Role of Protein Interactions in Stabilizing Canonical DNA Features in Simulations of DNA in Crowded Environments

Asli Yildirim (aslii.yildirim@gmail.com)[1], *Nathalie Brenner (Nathalie.Brenner@uni-duesseldorf.de)* [2,3], *Robert Sutherland (robert83sutherland@gmail.com)*[3], *Michael Feig (mfeiglab@gmail.com)*[3,*]

[1]Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA

[2]Faculty of Mathematics and Natural Sciences, University of Düsseldorf, 40225 Düsseldorf, Germany

[3]Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

Corresponding Author Contact Information

[*]603 Wilson Road

Room BCH 218

East Lansing, MI 48824

USA

Phone: 517-432-7439

E-mail: mfeiglab@gmail.com

ABSTRACT

Background

Cellular environments are highly crowded with biological macromolecules resulting in frequent non-specific interactions. While the effect of such crowding on protein structure and dynamics has been studied extensively, very little is known how cellular crowding affects the conformational sampling of nucleic acids.

Results

The effect of protein crowding on the conformational preferences of DNA (deoxyribonucleic acid) is described from fully atomistic molecular dynamics simulations of systems containing a DNA dodecamer surrounded by protein crowders. From the simulations, it was found that DNA structures prefer to stay in B-like conformations in the presence of the crowders. The preference for B-like conformations results from non-specific interactions of crowder proteins with the DNA sugar-phosphate backbone. Moreover, the simulations suggest that the crowder interactions narrow the conformational sampling to canonical regions of the conformational space.

Conclusions

The overall conclusion is that crowding effects may stabilize the canonical features of DNA that are most important for biological function. The results are complementary to a previous study of DNA in reduced dielectric environments where reduced dielectric environments alone led to a conformational shift towards A-DNA. Such a shift was not observed here suggested that the reduced dielectric response of cellular environments is counteracted by non-specific interactions with protein crowders under *in vivo* conditions.

KEYWORDS

Molecular dynamics, protein G, A-DNA, B-DNA, conformational sampling, solvent interactions

INTRODUCTION

Biological cells are highly crowded environments due to the presence of various macromolecules. The macromolecular crowding in cells plays a crucial role in biological processes as it may alter the structure and dynamics of biomolecules [1]. A typical biological cell has a concentration of biomolecules in the range of 300 – 400 mg/ml [2], corresponding to a macromolecular volume fraction of 20 – 30 % [3]. Such an environment is substantially different from dilute solutions, the frequently considered environment in most biological experiments. Recent studies have begun to consider the effects of cellular crowding and have shed light on its effects on the structure and function of biomolecules [4-9]. Three essential crowding effects have been reported from experiments [10] and simulations [11]: (1) the volume exclusion effect has been suggested to favor more compact conformations based on entropic arguments, thereby generally stabilizing more compact states[12, 13]; (2) non-specific interactions between biomolecules and surrounding protein crowders have led to the destabilization of native states [14-16] as well as reduced diffusion [17]; and (3) altered solvation properties including reduced dynamic and dielectric properties [18] have implied a reduced hydrophobic effect [19, 20].

While much attention so far has been on proteins, nucleic acids are also affected by macromolecular crowding [21, 22]. G-quadruplex DNA structure assumes a parallel-G quadruplex form under crowded environments due to the excluded volume effect as well as alterations in the hydration of DNA [23-25]. Long DNA duplexes undergo a collapsing transition in the presence of polyethylene glycol (PEG) in solution, which can also be explained by the volume exclusion effect

favoring states that are more compact [26, 27]. The negatively charged protein bovine serum albumin (BSA) similarly causes a compaction of large DNA molecules due to the volume exclusion effect and repulsive electrostatic interactions [28]. Short DNA duplexes, on the other hand, have been extensively investigated by both experimental techniques and computer simulations in terms of co-solvent, salt effects, and crystallization [29-39]. The DNA duplex is well-known to be most stable in the B-form [40] in aqueous solution and in A-form in environments with depleted water and for certain sequences [41]. Very high concentrations of salt can also induce the B- to A- form transition by bringing the negatively charged phosphate groups of DNA closer [34, 38, 42-44] while the addition of ethanol favors the A-form due to reduced electrostatics [32, 33, 35, 37, 45-47]. More recently, the effect of reduced dielectric environments on DNA as one aspect of cellular crowding was investigated and has also been shown to favor non-canonical A-form structures in implicit solvent simulations [20]. On the other hand, a study based on a coarse-grained model has suggested that even in the presence of significant crowding there may be a solvent-rich region around DNA that is depleted in crowder molecules, which was found to have an impact on the kinetics of proteins diffusing along DNA [48]. However, to the extent that crowder proteins do interact non-specifically with DNA, the effect of explicit protein crowder molecules on DNA duplex structures is not well understood.

Here, we describe fully atomistic molecular dynamics (MD) simulations of DNA dodecamers in the presence of explicit protein crowders in order to investigate how DNA structure and stability may be affected under such conditions. We find a general tendency of the DNA to favor the B-form in crowded environments, which is in contrast to the shift towards A-form DNA observed in the simpler reduced dielectric environments [20]. The stabilization of B-DNA appears to be due to non-specific protein-DNA interactions. We also observe, some alterations in the hydration

structure and ion distributions around DNA under crowded conditions. The results are described in detail and discussed in the following after outlining the computational methods used in this study.

Table 1. Overview of Simulations

DNA Sequence	Box Size (Å)	Protein Vol (%)	Protein Conc. (g/L)	Ion Molarity (M)	Ion Molality (mol/kg)	Number of atoms	Simulation Length (μs)	Number of replicates
Drew-Dickerson	54.62	0	0.00	0.45	0.45	16,685-16,720	1	3
Drew-Dickerson	61.02	20	362.49	0.32	0.45	23,958-24,087	1	5
Drew-Dickerson	57.90	30	424.31	0.29	0.45	20,384-20,483	1	5
Drew-Dickerson	53.21	40	546.68	0.24	0.45	15,824-16,019	1	5
GC-rich	56.26	0	0.00	0.43	0.45	18,028-18,059	1	3
GC-rich	61.31	20	357.37	0.32	0.45	24,125-24,445	1	5
GC-rich	57.74	30	427.84	0.29	0.45	20,353-20,377	1	5
GC-rich	53.18	40	547.61	0.24	0.45	15,499-15,904	1	5

METHODS

MD simulations of Drew-Dickerson $((CGCGAATTCGCG)_2)$ and GC-rich $(CGCCCCGCGGGCG)_2$) dodecamers in crowded protein environments were carried out using the CHARMM (Chemistry at Harvard Molecular Mechanics) program package (version 41a1) [49] with the CHARMM36 force field [50, 51]. The initial Drew-Dickerson dodecamer structure was obtained from the X-ray structure (PDB: 1BNA) [52], and the initial GC-rich dodecamer structure was obtained by mutating the base sequence in the X-ray structure of the Drew-Dickerson dodecamer using the MMTSB (Molecular Modeling Tools in Structural Biology) Tool Set [53]. In experiment, the Drew-Dickerson dodecamer is very stable in B-form, in the crystal as well as

in solution [52]. The crystal structure of the GC-rich dodecamer is in A-form [54] but there is less known about the conformation of its conformation in solution. Generally there is little evidence for A-DNA conformations in solution unless the salt concentration is much higher than typical physiological conditions [34, 38, 42-44] and/or when co-solvents such as ethanol are present in significant fractions [32, 33, 35, 37, 45-47]. Therefore, we setup both systems in B-form as the likely conformation of both sequences in dilute aqueous solvent. The choice of sequences and initial structures also allows a direct comparison with our previous continuum dielectric study [20].

For each dodecamer, a dilute system without crowders (0% crowder fraction) and three systems with different protein crowder volume fractions (20%, 30%, 40%) were prepared. Protein G (PDB: 1PGB) [55] was selected as the crowder protein due to its small size and stability in computer simulations [15]. We used neutral protein G models molecules introduced in previous work [56], where D36, D40, E19 and E42 are protonated. In the previous study, both, the charged and neutralized variants of protein G were studied under crowded conditions similar to the systems studied here but without DNA and both were found to be stable in simulations [56]. Protein G is not known to specifically interact with DNA and we chose the net-neutral form to reduce electrostatic interactions with the highly charged DNA to focus on more general crowding effects while still maintaining protein-like crowders. The crowded systems (20%, 30%, 40%) consisted of one dodecamer and 8 protein G molecules, whereas the dilute systems only contained one dodecamer. Simulation box sizes were varied between 53.2 – 61.3 Å to obtain the abovementioned crowder volume fractions. The box sizes were varied instead of the number of protein copies to achieve exactly the target crowder fractions and minimize computational costs at the higher concentrations as in previous work [15, 57]. Simulation conditions of the systems are given in Table 1. There is no experimental evidence for a specific DNA-protein G complex that is stable

over long time and consequently the system is assumed to be fully dynamic in the liquid state with molecular interactions varying transiently. To avoid biasing towards any specific initial protein G-DNA interaction, the initial crowded systems were set up by randomly rotating and placing the DNA dodecamer and the crowder proteins in the simulation box using a protocol developed previously [58]. Different replicates of each system had different initial orientations and placements of the DNA and the surrounding crowders. All systems were solvated with explicit TIP3P (three-site transferable intermolecular potential) [59] water molecules. To neutralize the DNA dodecamer, 22 sodium ions were added to the systems. In order to keep the ion molality of all systems the same, 6 and 12 additional pairs of sodium and chloride ions were added to 30 % and 0/20 % systems, respectively. Therefore, all systems had 0.45 mol/kg ion molality.

The initial systems were minimized for 1,000 steps using the adopted bases Newton Raphson (ABNR) algorithm and were subsequently heated by running simulations without using any restraints at 50K, 100K, 150K, 200K, 250K for 4 ps and at 298K for 10 ps. Productions runs were carried out at 298 K in the NVT ensemble for 1 μs with a 2 fs time step. The SHAKE algorithm [60] was used to constrain bond lengths involving hydrogen atoms. Temperature control was obtained by a Langevin thermostat with a 0.01 ps^{-1} friction coefficient. Lennard-Jones and direct electrostatic interactions were cut off at 12 Å with a switching function becoming effective at 10 Å. Electrostatic interactions were calculated from particle-mesh Ewald [61] summation using 1 Å grid spacing. All simulations were performed using periodic boundary conditions. For the crowded systems, five independent simulations were carried out starting from different initial orientations. For the dilute systems without protein crowders, simulations were replicated three times starting from different initial velocities for the atoms.

The analysis of the helicoidal and backbone parameters of the dodecamers (see Figure S1) were performed by using the 3DNA program package [62]. The reported values are averages over snapshots. Radial distribution functions and 3D volume densities were analyzed by using in-house scripts. All the other analysis was carried out using the MMTSB Tool Set [53] in combination with CHARMM [49]. Clustering analysis was performed by applying the k-means clustering algorithm by using the *kclust* program in MMTSB [53]. For each dodecamer, all snapshots from the simulations with different protein concentrations were aggregated and clustered by using a 3 Å clustering radius. Only the last 700 ns of the simulations were analyzed because of larger variations in the helicoidal parameters during the first 300 ns (see Fig. S2). Only the inner eight base-pairs were taken into consideration to ignore structural distortions due to base fraying. VMD (visual molecular dynamics) [63] and PyMOL [64] were used for visualization.

RESULTS

Microsecond-scale molecular dynamics simulations of DNA dodecamers with and without protein crowders were carried to study the effect of crowding on DNA structure. We focused our analysis on helical properties including base geometries, groove widths and DNA bending, backbone torsions, interactions with crowder proteins, correlations between protein contacts and helical properties, and water and ion distributions around DNA.

Helical properties

Snapshots from the simulations were clustered to identify major conformations. Representative structures for each of the major clusters (with more than 5% population) are depicted in Fig. 1. Generally, the helices stayed intact with base fraying at the termini, which is common in

simulations of short DNA fragments [65]. The structures generally resemble B-DNA structures for both sequences. Average root mean square deviation (RMSD) values of different clusters from the initial canonical B-DNA structures vary between 1.4 and 2.0 Å for the Drew-Dickerson dodecamer and between 1.6 to 2.6 Å except for one cluster with an RMSD of 3.7 Å for the GC-rich dodecamer (see Table S1). There is no clear pattern of increasing or decreasing RMSD values for the clusters most populated at different crowder concentrations.

Table 2. Average Helical Parameters for the Drew-Dickerson Dodecamer

		Canonical		**Simulations in crowded environment**			
	X-ray (1BNA)	A-DNA	B-DNA	0 %	20 %	30 %	40 %
Slide (Å)	0.07 (*0.20*)	-1.62 (*0.06*)	0.16 (*0.08*)	0.26 (*0.04*)	0.25 (*0.06*)	0.21 (*0.04*)	0.25 (*0.05*)
Twist (deg)	34.22 (*2.13*)	30.34 (*0.58*)	34.70 (*0.70*)	33.42 (*0.18*)	34.06 (*0.25*)	33.96 (*0.38*)	33.43 (*0.31*)
X-displacement (Å)	-0.23 (*0.20*)	-4.50 (*0.18*)	-0.20 (*0.13*)	-0.59 (*0.10*)	-0.60 (*0.14*)	-0.72 (*0.11*)	-0.79 (*0.12*)
Helical rise (Å)	3.29 (*0.05*)	2.68 (*0.08*)	3.25 (*0.02*)	3.24 (*0.00*)	3.27 (*0.01*)	3.25 (*0.02*)	3.27 (*0.02*)
Inclination (deg)	4.02 (*2.73*)	17.78 (*1.56*)	4.34 (*0.77*)	10.87 (*0.22*)	10.62 (*0.44*)	11.15 (*0.59*)	12.59 (*0.77*)
z_p (Å)	-0.23 (*0.07*)	2.06 (*0.07*)	-0.33 (*0.04*)	-0.07 (*0.03*)	0.12 (*0.05*)	0.14 (*0.05*)	0.22 (*0.04*)
Minor groove (Å)	10.32 (*0.46*)	15.72 (*0.12*)	10.77 (*0.12*)	13.50 (*0.05*)	13.42 (*0.18*)	13.08 (*0.14*)	13.73 (*0.19*)
Major groove (Å)	17.34 (*0.33*)	12.94 (*0.39*)	17.14 (*0.12*)	16.47 (*0.11*)	16.54 (*0.10*)	16.39 (*0.15*)	16.31 (*0.12*)

Averages over all base-pairs excluding the first and last two terminal base-pairs with errors given in parentheses based on the variations in the independent simulations. Canonical values were averaged over the A-form structures 3V9D, 3QK4, 2B1B, 1ZEX, 1ZEY, 1ZF1, 1ZF8, 1ZF9, 1ZFA and the B-form structures 2M2C, 4AGZ, 4H0, 4AH1, 3U05, 3U08, 1VTJ,3U2N, 3OIE, 3BSE.

Table 3. Average Helical Parameters for the GC-rich Dodecamer

		Canonical		**Simulations in crowded environment**			
	X-ray (399D)	A-DNA	B-DNA	0 %	20 %	30 %	40 %

Slide (Å)	-1.71 (*0.16*)	-1.62 (*0.06*)	0.16 (*0.08*)	0.32 (*0.19*)	0.02 (*0.10*)	0.02 (*0.10*)	0.00 (*0.07*)
Twist (deg)	29.59 (*1.34*)	30.34 (*0.58*)	34.70 (*0.70*)	32.47 (*0.85*)	32.93 (*0.42*)	33.34 (*0.41*)	33.11 (*0.36*)
X-displacement (Å)	-5.01 (*0.41*)	-4.50 (*0.18*)	-0.20 (*0.13*)	-0.46 (*0.28*)	-1.03 (*0.20*)	-1.09 (*0.22*)	-1.10 (*0.16*)
Helical rise (Å)	2.66 (*0.22*)	2.68 (*0.08*)	3.25 (*0.02*)	3.26 (0.*03*)	3.23 (*0.04*)	3.26 (*0.02*)	3.27 (*0.02*)
Inclination (deg)	20.71 (*4.33*)	17.78 (*1.56*)	4.34 (*0.77*)	10.01 (*0.33*)	10.70 (*0.38*)	11.34 (*0.49*)	11.34 (*1.19*)
z_p (Å)	1.56 (*0.35*)	2.06 (*0.07*)	-0.33 (*0.04*)	-0.21 (*0.12*)	-0.20 (*0.09*)	-0.23 (*0.07*)	-0.18 (*0.05*)
Minor groove (Å)	16.22 (*0.47*)	15.72 (*0.12*)	10.77 (*0.12*)	14.91 (*0.23*)	14.54 (*0.08*)	14.52 (*0.10*)	14.76 (*0.21*)
Major groove (Å)	13.14 (*2.63*)	12.94 (*0.39*)	17.14 (*0.12*)	16.26 (*0.16*)	16.53 (*0.13*)	16.22 (*0.14*)	16.27 (*0.23*)

see Table 2.

Helicoidal parameters for both, Drew-Dickerson and GC-rich dodecamers were averaged from the simulations. They are summarized in Tables 2 and 3, respectively. Helicoidal parameters for crystal structures of the respective dodecamers as well as canonical A- and B-forms of DNA, averaged over ten A- and B-form crystal structures each, are provided for comparison. Average properties for each of the clusters shown in Fig. 1 are given in Tables S1 and S2. The more detailed analysis of the base geometries also indicates that both dodecamers remained close to B-DNA. The Drew-Dickerson dodecamer also remained reasonably close to the respective crystal structure (1BNA), but there are larger deviations between the simulation results and the crystal structure of the GC-rich dodecamer. The crystal structure for the GC-rich dodecamer is predominantly in A-form, presumably as a result of salt concentrations above 1 M and/or the crystal environment [54]. As mentioned above, although the crystal structure of the GC-rich dodecamer has been reported in A-form, there is no evidence that this sequence (or any other DNA sequence) assumes an A-DNA conformation in solution at sub-molar salt concentrations and in the absence of co-solvents. Therefore we expected the GC-dodecamer to remain in B-form. In the presence of the protein

196 crowders, the helical parameters generally did not change much. We found increased X-
197 displacement (p-values: 0.91 (Drew-Dickerson 20%), 0.16 (Drew-Dickerson 30%), 0.05 (Drew-
198 Dickerson 40%), 0.05 (GC-rich 20%), 0.04 (GC-rich 30%), 0.07 (GC-rich 40%) and base
199 inclination (p-values: 0.33 (Drew-Dickerson 20%), 0.38 (Drew-Dickerson 30%), 0.01 (Drew-
200 Dickerson 40%), 0.05 (GC-rich 20%), 0.01 (GC-rich 30%), 0.08 (GC-rich 40%) for both
201 dodecamers as a function of crowding. The increased x-displacement and base inclination point
202 towards A-DNA but the values upon crowding still remained much closer to canonical B-DNA
203 than A-DNA.

204 We further analyzed the displacement of phosphorus atoms relative to the horizontal plane
205 passing between base-pairs in a base-pair step (z_p) and major/minor grooves (Tables 2 & 3). The
206 z_p parameter is very different between the two forms of DNA. While B-DNA has values near -0.3
207 Å, the parameter is mostly larger than 2.0 Å for A-DNA. This parameter does not show a trend
208 upon crowding for the GC-rich dodecamer, while the Drew-Dickerson dodecamer had larger
209 values in crowded environments (p-values: 0.0011 (20%), 0.0007 (30%), 0.0001 (40%)). Again
210 this indicates a slight tendency towards A-DNA geometries while still remaining much closer to
211 canonical B-DNA values. Minor and major groove widths also did not change significantly upon
212 crowding, but we note that minor groove widths were generally overestimated compared to
213 canonical B-DNA values. This is a general feature of the CHARMM force field that was used here
214 [66]. Finally, we analyzed the helical bending angles (see Table S3) which also did not show a
215 significant change upon crowding.

216 ***Sugar conformations and backbone torsions***

217 A key feature of nucleic acid backbone is the ribose pucker conformation. A-form DNA is known
218 to prefer C3$^{'}$-endo and C2$^{'}$-exo conformations whereas B-form DNA is characterized by C3$^{'}$-

exo and C2'-endo conformations. As shown in Fig. 2, the sugars of both dodecamers generally remained in C3'-exo and C2'-endo conformations. As expected, C3'-endo and C2'-exo sugar conformations are more prominent for the GC-rich dodecamer (see Fig. 2B). Again, there is no major change upon crowding, but in the GC-rich dodecamer, sugars shift slightly to C3'-exo and C2'-endo sugar conformations up to 30% crowding, but then revert back to more A-form conformations at 40% crowder concentrations.

We further analyzed torsion angles along the phosphate backbone. χ and δ angles are the most distinctive backbone angles to distinguish between A- and B- form DNA. We constructed potentials of mean force (PMF) as a function of δ and χ from the simulations (Fig. 3). The separation between A- and B-DNA torsion angles is readily apparent. Consistent with the ribose puckers and helical geometries, there is more sampling of B-DNA torsion angles for both dodecamers. While there is little change in the sampling of the major A- and B-form, the presence of crowders appears to affect the sampling of minor conformations with A-like δ values around 80 degrees and B-like χ values around -100 degrees. Sampling in this region is significantly reduced in both dodecamers upon crowding (see Fig. 3). This region corresponds to a conformation where bases stay in the same orientation relative to the sugar as in B-form, but they are slightly more exposed to the environment, and apparently, this conformation is largely prevented by crowder proteins. The sampling of ε and ζ torsion angles distinguishes between BI/BII forms. A similar trend is observed where crowding reduces the sampling of minor states outside the major BI/BII basins (see Fig. S3). Based on this analysis, it appears that one effect of protein crowders may be to focus the sampling of DNA conformations on the major conformations.

Protein crowder conformations

In previous simulation studies involving protein G under crowded conditions, the protein G conformations remained close to the experimental structure and were not affected strongly by the concentrated environment [15]. In the systems studied here, protein G also remains highly stable and close to the experimental structure (see Fig. S4). The overall average Cα RMSD value is 0.91 Å with a standard deviation of 0.24 Å between individual protein G molecules and the experimental structure (PDB ID: 3GB1 [55, 67]). The average radius of gyration is 10.76 Å with a standard deviation of 0.1 A, compared to a value of 10.65 Å for the experimental structure. A few conformations deviated slightly further from the native (as much as 2.5 Å Cα RMSD) and with slightly increased radii of gyration, especially at the highest crowder concentration (Fig. S4B). Further analysis via clustering revealed minor substates with slightly increased RMSD values that correlate with closer contacts to the DNA (see Table S4). This suggests that the conformational sampling of protein G may be affected slightly when interacting with the DNA. Almost all of the variations are in the flexible loop involving residues 9 to 13 (see Fig. S4D) where root mean square fluctuations (RMSF) are largest (see Fig. S4C).

DNA-protein interactions

Protein G is not known to interact specifically with DNA but under highly crowded conditions, interactions are unavoidable. Fig. 4 shows where contacts between protein G and DNA occur based on minimum distances between the major/minor grooves and sugar/phosphate groups of the DNA with different residues of protein G. More detailed contact analysis between individual base-pairs and protein G residues is shown in Figures S5 and S6 for the Drew-Dickerson and GC-rich dodecamers, respectively. Most of the contacts are between the DNA sugar-phosphate backbone and protein residues 15-30, mostly in the α-helix of protein G, as well as residues at the N-terminus and near the C-terminus. Contacts involving the DNA grooves, a typical mode of interaction for

DNA-binding proteins were not common with protein G. The interactions partially involve electrostatic attraction between the DNA phosphate and certain lysine residues (K4, K28, K31, and K50), but sugar oxygens O3' and O4' as well as phosphate oxygens also form hydrogen bonds with other polar protein residues. Representative snapshots of protein G-DNA interactions are shown in Fig. 5. As would be expected, the contacts between the proteins and DNA increase with crowder concentration and crowding seems to increase sugar-phosphate-protein contacts more for the Drew-Dickerson dodecamer than for the GC-rich dodecamer.

Correlations between DNA-protein contacts and DNA helix properties

To investigate in more detail whether the close contacts of the crowder proteins with the DNA have the potential to perturb DNA structure, we analyzed correlations between DNA-protein contacts and helicoidal properties of DNA as well as backbone torsion and pseudorotation phase angles. First, we examined the effect of close contacts on the helicoidal parameters listed in Tables 2 and 3. We found that a higher number of close protein contacts corresponded to a more narrow range of sampled values for all of the helicoidal parameters (Figures 6-9, Figures S7-S10). Among these parameters, slide (Fig. 6), x-displacement (Fig. 7), helical rise (Fig. 8) and z_p (Fig. 9) values showed a clear shift towards B-form values with increasing number of contacts. These parameters focus on the displacement of bases along the x- (x-displacement) and y- (slide) axes and of phosphates along the base-pair axis (z_p). All of the values approach zero with crowding. This suggests that DNA bases and phosphates undergo less displacement as a result of crowding. On the other hand, rotations of base-pairs about helical (twist) or base-pair axes (inclination) do not show a distinct shift towards any canonical values (Fig. S7, S8). Major and minor groove widths do not seem to be affected by contacts except for the GC-rich dodecamer, where there appears to

be a clear tendency towards larger minor groove values, i.e. values more similar to A-DNA (Fig. S9, S10).

Similar to the helicoidal parameters, backbone torsion angles also fluctuate in a more narrow range upon crowding (Fig. S11–S17). This suggests that non-specific protein-DNA interactions may limit the conformational fluctuations of the DNA backbone. Particularly, δ and χ angles shift towards B-form values upon higher number of protein contacts, explaining a decrease in the sampling of non-canonical conformations shown in Fig. 3. Finally, pseudorotation angles move to B-form values with protein contacts which lead to C3'-exo and C2'-endo sugar pucker conformations (Fig. S18).

The results discussed here are most pronounced for the Drew-Dickerson dodecamer. In the GC-rich dodecamer, the fluctuations of helicoidal parameters and backbone angles are reduced less and a tendency to sample A-form values further complicates the picture. Overall, our results suggest that the interactions of protein crowders with DNA sugar/phosphate backbone shown in the previous section result in a stiffer DNA backbone. The stiffer backbone also prevents larger base/base-pair displacements and, therefore, restricts the conformational space of DNA. Although it appears that there is not a specific tendency towards one of the major forms of DNA upon crowding, there is a distinct effect of protein crowders on DNA structure by narrowing the conformational sampling to canonical structures.

Hydration and ion distributions around DNA

Water and ions are integral parts of DNA structures. We analyzed hydration patterns and sodium ion distributions around DNA as a function of crowding. Conditional water radial distribution functions (RDF) were obtained for water oxygen distances to the closest heavy atoms in DNA, normalized by the corresponding accessible volume at each distance and the bulk water density

(0.034 $Å^{-3}$) (see Figure 10A). The analysis shows that the first hydration shell is almost unaffected by the level of crowding, but the RDF decreases beyond the hydration shell significantly as a function of crowding. This observation is similar to what has been reported previously for the hydration around proteins under crowded environments [18].

Sodium RDFs were calculated in the same way as the water RDFs but normalized by the ion density of the system (0.002 $Å^{-3}$). There are two peaks in the sodium RDFs corresponding to ions in direct contact with the DNA (around 2.5 Å and largely in the minor groove) and ions interacting with the DNA through water (around 4.5 Å) [68-70]. While the direct contact peak is not affected significantly by crowding, the second peak shows a greater dependence on crowding. At the highest crowder fractions, the second peak is significantly reduced in both dodecamers (see Fig. 10B) and the ion density is reduced further at larger distances similar to the reduction in hydration upon crowding. The effect of crowding on the ion distributions also impacts the DNA neutralization as a function of distance (Figure 10C). 76 % of the DNA phosphate groups are neutralized as suggested by counterion condensation theory at around 9 Å for the dilute system, however, it takes up to 11 – 12 Å to reach 76 % DNA neutralization under crowding conditions. It is interesting, that despite the impact of crowding on the second peak of the ion distribution, the counterion condensation is affected less for distances less than 6 Å. As this may seem counterintuitive, the reader is reminded that the RDF is normalized by the available volume and the overall ion density, at constant ion molality, whereas Fig. 10C simply describes the net neutralization of the DNA by the ions. The extended distance to reach 76% charge neutralization upon crowding may seem to challenge counterion condensation theory. However, the protein crowders, despite being net neutral, can provide additional charge neutralization by orienting basic

lysines near the DNA surface as described above to compensate for the reduced neutralization by the sodium ions.

Finally, the 3D distributions of sodium ions around the Drew-Dickerson and GC-rich dodecamers are compared in Fig. 11. The sodium ion networks in the major and minor grooves of DNA are largely preserved for both dodecamers with little changes upon crowding. However, additional densities become apparent further away from the DNA at different locations upon crowding. Additional ordering of ions could be a result of crowder proteins interacting with the DNA and coordinating ions near the DNA. A snapshot showing a crowder protein interacting with the DNA and orienting a sodium ion at the same time is shown in Figure S19.

DISCUSSION

In this study, we investigated the effect of protein crowding on the conformational preferences of DNA duplexes. In a previous study, we examined one aspect of cellular crowding, namely a reduced dielectric response of the environment due to the less available water and its slowed dynamics. Using continuum models, we found an overall shift towards A-like conformations for DNA as a result of a reduced dielectric response of its environment [20]. Here, we included protein crowders and solvent explicitly to test whether the same conclusions would be found. In the earlier work, we compared environments with ε=20, ε=40, and ε=80. Past work suggests that water under crowded conditions exhibits a reduced dielectric response of about 40 (with uncertainties) at a protein crowder volume fraction of 0.3 [18]. If one makes the further assumption that proteins have an interior dielectric of around 10, one can estimate an average effective dielectric for the entire medium surrounding the DNA at this crowder fraction as ε_{eff}=0.3*10+0.7*40=31 with even lower values at 40% crowder fraction. However, although some of the base parameters moved slightly

towards A-like values upon crowding, B-DNA was largely maintained with the explicit crowder environment in contrast to our previous findings. This suggests that a reduced dielectric response of crowded environments and interactions with crowder proteins have different effects on DNA conformational preferences with a net effect of not altering canonical B-DNA structures much.

We found that the crowder proteins mostly interact with DNA via its phosphate-sugar backbone as previously observed in non-specific binding of proteins to DNA [71]. These interactions arise from the electrostatic interactions between negatively charged phosphate oxygens and positively charged amino acid residues as well as the polar interactions between phosphate and/or sugar oxygens and side chains of polar amino acid residues. Previous studies have shown that DNA can undergo structural deformations from its B-form towards A-type helix as a result of forming complexes with specific DNA binding proteins [72-76], but we did not see such an effect here. It does appear, however, that for the system studied here, the presence of the protein crowders limits the conformational space of DNA to more canonical structures, mostly in B-form, both for the backbone torsions and the helical parameters. However, the narrowed conformational sampling appears to have little effect on the overall structural averages. Such a crowding effect on DNA structure may be understood in similar ways as protein native state stabilization due to the volume exclusion effect [1, 13, 77, 78], where the reduced space due to crowders limits the ability to widely sample conformational space. This would mean that protein crowding *in vivo* helps stabilize the biologically most relevant forms of DNA.

We also studied hydration patterns and ion densities around DNA in protein crowding. The first hydration shell around DNA is largely unaffected by crowding, while the water densities beyond the first solvation shell significantly reduced compared to the bulk water density under crowding effect. This result is very similar to the hydration shell around proteins upon crowding

[18]. This further confirms that, protein crowding in cells generally does not alter the first hydration shell around biomolecules. However, sodium densities around DNA are affected already when interacting with DNA through water. Only the direct-contact first peak in the sodium-DNA RDF appears to be unaffected by crowding. Moreover, the charge neutralization by ions is altered upon crowding with the classical counter-ion condensation threshold reached at larger distances from the DNA than under dilute conditions. This suggests that proteins have to play an increasing role in neutralizing DNA under highly crowded conditions.

CONCLUSION

The results obtained here shed light on the effect of protein crowding on DNA structure. We found that the crowder proteins mostly assist DNA to stay in canonical B-like conformations, limiting excursions to non-canonical conformations rather than a clear shift in the overall, average structure as suggested by a simple dielectric model of cellular environments. We hope that this hypothesis will motivate new experimental efforts to characterize DNA structure under crowded conditions. We expect that reduced conformational dynamics upon crowding would be observable via NMR spectroscopy. Another testable hypothesis is the altered ion distribution predicted by our simulations, which could be amenable to the ion-counting experiments recently carried out by the Herschlag group [79-82].

ABBREVIATIONS

ABNR – Adopted-bases Newton-Raphson; BSA – bovine serum albumin; CHARMM – Chemistry at Harvard Molecular Mechanics; DNA – deoxyribonucleic acid; MD – molecular

dynamics; MMTSB – Molecular Modeling Tools in Structural Biology; PEG – polyethylene glycol; PMF – potential of mean force; RDF – radial distribution function; RMSD – root mean square deviation; RMSF – root mean square fluctuations; TIP3P – three-site transferable intermolecular potential; VMD – visual molecular dynamics

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Initial conformations, simulation data, and results are available from the authors upon request.

Competing interests

None.

Funding

Funding from NSF (MCB 1330560 and MCB 1817307) and NIH (R01 GM092949 and R35 GM126948) to carry out all aspects of the research, analysis, and writing is acknowledged. The funding bodies did not play any direct role in the design of the study, the collection, analysis, or interpretation of data, or in writing the manuscript.

Author's contributions

AY and MF designed the research. AY, NB, RS, and MF carried out simulations and analysis. AY and MF wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

We thank Prof. Charles Hoogstraten for discussions and exploratory analyses of the GC-rich dodecamer under different solution conditions.

Author's Information

Corresponding author: Michael Feig; 603 Wilson Road, Room BCH 218; East Lansing, MI 48824; USA; Phone: 517-432-7439; E-mail: mfeiglab@gmail.com

Supporting Information

Definition of backbone torsions and helicoidal parameters (Figure S1). Time series of helicoidal parameters (Figure S2). Potential of mean force (kcal/mol) as a function of backbone angles

(Figure S3). Analysis of the protein G crowder conformational sampling (Figure S4). Average minimum distances between the crowder protein residues and DNA (Figure S5-S6). Potentials of mean force as a function of helicoidal parameters and protein contacts (Figure S7-S18). A snapshot for the crowder protein interacting with the DNA and sodium (Figure S19). Helicoidal parameters for the clusters (Table S1 and S2), bending angles for the dodecamers (Table S3), clustering analysis of protein G crowders (Table S4), and PMF error analysis (Table S5).

REFERENCES

[1] Zhou H-X, Rivas G and Minton A P 2008 Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences *Annu. Rev. Biophys.* **37** 375-97

[2] Zimmerman S B and Trach S O 1991 Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of *Escherichia coli J. Mol. Biol.* **222** 599-620

[3] Ellis R J 2001 Macromolecular Crowding: an Important but Neglected Aspect of the Intracellular Environment *Curr. Opin. Struct. Biol.* **11** 114-9

[4] Minton A P and Wilf J 1981 Effect of Macromolecular Crowding Upon the Structure and Function of an Enzyme - Glyceraldehyde-3-Phosphate Dehydrogenase *Biochemistry* **20** 4821-6

[5] McPhie P, Ni Y S and Minton A P 2006 Macromolecular Crowding Stabilizes the Molten Globule Form of Apomyoglobin with Respect to Both Cold and Heat Unfolding *J. Mol. Biol.* **361** 7-10

[6] Dix J A and Verkman A S 2008 Crowding Effects on Diffusion in Solutions and Cells *Annu. Rev. Biophys.* **37** 247-63

[7] Minton A P 1993 Macromolecular Crowding and Molecular Recognition *J. Mol. Recognit.* **6** 211-4

[8] Zimmerman S B and Minton A P 1993 Macromolecular Crowding - Biochemical, Biophysical, and Physiological Consequences *Annu. Rev. Biophys. Biomol. Struct.* **22** 27-65

[9] Elcock A H 2010 Models of Macromolecular Crowding Effects and the Need for Quantitative Comparisons with Experiment *Curr. Opin. Struct. Biol.* **20** 196-206

[10] Rivas G and Minton A P 2016 Macromolecular Crowding In Vitro, In Vivo, and In Between *Trends Biochem. Sci.* **41** 970-81

[11] Feig M, Yu I, Wang P-h, Nawrocki G and Sugita Y 2017 Crowding in Cellular Environments at an Atomistic Level from Computer Simulations *J. Phys. Chem. B* **121** 8009-25

[12] Zhou H X 2004 Protein Folding and Binding in Confined Spaces and in Crowded Solutions *J. Mol. Recognit.* **17** 368-75

[13] Cheung M S, Klimov D and Thirumalai D 2005 Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins *Proc. Natl. Acad. Sci. U.S.A.* **102** 4753-8

[14] Senske M, Tork L, Born B, Havenith M, Herrmann C and Ebbinghaus S 2014 Protein Stabilization by Macromolecular Crowding through Enthalpy rather than Entropy *J. Am. Chem. Soc.* **136** 9036-41

[15] Harada R, Tochio N, Kigawa T, Sugita Y and Feig M 2013 Reduced Native State Stability in Crowded Cellular Environment Due to Protein-Protein Interactions *J. Am. Chem. Soc.* **135** 3696-701

[16] Feig M and Sugita Y 2012 Variable Interactions between Protein Crowders and Biomolecular Solutes are Important in Understanding Cellular Crowding *J. Phys. Chem. B* **116** 599-605

[17] Wang Y, Li C and Pielak G J 2010 Effects of Proteins on Protein Diffusion *J. Am. Chem. Soc.* **132** 9392-7

[18] Harada R, Sugita Y and Feig M 2012 Protein Crowding Affects Hydration Structure and Dynamics *J. Am. Chem. Soc.* **134** 4842-9

[19] Tanizaki S, Clifford J, Connelly B D and Feig M 2008 Conformational Sampling of Peptides in Cellular Environments *Biophys. J.* **94** 747-59

[20] Yildirim A, Sharma M, Varner B M, Fang L and Feig M 2014 Conformational Preferences of DNA in Reduced Dielectric Environments *J. Phys. Chem. B* **118** 10874-81

[21] Nakano S-i, Miyoshi D and Sugimoto N 2014 Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids *Chem. Rev.* **114** 2733-58

[22] Nakano S-i and Sugimoto N 2017 Model Studies of the Effects of Intracellular Crowding on Nucleic Acid Interactions *Mol. Biosyst.* **13** 32-41

[23] Xue Y, Kan Z Y, Wang Q, Yao Y, Liu J, Hao Y H and Tan Z 2007 Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K+ Solution Under Molecular Crowding Condition *J. Am. Chem. Soc.* **129** 11185-91
[24] Miyoshi D, Nakao A and Sugimoto N 2002 Molecular Crowding Regulates the Structural Switch of the DNA G-Quadruplex *Biochemistry* **41** 15017-24
[25] Heddi B and Phan A T 2011 Structure of Human Telomeric DNA in Crowded Solution *J. Am. Chem. Soc.* **133** 9824-33
[26] Livolant F and Leforestier A 1996 Condensed phases of DNA: Structures and phase transitions *Progr. Polym. Sci.* **21** 1115-64
[27] Bloomfield V A 1996 DNA Condensation *Curr. Opin. Struct. Biol.* **6** 334-41
[28] Yoshikawa K, Hirota S, Makita N and Yoshikawa Y 2010 Compaction of DNA Induced by Like-Charge Protein: Opposite Salt-Effect against the Polymer-Salt-Induced Condensation with Neutral Polymer *J. Phys. Chem. Lett.* **1** 1763-6
[29] Kulkarni M and Mukherjee A 2017 Understanding B-DNA to A-DNA Transition in the Right-Handed DNA Helix: Perspective from a Local to Global Transition *Progr. Biophys. Mol. Biol.* **128** 63-73
[30] Ivanov V I, Minchenkova L E, Minyat E E, Frank-Kamenetskii M D and Schyolkina A K 1974 The B to A Transition of DNA in Solution *J. Mol. Biol.* **87** 817-33
[31] Jose D and Porschke D 2004 Dynamics of the B-A Transition of DNA Double Helices *Nucleic Acids Res.* **32** 2251-8
[32] Zimmerman S B and Pheiffer B H 1979 A Direct Demonstration that the Ethanol-Induced Transition of DNA is Between the A and B Forms: An X-ray Diffraction Study *J. Mol. Biol.* **135** 1023-7
[33] Cheatham T E, Crowley M F, Fox T and Kollman P A 1997 A Molecular Level Picture of the Stabilization of A-DNA in Mixed Ethanol-Water Solutions *Proc. Natl. Acad. Sci. U.S.A.* **94** 9626-30
[34] Cheatham T E and Kollman P A 1997 Insight into the Stabilization of A-DNA by Specific Ion Association: Spontaneous B-DNA to A-DNA Transitions Observed in Molecular Dynamics Simulations of d(ACCCGCGGGT)$_2$ in the Presence of Hexaamminecobalt(III) *Structure* **15** 1297-311
[35] Noy A, Perez A, Laughton C A and Orozco M 2007 Theoretical Study of Large Conformational Transitions in DNA: The B<-->A Conformational Change in Water and Ethanol/Water *Nucleic Acids Res.* **35** 3330-8
[36] Pastor N 2005 The B- to A-DNA Transition and the Reorganization of Solvent at the DNA Surface *Biophys. J.* **88** 3262-75
[37] Gu B, Zhang F S, Wang Z P and Zhou H Y 2008 Solvent-Induced DNA Conformational Transition *Phys. Rev. Lett.* **100** 088104
[38] Arscott P G, Ma C, Wenner J R and Bloomfield V A 1995 DNA Condensation by Cobalt Hexaammine (III) in Alcohol-Water Mixtures: Dielectric Constant and Other Solvent Effects *Biopolymers* **36** 345-64
[39] Kuzmanic A, Dans P D and Orozco M 2018 An in-depth look at DNA crystals through the prism of molecular dynamics simulations *bioRxiv*
[40] Watson J D and Crick F H 1953 The Structure of DNA *Cold Spring Harbor Symp. Quant. Biol.* **18** 123-31
[41] Franklin R E and Gosling R G 1953 Molecular Configuration in Sodium Thymonucleate *Nature* **171** 740-1
[42] Gao Y-G, Robinson H, Boom J H v and Wang A H-J 1995 Influence of Counter-Ions on the Crystal Structures of DNA Decamers: Binding of [Co(NH3)6]3+ and Ba2+ to A-DNA *Biophys. J.* **69** 559-68

[43] Robinson H and Wang A H 1996 Neomycin, Spermine and Hexaamminecobalt (III) Share Common Structural Motifs in Converting B- to A-DNA *Nucleic Acids Res.* **24** 676-82
[44] Xu Q W, Shoemaker R K and Braunlin W H 1993 Induction of B-A Transitions of Deoxyoligonucleotides by Multivalent Cations in Dilute Aqueous-Solution *Biophys. J.* **65** 1039-49
[45] Fang Y, Spisz T S and Hoh J H 1999 Ethanol-Induced Structural Transitions of DNA on Mica *Nucleic Acids Res.* **27** 1943-9
[46] Young M A and Beveridge D L 1998 Molecular Dynamics Simulations of an Oligonucleotide Duplex with Adenine Tracts Phased by a Full Helix Turn *J. Mol. Biol.* **281** 675-87
[47] Srinivasan J, Cheatham T E, Cieplak P, Kollman P A and Case D A 1998 Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices *J. Am. Chem. Soc.* **120** 9401-9
[48] Dey P and Bhattacherjee A 2018 Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins *Sci. Rep.-Uk* **8** 844
[49] Brooks B R, Brooks C L, Mackerell A D, Nilsson L, Petrella R J, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner A R, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor R W, Post C B, Pu J Z, Schaefer M, Tidor B, Venable R M, Woodcock H L, Wu X, Yang W, York D M and Karplus M 2009 CHARMM: The Biomolecular Simulation Program *J. Comput. Chem.* **30** 1545-614
[50] Hart K, Foloppe N, Baker C M, Denning E J, Nilsson L and Mackerell A D, Jr. 2012 Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium *J. Chem. Theory Comput.* **8** 348-62
[51] Best R B, Zhu X, Shim J, Lopes P, Mittal J, Feig M and MacKerell Jr A D 2012 Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **8** 3257-73
[52] Drew H R, Wing R M, Takano T, Broka C, Tanaka S, Itakura K and Dickerson R E 1981 Structure of a B-DNA Dodecamer: Conformation and Dynamics *Proc. Natl. Acad. Sci. U.S.A.* **78** 2179-83
[53] Feig M, Karanicolas J and Brooks III C L 2004 MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology *J. Mol. Graph. Modell.* **22** 377-95
[54] Malinina L, Fernandez L G, Huynh-Dinh T and Subirana J A 1999 Structure of the d(CGCCCGCGGGCG) Dodecamer: A Kinked A-DNA Molecule Showing Some B-DNA Features *J. Mol. Biol.* **285** 1679-90
[55] Gronenborn A M, Filpula D R, Essig N Z, Achari A, Whitlow M, Wingfield P T and Clore G M 1991 A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G *Science* **253** 657-61
[56] Wang P-H, Yu I, Feig M and Sugita Y 2017 Influence of Protein Crowder Size on Hydration Structure and Dynamics in Macromolecular Crowding *Chem. Phys. Lett.* **671** 63-70
[57] Nawrocki G, Wang P-h, Yu I, Sugita Y and Feig M 2017 Slow-Down in Crowded Protein Solutions Correlates with Transient Oligomer Formation *J. Phys. Chem. B* **121** 11072-84
[58] Feig M, Harada R, Mori T, Yu I, Takahashi K and Sugita Y 2015 Complete Atomistic Model of a Bacterial Cytoplasm Integrates Physics, Biochemistry, and Systems Biology *J. Mol. Graph. Modell.* **58** 1-9
[59] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 Comparison of Simple Potential Functions for Simulating Liquid Water *J. Chem. Phys.* **79** 926-35
[60] Ryckaert J P, Ciccotti G and Berendsen H J C 1977 Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes *J. Comput. Phys.* **23** 327-41
[61] Darden T A, York D and Pedersen L G 1993 Particle-Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems *J. Chem. Phys.* **98** 10089-92

[62] Lu X J and Olson W K 2003 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures *Nucleic Acids Res.* **31** 5108-21
[63] Humphrey W, Dalke A and Schulten K 1996 VMD: Visual Molecular Dynamics *J. Mol. Graph.* **14** 33-8
[64] Amadei A, Apol M E F and Berendsen H J C 1998 On the Use of the Quasi-Gaussian Entropy Theory in Noncanonical Ensembles. I. Prediction of Temperature Dependence of Thermodynamic Properties *J. Chem. Phys.* **109** 3004-16
[65] Zgarbová M, Otyepka M, Šponer J, Lankaš F and Jurečka P 2014 Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA *J. Chem. Theory Comput.* **10** 3177-89
[66] Cheatham T E and Young M A 2001 Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise *Biopolymers* **56** 232-56
[67] Kuszewski J, Gronenborn A M and Clore G M 1999 Improving the Packing and Accuracy of NMR Structures with a Pseudopotential for the Radius of Gyration *J. Am. Chem. Soc.* **121** 2337-8
[68] Feig M and Pettitt B M 1999 Sodium and Chlorine Ions as Part of the DNA Solvation Shell *Biophys. J.* **77** 1769-81
[69] Korolev N, Lyubartsev A P, Laaksonen A and Nordenskiold L 2002 On the Competition Between Water, Sodium Ions, and Spermine in Binding to DNA: A Molecular Dynamics Computer Simulation Study *Biophys. J.* **82** 2860-75
[70] Bonvin A M 2000 Localisation and Dynamics of Sodium Counterions Around DNA in Solution from Molecular Dynamics Simulation *Eur. Biophys. J.* **29** 57-60
[71] Kalodimos C G, Biris N, Bonvin A M, Levandoski M M, Guennuegues M, Boelens R and Kaptein R 2004 Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes *Science* **305** 386-9
[72] Nekludova L and Pabo C O 1994 Distinctive DNA Conformation with Enlarged Major Groove Is Found in Zn-Finger-DNA and Other Protein-DNA Complexes *Proc. Natl. Acad. Sci. U.S.A.* **91** 6948-52
[73] Shakked Z, Guzikevich-Guerstein G, Frolow F, Rabinovich D, Joachimiak A and Sigler P B 1994 Determinants of Repressor/Operator Recognition from the Structure of the Trp Operator Binding Site *Nature* **368** 469-73
[74] Guzikevich-Guerstein G and Shakked Z 1996 A Novel Form of the DNA Double Helix Imposed on the TATA-box by the TATA-binding Protein *Nat. Struct. Biol.* **3** 32-7
[75] Olson W K, Gorin A A, Lu X J, Hock L M and Zhurkin V B 1998 DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes *Proc. Natl. Acad. Sci. U.S.A.* **95** 11163-8
[76] Lu X J, Shakked Z and Olson W K 2000 A-Form Conformational Motifs in Ligand-Bound DNA Structures *J. Mol. Biol.* **300** 819-40
[77] Stagg L, Zhang S Q, Cheung M S and Wittung-Stafshede P 2007 Molecular Crowding Enhances Native Structure and Stability of α/β Protein Flavodoxin *Proc. Natl. Acad. Sci. U.S.A.* **104** 18976-81
[78] Minton A P 2000 Implications of Macromolecular Crowding for Protein Assembly *Curr. Opin. Struct. Biol.* **10** 34-9
[79] Gebala M, Bonilla S, Bisaria N and Herschlag D 2016 Does Cation Size Affect Occupancy and Electrostatic Screening of the Nucleic Acid Ion Atmosphere? *J. Am. Chem. Soc.* **138** 10925-34
[80] Gebala M, Giambasu G M, Lipfert J, Bisaria N, Bonilla S, Li G, York D M and Herschlag D 2015 Cation-Anion Interactions within the Nucleic Acid Ion Atmosphere Revealed by Ion Counting *J. Am. Chem. Soc.* **137** 14705-15
[81] Giambasu G M, Luchko T, Herschlag D, York D M and Case D A 2014 Ion Counting from Explicit-Solvent Simulations and 3D-RISM *Biophys. J.* **106** 883-94

638 [82] Allred B E, Gebala M and Herschlag D 2017 Determination of Ion Atmosphere Effects on the
639 Nucleic Acid Electrostatic Potential and Ligand Association Using AH+.C Wobble Formation in
640 Double-Stranded DNA *J. Am. Chem. Soc.* **139** 7540-8

641

642

FIGURE LEGENDS

Figure 1. Representative conformations from clustering simulation snapshots for the Drew-Dickerson (A) and GC-rich (B) dodecamers at 0%, 20%, 30% and 40% protein concentrations. The structures are the structures in each cluster closest to the closest center based on RMSD. Cluster populations are given in parentheses.

Figure 2. Sugar pucker conformations for each base of the Drew-Dickerson dodecamer (A) and the GC-rich dodecamer (B) from simulations at different protein concentrations.

Figure 3. Potentials of mean force (kcal/mol) as a function of δ and χ backbone angles (see Fig. S1) for the Drew-Dickerson dodecamer at 0% (A), 20% (B), 30 % (C) and 40% (D) protein concentrations, and for the GC-rich dodecamer at 0% (E), 20% (F), 30% (G) and 40% (H) protein concentrations from the simulations. See Table S5 for uncertainties.

Figure 4. Average minimum heavy atom distances between crowder protein residues and DNA major groove, minor groove, sugar and phosphate backbone for the Drew-Dickerson dodecamer (left) and the GC-rich dodecamer (right) at different protein concentrations. The secondary structure of protein G is indicated on top for reference.

Figure 5. Representative structures showing interactions between crowder proteins and the Drew-Dickerson dodecamer (A, B) and the GC-rich DNA (C, D). Specific interactions are shown between lysines and the DNA phosphate groups (A), between a threonine residue and a ribose sugar (B), between protonated glutamate and a phosphate (C), and between asparagine and the phosphate (D). Interacting residues are shown in licorice (DNA) and ball-and-stick (protein) representation. The chosen structures have minimal distances between the DNA and the protein.

664 **Figure 6.** Potentials of mean force (kcal/mol) as a function of slide (see Fig. S1) and number of
665 protein contacts for the Drew-Dickerson dodecamer at 20% (A), 30% (B), 40% (C) protein
666 concentrations, and for the GC-rich dodecamer at 20% (D), 30% (E), 40% (F) protein
667 concentrations from the simulations. A contact is defined when the minimum distance between the
668 heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed
669 lines indicate the slide and x-displacement values for canonical B- and A-forms, respectively. See
670 Table S5 for uncertainties.

671 **Figure 7.** Potentials of mean force (kcal/mol) as a function of x-displacement (see Fig. S1) and
672 number of protein contacts for the Drew-Dickerson dodecamer at 20% (A), 30% (B), 40% (C)
673 protein concentrations, and for the GC-rich dodecamer at 20% (D), 30% (E), 40% (F) protein
674 concentrations from the simulations. See Table S5 for uncertainties.

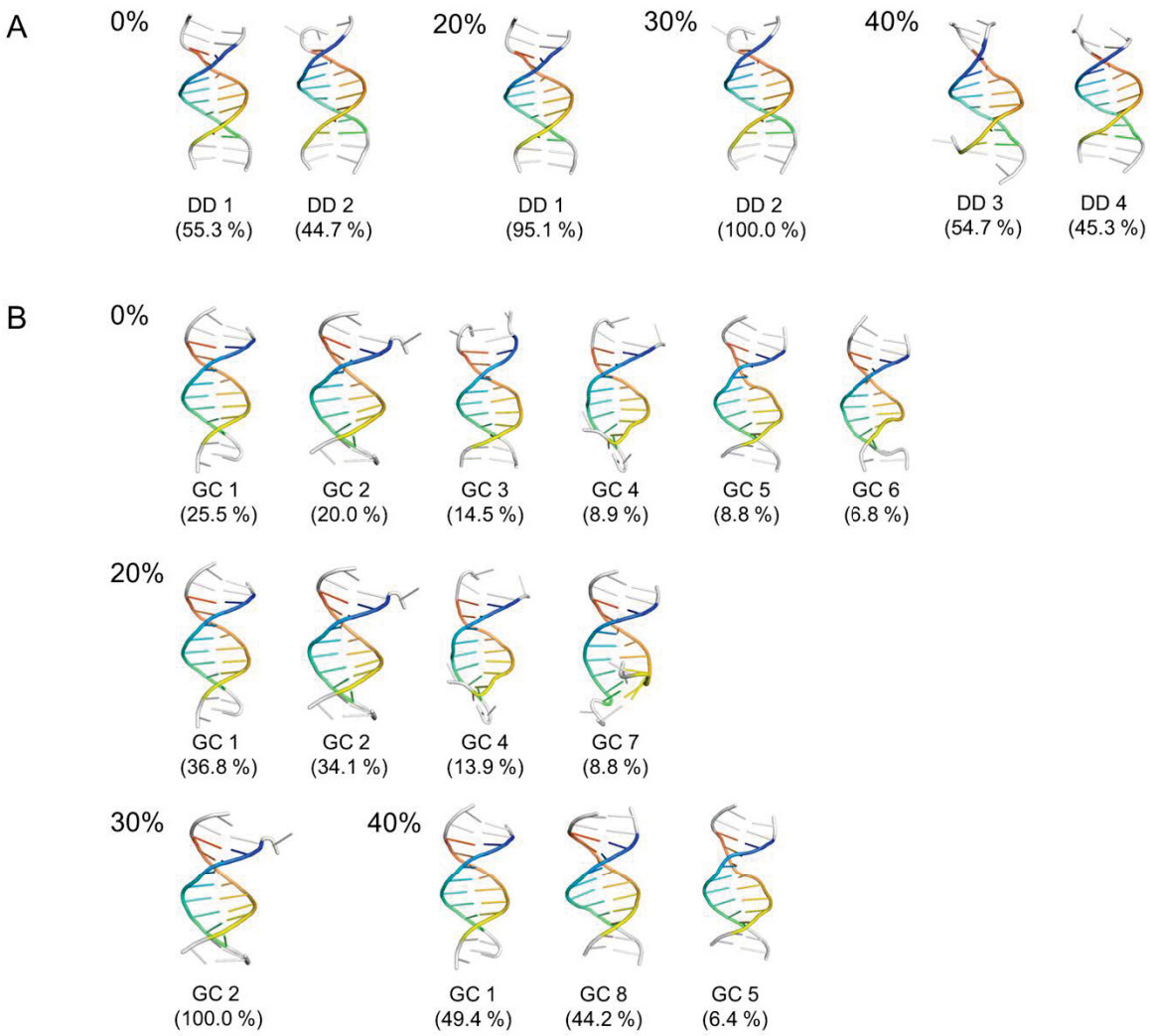
675 **Figure 8.** Potentials of mean force (kcal/mol) as a function of helical rise (see Fig. S1) and number
676 of protein contacts for the Drew-Dickerson dodecamer at 20% (A), 30% (B), 40% (C) protein
677 concentrations, and for the GC-rich dodecamer at 20% (D), 30% (E), 40% (F) protein
678 concentrations from the simulations. See Table S5 for uncertainties.

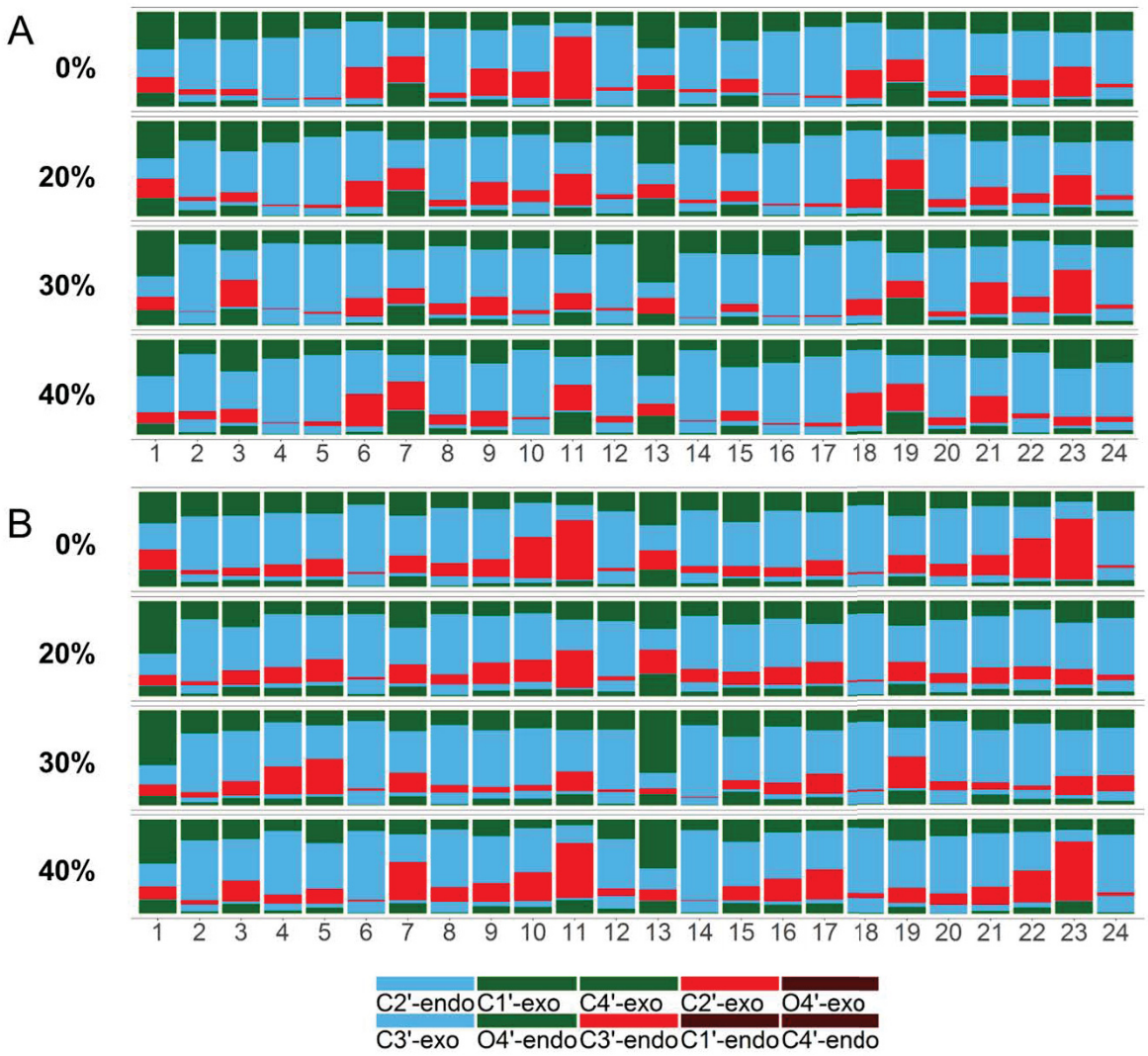
679 **Figure 9.** Potentials of mean force (kcal/mol) as a function of Zp (see Fig. S1) and number of
680 protein contacts for the Drew-Dickerson dodecamer at 20% (A), 30% (B), 40% (C) protein
681 concentrations, and for the GC-rich dodecamer at 20% (D), 30% (E), 40% (F) protein
682 concentrations from the simulations. See Table S5 for uncertainties.

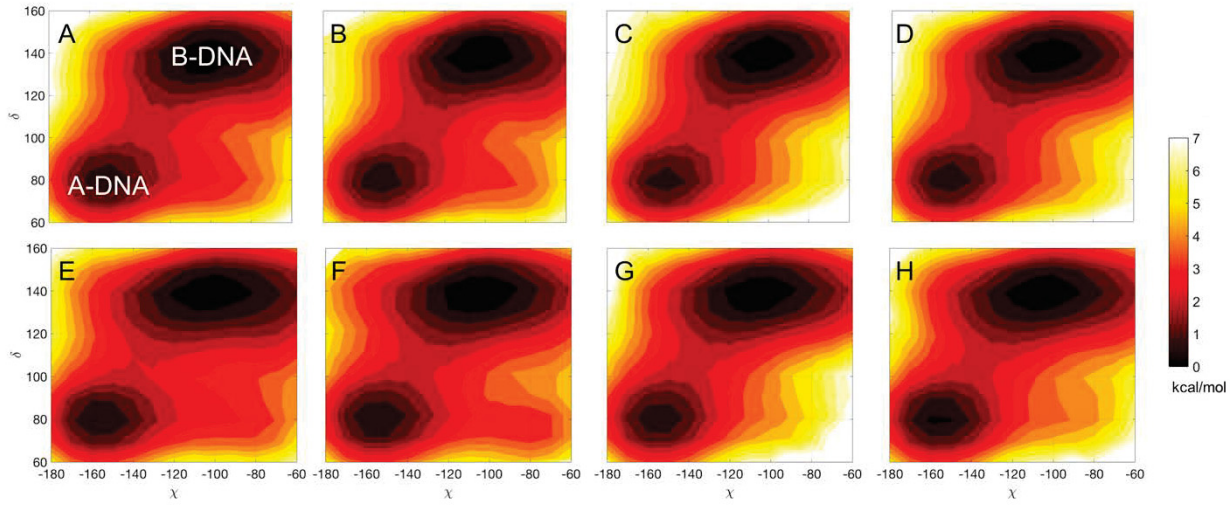
683 **Figure 10.** Radial distribution functions for water (A), sodium ions (B) and DNA neutralization
684 fractions (C) as a function of distance from the closest heavy atoms of the Drew-Dickerson
685 dodecamer (left) and the GC-rich dodecamer (right). Line colors indicate different protein

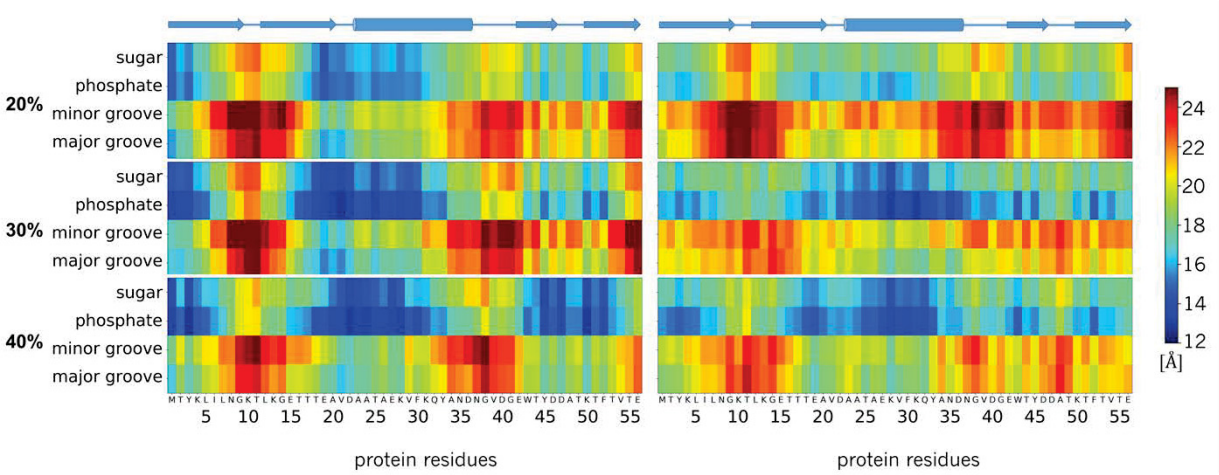
686 concentrations (black: 0%, red: 20%, green: 30%, blue: 40%). The horizontal black line in C
687 indicates the counterion condensation value of 76% of the ions to be condensed on the surface of
688 the DNA. Error bars indicate the calculated standard errors from five independent replicate
689 simulations.

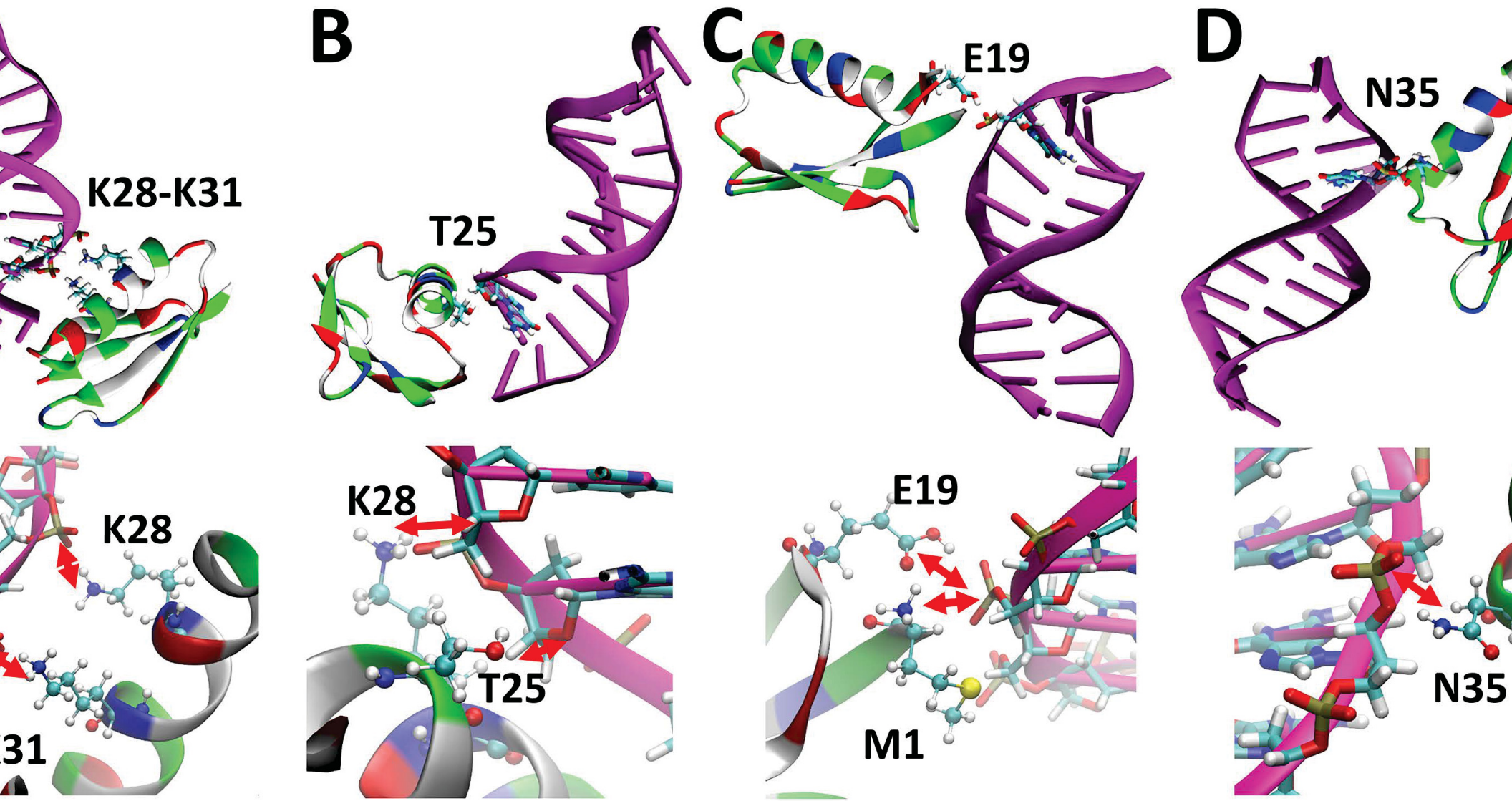
690 **Figure 11.** 3D sodium ion densities around the Drew-Dickerson dodecamer (A) and the GC-rich
691 dodecamer (B) at different protein concentrations. The ion density observed in 0% crowding is
692 shown with a transparent representation for comparison with ion densities (orange) in crowded
693 solutions. Density contours are shown at a level of 0.002 $Å^{-3}$. The top and bottom figures represent
694 front and top views of DNA.

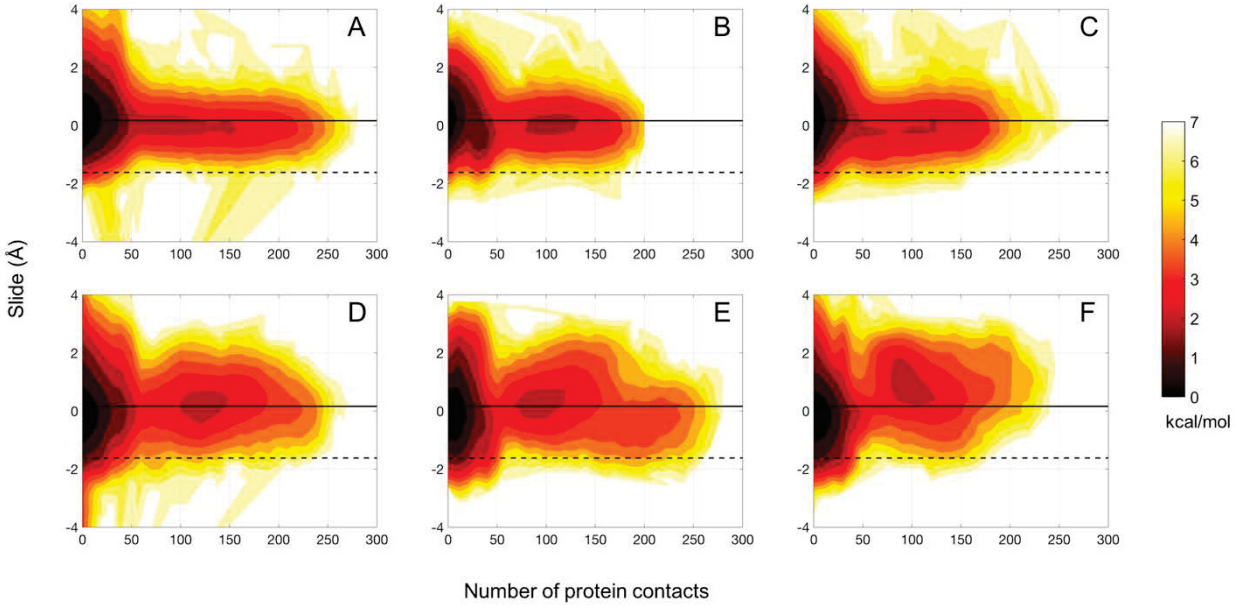

A
0%
DD 1
(55.3 %)
DD 2
(44.7 %)
20%
DD 1
(95.1 %)
30%
DD 2
(100.0 %)
40%
DD 3
(54.7 %)
DD 4
(45.3 %)
B
0%
GC 1
(25.5 %)
GC 2
(20.0 %)
GC 3
(14.5 %)
GC 4
(8.9 %)
GC 5
(8.8 %)
GC 6
(6.8 %)
20%
GC 1
(36.8 %)
GC 2
(34.1 %)
GC 4
(13.9 %)
GC 7
(8.8 %)
30%
GC 2
(100.0 %)
40%
GC 1
(49.4 %)
GC 8
(44.2 %)
GC 5
(6.4 %)

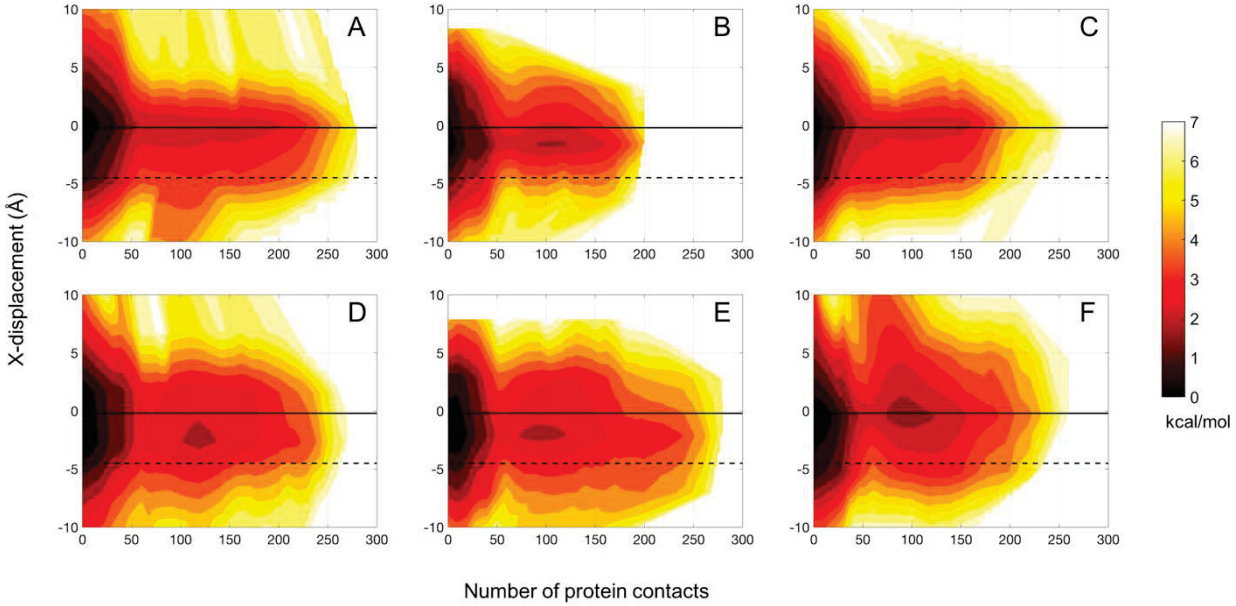

A
0%
20%
30%
40%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
B
0%
20%
30%
40%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
C2'-endo
C1'-exo
C4'-exo
C2'-exo
O4'-exo
C3'-exo
O4'-endo
C3'-endo
C1'-endo
C4'-endo

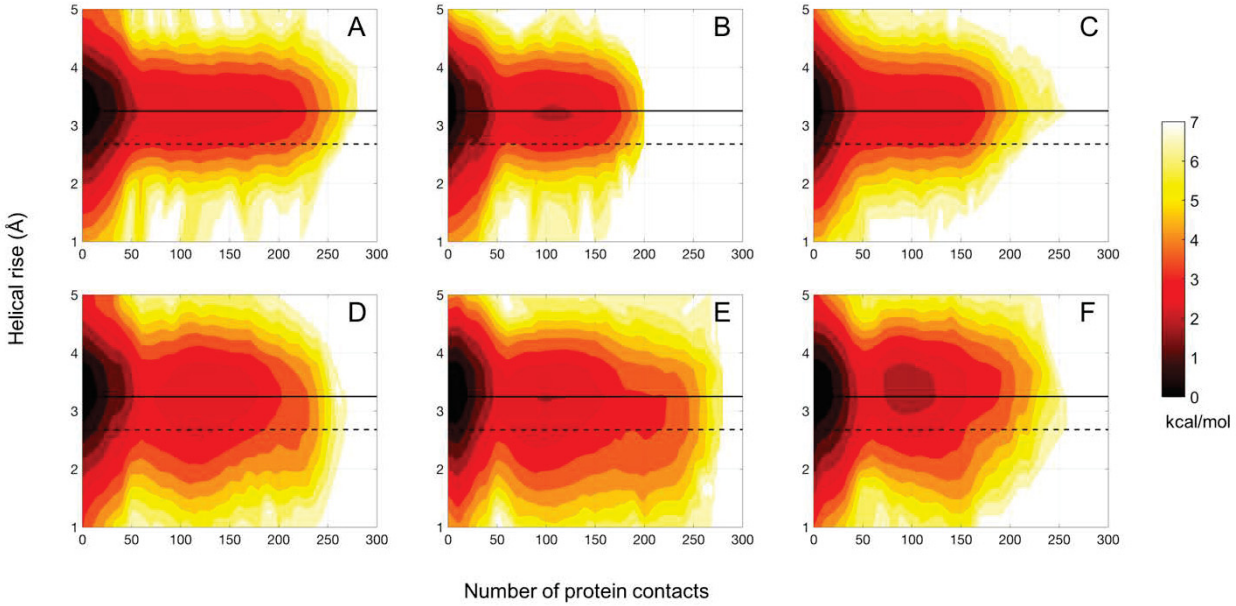

A
B-DNA
A-DNA
B
C
D
E
F
G
H
160
140
120
100
80
60
δ
-180 -160 -140 -120 -100 -80 -60
χ
7
6
5
4
3
2
1
0
kcal/mol

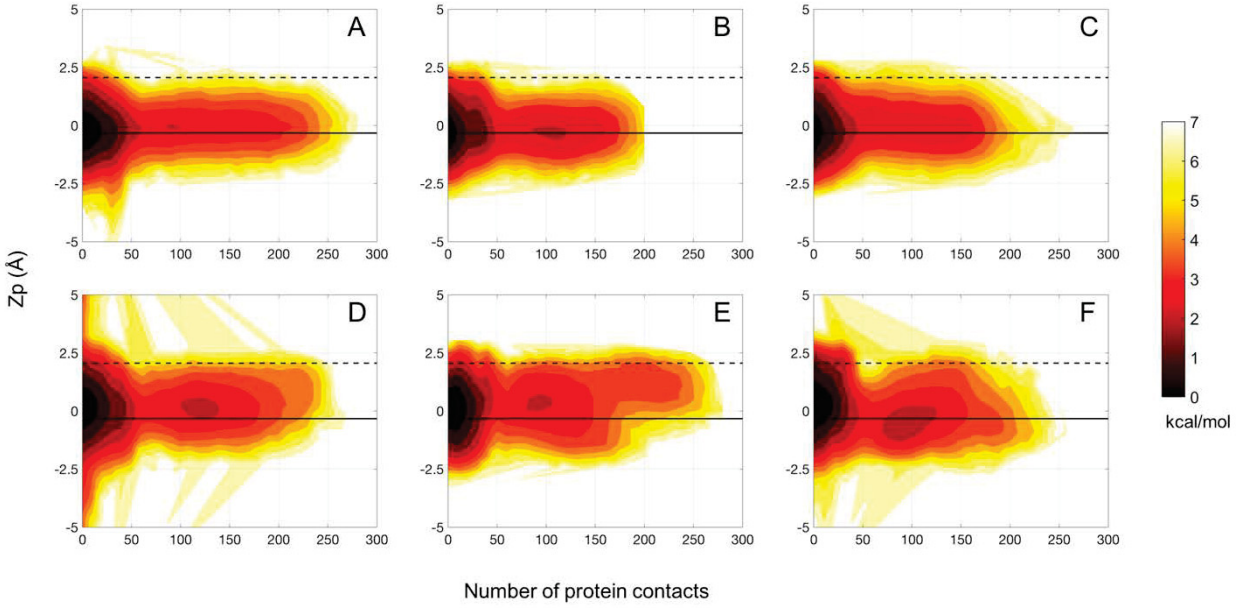

20%
30%
40%
sugar
phosphate
minor groove
major groove
MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE
5 10 15 20 25 30 35 40 45 50 55
protein residues
24
22
20
18
16
14
12
[Å]

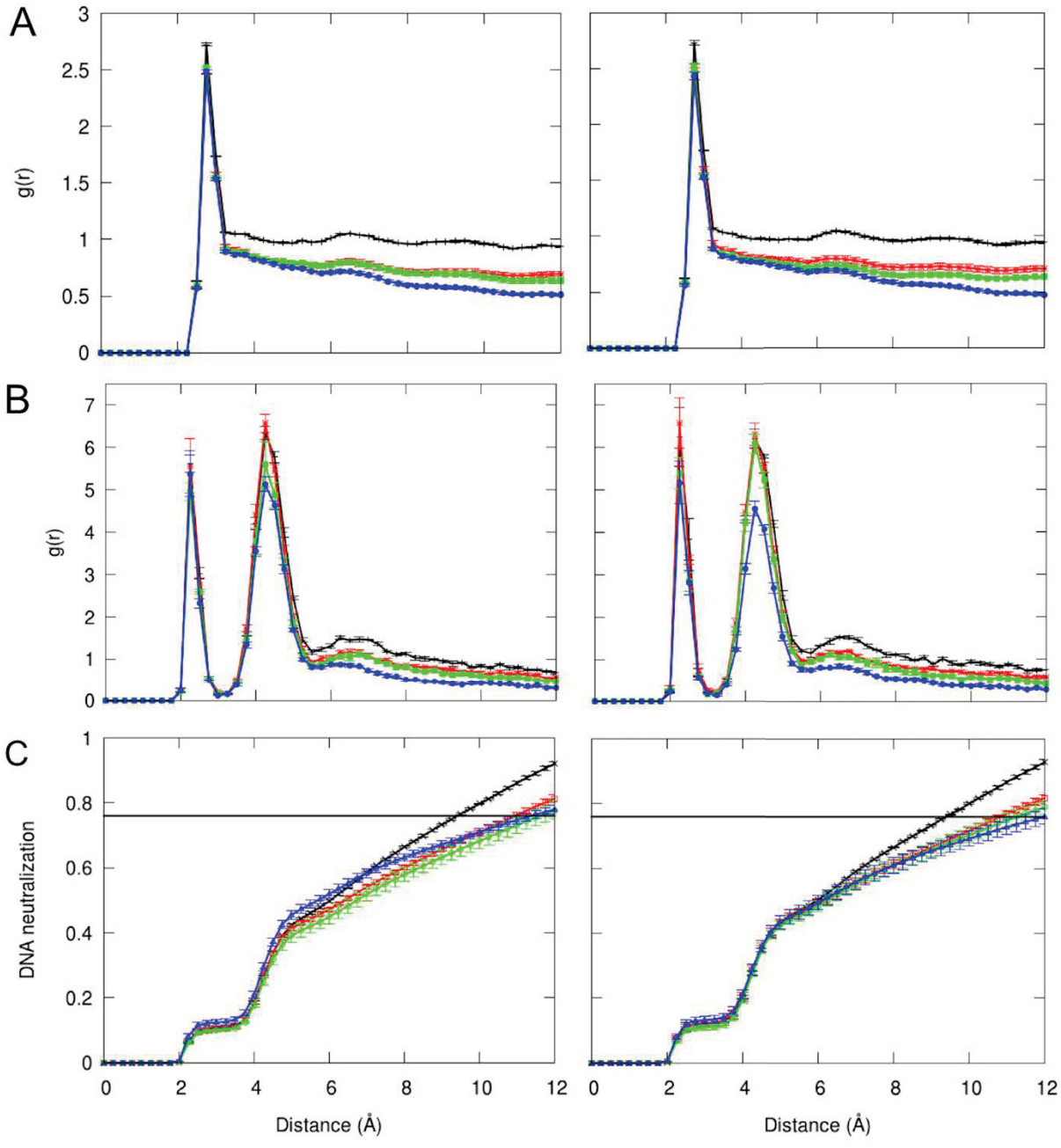

K28-K31
K28
31
B
T25
K28
T25
C
E19
E19
M1
D
N35
N35

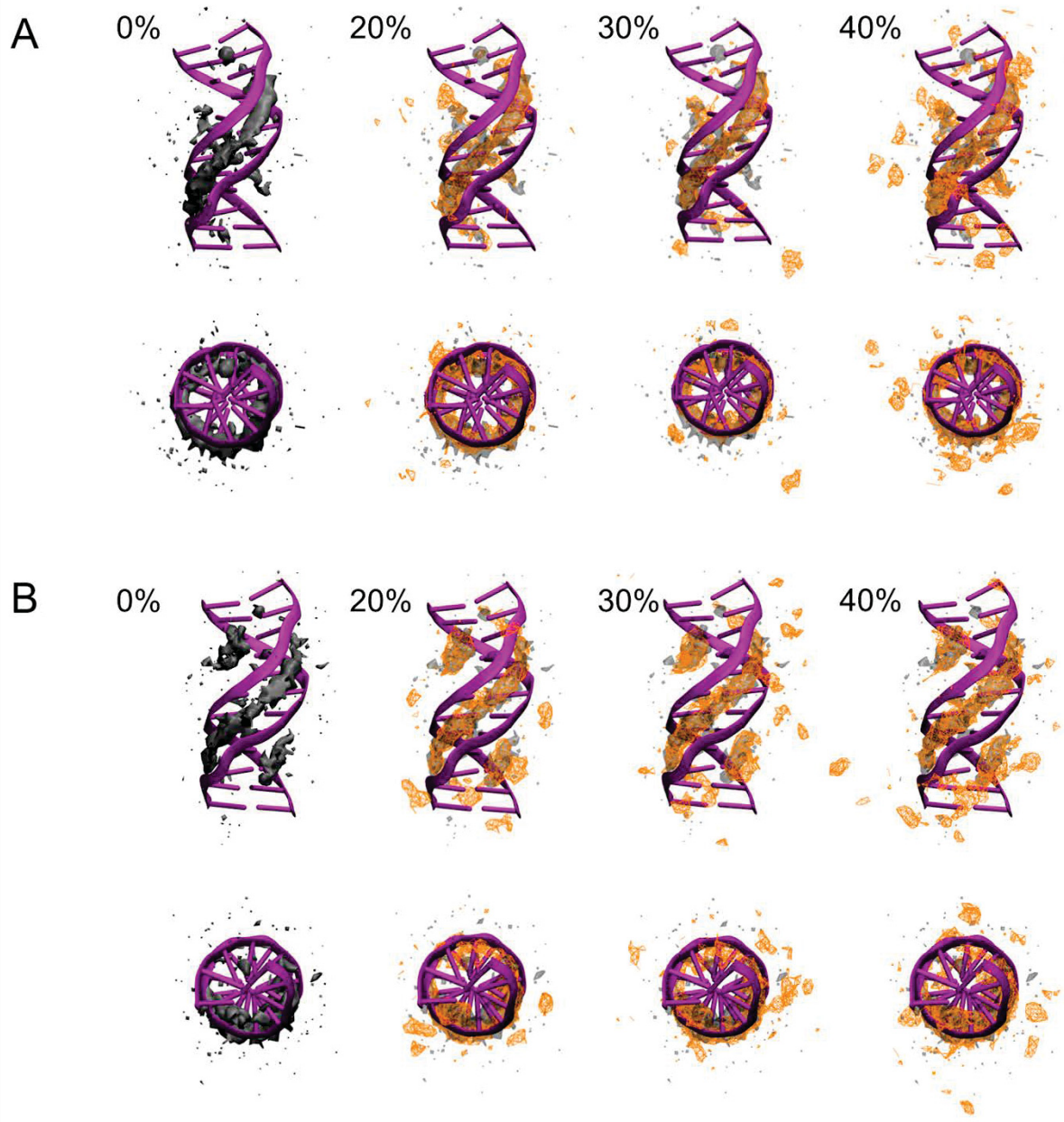

A
B
C
D
E
F
Slide (Å)
Number of protein contacts
kcal/mol
4
2
0
-2
-4
0
50
100
150
200
250
300
7
6
5
4
3
2
1
0


A
B
C
D
E
F
X-displacement (Å)
Number of protein contacts
kcal/mol
10
5
0
-5
-10
0
50
100
150
200
250
300
7
6
5
4
3
2
1
0


A
B
C
D
E
F
Helical rise (Å)
Number of protein contacts
kcal/mol
0
1
2
3
4
5
6
7
50
100
150
200
250
300


A
B
C
D
E
F
Zp (Å)
Number of protein contacts
kcal/mol
5
2.5
0
-2.5
-5
0
50
100
150
200
250
300
7
6
5
4
3
2
1
0


A
B
C
g(r)
g(r)
DNA neutralization
Distance (Å)
Distance (Å)


A
0%
20%
30%
40%
B
0%
20%
30%
40%

Supplementary Information

Role of Protein Interactions in Stabilizing Canonical DNA Features in Simulations of DNA in Crowded Environments

Asli Yildirim[1], Nathalie Brenner[2,3], Robert Sutherland[3], Michael Feig[3,]*

[1]Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA

[2]Faculty of Mathematics and Natural Sciences, University of Düsseldorf, 40225 Düsseldorf, Germany

[3]Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

Corresponding Author Contact Information

[*]603 Wilson Road
Room BCH 218
East Lansing, MI
48824
USA
Phone: 517-432-7439
E-mail: feig@msu.edu

Table S1. Parameters and RMSD values from the canonical B-form structure for the individual clusters of the Drew-Dickerson dodecamer.

	DD1	DD2	DD3	DD4
Slide (Å)	0.07 (*0.00*)	0.34 (*0.00*)	0.19 (*0.00*)	0.38 (*0.00*)
Twist (deg)	33.67 (*0.01*)	34.63 (*0.01*)	33.12 (*0.01*)	32.73 (*0.02*)
X-displacement (Å)	-1.09 (*0.00*)	-0.30 (*0.01*)	-0.94 (*0.01*)	-0.42 (*0.01*)
Helical rise (Å)	3.21 (*0.00*)	3.30 (*0.00*)	3.24 (*0.00*)	3.29 (*0.00*)
Inclination (deg)	12.89 (*0.03*)	8.95 (*0.03*)	13.14 (*0.04*)	11.07 (*0.05*)
z_p (Å)	-0.06 (*0.00*)	-0.32 (*0.00*)	-0.14 (*0.00*)	-0.29 (*0.00*)
Minor groove (Å)	13.63 (*0.01*)	12.77 (*0.01*)	13.86 (*0.01*)	13.69 (*0.01*)
Major groove (Å)	16.10 (*0.01*)	16.52 (*0.01*)	16.27 (*0.01*)	16.97 (*0.01*)
RMSD (Å)	1.82 (*0.00*)	1.37 (*0.00*)	2.00 (*0.00*)	1.70 (*0.00*)

All values are averaged over all base-pairs excluding the first and last two terminal base-pairs with standard errors given in the parentheses.

Table S2. Parameters and RMSD values from the canonical B-form structure for the individual clusters of the GC-rich dodecamer.

	GC1	**GC2**	**GC3**	**GC4**
Slide (Å)	0.12 (*0.00*)	-0.30 (*0.00*)	0.69 (*0.01*)	0.44 (*0.01*)
Twist (deg)	34.36 (*0.01*)	33.13 (*0.01)*	34.63 (*0.03*)	33.02 (*0.03*)
X-displacement (Å)	-0.67 (*0.01*)	-1.69 (*0.01*)	0.53 (*0.01*)	-0.20 (*0.01*)
Helical rise (Å)	3.30 (*0.00*)	3.20 (*0.00)*	3.37 (*0.00*)	3.33 (*0.00*)
Inclination (deg)	8.80 (*0.03*)	12.12 (*0.03*)	5.90 (*0.08*)	9.42 (*0.07*)
z_p (Å)	0.05 (*0.00*)	0.34 (*0.00*)	-0.21 (*0.01*)	-0.08 (*0.01*)
Minor groove (Å)	14.05 (*0.01*)	14.55 (*0.01*)	14.00 (*0.02*)	14.78 (*0.01*)
Major groove (Å)	16.34 (*0.01*)	16.13 (*0.01*)	16.47 (*0.01*)	16.82 (*0.02*)
RMSD (Å)	1.62 (*0.00*)	2.28 (*0.00*)	1.68 (*0.01*)	2.05 (*0.01*)
	GC5	**GC6**	**GC7**	**GC8**
Slide (Å)	0.44 (*0.01*)	-0.06 (*0.01*)	-0.11 (*0.02*)	0.04 (*0.00*)
Twist (deg)	30.98 (*0.06*)	31.97 (*0.03*)	30.90 (*0.41*)	32.72 (*0.02*)
X-displacement (Å)	-0.55 (*0.02*)	-1.41 (*0.01*)	-0.94 (*0.03*)	-1.19 (*0.01*)
Helical rise (Å)	3.36 (*0.01*)	3.24 (*0.00*)	3.07 (*0.02*)	3.25 (*0.00*)
Inclination (deg)	12.55 (*0.14*)	13.05 (*0.06*)	12.00 (*0.30*)	12.99 (*0.04*)
z_p (Å)	0.03 (*0.01*)	0.19 (*0.01*)	-0.20 (*0.03*)	0.17 (*0.01*)
Minor groove (Å)	15.68 (*0.02*)	15.11 (*0.01*)	14.94 (*0.03*)	15.09 (*0.01*)
Major groove (Å)	16.82 (*0.02*)	16.69 (*0.02*)	15.47 (*0.05*)	16.25 (*0.01*)
RMSD (Å)	2.58 (*0.01*)	2.47 (*0.01*)	3.70 (*0.02*)	2.25 (*0.00*)

All values are averaged over all base-pairs excluding the first and last two terminal base-pairs with standard errors given in the parentheses.

Table S3. Bending angles for both DNA dodecamers (degrees).

		Canonical		Simulations in crowded environment			
	X-ray	A-DNA	B-DNA	0%	20%	30%	40%
Drew-Dickerson	170.11	122.72 (*3.57*)	163.05 (*5.94*)	155.24 (*0.47*)	155.55 (*0.65*)	155.50 (*0.76*)	152.80 (*0.87*)
GC-rich	100.77	122.72 (*3.57*)	163.05 (*5.94*)	156.85 (*1.09*)	153.00 (*1.12*)	153.03 (*0.91*)	152.18 (*1.04*)

Bending angle is defined as the angle between the center of masses of three sections of base-pairs: Section 1: 3 – 5, Section 2: 6 – 7, and Section 3: 8 – 10. Standard errors given in the parentheses. Statistical errors of the averages over the simulations are estimated from block averaging by comparing results for 100 ns segments from the simulations. Canonical values are averaged over the A-form structures 3V9D, 3QK4, 2B1B, 1ZEX, 1ZEY, 1ZF1, 1ZF8, 1ZF9, 1ZFA and the B-form structures 2M2C, 4AGZ, 4H0, 4AH1, 3U05, 3U08, 1VTJ,3U2N, 3OIE, 3BSE. For X-ray structure values; 1BNA and 399D are used for Drew-Dickerson and GC-rich dodecamers, respectively.

Table S4. Conformational clustering of protein G crowder molecules

		Cluster	Population	Cluster center		Avg. minimum DNA distance [Å]
				Cα RMSD [Å]	Radius of gyration [Å]	
Drew-Dickerson	20%	1	76.9 %	0.57	10.53	7.9
		2	21.6 %	0.73	10.67	6.9
		3	1.2 %	1.18	10.61	4.2
		4	0.3%	2.32	10.49	4.3
	30%	1	68.8 %	0.57	10.57	7.2
		2	28.1 %	0.76	10.56	7.1
		3	3.1 %	1.45	10.71	5.3
	40%	1	63.1 %	0.63	10.59	6.4
		2	36.9 %	0.91	10.76	6.5
GC-rich	20%	1	95.6 %	0.64	10.59	8.0
		2	4.4 %	2.07	10.62	3.8
	30%	1	100 %	0.65	10.63	6.0
	40%	1	94.2 %	0.59	10.59	6.3
		2	5.8 %	1.14	10.70	3.7

Clustering analysis of snapshots extracted at 10 ns intervals via the kclust program of the MMTSB Tool Set based on Cα atoms using a clustering radius of 2 Å. Cα RMSD is with respect to experimental structure (3GB1). Minimum DNA distance is calculated between heavy atoms of protein and DNA.

Table S5. PMF error analysis

	Figure	Drew-Dickerson			GC-rich		
		20%	30%	40%	20%	30%	40%
slide	6	0.21 [-0.8; 0.5]	0.30 [-0.5; 0.8]	0.34 [-0.8; 0.7]	0.39 [-0.6; 1.3]	0.41 [-1.3; 1.0]	0.41 [-1.3; 1.0]
x-displacement	7	0.20 [-0.6; 0.3]	0.25 [-0.2; 0.7]	0.32 [-0.6; 0.5]	0.41 [-0.6; 1.2]	0.40 [-1.1; 1.0]	0.39 [-0.7; 1.3]
helical rise	8	0.20 [-0.4; 0.4]	0.22 [-0.5; 0.6]	0.32 [-0.7; 0.6]	0.35 [-0.7; 1.0]	0.30 [-0.9; 0.9]	0.40 [-0.7; 1.5]
z_p	9	0.22 [-0.9; 0.6]	0.23 [-0.5; 0.6]	0.31 [-0.8; 0.6]	0.42 [-1.1; 1.2]	0.38 [-1.0; 1.2]	0.41 [-0.8; 1.4]
twist	S7	0.22 [-0.4; 0.8]	0.22 [-0.7; 0.6]	0.32 [-0.7; 0.7]	0.36 [-0.7; 1.0]	0.26 [-0.7; 0.8]	0.39 [-0.3; 1.3]
inclination	S8	0.23 [-0.8; 0.6]	0.29 [-0.8; 0.9]	0.34 [-1.1; 0.8]	0.38 [-0.8; 1.2]	0.34 [-1.2; 1.1]	0.37 [-0.6; 1.4]
minor groove	S9	0.41 [-1.3; 1.3]	0.35 [-0.6; 1.2]	0.42 [-1.0; 1.1]	0.43 [-1.2; 1.6]	0.43 [-1.5; 1.4]	0.45 [-1.2; 1.4]
major groove	S10	0.27 [-0.9; 0.9]	0.38 [-1.3; 1.3]	0.41 [-1.3; 1.1]	0.46 [-1.3; 2.0]	0.41 [-1.1; 1.5]	0.50 [-1.1; 1.6]
α	S11	0.22 [-0.7; 0.8]	0.23 [-0.3; 0.8]	0.31 [-0.7; 0.7]	0.35 [-0.8; 1.2]	0.25 [-0.6; 0.8]	0.55 [-0.5; 2.0]
β	S12	0.22 [-0.5; 0.9]	0.22 [-0.5; 0.7]	0.34 [-0.9; 0.7]	0.42 [-0.8; 1.7]	0.31 [-0.9; 1.0]	0.49 [-0.8; 2.2]
γ	S13	0.20 [-0.2; 0.5]	0.19 [-0.1; 0.5]	0.33 [-0.9; 0.6]	0.36 [-0.7; 1.3]	0.27 [-0.8; 0.7]	0.53 [-0.3; 2.1]
δ	S14	0.22 [-0.3; 0.6]	0.23 [-0.4; 0.8]	0.33 [-0.8; 0.7]	0.37 [-0.7; 1.2]	0.30 [-0.8; 0.7]	0.38 [-0.5; 1.5]
ε	S15	0.21 [-0.6; 0.6]	0.26 [-0.6; 0.8]	0.31 [-0.7; 0.6]	0.33 [-0.8; 1.4]	0.29 [-1.1; 0.6]	0.29 [-0.5; 1.2]
ζ	S16	0.25 [-1.0; 0.8]	0.30 [-0.6; 0.8]	0.29 [-0.9; 0.7]	0.39 [-1.0; 1.3]	0.30 [-0.9; 0.9]	0.35 [-1.0; 1.2]
χ	S17	0.25 [-0.9; 1.0]	0.26 [-0.2; 0.8]	0.32 [-0.7; 0.7]	0.40 [-0.6; 1.6]	0.37 [-1.1; 0.9]	0.39 [-1.0; 1.2]
pucker	S18	0.26 [-0.6; 0.9]	0.29 [-0.6; 1.1]	0.33 [-1.0; 0.6]	0.38 [-1.0; 1.2]	0.31 [-0.9; 0.9]	0.40 [-0.9; 1.5]
δ/χ	3	0.16 [-0.6; 0.1]	0.14 [-0.2; 0.4]	0.16 [-0.0; 0.4]	0.29 [-0.2; 1.1]	0.24 [-0.6; 0.3]	0.19 [-0.4; 0.5]
ε/ζ	S3	0.19 [-0.9; 0.5]	0.24 [-0.4; 1.4]	0.22 [-0.3; 0.7]	0.20 [-0.7; 0.7]	0.24 [-0.7; 0.2]	0.22 [-0.7; 1.0]

Errors in kcal/mol estimated from comparing PMFs generated from first (300-650 ns) and second (650-1000 ns) halves of production trajectories. The first value corresponds to the root mean square deviation across all PMF grid points where PMF energies are less than 4 kcal/mol. Values in square brackets are minimum and maximum deviations.

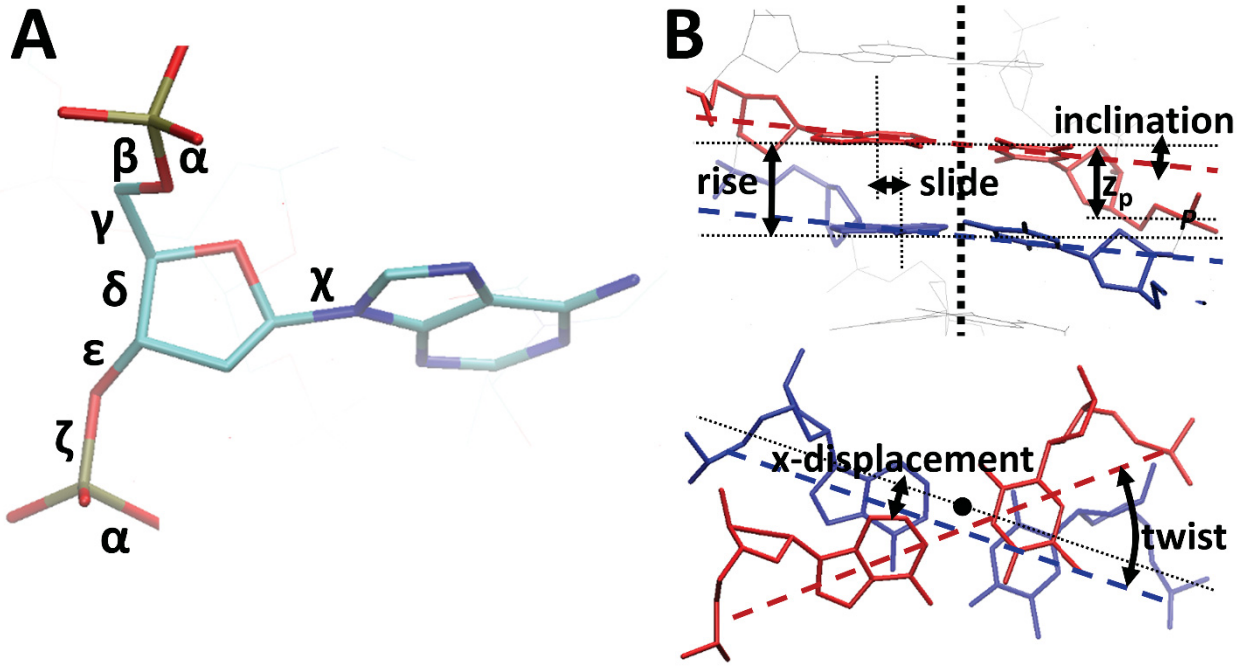


Figure S1. Definitions of DNA backbone torsion angles (A) and helical base pair parameters discussed in this study (B).

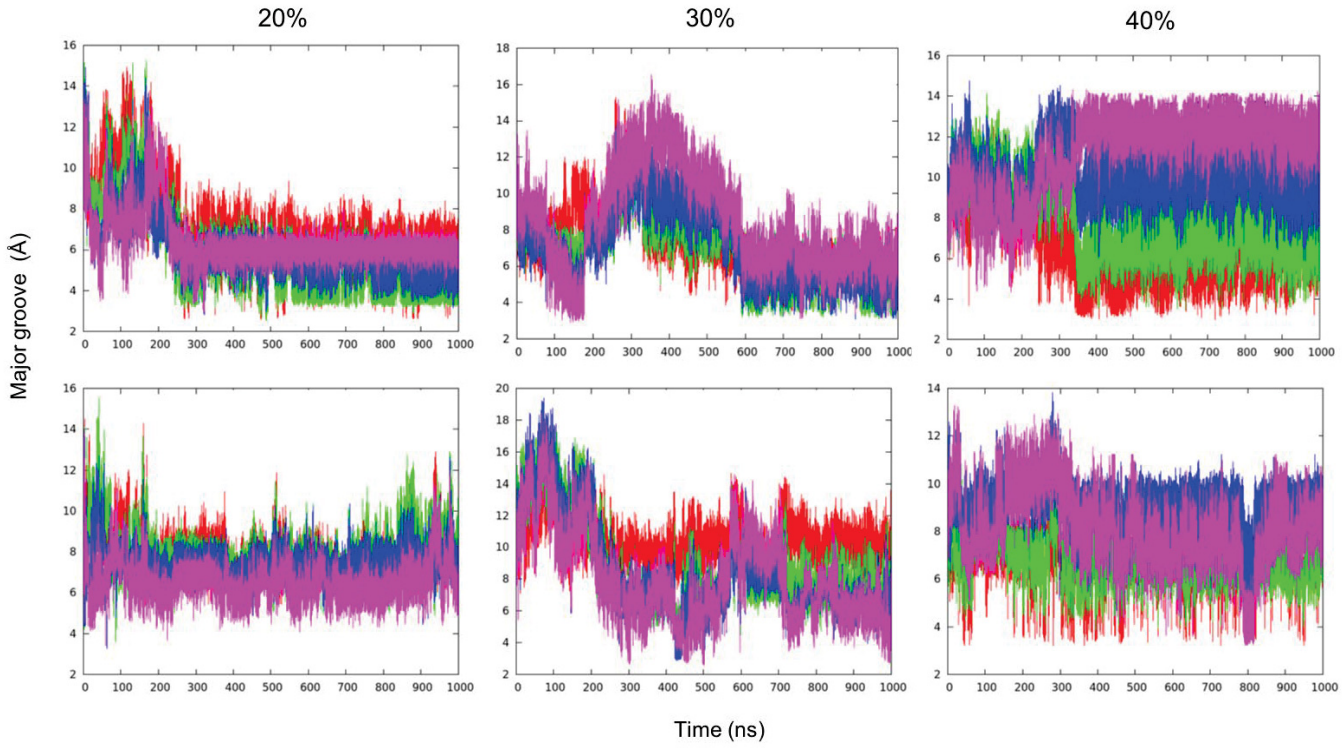


Figure S2. Time series of major groove widths for inner 5th (red), 6th (green), 7th (blue) and 8th (pink) basepairs for the Drew-Dickerson (top) and GC-rich dodecamers in crowded simulations.

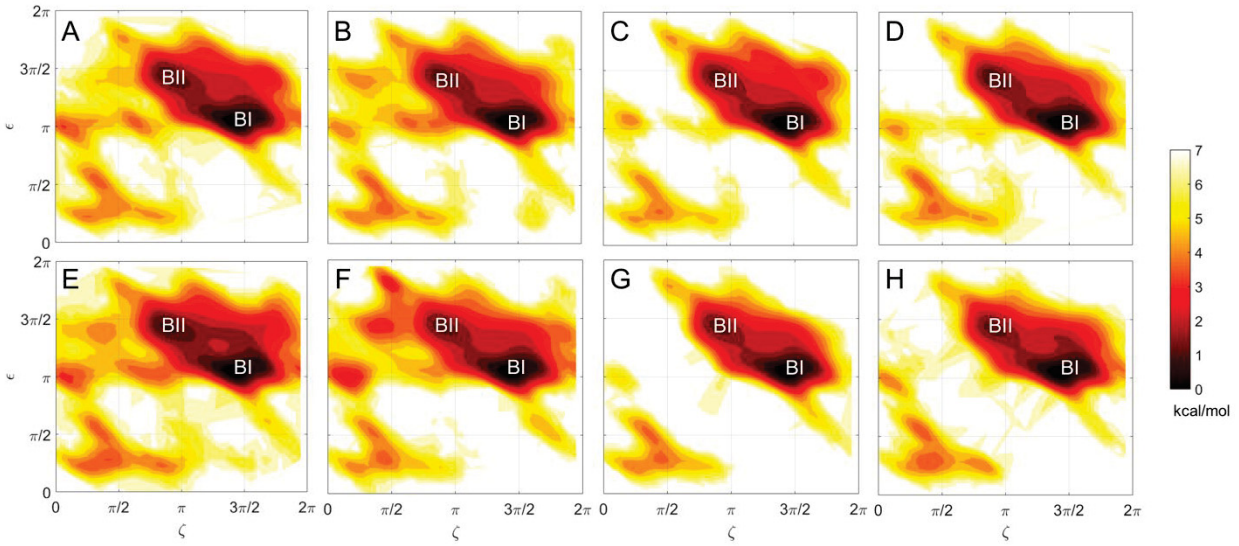


Figure S3. Potential of mean force (kcal/mol) as a function of ε and ξ backbone angles for the Drew-Dickerson dodecamer at 0 % (A), 20 % (B), 30 % (C) and 40 % (D) protein concentrations, and for the GC-rich dodecamer at 0 % (E), 20 % (F), 30 % (G) and 40 % (H) protein concentrations. See Table S5 for uncertainties.

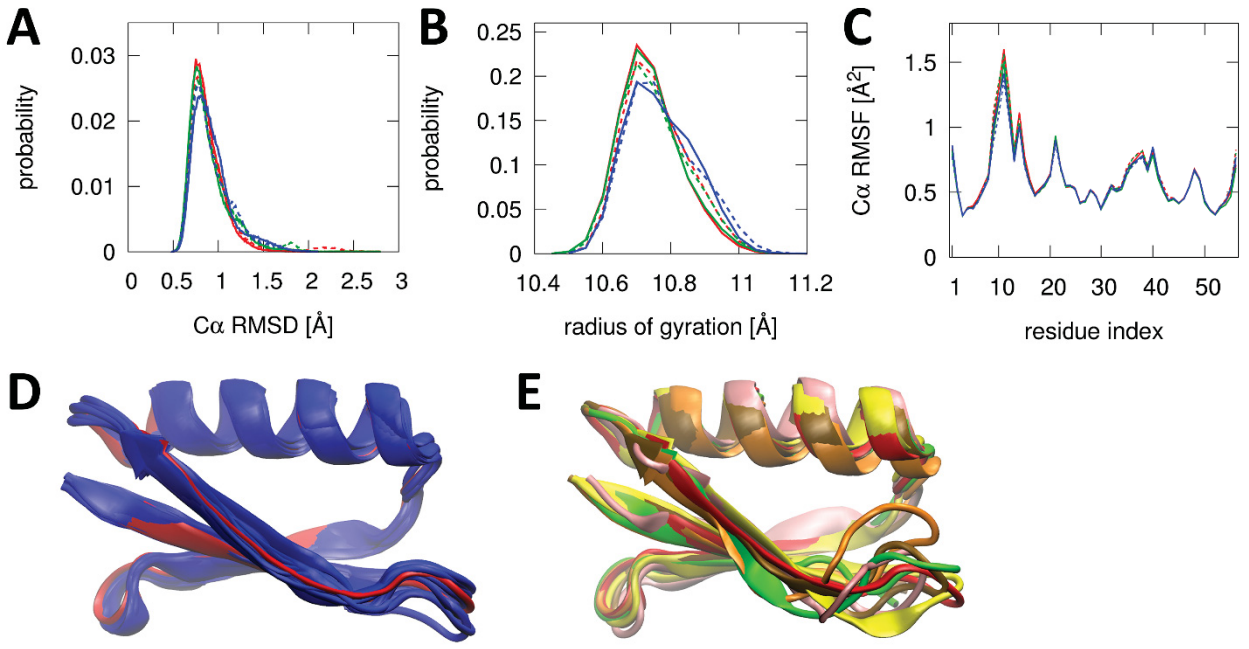


Figure S4. Conformational analysis of protein G crowders for 20% (red), 30% (green), and 40% (blue) crowder concentrations around the Drew-Dickerson dodecamer (solid lines) and the GC-rich DNA (dashed lines) (A-C). Distribution of Cα RMSD with respect to experimental structure (3GB1) (A); distribution of radius of gyration (B); averaged root mean square fluctuations for Cα atoms as a function of protein G residue. Representative structures from cluster analysis for highly populated clusters (blue, D) and minor substates (colors, E), compared with experimental structure (red).

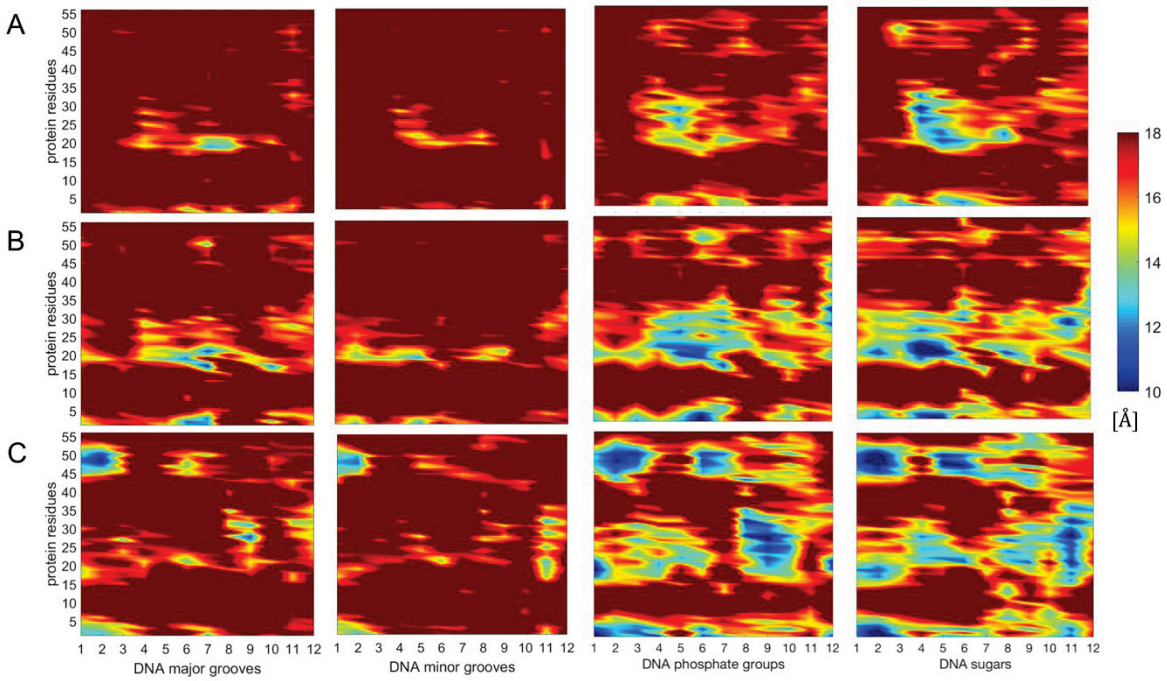


Figure S5. Average minimum distances between the crowder protein residues and the major groove, minor groove, sugar and phosphate backbone for the individual base-pairs of Drew-Dickerson dodecamer at 20% (A), 30% (B) and 40% (C) protein concentrations. Distances were calculated from the heavy atoms only.

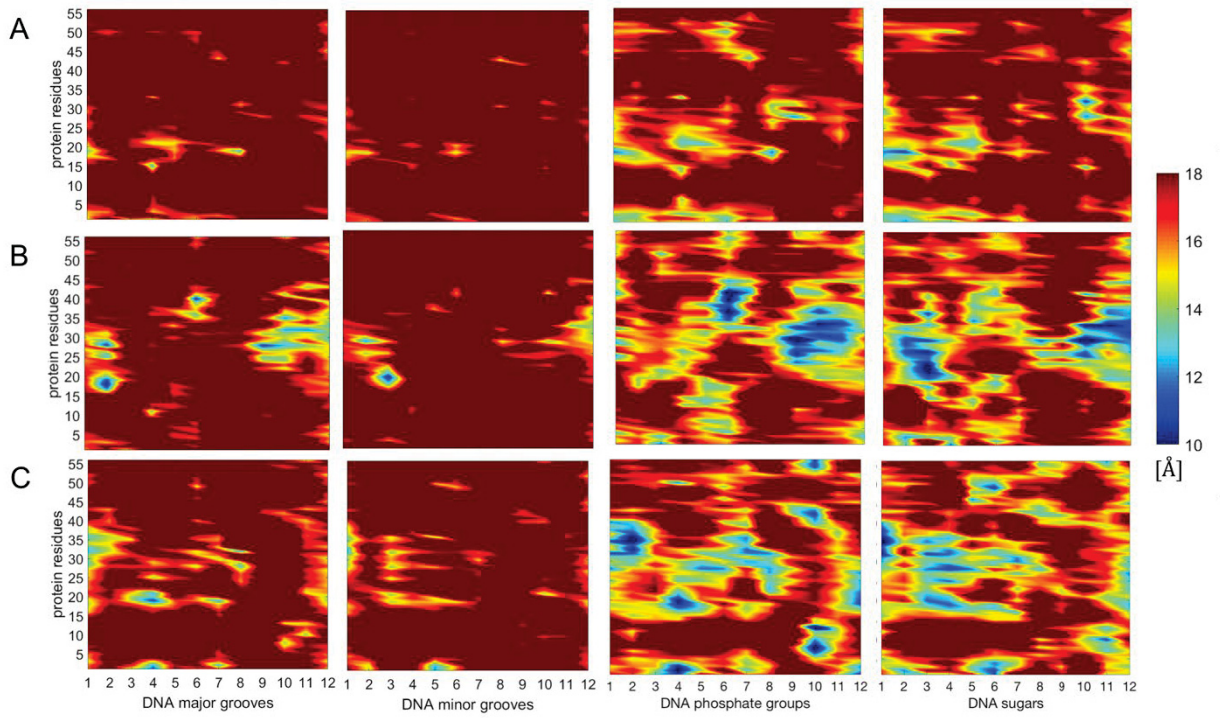


Figure S6. Average minimum distances between the crowder protein residues and the major grooves, minor grooves, sugar and phosphate backbone for the individual base-pairs of GC-rich dodecamer at 20% (A), 30% (B) and 40% (C) protein concentrations. Distances were calculated from the heavy atoms only.

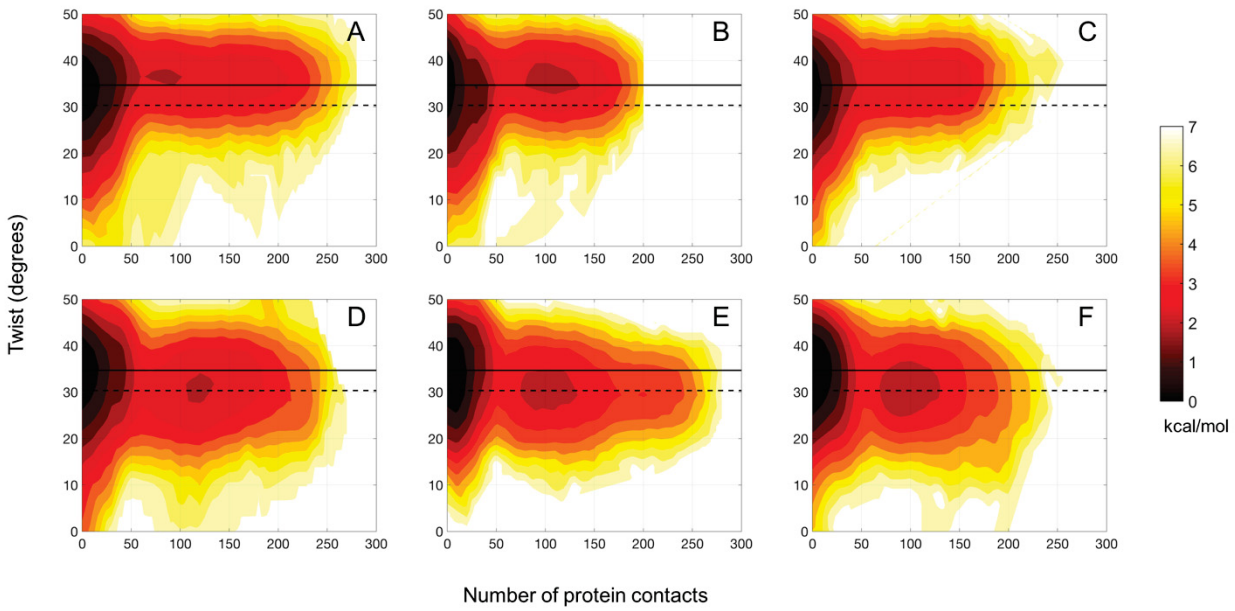


Figure S7. Potential of mean force (kcal/mol) as a function of twist angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

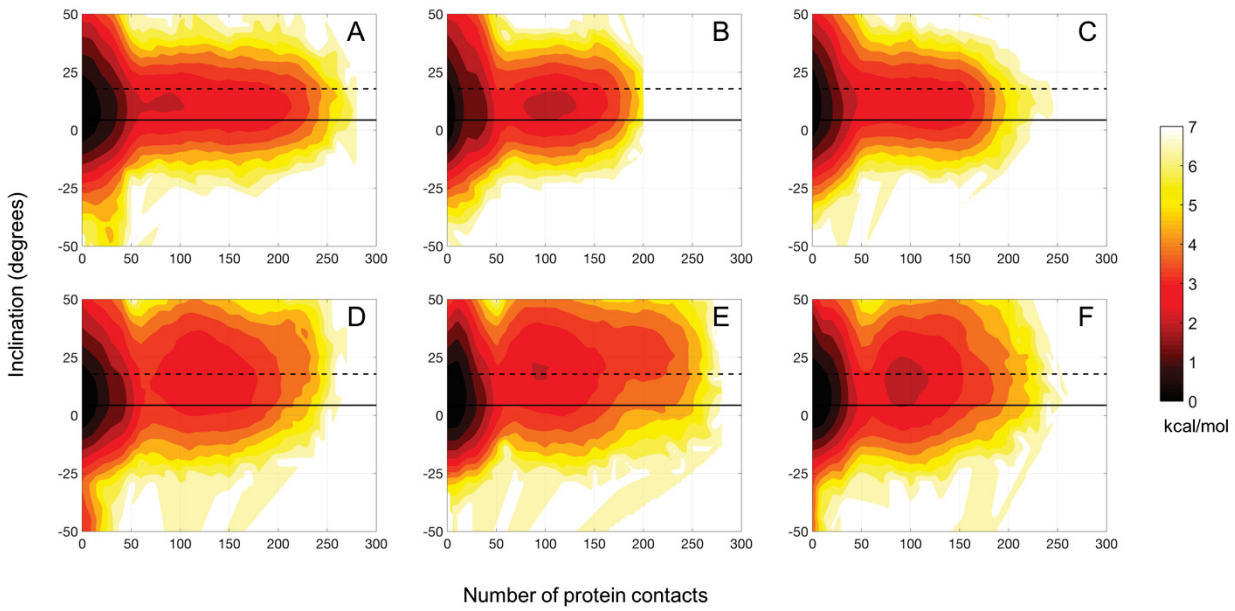


Figure S8. Potential of mean force (kcal/mol) as a function of inclination angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

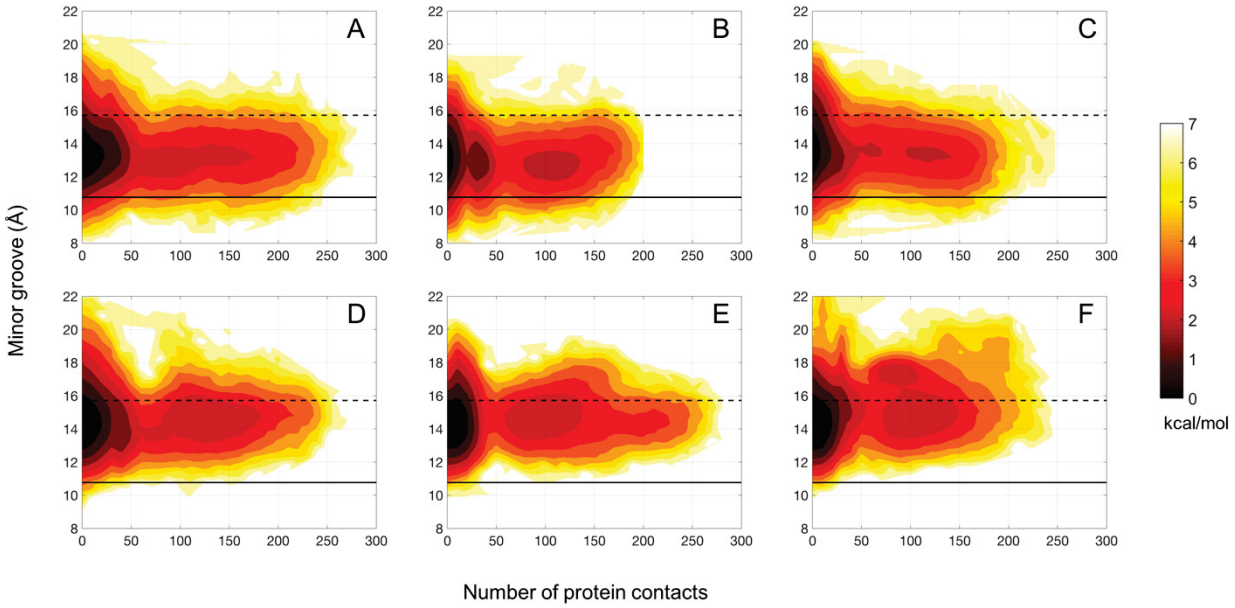


Figure S9. Potential of mean force (kcal/mol) as a function of minor groove width and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

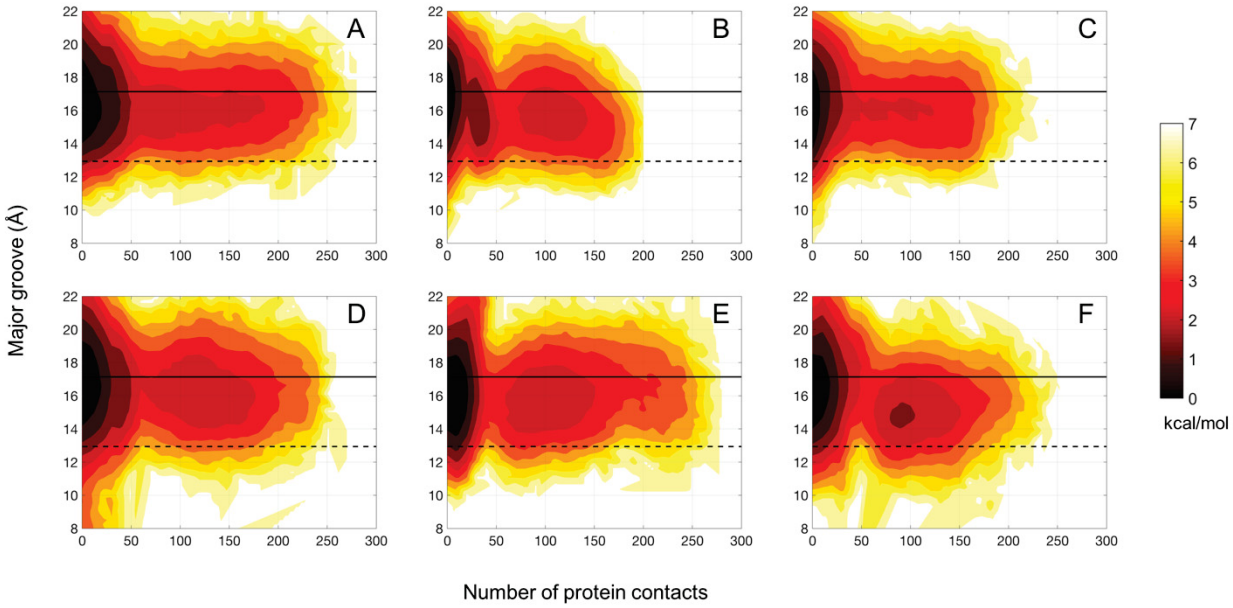


Figure S10. Potential of mean force (kcal/mol) as a function of major groove width and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

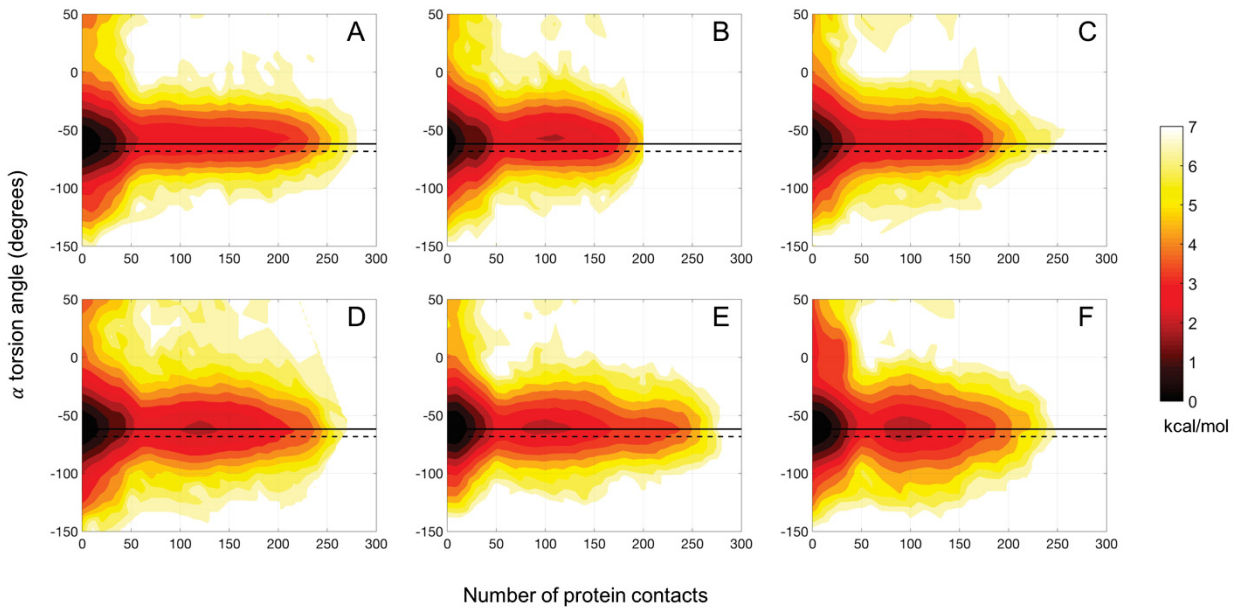


Figure S11. Potential of mean force (kcal/mol) as a function of α backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

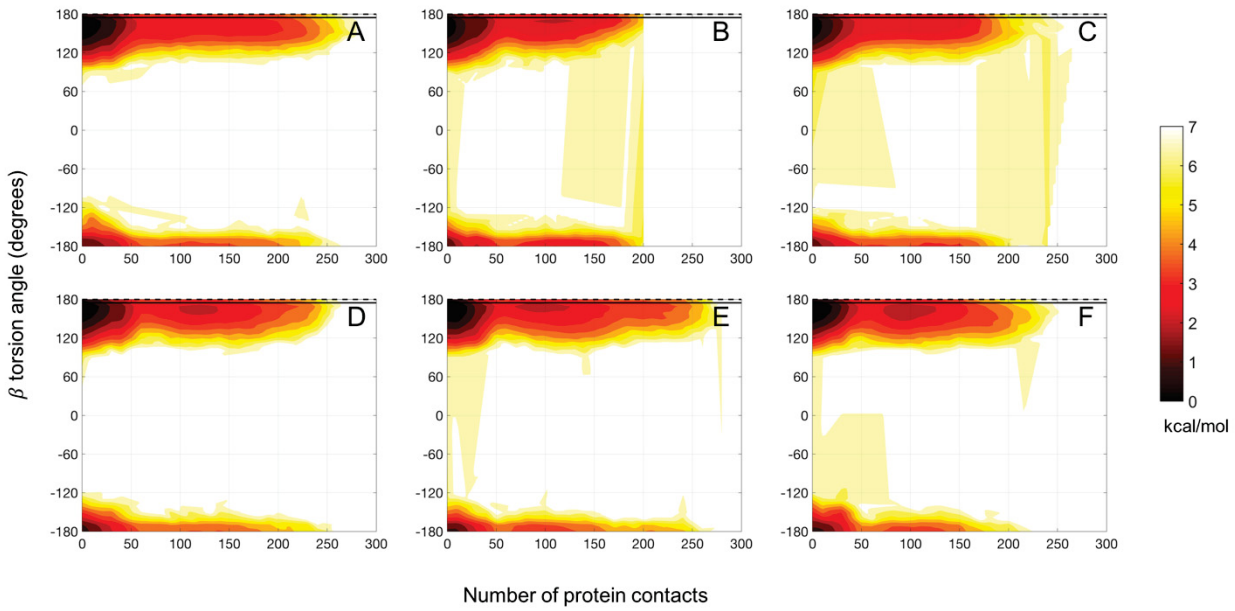


Figure S12. Potential of mean force (kcal/mol) as a function of β backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

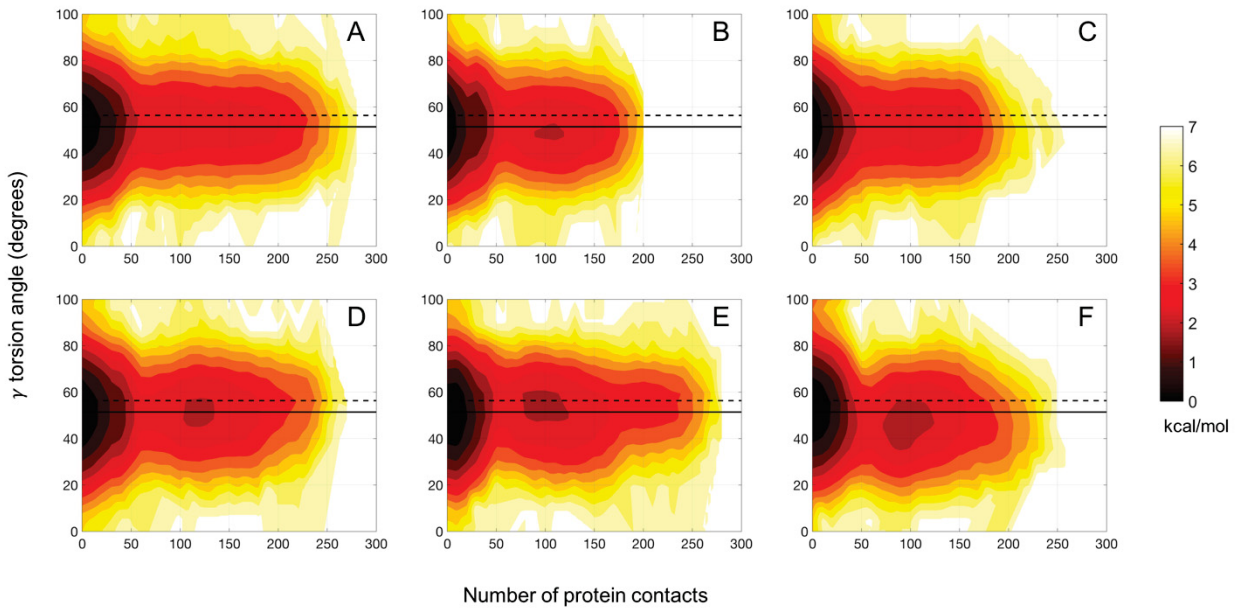


Figure S13. Potential of mean force (kcal/mol) as a function of γ backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

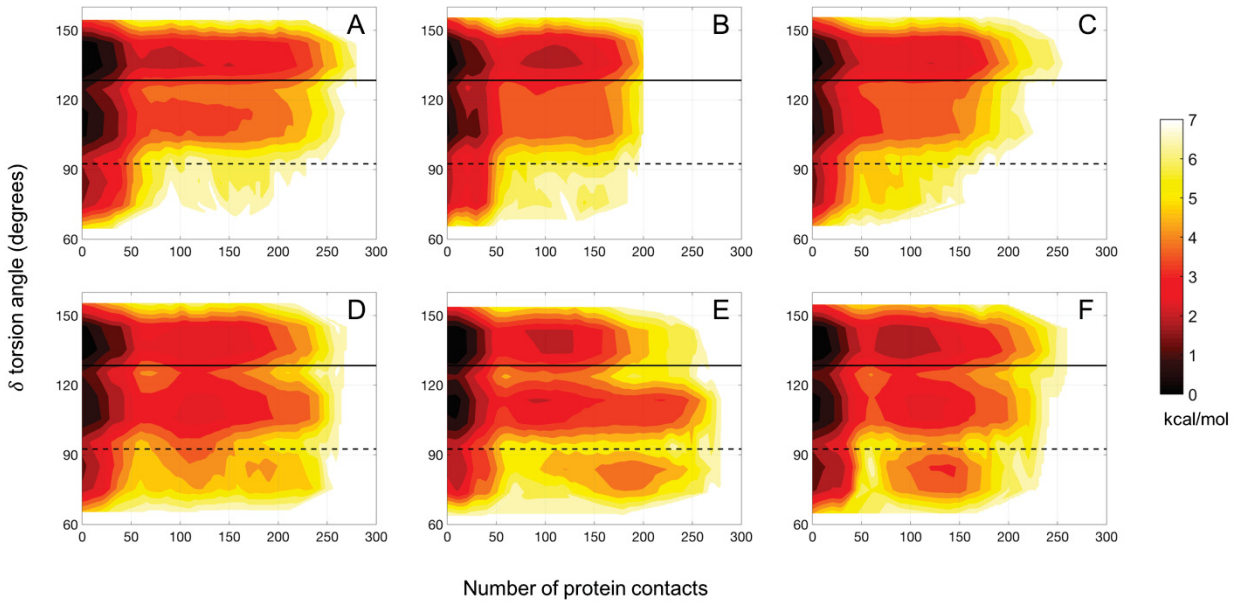


Figure S14. Potential of mean force (kcal/mol) as a function of δ backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

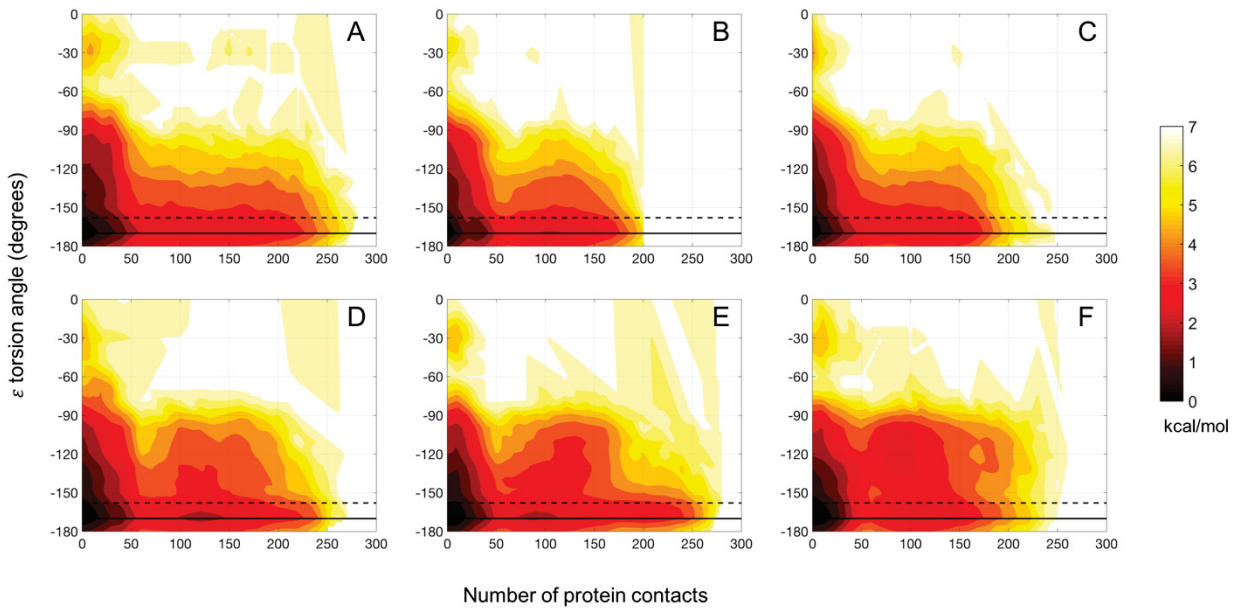


Figure S15. Potential of mean force (kcal/mol) as a function of ε backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

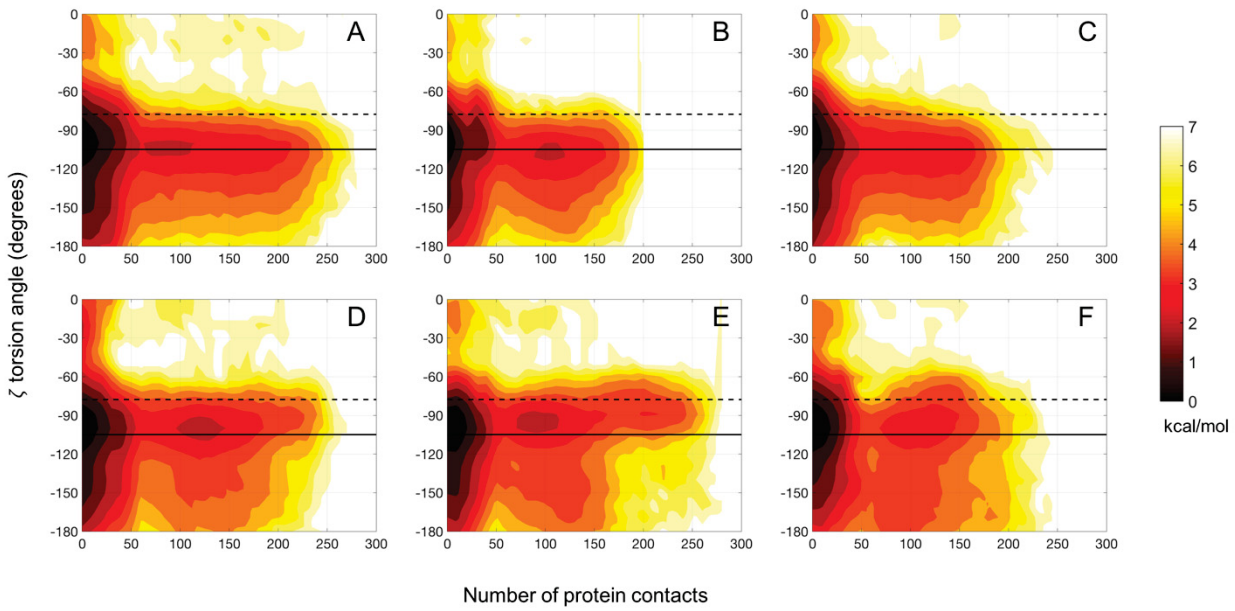


Figure S16. Potential of mean force (kcal/mol) as a function of ζ backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

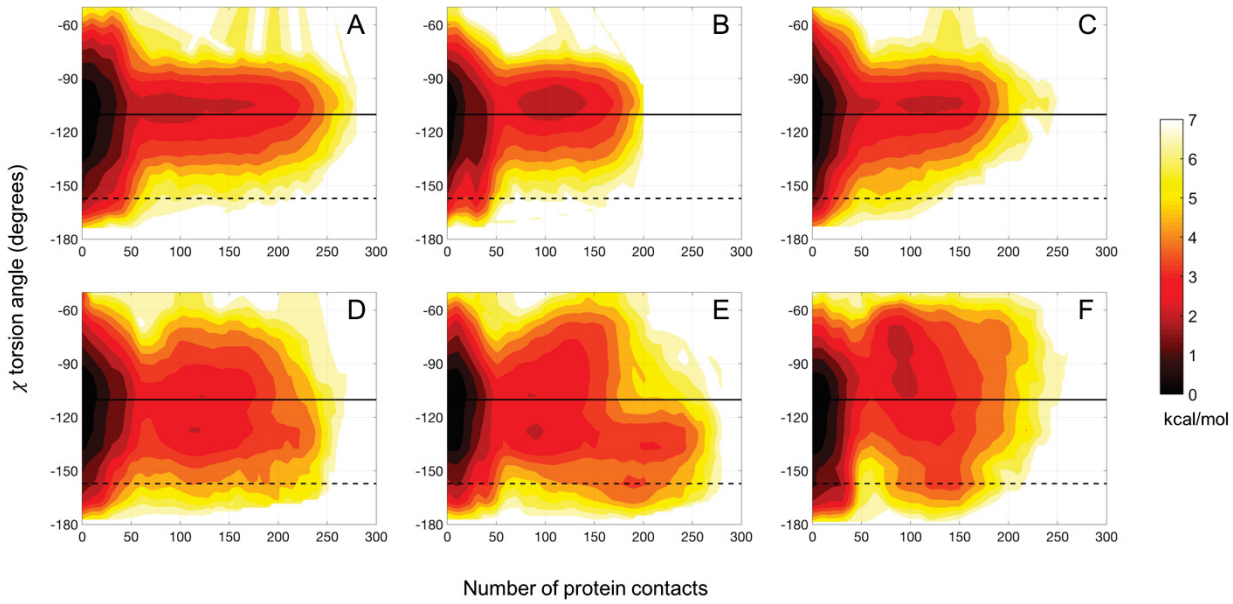


Figure S17. Potential of mean force (kcal/mol) as a function of χ backbone torsion angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

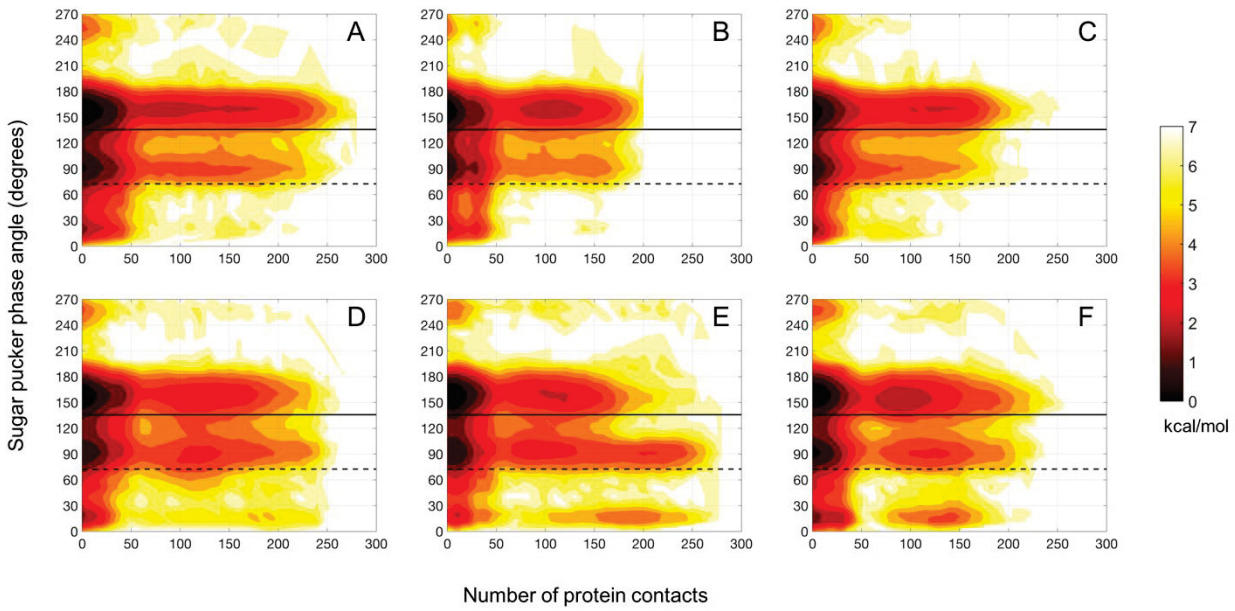


Figure S18. Potential of mean force (kcal/mol) as a function of sugar pucker phase angle and number of protein contacts for the Drew-Dickerson dodecamer at 20 % (A), 30 % (B), 40 % (C) protein concentrations, and for the GC-rich dodecamer at 20 % (D), 30 % (E), 40 % (F) protein concentrations. A contact is defined when the minimum distance between the heavy atoms of crowder proteins and DNA phosphate groups is less than 5 Å. Solid and dashed lines indicate the canonical B- and A-form values, respectively. See Table S5 for uncertainties.

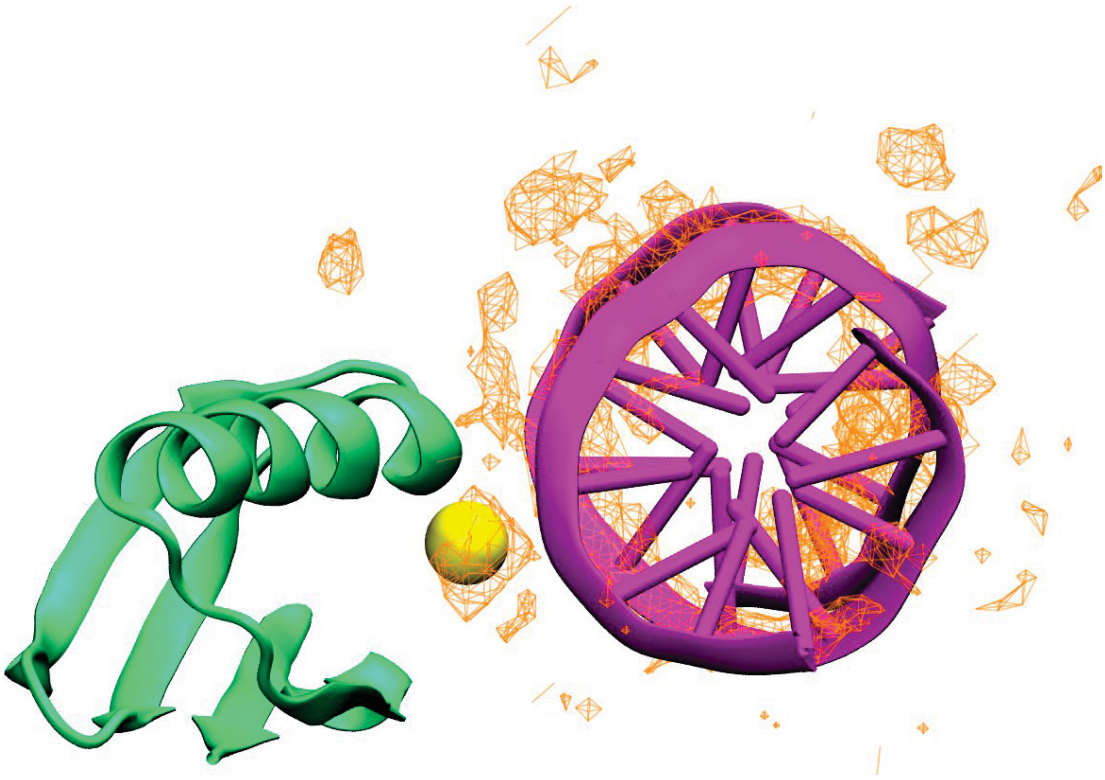


Figure S19. A snapshot showing a crowder protein interacting with the Drew-Dickerson DNA and orienting a sodium ion at the same time.