Supporting Information

Protamine Binding Site on DNA: Molecular Dynamics Simulations and Free Energy Calculations with Full Atomistic Details

Arnab Mukherjee,[1] Suman Saurabh,[1,±] Enrick Olive,[1] Yun Hee Jang[2,*], Yves Lansac,[1,2,3,*]

[1] GREMAN, UMR 7347, CNRS, Université de Tours, 37200 Tours, France
[2] Department of Energy Science and Engineering, DGIST, Daegu 42988, Korea;
[3] Laboratoire de Physique des Solides, CNRS UMR 8502, Université Paris-Saclay, 91405 Orsay, France

[±] Present address: Department of Chemistry, Imperial College London, London UK

*E-mail: lansac@univ-tours.fr, yhjang@dgist.ac.kr

S1. DNA-protamine adsorption statistics over 20 assemblies (S1-S20)

From MD simulations on the four complexes (**C1**, **C2**, **H-MAJ**, and **H-MIN**) shown in Figures 4-6 of the main text, we see that adsorption of protamine on DNA proceeds extremely fast (with the initial separation of 4-5 nm used between DNA and protamine). Once adsorbed on DNA, protamine barely moves during the course of the simulation due to strong electrostatics. One may criticize that the 300-ns dynamics of only four complexes, which are dependent in a great deal of the initial configuration, may not be sufficient to represent the true dynamics of DNA-protamine complexes. Therefore, to extract better statistics of protamine binding site on DNA and protamine conformation upon binding from a larger amount of samples, we perform short (7-to-15-ns long) MD simulations on 20 additional DNA-protamine complexes (or rather assemblies) **S1-S20**. The initial structures of **S1-S20** are built by extracting a total of 20 protamine conformations from the trajectory of a free protamine and then placing them at 20 different positions and orientations around a DNA fragment. These 161,030-to-220,370-atom systems **S1-S20** are energy-minimized using the protocol described in the main text and then submitted to 7-15 ns NPT MD simulations. Their final snapshots are shown in Figure S1.

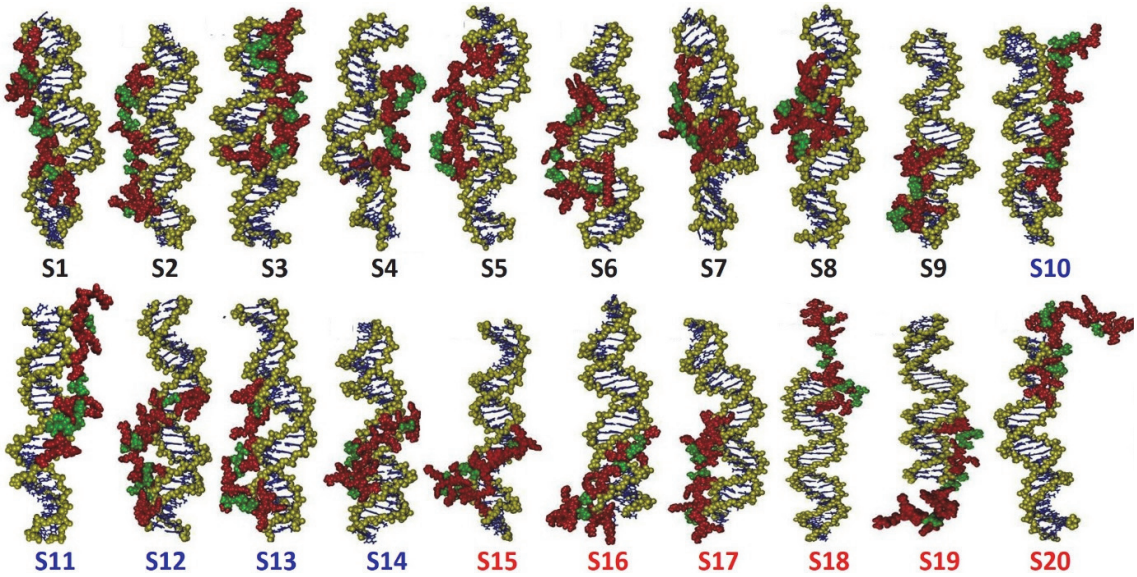

Figure S1. Final snapshots of **S1-S20**. ARG and non-ARG residues are shown in red and in green, respectively. The three different color labels (black, blue, and red) denote the four different groups (Groups 1-3) to which they belong.

The time evolution of the number of ARG C_Z atoms of protamine and Na^+ ions adsorbed within 6.5 Å from the P atoms of DNA in **S1**-**S20** is shown in Figure S2, as done in Figure 6 of the main text for the 300-ns simulations on **C1**, **C2**, **H-MAJ**, and **H-MIN**. From these profiles and also from the trajectories of the simulations (not shown), we notice a common feature in early-stage adsorption dynamics. The ARG residue near one of the protamine termini (residue 1 or 31) first binds to the DNA, followed by the adsorption of the rest of the protamine. However, the time scale of the subsequent binding of the whole protamine varies. Therefore, although the snapshots taken at 7 ns (or 14-15 ns for **S16**-**S17**) show incomplete protamine adsorption for some assemblies (**S18**-**S20** in particular; Figure S1), we expect that essentially all the assemblies would end up with complete protamine adsorption in the long run (see the increase of C_Z adsorption in **S16**-**S17**, which continues after 7 ns; Figure S2). Due to the various adsorption time scales of **S1**-**S20**, various sampling periods are used to calculate the average numbers of ARG C_Z atoms and Na^+ adsorbed on DNA. Each profile is integrated from the point where the adsorption becomes stable, as indicated by vertical dashed lines in Figure S2. The average numbers are listed in Table S1, as done in Table 2 of the main text for the 300-ns simulations on **C1**, **C2**, **H-MAJ**, and **H-MIN**.

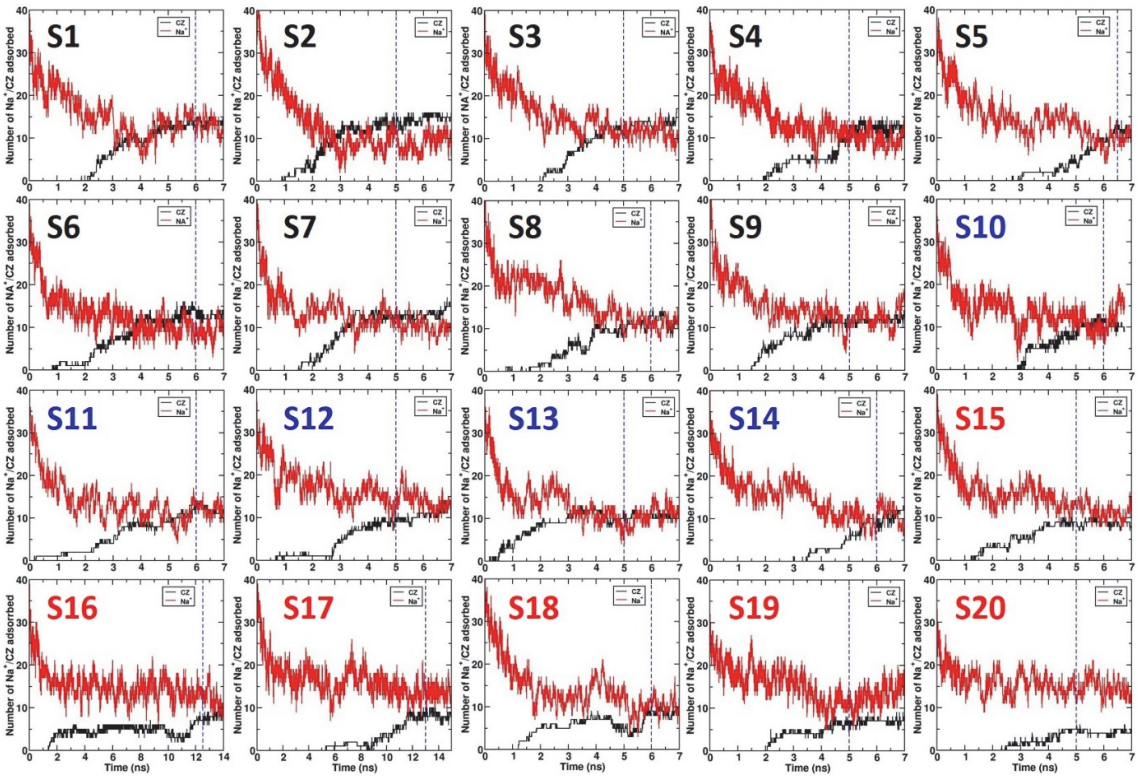

Figure S2. Time evolution of the numbers of ARG C_Z atoms of protamine (black) and Na^+ (red) adsorbed within 6.5 Å from P atoms of DNA in **S1**-**S20**. Note the longer simulation time needed for **S16**-**S17**. The beginning of the sampling period is indicated by a vertical dashed line. The three different color labels (black, blue, and red) denote the three different groups (Groups 1-3) to which they belong.

Table S1. Average number of ARG C_Z and Na^+ around P of DNA in **S1**-**S20**.[a]

Group [b]	1E	1E	1E	1E	1E	1G	1G	1G	1G	2E
Assembly	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
C_Z	13.2	13.6	13.1	12.2	11.7	12.9	12.8	11.7	11.7	10.7
Na^+	13.4	9.1	11.4	9.8	10.2	9.6	11.0	11.4	12.5	10.7
Group [b]	2E	2E	2E	2E	3G	3G	3G	3E	3E	3E
Assembly	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**	**S17**	**S18**	**S19**	**S20**
C_Z	11.5	10.4	10.7	11.3	8.7	7.3	7.9	8.2	6.8	4.4
Na^+	10.9	13.8	11.6	10.1	10.9	9.8	12.4	13.4	14.1	14.6

[a] Averaged during the sampling period after the vertical dashed line shown in Figure S2.

[b] E for extended and G for globular.

The **S1-S20** assemblies are divided into a few groups according to the adsorption profiles of C_Z and Na^+ as well as the average number of C_Z adsorbed on DNA. Assemblies such as **S2-S3** and **S6-S7** exhibit fast adsorption of protamine on DNA and fast release of Na^+ from DNA (Figure S2). The average numbers of C_Z and Na^+ on DNA in these assemblies reach 13-14 and 9-11 (Table S1), which match those of **C1** and **C2** in the early stage (10-30 ns) of the 300 ns simulations (13-14 and 9-11; Table 2 of the main text). These assemblies represent complete adsorption of protamine on DNA as seen in Figure S1. In fact, more assemblies, **S1-S9** (Group 1), share similar characteristics: the C_Z adsorption profile reaches a distinct flat region within 7 ns and the average C_Z adsorption number is high (12-13). Although the C_Z adsorption curve of **S5** is not flat, the average number over the last 0.5 ns is close to 12.

We also see that Group 1 exhibits the maximum value of the average buried surface area (BSA) (Table S2) among all the assemblies, although they are still much smaller than those from the 300 ns simulations on the four complexes (**C1**, **C2**, **H-MAJ**, and **H-MIN**; Table 4 of the main text). The contact maps (Figure S3) for Group 1, which show more than 30 DNA-protamine contacts involving most protamine residues, are also the closest to those exhibited by the 300-ns simulations on **C1**, **C2**, **H-MAJ**, and **H-MIN** (Figures 10 of the main text). The assemblies **S10-S14** belong to Group 2, where protamine is adsorbed to DNA except one or two terminal ARG residues dangling outside DNA. Their average numbers of C_Z adsorbed on DNA are around 10-11. Less DNA-protamine contacts (24-30) are found for the Group 2. The rest of the assemblies, **S15-S20**, belong to Group 3. The protamine adsorption is slow (Figure S2, black) and thus the average number of C_Z atoms adsorbed on DNA is still 4-9. The least amount of contacts (< 20) is found. In **S18-S20** in particular, more than a half of protamine residues make no contact with DNA. In the Group 3, protamine is partially adsorbed on DNA through loop formation or dangling.

Table S2. Average BSA (Å^2) of **S1-S20**.[a]

Group [b]	1E	1E	1E	1E	1E	1G	1G	1G	1G	2E
Assembly	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
BSA	983±3	930±4	865±5	954±3	940±4	945±3	977±4	937±3	932±4	778±3
Group [b]	2E	2E	2E	2E	3G	3G	3G	3E	3E	3E
Assembly	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**	**S17**	**S18**	**S19**	**S20**
BSA	712±2	674±5	784±7	708±5	686±3	740±8	653±2	367±1	426±1	397±3

[a] Averaged during the sampling period after the vertical dashed line shown in Figure S2.

[b] E for extended and G for globular.

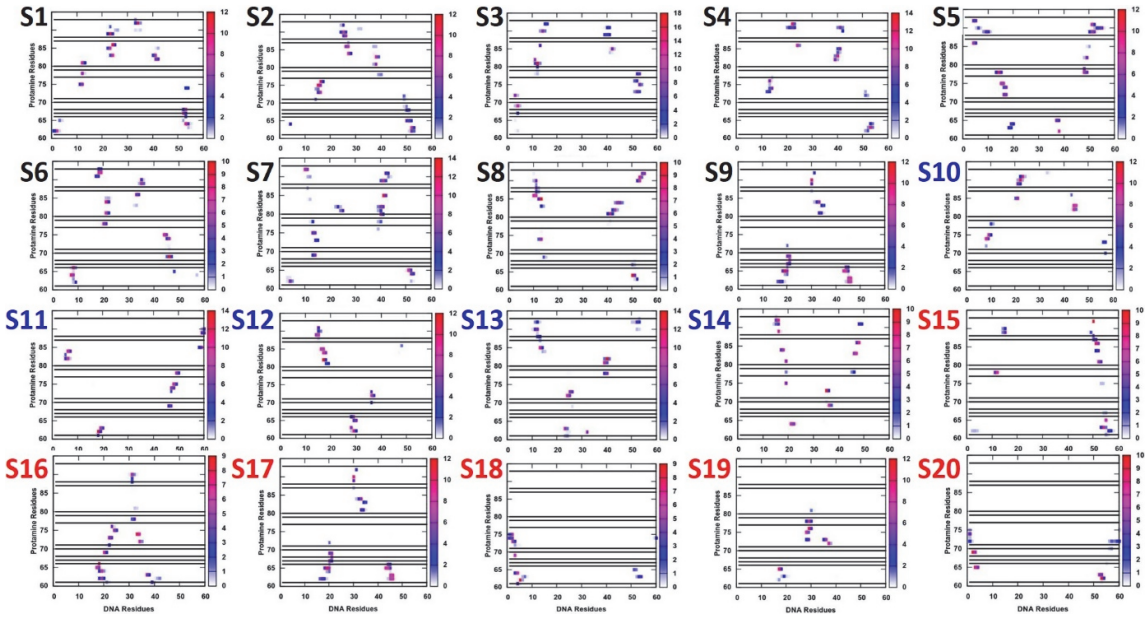

Figure S3. Contact map between DNA (x-axis) and protamine (y-axis) in **S1-S20**. Black horizontal lines indicate non-ARG uncharged residues of protamine. The three different color labels (black, blue, and red) denote the three different groups (Groups 1-3) to which they belong.

The average R_g and R_{ee} of the protamine at the start of the 20 short simulations on **S1-S20** are 21±2 and 54±12 Å, which are essentially the same average values of free DNA (21.3 and 53 Å; Table 1 of the main text). The R_g and R_{ee} of the protamine bound to DNA in each assembly, which are averaged over the varying adsorption period indicated by the vertical dashed line in Figure S1, are shown in Table S3. Their averages over all the assemblies of **S1-S20** are now 19±2 and 48±12 Å, which are significantly smaller than at the beginning, as is the case for the **C1**, **C2**, **H-MAJ**, and **H-MIN** complexes (17±2 and 48±6 Å; Table 1 of the main text). Table S3 shows that Group 3 (**S15-S20**) is again divided into two subgroups. The Group 3G (G for globular; **S15-S17**) shows significantly shorter R_g (16-18 vs. 21-22 Å) than the Group 3E (E for extended; **S18-S20**). In Group 3G, a part of the protamine follows the DNA helix and then forms a loop with the ARG residues near the other end of the protamine attaching to the nearby phosphates of the DNA. On the other hand, in Group 3E, the protamine is partially adsorbed near the terminus of DNA and more than a half of the residues are dangling and extending in solution without making any contact with DNA phosphates, as clearly shown in Figures S1 and S3 as well as in Table S2. Table S3 and Figure S1 show that the assemblies in the Group 2 (**S10-S14**) are mostly extended but the Group 1 is again divided into two subgroups, Group 1E (**S1-S5** with R_g and R_{ee} of 19-23 and 49-67 Å) and Group 1G (**S6-S9** with 15-18 and 28-35 Å). The characteristics of Group 1G (**S6-S7** in particular) and Group 1E (**S2-S3** in particular) resemble those of **C1** and **C2**, respectively, in the early stage of 300 ns simulation, supporting our two choices of representative model complexes, **C1** and **C2**, for 300 ns long simulations.

Table S3. Average R_g and R_{ee} (Å) of protamine in **S1-S20**.[a]

Group [b]	1E	1E	1E	1E	1E
Assembly	**S1**	**S2**	**S3**	**S4**	**S5**
R_g (Å)	21.5±0.3	23.3±0.8	20.6±0.8	18.5±0.5	19.1±0.6
R_{ee} (Å)	67.0±1.8	57.9±1.7	48.9±2.7	49.7±2.1	49.4±2.0
Group [b]	1G	1G	1G	1G	2E
Assembly	**S6**	**S7**	**S8**	**S9**	**S10**
R_g (Å)	18.2±0.3	15.4±0.5	15.0±0.4	17.0±0.3	22.9±0.9
R_{ee} (Å)	27.5±1.4	29.8±2.8	30.5±3.8	35.2±2.6	64.8±4.0
Group [b]	2E	2E	2E	2E	3G
Assembly	**S11**	**S12**	**S13**	**S14**	**S15**
R_g (Å)	21.9±0.8	18.9±0.5	18.6±0.5	17.2±0.2	16.3±0.3
R_{ee} (Å)	62.1±3.8	59.8±3.1	49.0±2.6	42.7±2.7	34.9±1.9
Group [b]	3G	3G	3E	3E	3E
Assembly	**S16**	**S17**	**S18**	**S19**	**S20**
R_g (Å)	17.7±0.5	18.3±0.6	22.2±0.6	22.1±0.9	20.6±1.6
R_{ee} (Å)	35.6±2.3	58.0±2.2	58.7±3.0	50.7±4.8	54.2±3.5

[a] Averaged during the sampling period after the vertical dashed line shown in Figure S2.
[b] E for extended and G for globular.

The RDF(P-C_Z) curve averaged over **S1-S20** shown in Figure S4 (left) exhibits the same trend as those of the 300-ns-long simulations on **C1-C2** shown in Figure S4 (middle and right) and also in Figure 7 of the main text. The CN of P around each C_Z is only 0.6 when averaged over **S1-S20** because they are still in early stage (up to 15 ns) of protamine binding. It grows up to 0.8-0.9 in case of **C1-C2**.

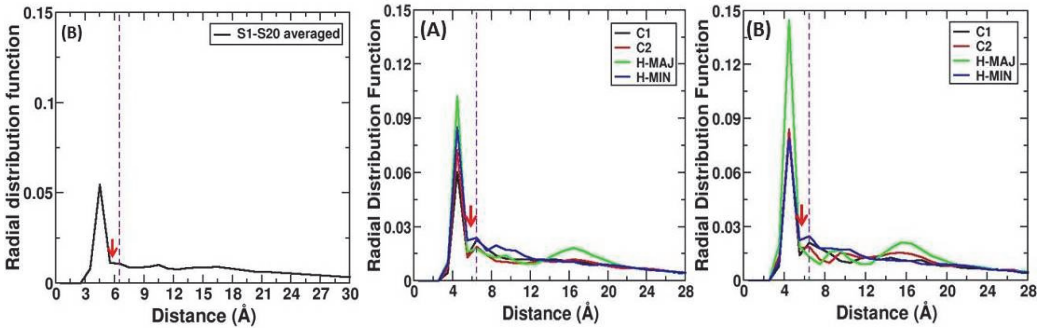

Figure S4. RDF(P-C_Z) averaged over **S1-S20** (up to 15 ns; left) and averaged during early 20 ns (10-30 ns; middle) and the last 20 ns (280-300 ns; right) on each of **C1**, **C2**, **H-MAJ**, and **H-MIN**.

S2. DNA-protamine adsorption at physiological salt concentration

Since the major driving forces for both the initial approach of the protamine to the DNA and the direct DNA-protamine contacts are electrostatic, the ionic environment becomes important. In the four DNA-protamine complex systems (**C1**, **C2**, **H-MAJ**, and **H-MIN**), there are ~200 000 atoms, mostly water, with 58 Na^+ and 21 Cl^- ions. Since there are ~200 Na^+ ions present physiologically (~150 mM Na^+), the consequences of using only neutralizing salt is discussed in this section. We select one of these systems (**C1**), add 202 Na^+ and 202 Cl^- ions to its initial configuration (Figure 1A of the main text and Figure S5, far left) to mimic the physiological salt concentration of 150 mM, call this new system **C1-salt**, and carry out the same type of 300-ns simulation as for **C1**.

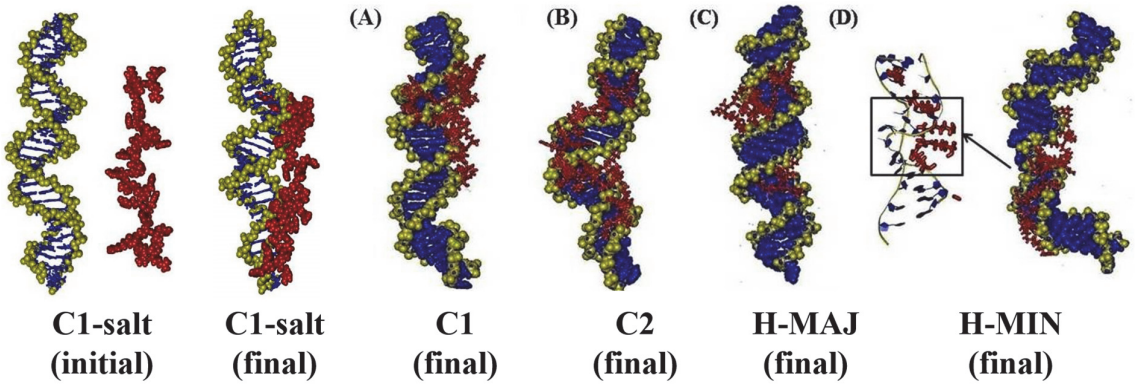

Figure S5. Salt dependence. [Left] Initial and final structures of **C1-salt** at salt concentration of 150 mM in comparison with [right] final structures of no-extra-salt systems (A) **C1**, (B) **C2**, (C) **H-MAJ**, and (D) **H-MIN** (Figure 3 of the main text).

We compare **C1-salt** with **C1** (and the other no-extra-salt systems **C2**, **H-MAJ**, and **H-MIN**) in their complexation dynamics and DNA-protamine interaction. Figure S6 shows the time evolution of RMSD and DNA-protamine CM-CM distance of the complex as well as RMSD and radius of gyration (R_g) of only protamine in **C1-salt** (left panel, black) as well as **C1**, **C2**, **H-MAJ**, and **H-MIN** (right panel, black, red, green, and blue, respectively). Both **C1-salt** and **C1** systems show little time-dependent variation in all these quantities. Once protamine binds to DNA with strong electrostatic interaction, little deviation in the structure of the complex (DNA-protamine CM-CM distance) and in the conformation of protamine (RMSD), i.e., slow dynamics and stable binding, is observed for both systems. However, a prominent difference between the two systems is that protamine is more extended in **C1-salt** than in **C1** (Figure S6C) by keeping its initial extended conformation of free protamine (R_g ~ 21 Å and R_{ee} ~ 67 Å) instead of converting to rather globular or folded conformation of **C1** (R_g ~ 15.9 Å and R_{ee} ~ 41 Å) and others (R_g 15~19 Å and R_{ee} 45~53 Å) (Figure S6B-C and Table S4).

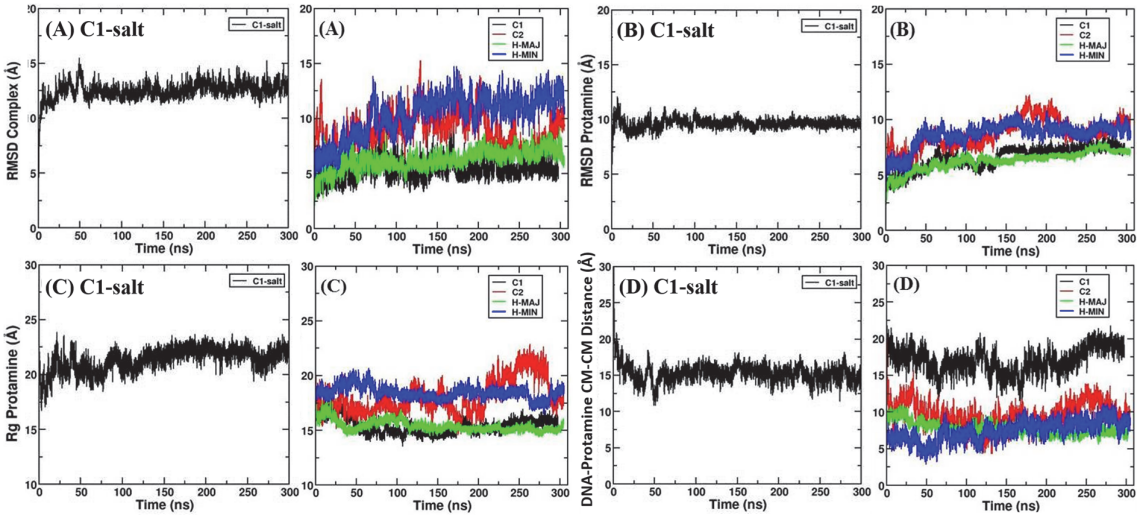

Figure S6. Salt dependence. Time evolution in (A) RMSD and (D) DNA-protamine CM-CM distance of the DNA-protamine complex as well as (B) RMSD and (C) R_g of protamine of [left] **C1-salt** (black) at physiological salt concentration 150 mM and [right] no-extra-salt systems **C1**, **C2**, **H-MAJ**, and **H-MIN** (black, red, green, and blue, as in Figure 4 in the main text).

Table S4. R_g and R_{ee} for protamine in DNA-protamine complex averaged over the last 20 ns.

/Å	(free)	**C1-salt**	**C1**	**C2**	**H-MAJ**	**H-MIN**
R_g	(21.3±1.9)	**~21**	15.9±0.2	18.6±1.3	15.1±0.1	18.1±0.6
R_{ee}	(53±10)	**~67**	41±2	53±6	45±1	51±4

Figure S7 shows the time evolution of the number of ARG CZ (black curve) and Na^+ (red curve) adsorbed on DNA in **C1-salt** as well as **C1** and the other no-extra-salt systems **C2**, **H-MAJ**, and **H-MIN**. Adsorption of protamine (i.e., ARG CZ) on DNA leads to ejection of Na^+ ions from it. The average numbers of adsorbed ARG CZ and Na^+ averaged during the last 20 ns (280-300 ns) of the simulation is listed in Table S5. Owing to the additional Na^+ ions in the solution, the number of Na^+ adsorbed on DNA is significantly larger in **C1-salt** (16.5) than in **C1** (10.6) and other no-extra-salt systems (~9.6). However, the number of ARG CZ adsorbed on DNA in **C1-salt** (16.7) is similar to those in **C1** (15.1) and other no-extra-salt systems (15.7~19.0), indicating the stability of the DNA-protamine binding at the physiological salt concentration. In addition, the number of contacts made between any atom of the DNA and any atom of the protamine within a cutoff distance of 3 Å (i.e., mostly H-bond pairs and H--H contacts) averaged during the last 100 ns of simulation is 270 ± 23 for **C1-salt** (Figure S8, left panel), which is close to 263, 275, and 288 of **C1**, **C2**, and **H-MAJ**, respectively (Figure S8, right panel).

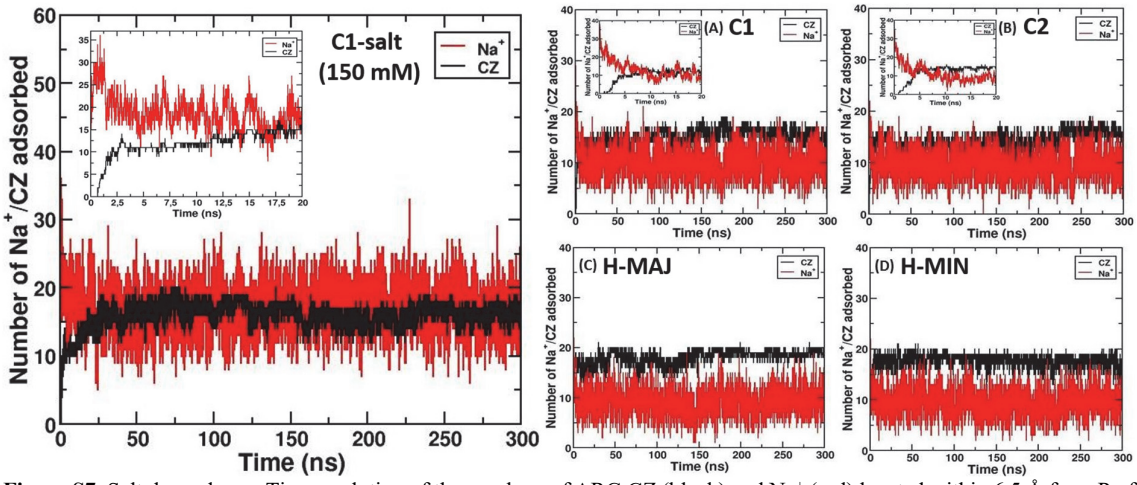

Figure S7. Salt dependence. Time evolution of the numbers of ARG CZ (black) and Na^+ (red) located within 6.5 Å from P of DNA. The left panel is for **C1-salt**. and the right panel is for (A) **C1**, (B) **C2**, (C) **H-MAJ**, and (D) **H-MIN**, which is the same as Figure 6 of the main text. Zoomed-in profiles of the first 20 ns are shown as insets.

Table S5. Number of ARG CZ or Na around P of DNA averaged over the last 20 ns.

Complex	(DNA)[b]	**C1-salt**	**C1**	**C2**	**H-MAJ**	**H-MIN**
C_Z (280-300 ns)[a]	-	**16.7 ± 0.9**	15.1	15.72	19.03	17.15
Na^+ (280-300 ns)[a]	(16.3)[b]	**16.5 ± 3.1**	10.6	9.61	9.62	9.56

[a]Located within 6.5 Å; [b]Averaged over the last 20 ns (160-180 ns) for free DNA.

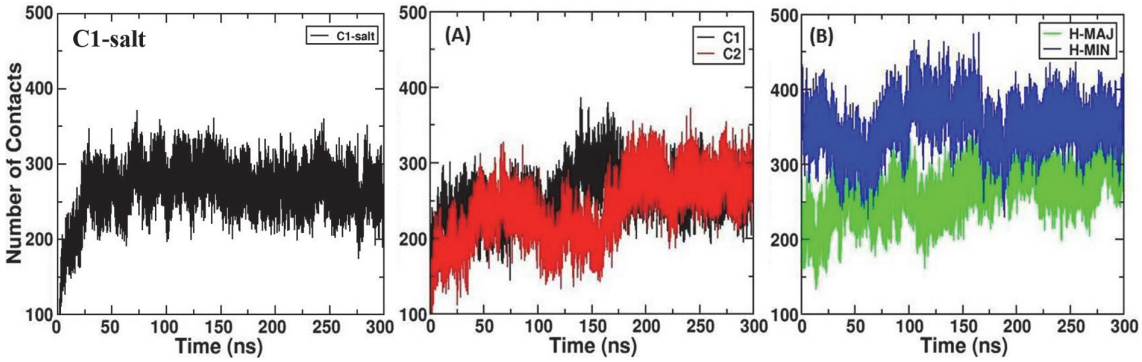

Figure S8. Salt dependence. Time evolution of the number of DNA-protamine contacts within a cutoff of 3 Å (H-bonding pairs and H--H contacts) in (left) **C1-salt** and (right) **C1**, **C2**, **H-MAJ**, and **H-MIN** (taken from Figure 5 of the main text).

Table S6. Number of DNA-protamine (H-bonding and H--H) contacts within a cutoff of 3 Å.

Complex	**C1-salt**	**C1**	**C2**	**H-MAJ**	**H-MIN**
Contact number (H--H)[a]	**270 ± 23**	263	275	288	355

[a]Averaged over the last 20 ns (280-300 ns).

Based on these results, we conclude that the strong electrostatic interaction (including hydrogen bonding) between DNA and protamine is little affected by salt concentration, as also proposed by previous studies.[1-2]

REFERENCES

(1) Mascotti, D. P.; Lohman, T. M. Thermodynamics of Oligoarginines Binding to RNA and DNA. *Biochemistry* **1997,** *36*, 7272-7279.

(2) Chen, A. A.; Pappu, R. V.; Quantitative Characterization of Ion Pairing and Cluster Formation in Strong 1:1 Electrolytes. *J. Phys. Chem. B* **2007,** *111*, 6469-6478.

(3) DeRouchey, J. E.; Rau, D. C. Salt Effects on Condensed Protamine-DNA Assemblies: Anion Binding and Weakening of Attraction. *J. Phys. Chem. B* **2011**, *115*, 11888-11894.