no added salt other than the monovalent Na^+ counterions added to neutralize DNA, i.e., at a very low salt concentration close to the zero-salt or salt-free limit (denoted as **no added salt**).

Please note that the term "no added salt" here does not mean that no ions are present in the system. The counterions of dsDNA already contribute an ionic concentration of 3.6 mM, which corresponds to a Debye screening length λ_D of about 5 nm.

Several definitions for calculation of persistence lengths are possible.^{58,59} Since the DNA chain is short and the bending constant is large, i.e., the chain is relatively rigid, we use an approach based on the wormlike-chain (WLC) model describing the decay of the correlation $C(s)$ in the orientation of two segments of the chain separated by a distance s along the contour length. For the WLC model where the only interaction is the bending potential, $C(s)$ is expressed as

$$C(s) = \langle \cos \theta(s) \rangle = \exp(-s/L_{oc}),$$

where the bending angle $\theta(s)$ is the angle between the two directions of the chain separated by the contour length s , L_{oc} is the orientation correlation persistence length, and $\langle \dots \rangle$ denotes an average over all the orientations of segments along the chain separated by s and over all the sampled conformations of DNA. For sufficiently small fluctuations, i.e., for small θ , and small contour lengths s , it is possible to directly sample the WLC orientation probability distribution approximated by a Gaussian distribution

$$P(s) \approx \exp[-L_p \theta^2 / 2s] \approx \exp[-L_p(1 - \cos \theta) / s],$$

where L_p is the persistence length and s is the contour length of a short fragment of bending angle θ .^{S10,S11}

The probability distribution P of the bending angle $\theta(s)$ of a DNA short segment whose length s is 34 Å (Fig. S3-1) is computed for the last 5×10^6 time steps of the ensemble of the 10 independent runs. The five last end beads of the chain are excluded to minimize the chain end effects. Our results are not affected by the number of beads excluded. The L_p values calculated after excluding 5 or 30 beads differ by less than 2%. We first use the bending constant K_{bend} of $700 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, which is used for our other simulations. From the linear fits to the initial decrease of $\ln(P)$ down to e^{-3} (Fig. S3-1a), we obtain a persistence length L_p of 2016 ± 7 Å (**neutral**), 2050 ± 2 Å (**150 mM**), and 2040 ± 1 Å (**no added salt**) Å. The dependence of bending angle $\theta(s)$ on the chain contour length s , which is fitted up to a contour length s of 30 Å, leads to a persistence length of 1994 ± 3 Å (**neutral**), 2038 ± 2 Å (**150 mM**), and 2039 ± 3 Å (**no added salt**). These L_p values are about four times longer than typical literature values (~ 500 Å), but they are in fact comparable to an L_p value measured at the lowest salt concentration that we found ($\sim 1800 \pm 200$ Å at 0.0001 M NaCl, Fig. S3-2, top left).^{S12}

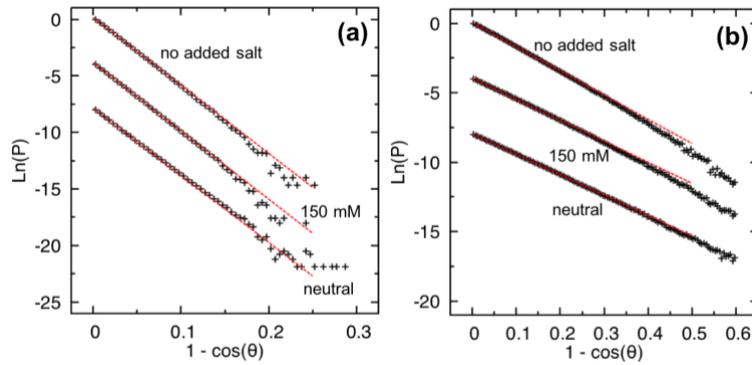


Figure S3-1. Probability distribution P of the bending angle θ of a segment, whose contour length s is 34 \AA , along a single DNA made of 240 positively-charged beads, which is modelled with a bending constant K_{bend} of (a) 700 and (b) $175 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, in three different cases of **no added salt** (without added salt), **150 mM** (in presence of 150-mM salt), and **neutral** (with electrostatic interactions turned off). The plots from the three different cases are vertically shifted by 4 consecutively for clarity. Data are accumulated over ten independent runs. The red dashed lines correspond to the linear fits of the initial decrease of the distribution down to e^{-3} .

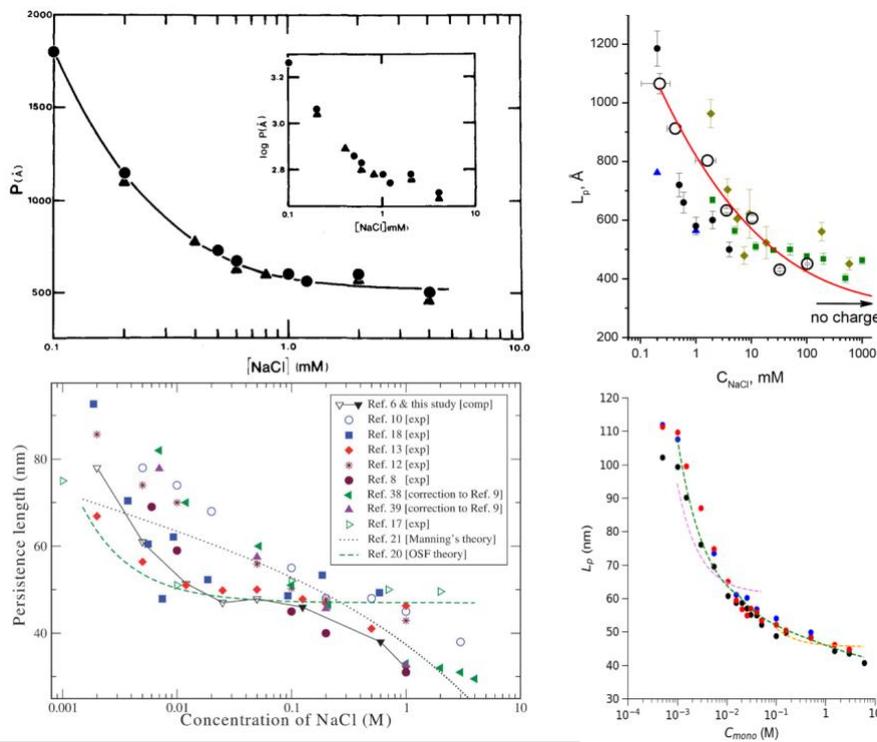


Figure S3-2. Persistence length (P or L_p ; nm or Å) of DNA as a function of monovalent salt (NaCl) concentration (ionic strength; M or mM) from Refs. S12,S13,S14,S15. **[Top left]** Reproduced with permission from Ref. S12 (copyright 1981 John Wiley and Sons). Experiments on 587-bp (circle) and 434-bp (triangle) DNA with $\sim 10\%$ of uncertainty. **[Top right]** Reproduced from Ref. S13. Coarse-grained simulations (circles with red line to guide the eye) and experiments from refs. S12 (black circle), S16 (khaki rhombi), S17 (green square), and S18 (blue triangle). **[Bottom left]** Reproduced from Ref. S14 (copyright 2012 Royal Society of Chemistry). Simulations, experiments, theoretical corrections to some experimental data, and predictions from the Odjik-Skolnick-Fixman (refs. S19,S20) and Manning (ref. S21) theories. **[Bottom right]** Ref. S15. Experiments on a 1200-bp DNA with monovalent (black; Li^+ , blue; Na^+ , red; K^+) salts. Corresponding fits from the Trizac-Shen (green dashed line), OSF (orange dashed line) and OSFM (violet dashed line) theories are shown together.

The persistence length estimated by our simulations increases by $\sim 40 \text{ \AA}$ from **neutral** to **150 mM** or to **no added salt**, i.e., when the electrostatic interactions are considered with or without additional 150-mM monovalent salts. The Odjik-Skolnick-Fixman (OSF) theory,^{S15,S19,S20} which is valid for an ionic strength

greater than 0.1 M, calculates the persistence length of a polyelectrolyte as the sum of a bare persistence length of a neutral chain and an electrostatic contribution L_e given by

$$L_e = l_B / (4 A^2 \kappa^2),$$

where l_B is the Bjerrum length, A is the distance between charges along the polyelectrolyte, and κ is the inverse Debye-Hückel length given by

$$\kappa = [4\pi l_b (c_+ + c_-)]^{1/2}$$

with c_{\pm} the ion concentrations. At lower ionic strengths of monovalent counterions, we can use the Odjik-Skolnick-Fixman-Manning (OSFM) model,^{S15,S22} which estimates the electrostatic contribution by

$$L_e = 1 / (4 l_b \kappa^2).$$

Alternatively, to embrace the whole ionic strength range, we can use the Trizac-Shen (TS) model which is valid for monovalent counterions.^{S23} With the parameters of our system, the electrostatic contribution to the persistence length, L_e , is predicted as ~ 40 Å by the OSF model, ~ 100 Å by the OSFM model, and ~ 10 Å at 150 mM and ~ 100 Å at 3.36 mM by the TS model. Our estimations of ~ 40 Å (**150 mM**) as the electrostatic contribution at 150 mM indeed agree semi-quantitatively with the OSF theory (~ 40 Å), but to a lesser extent with the TS theory (~ 10 Å). Moreover, our estimations show essentially no dependence on the ionic strength and thus our estimations of ~ 40 Å at a low ionic strength (< 3.6 mM; **no added salt**) are not in agreement with the OSFM and TS theories (~ 100 Å). Our parameters lead to persistence lengths about 3-4 times higher than the ones reported experimentally for DNA in the salt range considered.

For stiff polymers, the bending angles θ between two consecutive beads separated by a distance A (monomer size) are small and we can derive an approximate expression from the WLC model with only bending interactions,

$$\langle \cos \theta \rangle = \exp(-A/L_{oc}) = \frac{\int_0^\pi \exp(-\beta \frac{K_b}{2} \theta^2) \cos \theta \sin \theta d\theta}{\int_0^\pi \exp(-\beta \frac{K_b}{2} \theta^2) \sin \theta d\theta}$$

With $\beta = k_B T$, this expression lead to

$$L_{oc} = k_{bend} A / k_B T.$$

With $A = 1.7$ Å and $k_B T = 0.6$ kcal mol⁻¹ at room temperature, we find a persistence length L_{oc} of ~ 1984 Å for the neutral polymer chain, which is in good agreement with our simulation. Moreover, such a linear relationship between the bare persistence length and the bending constant K_{bend} leads us to use a four times smaller bending constant K_b , 175 kcal mol⁻¹ rad⁻², to obtain a bare persistence length around 500 Å. Indeed, from the linear fits to the initial decrease of the probability distribution $P(s)$ down to e^{-3} (Fig. S3-1b), we obtain a persistence length L_p of 500 ± 1 Å (**neutral**), 516 ± 1 Å (**150 mM**), and 593 ± 2 Å (**no added salt**). The dependence of bending angles $\theta(s)$ on the chain contour length s , which is fitted up to a contour length of 30 Å, also leads to a persistence length of 490 ± 2 Å (**neutral**), 506 ± 1 Å (**150 mM**), and 574 ± 5 Å (**no added salt**). Both results are in good agreement with the typical experimental data. Moreover, the electrostatic contribution to the persistence length (16 and ~ 90 Å) are in semi-quantitative agreement with the OSF and OSFM theories and in excellent quantitative agreement with the TS model.

Comparing our results obtained with the two bending constants K_{bend} , we hypothesize that the generic bending constant of 700 kcal mol⁻¹ rad⁻² might be too large in the range of ionic strength greater than 3 mM, although the persistence lengths obtained with our K_{bend} value of 700 kcal mol⁻¹ rad⁻² are certainly

adequate to reproduce the behavior of short stiff DNA fragments. Indeed, a series of 100-ns unbiased simulations using the DNA fragments modeled with the four times smaller bending constant K_b , 175 kcal mol⁻¹ rad⁻², in presence of 21-bead-long protamine at different $R_{+/-}$ ratios around the isoelectric point present no significant difference in the aggregation behavior from the initial simulations performed with stiffer DNA fragments (Fig. S3-3).

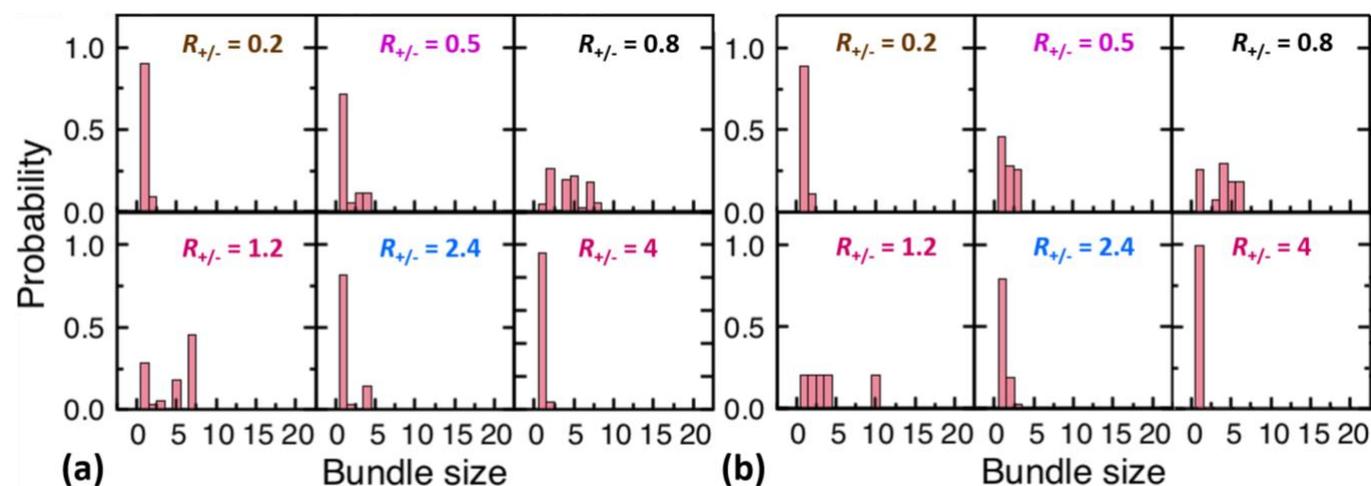


Figure S3-3. Distributions of DNA bundle size, which are averaged over the last 20 ns or 20×10^6 steps of 100 ns of unbiased MD simulations of a homogeneous mixture of protamine and DNA fragments at different $R_{+/-}$ ratios. Protamine is modelled by 21-positively charged beads connected with no angle bending interactions, and DNA is modelled by a 17-nm-long chain made of 100 negatively-charged beads connected with a bending constant K_b of (a) 175 and (b) 700 kcal mol⁻¹ rad⁻².

S4. 20 Independent simulations

The aggregation onset R_{CA} for formation of a DNA-protamine bundle upon adding protamine and the redissolution onset R_{CR} of the bundle upon adding protamine are both shifted towards lower $R_{+/-}$ as the protamine length n_{PRO} is increased from 7 to 21. The 21-bead-long protamine with a total charge of $+21|e|$ ($n_{PRO} = 21$) induce the redissolution at a much lower $R_{+/-}$ than three-times-shorter 7-bead protamine with a total charge of $+7|e|$ ($n_{PRO} = 7$). The redissolution onset R_{CR} significantly depends on the protamine length n_{PRO} , most likely due to the translation entropy of free protamine in solution, which is higher for 7-bead protamine than for 21-bead protamine at the same $R_{+/-}$. On the other hand, the aggregation onset R_{CA} is rather insensitive to the protamine length, although R_{CA} is slightly lower with longer protamine. These conclusions on the protamine-length dependence of R_{CR} and R_{CA} are solely based on a collection of a single simulation performed at each $R_{+/-}$. Our system is strongly charged and thus presents a two-stage aggregation process with a fast first-stage DNA complexation with protamine, followed by a slow second-stage bundle formation and evolution. It is therefore important to verify that our simulations of the system are long enough to remain independent of the initial conditions chosen, i.e., the initial positions and orientations as well as the initial velocities of the molecules.

We therefore check the validity and the reproducibility of our simulation by performing a series of 20 independent simulations of 100 ns or 100×10^6 time steps each, starting from different random initial positions and velocities, for two critical charge ratios ($R_{+/-} = 0.5$ in excess of DNA and 2.4 in excess of protamine) and for the two protamine lengths investigated ($n_{PRO} = 7$ and 21). The charge ratio $R_{+/-}$ of 2.4

is chosen since, at this charge ratio, large bundles are formed when n_{PRO} is 7 while DNA remain mostly in solution or form only small bundles when n_{PRO} is 21. Similarly, when $R_{+/-}$ is 0.5, DNA fragments form slightly larger, albeit remaining small, bundles with 21-bead protamine than with 7-bead protamine.

Fig. S4-1a shows the time evolutions of the numbers of DNA bundles (including complexes; top) and of only the DNA complexes (bottom), which are averaged over the 20 independent runs on these four cases of (0.5, 7), (0.5, 21), (2.4, 7), and (2.4, 21). Fig. S4-1b shows the size distributions of these bundles averaged over the last 20 ns of all the 20 independent runs. When $R_{+/-}$ is 0.5 (in excess of DNA, left panels), 21-bead protamine (red curves and bars) induce slightly stronger DNA aggregation (up to size of 3 vs. 2) and leave less complexes ($\sim 40\%$ vs. $\sim 70\%$) than 7-bead protamine (blue curves and grey bars). When $R_{+/-}$ is 2.4 (in excess of protamine, right panels), 7-bead protamine induce a strong (even complete) aggregation process forming large (up to size 20) DNA bundles and almost no ($< 5\%$) DNA complexes (blue curves and grey bars), while 21-bead protamine induce a very limited aggregation process where only small (up to size 3) DNA bundles coexist with a significant amount of ($> 70\%$) DNA complexes (red curves and bars).

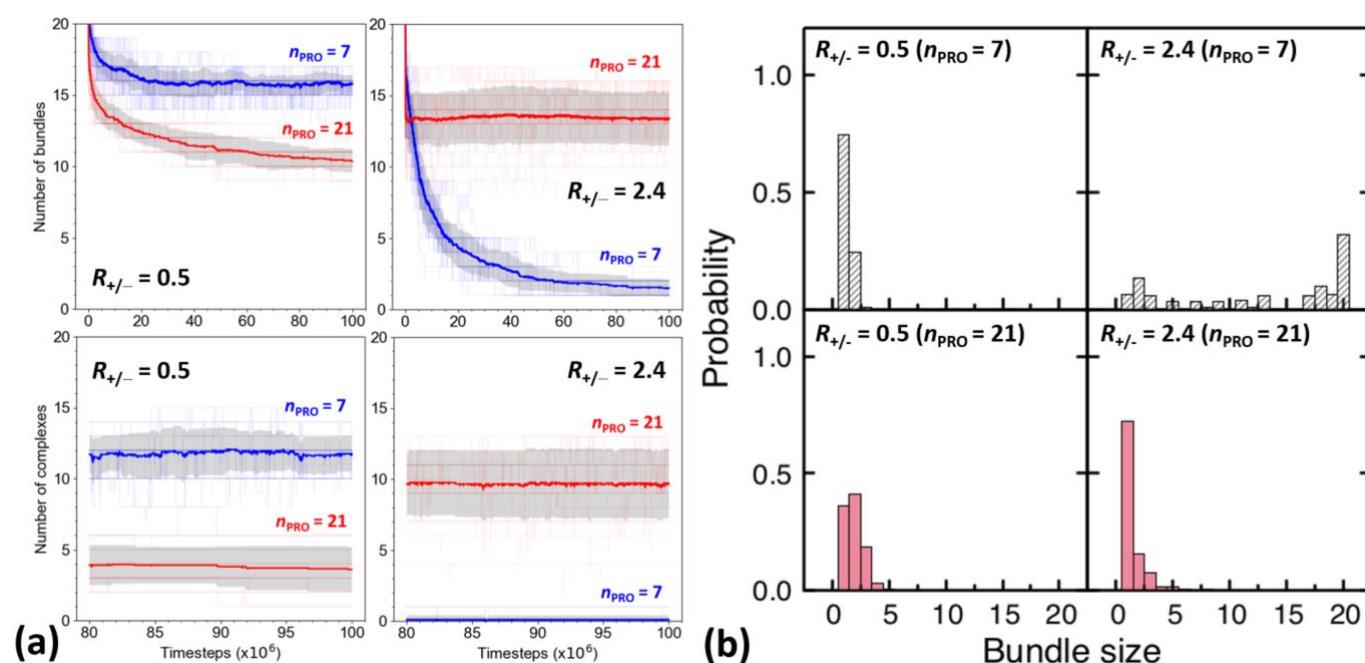


Figure S4-1. (a) Time evolution of the total number of DNA bundles including DNA complexes counted as a bundle of size 1 (top) and the number of DNA complexes (bottom) which are formed during 100-ns or 100×10^6 -step MD simulations at $R_{+/-}$ of 0.5 (left) and 2.4 (right) in presence of 7-bead (blue) or 21-bead (red) protamine. The bold curves (blue/red) and the shaded area (grey) represent the average and the standard deviation over the 20 independent time evolutions shown as transparent curves. (b) Bundle size distributions when $R_{+/-}$ is 0.5 (left) or 2.4 (right) and when n_{PRO} is 7 (top) and 21 (bottom), each of which is averaged over the last 20 ns of the 20 independent simulations. Both panels show the same trend of shifts in the aggregation and resolution onsets towards lower $R_{+/-}$ of 0.5 and 2.4, respectively, observed as n_{PRO} increases from 7 to 21.

As shown side by side in Fig. S4-2, these trends are consistent with those drawn from a set of four simulations, a single simulation on each case, whose representativeness is now confirmed by this 20-fold larger set of 80 simulations. However, in all the four cases, the extended sampling realized by the 20 independent simulations shifts the size distribution slightly towards higher sizes, allowing slightly larger bundles. In particular, at $R_{+/-}$ of 2.4 in presence of 7-bead protamine, a significant population ($\sim 30\%$) of a single large bundle of size 20 is newly observed from the 20 independent simulations, supporting the nature of small charged bundles as kinetically-trapped structures that eventually coalesce.

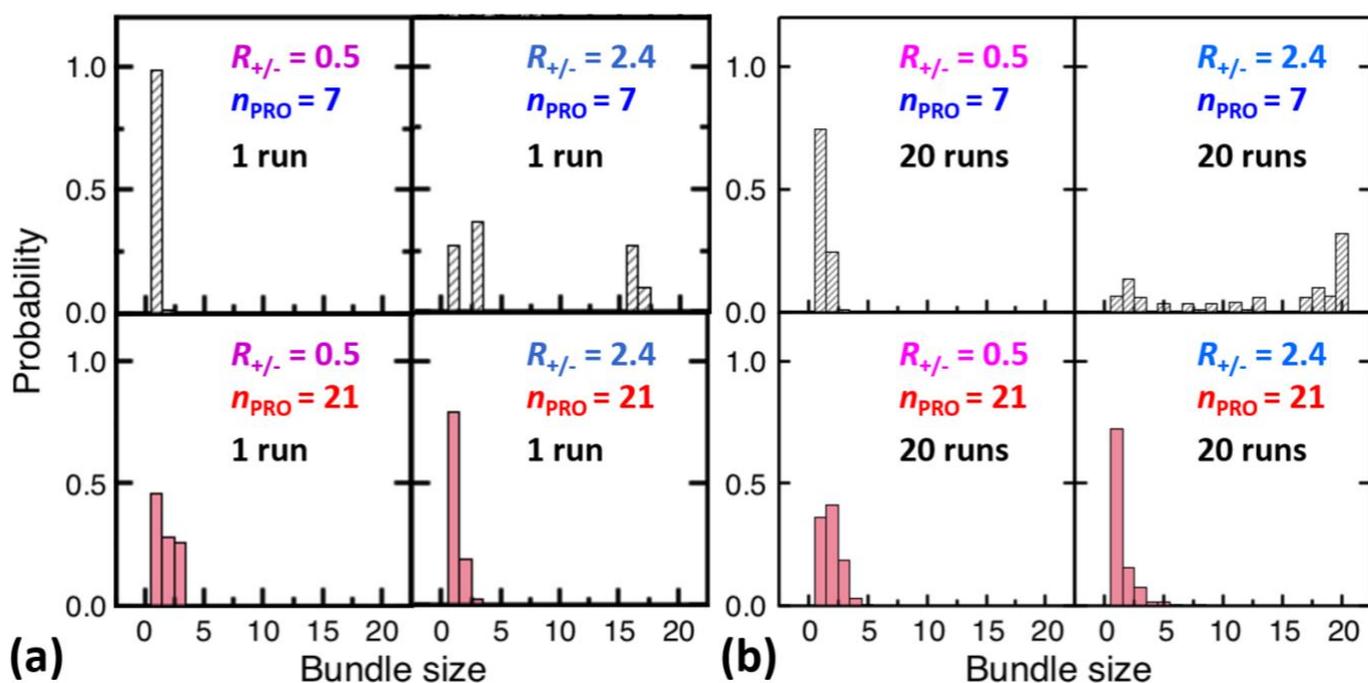


Figure S4-2. Side-by-side comparison of average bundle size distributions extracted from (a) a single run or (b) 20 independent runs on each of the four cases where $R_{+/-}$ is 0.5 (left) or 2.4 (right) and n_{PRO} is 7 (top) and 21 (bottom). Each distribution is averaged over the last 20 ns of (a) the single run or (b) all the 20 independent runs. Both panels show the same trend that the aggregation and resolution onsets, R_{CA} and R_{CR} , shift downwards towards lower $R_{+/-}$ as n_{PRO} increases from 7 to 21. In all the four cases, in particular at $R_{+/-}$ of 2.4 in presence of 7-bead protamine, the extended sampling realized by the 20 independent simulations shifts the size distribution slightly towards higher sizes, allowing slightly larger bundles, but otherwise the trends found the two types of sampling are similar to each other.

Fig. S4-3a shows the average charge ratio per DNA exhibited by all the bundles (including complexes, size $N_{\text{DNA}} \geq 1$), which is calculated either by considering only protamine in bundles,

$$|q_{\text{PRO}}/q_{\text{DNA}}| = (N_{\text{PRO}} \times n_{\text{PRO}}) / (N_{\text{DNA}} \times n_{\text{DNA}}),$$

or by including Na^+ and Cl^- counterions in addition of protamine (not shown here),

$$|q_{\text{PRO}^*}/q_{\text{DNA}}| = (N_{\text{PRO}} \times n_{\text{PRO}} + N_{\text{Na}} - N_{\text{Cl}}) / (N_{\text{DNA}} \times n_{\text{DNA}}),$$

when N_{PRO} protamine with n_{PRO} of 7 or 21 and N_{DNA} DNA with n_{DNA} of 100 (the number of phosphate groups in a DNA) form the bundle along with N_{Na} Na and N_{Cl} Cl counterions. Whether with 7-bead protamine (blue) or with 21-bead protamine (red), as the bundle size N_{DNA} increases (left to right), the bundle charge ratios $|q_{\text{PRO}}/q_{\text{DNA}}|$ increase when undercharged in excess of DNA ($R_{+/-} = 0.5$, lower branch) and decrease when overcharged in excess of protamine ($R_{+/-} = 2.4$, upper branch), both towards 1, the charge neutrality. This means that more protamine is added to undercharged DNA complexes to assemble as bundles, and conversely, overcharged complexes lose protamine to form stable bundles, leading to complexes more negatively-undercharged (i.e., bound to less protamine per DNA) than bundles in excess of DNA and complexes more positively-overcharged (i.e., bound to more protamine per DNA) than bundles in excess of protamine, demonstrating the theoretically-predicted mechanism of charge disproportionation.

In excess of DNA at $R_{+/-}$ of 0.5 (Fig. S4-3a, lower branch), the DNA complexes ($N_{\text{DNA}} = \text{bundle size} = 1$) with 21-bead protamine are more negatively charged with the charge ratios farther from 1 ($|q_{\text{PRO}}/q_{\text{DNA}}| = 0.27 \pm 0.15$, red; $|q_{\text{PRO}^*}/q_{\text{DNA}}| = 0.64 \pm 0.06$, not shown) than those with 7-bead protamine ($|q_{\text{PRO}}/q_{\text{DNA}}| = 0.45 \pm 0.08$, blue; $|q_{\text{PRO}^*}/q_{\text{DNA}}| = 0.69 \pm 0.04$). The DNA complex is defined as an isolated DNA irrespective

of the number of protamine adsorbed on it. A significant quantity ($\sim 14\%$) of those complexes formed with 21-bead protamine are in fact naked DNA with no protamine adsorbed, while no DNA is naked in presence of 7-bead protamine. This mainly contributes to the difference between the two charge ratios (0.27 ± 0.15 vs. 0.45 ± 0.08) of the complexes made with the two protamine models, because the charge ratios including the counterions as well, $|q_{\text{PRO}^*}/q_{\text{DNA}}|$, are quite similar (0.64 ± 0.06 vs. 0.69 ± 0.04) in both models. Indeed, no strong difference is observed between the charge ratios of larger bundles ($N_{\text{DNA}} \geq 2$) whether they are made with 21-bead protamine ($|q_{\text{PRO}}/q_{\text{DNA}}| \approx 0.54 \pm 0.12$, red; $|q_{\text{PRO}^*}/q_{\text{DNA}}| \approx 0.77 \pm 0.06$) or 7-bead protamine ($|q_{\text{PRO}}/q_{\text{DNA}}| \approx 0.58 \pm 0.05$, blue; $|q_{\text{PRO}^*}/q_{\text{DNA}}| \approx 0.78 \pm 0.03$). In addition, there is a tendency to form slightly larger bundles with 21-bead protamine (up to size 5) than with 7-bead protamine (up to size 3) via slightly stronger charge disproportionation between complexes and bundles.

On the other hand, in excess of protamine at $R_{+/-}$ of 2.4 (Fig. S4-3a, upper branch), both DNA complexes ($|q_{\text{PRO}}/q_{\text{DNA}}| = 1.54 \pm 0.10$, red; $|q_{\text{PRO}^*}/q_{\text{DNA}}| = 1.40 \pm 0.07$) and larger bundles ($|q_{\text{PRO}}/q_{\text{DNA}}| \approx 1.32 \pm 0.07$, red; $|q_{\text{PRO}^*}/q_{\text{DNA}}| \approx 1.23 \pm 0.06$) made with 21-bead protamine are significantly more overcharged than the complexes ($|q_{\text{PRO}}/q_{\text{DNA}}| = 1.34 \pm 0.05$, blue; $|q_{\text{PRO}^*}/q_{\text{DNA}}| = 1.28 \pm 0.04$) and larger bundles ($|q_{\text{PRO}}/q_{\text{DNA}}| \approx 1.09 \pm 0.07$, blue; $|q_{\text{PRO}^*}/q_{\text{DNA}}| \approx 1.07 \pm 0.06$) made with 7-bead protamine, whether excluding or including the counterions in their estimation. Fig. S4-3 shows the average distribution of protamine, Na^+ and Cl^- for overcharged complexes ($N_{\text{DNA}} = 1 = \text{bundle size}$) at $R_{+/-} = 2.4$ for 7-bead (top) and 21-bead (bottom) protamine. We noticed that, since protamine is also a strong polyelectrolyte, a significant amount of Cl^- counterions are also present in the complexes (as well as in other bundles formed at $R_{+/-} > 1$, data not shown) in order to bring back the overall charge of the complexes closer to neutrality.

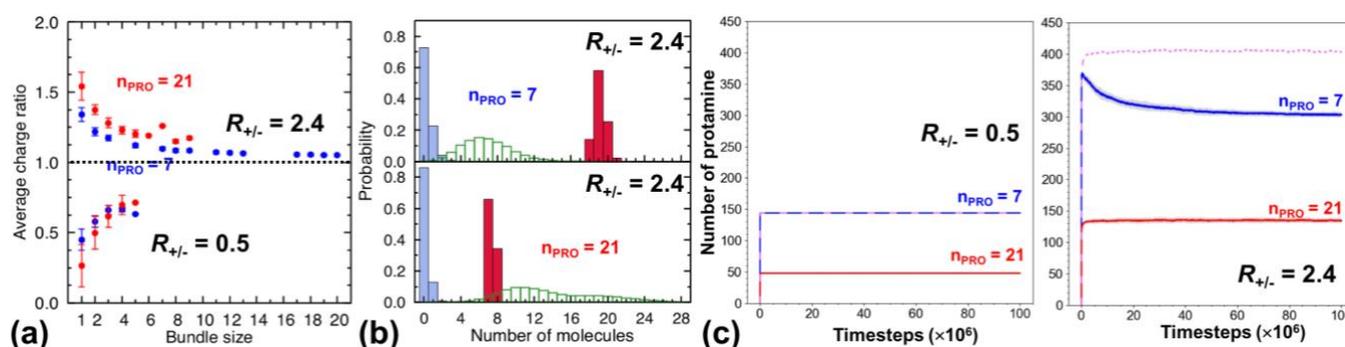


Figure S4-3. (a) Average bundle charge ratios $|q_{\text{PRO}}/q_{\text{DNA}}|$ as a function of the bundle size at $R_{+/-}$ of 0.5 (lower branch) and 2.4 (upper branch) in presence of 7-bead (blue) or 21-bead (red) protamine. The averages are over the last 20 ns of 20 independent simulations. (b) Distribution of species (protamine (red boxes), Na^+ (blue boxes) and Cl^- (open green boxes)) for complexes formed at $R_{+/-} = 2.4$ in presence of 7-bead (top) or 21-bead protamine (bottom), averaged over the last 20ns of 20 independent runs. (c) Time evolution of the total number of protamine in the DNA bundles at $R_{+/-} = 0.5$ (left) and 2.4 (right), which are averaged over the 20 independent simulations. Also shown as violet dashed line is simply the evolution of the system with 21-bead protamine (red) multiplied by three, considering that a 21-bead protamine corresponds to three 7-bead protamine. At $R_{+/-}$ of 0.5, all the protamine in the systems participate in the formation of complexes and bundles and therefore the evolution of systems made of 7-bead and 21-bead protamine are identical, while it is not the case at $R_{+/-}$ of 2.4.

Fig. S4-3c presents the time evolution of the number of protamine in all the DNA bundles (including complexes), which are averaged over the 20 independent simulations. All the protamine in the system are bound to DNA in the bundles in excess of DNA at $R_{+/-}$ of 0.5 (blue = dashed violet), while only a fraction of the protamine, particularly 7-bead protamine (blue < dashed violet), are bound to DNA in the bundles in excess of protamine at $R_{+/-}$ of 2.4. The DNA complexes at $R_{+/-}$ of 2.4 contain less 7-bead protamine than three times of 21-bead protamine, forming less overcharged complexes with 7-bead protamine than with

21-bead protamine. Therefore, a significant amount (55%) of 20 independent simulations with 7-bead protamine evolve to a complete aggregation into a single bundle of size 20, while small charged bundles, mostly the isolated complexes, are present in the system with 21-bead protamine, demonstrating the downward shift of the redissolution onset R_{CR} to lower $R_{+/-}$ by increasing the protamine length and also resolving the discrepancy between our simulations performed with two different protamine models.

S5. Simulations from a pre-assembled bundle

Snapshots of a preformed DNA bundle taken after 5-ns equilibration of protamine and counterions (Fig. S5a) show that more protamine (red) replace Na^+ counterions (blue) as $R_{+/-}$ increases (top to bottom). In excess of protamine ($R_{+/-} = 2$), the positively-charged bundle allows the condensation of Cl^- (green) on its surface. Final snapshots taken 20 ns after DNA are also relaxed (Fig. S5b) show that the single bundle stays mechanically stable in excess of protamine ($R_{+/-} = 2$), while it breaks into several smaller bundles in moderate excess of DNA ($R_{+/-} = 0.8$) or further into a dispersion of DNA complexes in large excess of DNA far from the isoelectric point ($R_{+/-} = 0.5$). The simulations confirm that the states formed in large excess of DNA ($R_{+/-} = 0.2\sim 0.5$) are equilibrium phases kept stable by strong electrostatic repulsion and partially support the scenario that bundles formed by charge disproportionation at $R_{+/-}$ of 0.8 constitute an equilibrium phase.

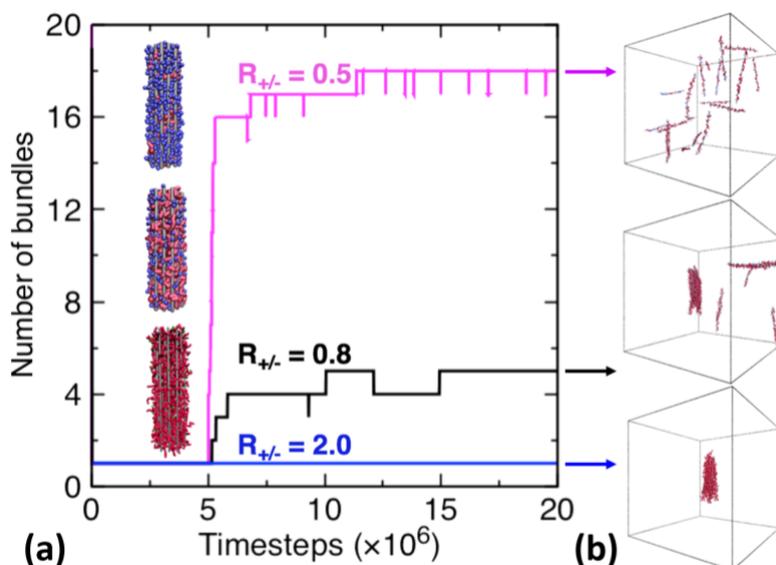


Figure S5. (a) Time evolution of a single preformed bundle composed of 19 DNA (grey), protamine (red), and Na^+/Cl^- counterions (blue/green) at $R_{+/-}$ of 0.5 (top, magenta curve), 0.8 (middle, black curve) and 2 (bottom, blue curve). Snapshots of the single bundle equilibrated with fixed DNA for 5 ns or 5×10^6 steps are shown as insets. (b) Snapshots of the states reached after 20 ns or 20×10^6 steps of production-period simulations with relaxed DNA at $R_{+/-}$ of 0.5 (top), 0.8 (middle) and 2 (bottom). Counterions and protamine ($R_{+/-} > 1$) remaining in the solution are not shown for clarity.

References

- S1 A. Mukherjee, A. de Izarra, J. Degrouard, E. Olive, P. K. Maiti, Y. H. Jang and Y. Lansac, *ACS Nano*, 2021, **15**, 13094-13104.
- S2 V. B. Teif and K. Bohinc, *Prog. Biophys. Mol. Biol.*, 2011, **105**, 208-222.
- S3 A. Y. Grosberg, T. T. Nguyen and B. I. Shklovskii, *Rev. Mod. Phys.*, 2002, **74**, 329-345.
- S4 Y. Levin, *Rep. Prog. Phys.*, 2002, **65**, 1577-1632.
- S5 G. L. Lukacs, P. Haggie, O. Seksek, D. Lechardeur, N. Freedman and A. S. Verkman, *J. Biol. Chem.*, 2000, **275**, 1625-1629.
- S6 D. Lochmann, J. Weyermann, C. Georgens, R. Prassl and A. Zimmer, *Eur. J. Pharm. Biopharm.*, 2005, **59**, 419-429.
- S7 D. R. Lide, *CRC Handbook of Chemistry and Physics, 87th Ed.*, CRC Press, Boca Raton, 2006-2007.
- S8 M. Ullner, *J. Phys. Chem. B*, 2003, **107**, 8097-8110.
- S9 M. Ullner, B. Jönsson, C. Peterson, O. Sommelius and B. Söderberg, *J. Chem. Phys.*, 1997, **107**, 1279-1287.
- S10 A. K. Mazur, *Phys. Rev. Lett.*, 2007, **98**, 218102.
- S11 A. Garai, S. Saurabh, Y. Lansac and P. K. Maiti, *J. Phys. Chem. B*, 2015, **119**, 11146-11156.
- S12 P. J. Hagerman, *Biopolymers*, 1981, **20**, 1503-1535.
- S13 N. Korolev, D. Luo, A. P. Lyubartsev and L. Nordenskiöld, *Polymers*, 2014, **6**, 1655-1675.
- S14 A. Savelyev, *Phys. Chem. Chem. Phys.*, 2012, **14**, 2250-2254.
- S15 S. Guilbaud, L. Salomé, N. Destainville, M. Manghi and C. Tardin, *Phys. Rev. Lett.*, 2019, **122**, 028102.
- S16 C. G. Baumann, S. B. Smith, V. A. Bloomfield and C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 6185-6190.
- S17 V. Rizzo and J. Schellman, *Biopolymers*, 1981, **20**, 2143-2163.
- S18 Y. Lu, B. Weers and N. C. Stellwagen, *Biopolymers*, 2002, **61**, 261-275.
- S19 T. Odijk, *J. Polym. Sci.*, 1977, **15**, 477-483.
- S20 J. Skolnick and M. Fixman, *Macromolecules*, 1977, **10**, 944-948.
- S21 G. S. Manning, *Biophys. J.*, 2006, **91**, 3607-3616.
- S22 G. S. Manning, *Biopolymers*, 1981, **20**, 1751-1755.
- S23 E. Trizac and T. Shen, *Europhys. Lett.*, 2016, **116**, 18007.