

# Lessons learned in atomistic simulation of double-stranded DNA: Solvation and salt concerns. [Article v1.0]

Rodrigo Galindo-Murillo<sup>1</sup> and Thomas E. Cheatham III<sup>1\*</sup>

<sup>1</sup>Department of Medicinal Chemistry, L. S. Skaggs Pharmacy Institute, University of Utah, Salt Lake City, UT 84112

This LiveCoMS document is maintained online on GitHub at [https://github.com/rodrigogalindo/DNA\\_salt\\_livecoms](https://github.com/rodrigogalindo/DNA_salt_livecoms); to provide feedback, suggestions, or help improve it, please visit the GitHub repository and participate via the issue tracker.

This version dated December 2, 2019

**Abstract** Nucleic acids are highly charged macromolecules sensitive to their surroundings of water, salt, and other biomolecules. Molecular dynamics simulations with accurate biomolecular force fields provide a detailed atomistic view into DNA and RNA that has been useful to study the structure and dynamics of these molecules and their biological relevance. In this work we study the Drew-Dickerson dodecamer duplex with the sequence d(GCGCAATTGCGC)<sub>2</sub> in three different salt concentrations and using different monovalent salt types to detect possible structural influence. Overall, the DNA shows no major structural changes regardless of amount or type of monovalent ions used. Our results show that only at very high salt conditions (5M) is a small structural effect observed in the DNA duplex, which mainly consist of narrowing of the grooves due to increased residence of ions. We also present the importance of sampling time to achieve a converged ensemble, which is of major relevance in any simulation to avoid biased or non-meaningful results.

**\*For correspondence:**

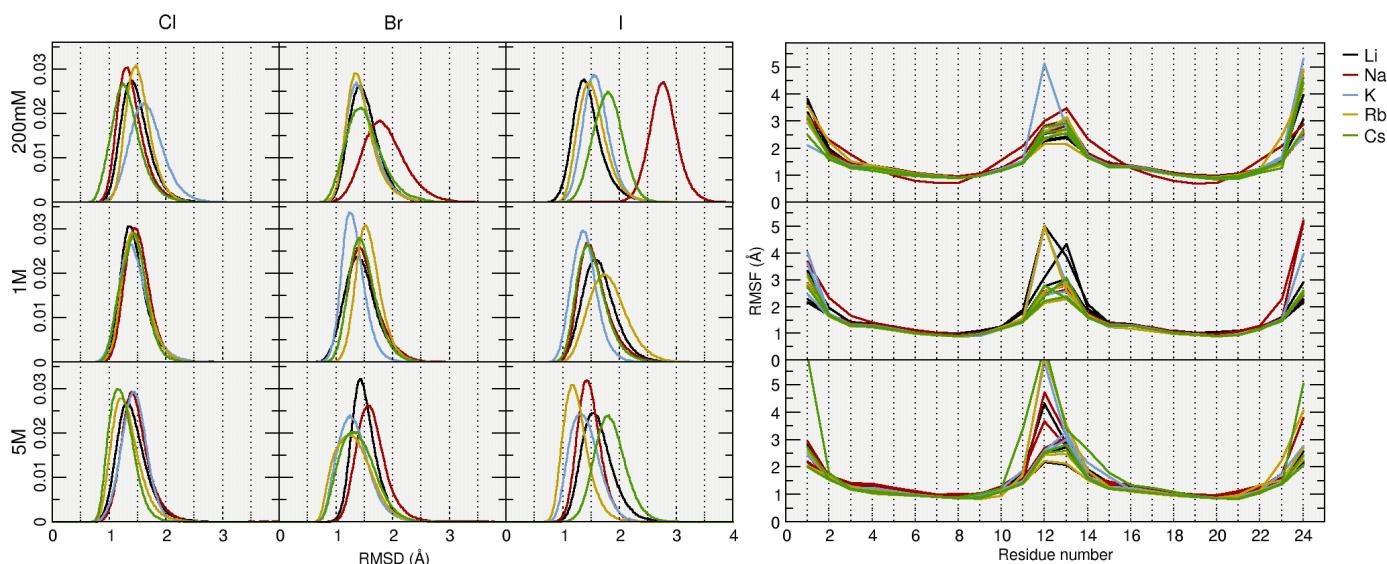
[rodrigo.galindo@utah.edu](mailto:rodrigo.galindo@utah.edu) (RG-M); [tec3@utah.edu](mailto:tec3@utah.edu) (TEC3)

## 1 Introduction

Monovalent and divalent cation interactions with nucleic acids play a major role in their conformational structure, dynamics and function[1]. The highly charged nature of nucleic acids makes them highly sensitive to the ionic conditions in the environment[2]. This environment is responsible for several DNA structural rearrangements that have been experimentally observed, changes that influence its biological function and binding to proteins and other ligands[3]. The most significant, and arguably the most biologically relevant, DNA helical conformational transition is the B→A transition that is related to both salt conditions and water activity[4, 5].

In low humidity – high salt conditions, interstrand phosphate repulsion increases resulting the formation of A-form DNA, which is stabilized by the presence of monovalent cations and phosphate bridging waters in an economy of hydration[4, 6, 7].

Other examples of conformational changes induced by monovalent ionic and solvent environment include the conversion of B to Z-DNA in a GC-rich sequence at high salt concentration[2], stabilization of guanine quadruplexes by a central Na<sup>+</sup> or K<sup>+</sup> ion[8] and formation of the C-DNA fiber stabilized by lithium[9, 10]. RNA biopolymers are also highly sensitive to ions and ionic conditions. Folding, biological activity and functionality of RNA has been found to be highly



**Figure 1.** Left: normalized populations of RMSD values at three different concentrations considering only the 10 internal base pairs. Using the net-neutralizing (NaCl) simulation of the Dickerson dodecamer with a 1  $\mu$ s duration, we calculated and extracted an average structure to use as a common reference for the RMSD calculations. Right: root mean square fluctuations per residue. Top to bottom is 200mM, 1M and 5M respectively.

dependent on the type and concentration of mono- and divalent ions[11–14].

Molecular dynamics simulations have provided extensive insight over the last three decades about the interactions of mono-[15–17] and divalent ions with nucleic acids[15, 18] including the B $\rightarrow$ A transition[4, 7, 19–21]. One of our research directions over the last decade has been to utilize molecular dynamics simulation methodologies to probe the influence of ion concentration and identity on the structure and dynamics of double-stranded DNA. An ambition was to spontaneously model the high salt concentration B $\rightarrow$ A helical transitions observed experimentally and also to understand the impact of different monovalent ions on DNA structure and dynamics. Multiple monovalent ion additive force field models have been developed[22–24], and these include the Åqvist cation set[25], Dang ion set[26], Jensen—Jorgensen OPLS set[27], Merz—Li[28], and the Joung—Cheatham ion parameters[29, 30]. Previously published research has focused on the influence of salt types and salt conditions on DNA structure and dynamics, however, most of this work has involved very short simulations which are unlikely to be converged. Orozco and co-workers investigated the influence of various ion models on a small 6-mer DNA duplex using 500 mM of NaCl and KCl, in independent simulations, with 50 ns of unrestrained fully solvated MD. The results show a lack of significant difference in the structural properties of the tested DNA model, regardless of the ion model. Additionally, using statistical methodologies to assess deviations between data sets, they evaluated the DNA dynamics observed in

the DNA trajectories, concluding that the simulations are indistinguishable regardless of the ion model used[31]. Also from the Orozco group, a series of simulations of DNA were performed to study the Na<sup>+</sup> distribution and lifetime of the ions within the minor groove[32]. They report no influence on the structure of DNA due to the presence of 200mM of salt concentration using 10 ns of sampling time. Yin and collaborators performed a series of simulations at 0.46 M, 1.86 M and 3.27 M of NaCl salt concentration to study the influence of salt concentration on the A $\rightarrow$ B transition[33]. In each set of simulations the A-DNA starting structures converted to B-DNA like structures within a sampling time of 1.5 ns with no drastic influence within the B-like conformation other than at high salt where the transition is slower. Regarding the use of a polarizable force field to study the effects of salt concentration in DNA, Savelyev and MacKerell report that the inclusion of polarization effects does indeed produce differences in solution scattering profiles when compared with both CHARMM 36 and AMBER bsc0 force fields[32]. Their work includes both  $\sim$ 100 mM of NaCl and independently  $\sim$ 100 mM NaCl with the addition of  $\sim$ 50 mM Tris-HCl aqueous buffer to compare directly to experimental X-ray scattering data. From their simulations, they calculate the solution scattering profiles and detect spectral variability when using the Drude force field which is not detected for the fixed charge counterparts. They also report structural influence of the minor groove between the polarizable and non-polarizable when comparing Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup> and Rb<sup>+</sup> over a sampling time of 150 ns. However, the statistical

significance of the observed differences is unclear, as will be further discussed in this work, for simulations on the sub-microsecond time scale.

In order to further study the effect of ion concentration on the structural properties of helical double-stranded DNA, we make use of the Drew-Dickerson[34] dodecamer sequence with different combinations of salts and concentrations. The monovalent ions  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  and their halides formed with the anions  $\text{Cl}^-$ ,  $\text{Br}^-$  and  $\text{I}^-$  have been tested using the Joung—Cheatham model[29, 30]. The different ion combinations included net neutralizing salt, 0.2M (representative of physiological conditions), ~1M and ~5M to determine the influence on DNA structure and dynamics from MD simulations of > 1  $\mu\text{s}$ . Additionally, a series of control MD simulations were also performed: a 'no-salt' simulation, an 'in-vacuo' simulation using the generalized Born implicit solvation model (referred as GB), and three simulations using the ff94, ff98 and ff99 versions of the AMBER nucleic acid force fields, all including the parmbsc0[35] modifications. Our initial simulations on the 1  $\mu\text{s}$  time scale appeared to suggest differences in structure populations and structure in the different conditions, however this was largely an artifact of lack of convergence. After this initial work, we were able to perform significantly longer MD simulations and clearly demonstrated that as sampling time is increased beyond the 5  $\mu\text{s}$  time scale, we are able to obtain a converged DNA duplex structure[36, 37]. This lead us to redo the DNA duplex simulations with the various salt combinations on the time scales required for convergence which provided structures that were very similar independent of the amount and type of ions used in the simulation". Structural properties are identical.

Our results show that as sampling time is increased, we obtain a *converged* DNA structure that is independent of the amount and type of ions used in the simulation. Structural properties are identical at a sub-angstrom resolution or difference in the central region of the DNA (base pairs 4 to 9) between different salt combinations and concentrations. As we increase the concentration of the salt, fraying events are increased due to increased lifetime of ions around the frayed bases at both ends of the duplex. The only significant differences in the structures are narrowing of the grooves and lower values of helical twist at the highest salt concentrations.

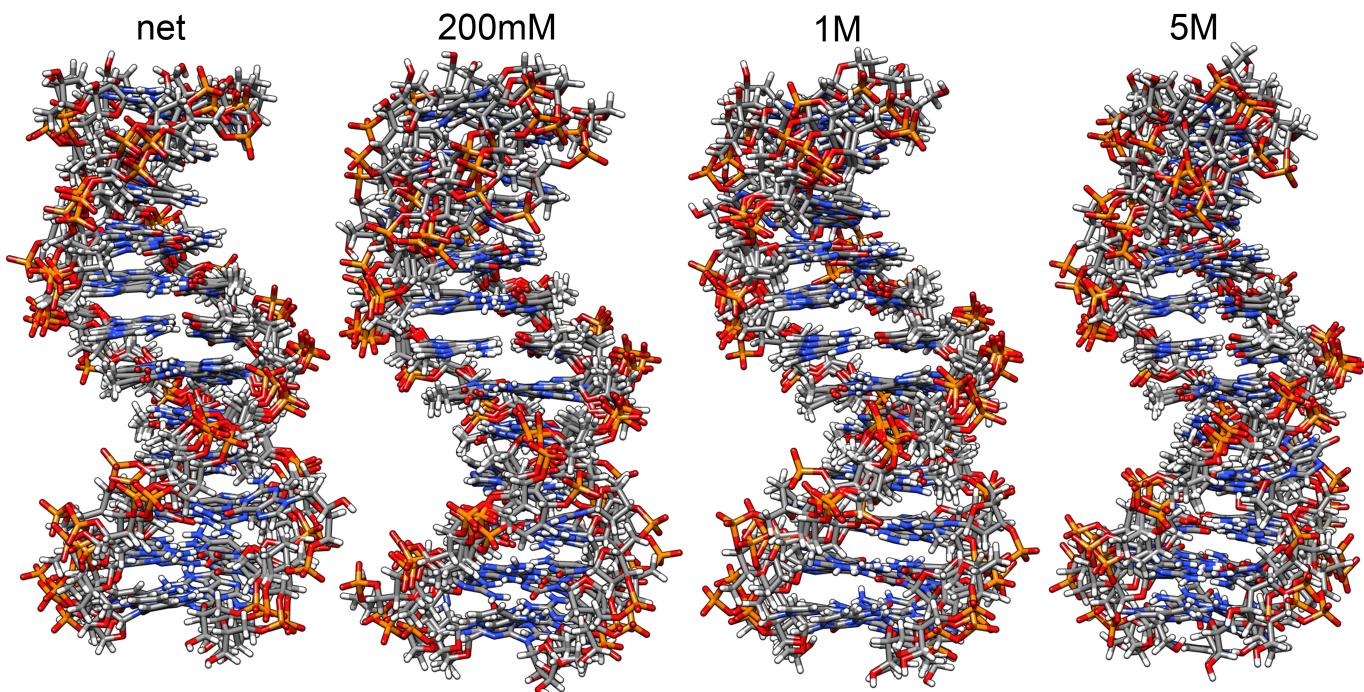
## 2 Theory and Methods

The initial model B-DNA structure used for this study was the Dickerson dodecamer d(CGCGAATTCTGCG)<sub>2</sub>, PDB code 1BNA[34]. A total of 55 simulations were performed starting from this structure using the various combinations of

cations, anions and concentrations, and each MD simulation was run for at least 1  $\mu\text{s}$ . The parm99[38] force field with the  $\alpha/\gamma$  parmbsc0 modifications[35] were used with the Joung-Cheatham[29, 30] ion parameters for  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$ ,  $\text{Cs}^+$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$  and  $\text{I}^-$ . Ions were added initially to neutralize the negative charge of the DNA backbone (net neutralizing ions). Then, additional ions were added to reach a concentration of either ~0.2M, ~1M or ~5M based on the initial volume of the constructed periodic cell. To provide further baseline sets of simulations, single runs were performed using the force fields ff94[39] + parmbsc0, ff98[7] + parmbsc0 and ff99[38] + parmbsc0 with net neutralizing NaCl ions. All the systems were solvated with the TIP3P[40] water model in a truncated octahedral box using a 15  $\text{\AA}$  water shell between the solute and the edge of the box. To avoid any biases, the initial ion positions were swapped by a random water molecule using the randomizeions command in CPPTRAJ or ptraj[41, 42].

This starting structure was then used to perform an initial equilibration phase where the DNA molecule was positionally restrained using a 25 kcal/mol- $\text{\AA}^2$  force constant and 500 steps of steepest descent minimization, switching to 500 steps of conjugate gradient minimization. Heating was then performed using the same positional restraints starting at a temperature of 150 K and scaled up to 300 K using 500 ps of MD simulation using the Langevin[43] thermostat at constant volume. This was followed with a series of minimization and equilibrium steps slowly decreasing the positional restraint applied to the DNA from 5 to 0.5 kcal/mol- $\text{\AA}^2$  using 2 ns in each step in constant pressure MD simulations. A 2 fs time step was used with a direct space cut off of 9.0  $\text{\AA}$ . Trajectory information was saved every 1 picosecond and the particle mesh Ewald[44] method was used to treat long range electrostatic interactions. The size for the periodic box was set to 69.787 in each direction using a value of 109.471 for  $\alpha$ ,  $\beta$  and  $\gamma$ . The PME grid (NFFT1, NFFT2 and NFFT3) was set to 72 with an Ewald Coefficient of 0.34864. A total production simulation time of no less than 1  $\mu\text{s}$  was run for every system in the absence of restraints using combined CPU and GPU technology and the PMEMD module of AMBER12[45].

In addition to these simulations, due to lack of convergence and unclear statistical certainty regarding the observed structural differences from simulations on the 1  $\mu\text{s}$  time scale[36, 37], we extended our dataset to include trajectories up to 15  $\mu\text{s}$  of sampling time. This second round of data used the same minimization and equilibration procedure as described and all possible pair combinations of cations and ions from the set of  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$  with  $\text{Cl}^-$  and  $\text{Br}^-$  ions with the same ion model as studied previously. The optimum point charge (OPC[46]) water model and the OL15[47] force field were employed instead. All the



**Figure 2.** Overlay of the 1  $\mu$ s average structures for LiCl, NaCl, KCl, RbCl and CsCl systems at different salt concentrations.

MD simulations were performed using either AMBER12 and AMBER14[48] suite of programs. Trajectory analysis was performed using CPPTRAJ version 15.00b available in AmberTools14[41, 42]. Average structures of the DNA were created by a straight coordinate average over all MD trajectory frames after imaging and RMS fitting to an appropriate reference structure. Clustering analysis was performed using the average linkage clustering algorithm[49] with an epsilon value of 2.0 as implemented in CPPTRAJ. Only the 10 central base pairs were considered for the clustering analysis. DNA structural and ion analyses were performed using Curves+, Canal and Canion[50, 51].

### 3 Results and Discussion

Molecular dynamics simulations were performed to determine if and how the different salt combinations and concentrations influenced the structure and dynamics of an experimentally well-studied 12-mer duplex DNA. Our first observations, based on MD simulations of up to 1  $\mu$ s (which were representative of the longest timescale MD simulation studies published investigating the influence of salt identity and concentration on nucleic acid up to the 2018 time frame), generated initially inconclusive results. Specifically, structural differences were seen, however the differences in observed RMSD values of  $> 1 \text{ \AA}$  proved to be greater than are expected for converged DNA helix simulations where RMSD values of  $< 1 \text{ \AA}$  are observed[36, 37].

**Table 1.** RMS deviation values ( $\text{\AA}$ ) for the net neutralizing simulations and alternate conditions. A 1  $\mu$ s average structure was calculated for each case and used as reference. All frames in the trajectory were considered for the RMS calculation. Inner residue values correspond to the RMSD using residues 3 to 10 and 15 to 22 (i.e. omitting the two terminal base pairs on each end of the helix). The Hawking, Cramer and Truhlar pairwise generalized Born model was used for the GB calculations[52].

	All residues		Inner residues	
		Std. Dev.		Std. Dev.
No salt	2.28	0.59	1.40	0.20
Li	2.08	0.51	1.62	0.31
Na	1.94	0.36	1.39	0.23
K	1.98	0.58	1.43	0.28
Rb	1.93	0.38	1.41	0.26
Cs	2.02	0.38	1.41	0.27
ff94bsc0	3.18	0.52	1.57	0.38
ff98bsc0	2.09	0.35	1.84	0.22
ff99bsc0	2.18	0.46	1.46	0.31
GB	4.19	0.71	3.46	0.75

It is important to mention that achieving this amount of sampling time during the 2012-2013 years, when these simulations were first performed, included running MD simulations for ~20 days each on a dedicated M2090 NVIDIA GPU. The following years, new GPU code performance optimizations[45, 53] and increases in the performance of the GPU cards coupled with the availability of more up-to-date AMBER force fields both for nucleic acids and water models, inspired us to continue our study of ion dependency with a new set of better converged MD trajectories on the 15  $\mu$ s time scale. We present both our earlier and the newer results which allows a direct comparison between converged and non-converged MD simulations. The results also show the importance and influence of sampling time and how the results from shorter non-converged simulations can be misleading.

### 3.1 Simulations covering the 1 $\mu$ s timescale.

Our first dataset consisted of independent simulations of each of the cations  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  paired with each of the anions  $\text{Cl}^-$ ,  $\text{Br}^-$  and  $\text{I}^-$ . The TIP3P water model[40] was used with the Joung-Cheatham parameters[29, 30] in a truncated octahedral periodic box. These simulations represent our first attempt to study the influence of different salt concentrations and identities on duplex DNA and, as mentioned, date back as far as 2012. We present this dataset as a legacy example that will provide a starting point for the present work.

The RMSD values for the net-neutralizing monovalent salt simulations are presented in Table 1 showing values suggesting structural differences ranging between 1.93 and 2.08 Å. These values are reduced by ~0.5 Å when only the central 10 base pairs are considered indicating that a significant source for the deviation from the average is the dynamics of the DNA base pairs located at both ends or termini of the helix, as was been discussed previously in the literature[37, 54]. The no-salt condition, which is a MD simulation where no net-neutralizing salt is present, shows a slightly higher deviation of 2.28 Å. On the 1  $\mu$ s time scale, the DNA helix does not denature as might be expected, albeit the time scale for this process is not obvious. The largest RMSD (4.19 Å), a value higher than expected for a good fidelity simulation of a DNA duplex, was observed in analysis of the GB implicit solvent simulation. Visual inspection of this implicit solvent trajectory shows multiple fraying events on both sides of the DNA chain with a frequent number of backbone transitions to ladder like structures[55], untwisting, and sampling of non-canonical configurations throughout the simulation.

The MD simulations for each of the salts at the various concentrations (200mM, 1M and 5M) led to RMS deviations in

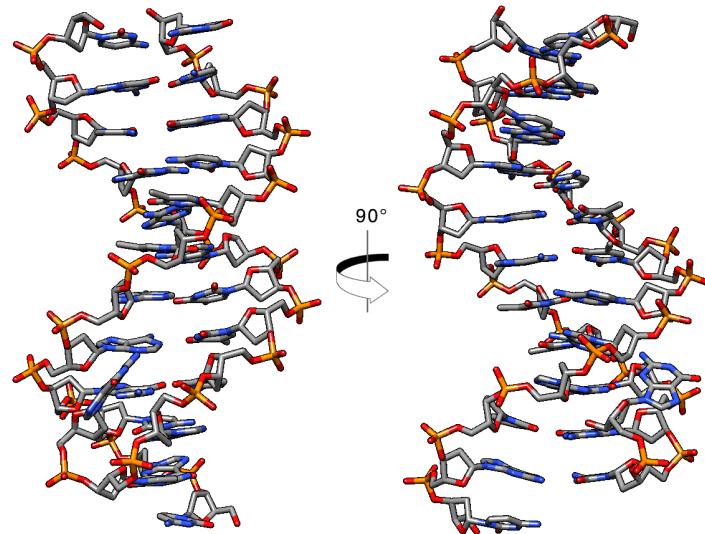
the range of ~0.8-2.5 Å (Figure 1). The RMSD calculation was performed with reference to the net-neutralizing ( $\text{NaCl}$ ) 1  $\mu$ s MD simulation average structure. Overlaying the population distributions (normalized to 1) in each test case we detect differences within 1 Å, except for the  $\text{Cl}^-$  case at 1 M, where the difference among the sampled trajectories is ~0.25 Å (inner residues). Differences in the RMSD histograms in 1 suggest sampling of different populations of structures or structural differences, however these differences disappear as the structures converge. Analysis performed using the 12 base pairs renders an increase in the RMS deviation among the simulations due to terminal base pair fluctuations. This effect can be observed in the RMS fluctuation plots (bottom, Figure 1) were we can detect increased dynamics at both ends of the DNA duplex (within the ~2-3 Å range) and a small (~0.25 Å) divergence in dynamics for the central base pairs. The 200mM  $\text{NaBr}$  and  $\text{NaI}$  outliers present are due to long lived fraying events where the terminal base flips inside the minor groove (Figure 3). As already mentioned, the source of the deviations among simulations is mainly due to base pair fraying events that occur at both ends of the DNA chain as observed in the RMS fluctuation measures in Figure 1. To confirm this fact, the 1  $\mu$ s average structure was compared among different systems and simulations to detect any noticeable structural deviation caused by the ions. An overlay of selected structures is presented in Figure 2 for the  $\text{LiCl}$ ,  $\text{NaCl}$ ,  $\text{KCl}$ ,  $\text{RbCl}$  and  $\text{CsCl}$  at different concentrations. It is evident from the structures that the central base pairs show less divergence between the selected system, whereas both ends of the DNA duplex show significant structural dynamics. Visual inspection also suggests that the divergence increases as we increase the ion concentration in the simulation.

In order to confirm the central base pair similarity between the systems presented in Figure 2, we calculated the RMS deviation among the inner base pairs considering residues 4 to 9 and 16 to 21 that corresponds to the six central base pairs of our duplex DNA test sequence. We employed the net neutralizing average structure as a reference and compared to the systems with the cations  $\text{Li}^+$ ,  $\text{Na}^+$  and  $\text{K}^+$  using the three simulated concentrations as our selected dataset (Table 2). Considering all the calculated values, the RMSD deviation ranges between ~0.75 and 1.80 Å with no discernable trend or preference for any combination of salt or concentration. The molecular graphics in Figure 2 and RMS deviations reported in Table 2 may be (and previously have been) interpreted to suggest structural differences, however the statistical significance of these differences is uncertain until longer simulations are performed that reach greater structural convergence.

**Table 2.** RMS deviations of selected systems. The comparison is using the 1  $\mu$ s average structures from the net neutralizing simulation and measuring the inner residues 4 to 9 and 16 to 21. "net" refers to net-neutralizing monovalent salt.

Reference structures are the net neutralizing conditions.

	Concentration	$\text{Li}^+$	$\text{Na}^+$	$\text{K}^+$	$\text{Cs}^+$	$\text{Rb}^+$
Li	net	-	1.53	1.09	1.09	1.06
LiCl	200 mM	1.21	1.21	1.17	0.99	0.95
	1 M	1.25	1.20	1.03	1.18	1.00
	5 M	1.15	1.29	1.20	0.74	0.98
LiBr	200 mM	1.34	0.96	1.42	1.05	1.13
	1 M	1.12	1.30	0.74	1.05	0.97
	5 M	1.40	1.36	1.22	1.18	1.10
LiI	200 mM	1.12	1.31	0.87	1.00	0.96
	1 M	1.54	1.19	1.57	1.45	1.34
	5 M	1.23	1.70	1.07	1.34	1.08
Na	net	1.53	-	1.41	1.23	1.23
NaCl	200 mM	1.21	0.99	1.26	0.97	0.97
	1 M	1.65	0.92	1.43	1.23	1.23
	5 M	1.58	1.04	1.33	1.21	1.21
NaBr	200 mM	1.15	1.48	1.33	1.21	1.21
	1 M	1.21	1.13	1.10	0.95	0.95
	5 M	1.57	1.22	1.30	1.26	1.26
NaI	200 mM	1.57	1.80	1.33	1.65	1.65
	1 M	1.20	1.09	1.02	0.94	0.94
	5 M	1.50	0.87	1.40	1.16	1.16
K	net	1.09	1.41	-	1.20	0.91
KCl	200 mM	1.48	1.18	1.50	1.51	1.37
	1 M	1.12	1.30	0.90	1.13	0.93
	5 M	1.55	1.17	1.13	1.54	1.21
KBr	200 mM	1.19	1.28	0.96	1.13	0.94
	1 M	1.41	1.28	1.14	1.36	1.12
	5 M	1.22	1.51	0.86	1.10	0.98
KI	200 mM	1.32	0.92	1.34	1.13	1.11
	1 M	1.22	1.27	0.80	1.30	0.96
	5 M	1.18	1.55	1.14	1.09	1.17

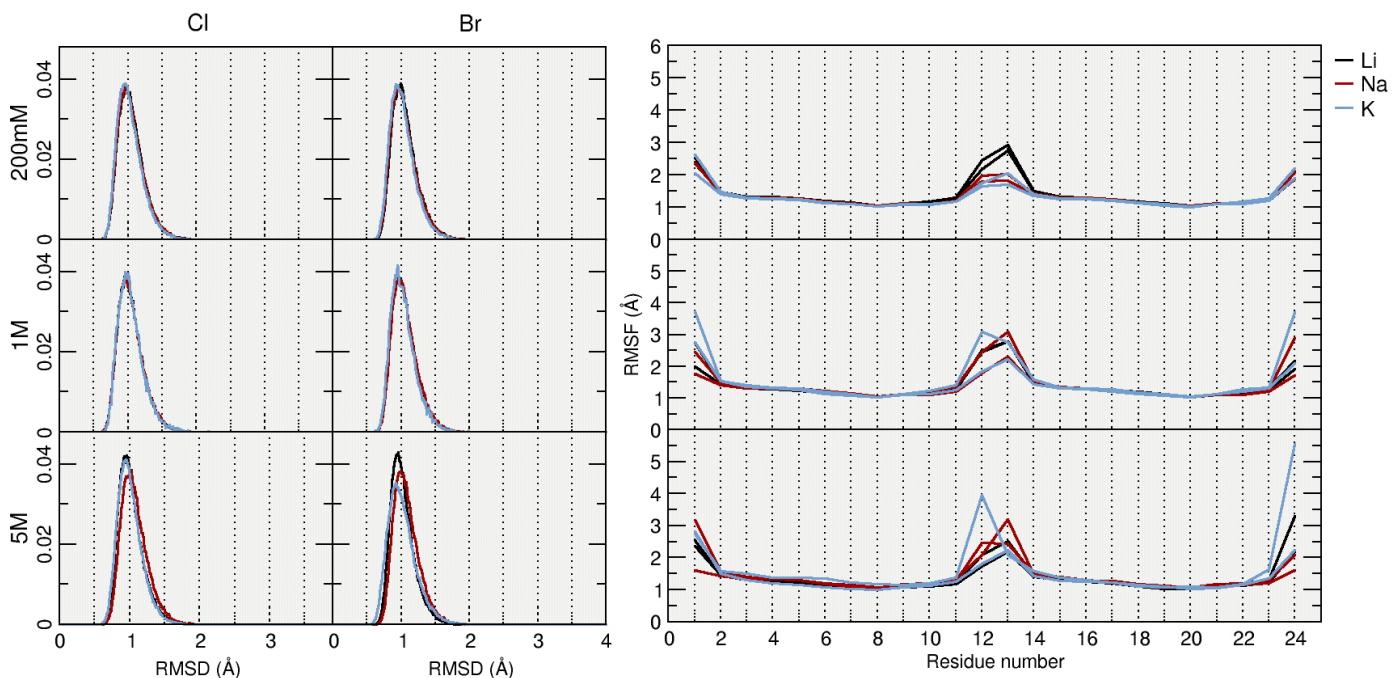


**Figure 3.** Most populated representative structure from the NaCl - 5M simulation showing a fraying event where the free dG nucleobase flips towards the minor groove and forms base-backbone interactions which stabilizes the mis-pair. Hydrogen atoms are hidden for clarity.

### 3.2 Simulations covering the 15 $\mu$ s timescale.

Our previous 1  $\mu$ s results showed no discernable difference or trend among the different salt combinations and/or concentrations, and it was not clear if the observed differences in structure and RMSD had any statistical significance. From previous work, we had observed that in longer simulations (beyond 2-5  $\mu$ s) RMSD differences between independent simulations were significantly less than 1  $\text{\AA}$  for the central base pairs[36, 37]. This led us to increase the sampling time up to 15  $\mu$ s with the same Dickerson dodecamer and the cations  $\text{Li}^+$ ,  $\text{Na}^+$  and  $\text{K}^+$  with the anions  $\text{Cl}^-$  and  $\text{Br}^-$ , maintaining the three concentrations so far tested (200 mM, 1 M and 5 M). These experiments were performed during the time when the improved OL15 DNA force field became available[47, 56]. We decided to test this more up-to-date force field and also to include the optimal point charge (OPC) water model[46] that was demonstrating promising improvements in simulations of RNA[57].

The RMS deviation histograms for all the simulated systems present tightly converged results using the 10  $\mu$ s average structures as reference (Figure 4). Overall, for the 200 mM and 1 M concentrations, the RMSD populations present sub-angstrom differences, regardless of cation or anion used. Some amount of discrepancy is observed for all the 5 M systems, although, still less than 1  $\text{\AA}$ . Fluctuations measured using the entire 15  $\mu$ s trajectory shows the familiar base pair fraying at both ends of the DNA duplex, with fluctuation values of  $\sim$ 1-1.5  $\text{\AA}$  for the inner residues. The simulations using



**Figure 4.** Left: normalized RMSD populations at three different concentrations, considering only the internal 10 base pairs. Using the net-neutralizing (NaCl) of the Dickerson dodecamer with a total sampling time of 10  $\mu$ s, we computed and extracted an average structure that was used as a reference for the RMSD calculations. Right: root mean square fluctuations per residue. Top to bottom is 200mM, 1M and 5M respectively. For each cation, there are two lines of the same color representing the Cl<sup>-</sup> and Br<sup>-</sup> anion values.

5 M salt concentration display the least observed differences among the used cations. To identify the main source of the structural deviation, we calculated a 10  $\mu$ s average structure for the Cl<sup>-</sup> system from the net, 200 mM, 1 M and 5 M simulations and RMS fit the structures (Figure 5). The similarity of the central base pairs is between values of  $\sim$ 0.05–0.3 Å whereas the terminal base pairs present increased fluctuations in the range of  $\sim$ 1–1.5 Å.

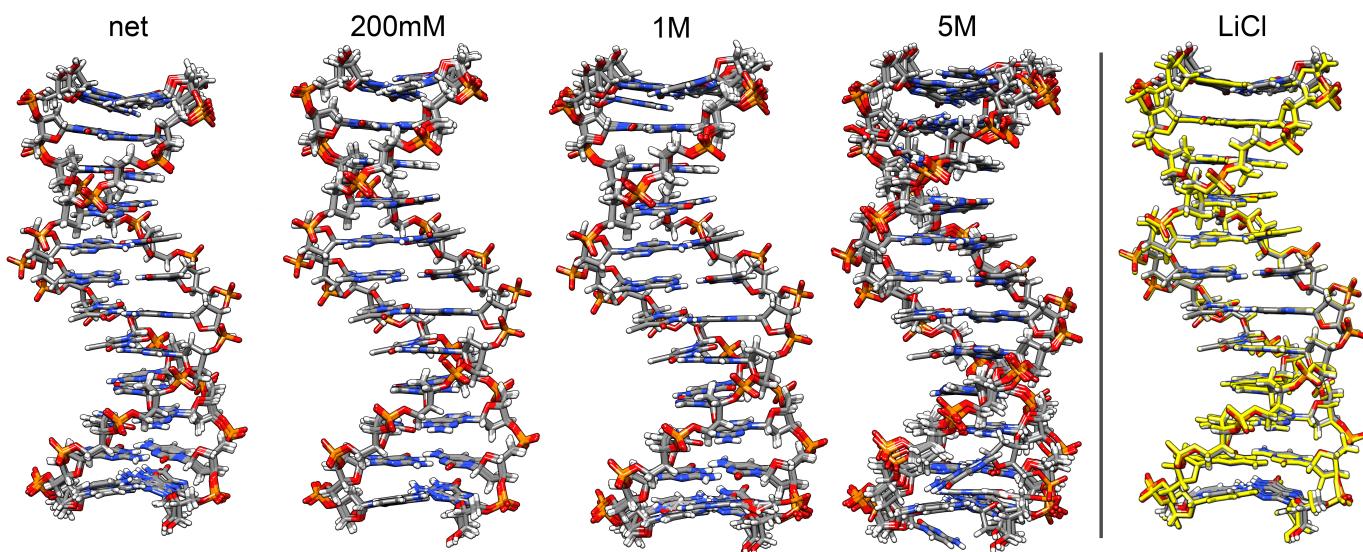
The magnitude of the difference between selected simulations is presented in Table 3 where it is remarkable to notice that the average RMS deviation is only 0.16, 0.21 and 0.16 Å when using the Li<sup>+</sup>, Na<sup>+</sup> and K<sup>+</sup> reference respectively. The only discernable trend observed is a slight increase in RMS deviation for the 5 M simulations.

### 3.3 Terminal base pair fraying events.

Fraying events are involved in multiple biological phenomena, including the recognition for enzymatic catalysis, DNA-enzyme recognition, duplex-DNA melting and nuclease activity[54]. The dynamic nature of terminal base pair fraying makes it difficult to study experimentally. Spin relaxation measurements on DNA with G:C terminal pairs report motions in the picosecond to nanosecond time scale[58, 59]. As a way to assess these events, we measured the distance between the C1' atoms of pairing bases. A matched pair is considered to have an average value of  $\sim$ 10.6 Å between

the C1' atoms. The normalized distribution of distances from our simulations are presented in Figure 4, where we can see for the base pairs at the end of the DNA chain (residues 1 and 24) that values other than 10.6 are typically observed. Visual inspection of the trajectories suggests that the larger values result from flipped or mis-paired bases. No clear fraying tendency regarding the specific type of ion is observed, potentially with the exception of Na<sup>+</sup> with the long-lived terminal base pair opening previously discussed. However, for the 5M salt concentration, terminal base pair fraying events are increased, generating multiple alternative conformations, conformations similar to those previously reported in analysis of DNA duplex MD simulations[54]. Our results also suggest that the DNA is not significantly affected by the type of ion used in the simulation, but, as mentioned, by the amount present. Our working hypothesis to explain the effect of the amount of salt with the DNA structure, and not the type, is that as salt concentration increases, the time of residence within the grooves increases, which in turn allow for an increase in shielding between the negatively charged phosphate groups.

The low RMS deviations obtained from the 10  $\mu$ s average structures present small, although, significant differences in some cases (Table 3). As we increase the concentration, fraying events become more frequent and RMS deviation increases. For example, KBr measured using Na<sup>+</sup> as a refer-



**Figure 5.** Overlay of the final 10  $\mu$ s average structures for the LiCl, NaCl, KCl, RbCl and CsCl systems at the different salt concentrations. The LiCl structure is an overlay of the 10  $\mu$ s average structure for each salt concentration. Shown in yellow is the 5M concentration average structure.

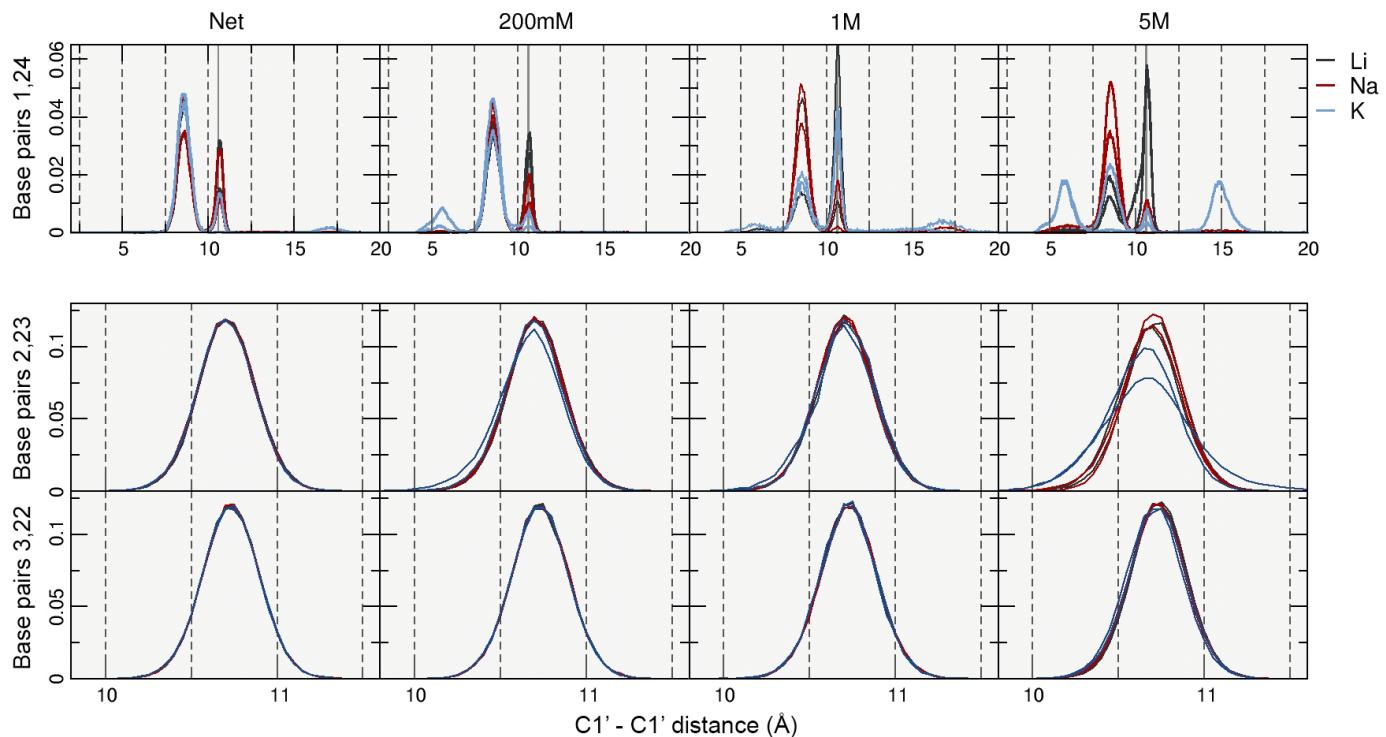
**Table 3.** RMS deviations of selected systems. The comparison is using the 10  $\mu$ s average structures and measuring the inner residues 4 to 9 and 16 to 21 (i.e. neglecting the terminal three base pairs on each end).

Reference used is the net neutralizing concentration.

		Li <sup>+</sup>	Na <sup>+</sup>	K <sup>+</sup>
LiCl	Li net	-	0.24	0.11
	200mM	0.05	0.23	0.11
	1M	0.05	0.23	0.12
	5M	0.23	0.16	0.20
LiBr	200mM	0.05	0.21	0.11
	1M	0.08	0.21	0.11
	5M	0.26	0.14	0.21
	Na net	0.24	-	0.15
NaCl	200mM	0.18	0.09	0.11
	1M	0.09	0.18	0.10
	5M	0.26	0.34	0.27
	NaBr	0.17	0.10	0.11
NaBr	200mM	0.17	0.10	0.11
	1M	0.19	0.14	0.13
	5M	0.19	0.20	0.17
	K net	0.11	0.15	-
KCl	200mM	0.14	0.20	0.11
	1M	0.13	0.25	0.15
	5M	0.27	0.46	0.34
	KBr	0.07	0.19	0.06
KBr	200mM	0.09	0.27	0.15
	1M	0.54	0.49	0.50

ence with RMS values of 0.19, 0.27 and 0.49 Å for 200mM, 1M and 5M respectively. This trend is consistent among the majority of the studied systems. When comparing the simulations using the net neutralizing Na<sup>+</sup> reference, a slight increase in the RMS deviation is observed, for example, 0.05 Å LiCl – 200mM with Li<sup>+</sup> as reference in contrast with the same system when using Na<sup>+</sup> as reference (0.21 Å). After visual inspection of the average structures and trajectories for the Na<sup>+</sup> net neutralizing simulation used for the calculations, we found multiple fraying trapped states consistent with the depiction in Figure 3 with one of the nucleobases tipping towards the minor groove which causes the slight increase in RMS deviations when using this frayed structure as reference.

As a measure of convergence within and between the simulations, we performed a pairwise RMSD combined clustering analysis using the 200mM, 1M and 5M salt concentrations from the NaCl MD simulations. The analysis algorithm (hierarchical agglomerative) was set to populate 10 clusters so we could have an even distribution among the different conditions. The normalized structure population for each cluster during the sampled time is presented in Appendix Figure 8. We observe from the plot and the line crossings that cluster populations have not stabilized or converged on the 1  $\mu$ s time scale. On the other hand, when the MD simulations on the 15  $\mu$ s are analyzed, it is clear that the populations are fairly well converged after ~1  $\mu$ s and remain almost unchanged for the remaining of the sampled time.



**Figure 6.** Normalized population for the distance between the C1' carbons for first base pair (residues 1 and 24, top) and the 2nd and 3rd base pairs (bottom). Reference value of 10.6 Å is depicted as a gray line in the top plot.

### 3.4 Analysis of the ion distribution.

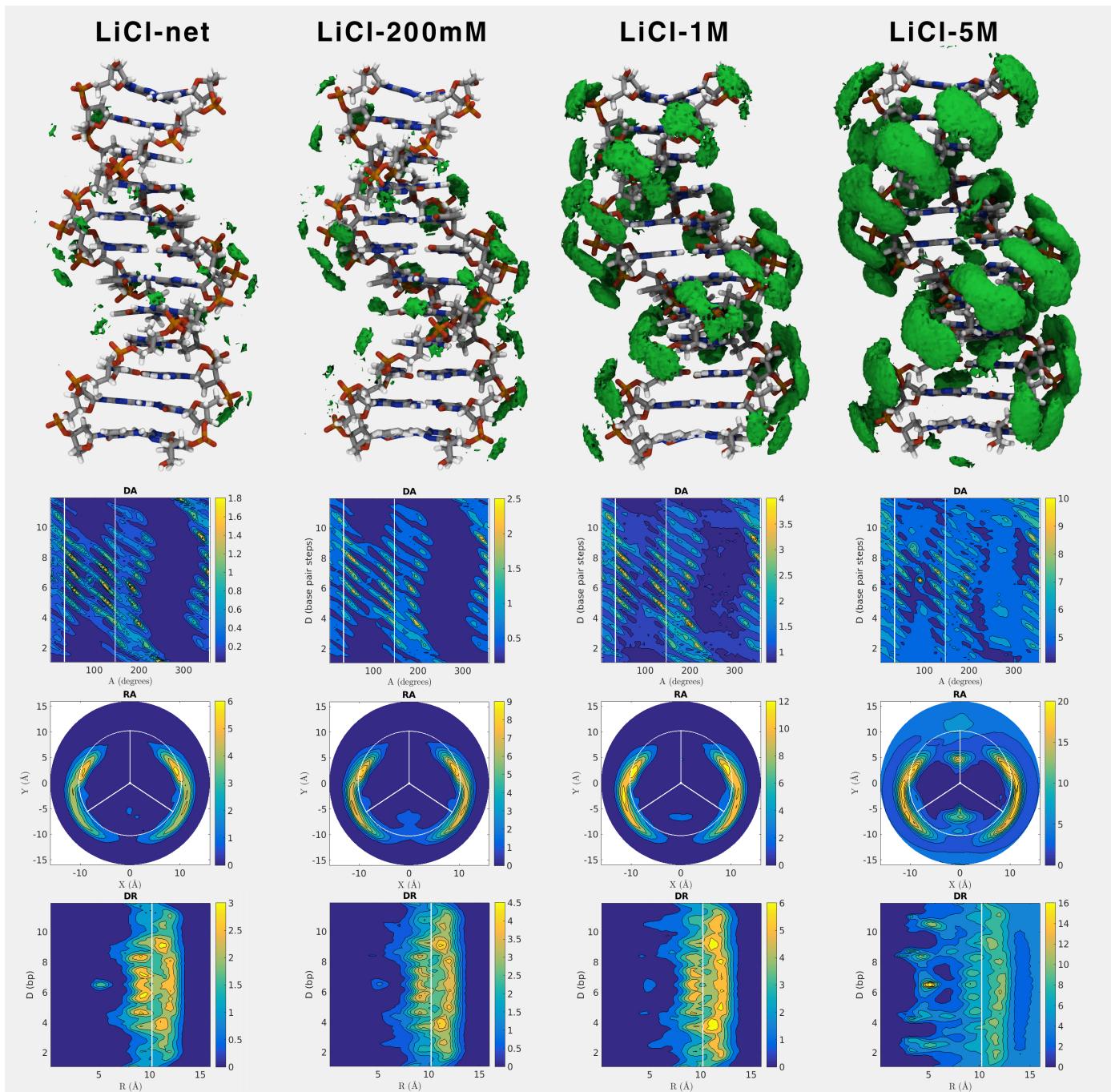
In order to study the effect of increased salt on the duplex DNA structure, we performed several ion distribution analyses on the 15  $\mu$ s trajectory of LiCl at net neutralizing, 200mM, 1M and 5M salt concentrations. Calculated Li<sup>+</sup> isodensity surfaces for each concentration show accumulation of the ions within the vicinity of the phosphate groups (Figure 7, top). As ion concentration is increased, the density becomes more pronounced and increased ion populations are clearly evident within both of the grooves.

By means of the cation ion distribution analysis software[50], we performed a study of the distribution of Li<sup>+</sup> ions at different concentrations (Figure 7). Briefly,  $D$  represent the distance along the helical axis,  $A$  is an angle that tracks the helical twist and  $R$  is the distance from the helical axis towards the solvent. In the  $DA$  plot, we include two vertical white lines located at 33deg and 147deg that delimit the minor groove region. The net neutralizing concentration shows an accumulation of ions within the minor groove with highest molarity values close to the central base pairs and little to no presence inside the major groove. As we increase the concentration, we observe a strong increase in molarity of ions within the major groove localized mainly within the phosphate groups as observed in the 3D grid density histogram. The presence of ions within the major groove

at 5M is confirmed with the aid of the  $RA$  and  $DR$  plots. The white circle in the  $RA$  plot located at a distance of 10.2 Å from the helical axis delimits the outer circumference of the DNA duplex and indicates the distance of the C1' carbon atoms. The white radial vectors indicate the groove limits and the center of the major groove (vertical vector). Similarly, for the  $DR$  plots, the white line at 10.2 Å delimits the C1' carbons. This analysis helps explain why some degree of modest structural difference is observed mostly only at the 5M salt concentration.

### 3.5 Structural analysis.

To further study the subtle structural influences of the different salt concentrations on the duplex DNA, noting that helicoidal parameters are extremely sensitive measures of small differences in structure, we calculated the helicoidal parameters using the 10  $\mu$ s average structure for each simulation condition (Appendix table). There is a reduction on the major groove width as salt concentration is increased. This reduction is independent of the type of salt and is observed consistently among the six systems. A similar reduction in width is observed for the minor groove except for the systems with the Li<sup>+</sup> cation where the width value remained similar for the lower salt concentrations and increased slightly in the 5M salt simulations. Lower twist values are also observed for the



**Figure 7.** Top: ion density of  $\text{Li}^+$  (green isosurface value of 0.54). Bottom: 2D ion distributions. The blue to yellow color indicate increasing values of molarity. Notice the difference in molarity scales across the net, 200mM, 1M and 5M concentrations. Using the same molarity scale across salt concentrations was not possible due to the fact that the scale that rendered a good image for the net/200mM system was completely saturated for the 5M and the scale that showed a balanced image for the 5M system would not allow to see any information for the net/200mM system. Allowing the molarity scale to adjust to its optimal value shows more detail in the analysis. The white lines in the DA plots represent the minor groove limits as defined by the C1' carbon positions, as with the DR plot. The white circle in the RA plots represent the radial vector limit for the C1' carbons; radial sampling limited to 15  $\text{\AA}$ . Refer to [50] for more information. Analysis calculated using the entire 15  $\mu\text{s}$  trajectory.

5M simulations in comparison with the rest of the concentrations (except for the KCl system). Similar decrease of rise distance between base pairs is observed as we increase ion concentration. The unwinding of the DNA in the 5M salt simulations couples with the reduction of the groove widths, with the exception of slight minor groove width increase in the Li<sup>+</sup> simulations which likely is due to the change in bending which is not observed in the Na<sup>+</sup> and K<sup>+</sup> 5M salt simulations.

## 4 Conclusions

The results presented highlight some of our recent experiences (~2012-2019) assessing and validating the capabilities of AMBER force fields parmbsc0 and OL15 to detect and be affected by different salt types and salt concentrations using the Joung—Cheatham ion parameters. Our observations suggest:

- Simulations of ~1 μs or less for a 12- to 18-mer duplex DNA are definitely not converged. MD simulations should likely be at least in the range of ~5-10 μs time range to be comparable and to assess subtle structural differences. As sampling time increases, the structural populations of different local-minima approach convergence. With more simulation time, the generated structures start to populate a consistent set of representative structures with consistent cluster populations for those structures. As the simulation time increases, the resulting average structure converges to a single representative structure.
- Regardless of the type of cation or anion used, with these additive force fields, the DNA is not significantly affected as measured by RMS deviations over averaged structures.
- As we increase salt concentration, increased fluctuations and fraying effects at both ends of the DNA chain increase.
- Decrease in major groove width, minor groove width (with the exception of Li<sup>+</sup> where the minor groove width increases and bending decreases), twist angle and inter base pairs rise as a function of increased ion concentration is observed for the 10 central base pairs.
- With these force fields, DNA helices in the absence of salt do not denature in simulations on the 1 μs time scale.

We take this opportunity of reporting our results in this version of the "Lessons Learned" category of the Living Journal of Computational Molecular Science to share with any research group that is interested in studying the effects of high salt concentration in DNA using AMBER in the hopes that they will find our experience useful. The next step for

updating our work will be incorporating the parmbsc1 variation of the AMBER DNA force field, exploring different water models, performing longer no-salt DNA simulations, and explorations of new nucleic acid polarizable force fields[60] in the hope of learning the source for, or to verify the, lack of sensitivity of DNA simulations to strong ionic environment.

Pre-processed trajectories and topology files are available for download at: <http://www.amber.utah.edu/DNA-dynamics/livecoms-salt/>

## 5 Author Contributions

RG-M and TEC3 contributed equally in this article. For a more detailed description of author contributions, see the GitHub issue tracking and changelog at [https://github.com/rodrigogalindo/DNA\\_salt\\_livecoms](https://github.com/rodrigogalindo/DNA_salt_livecoms).

## 6 Funding Information

TEC3 acknowledges the support of NIH grant GM081411.

## References

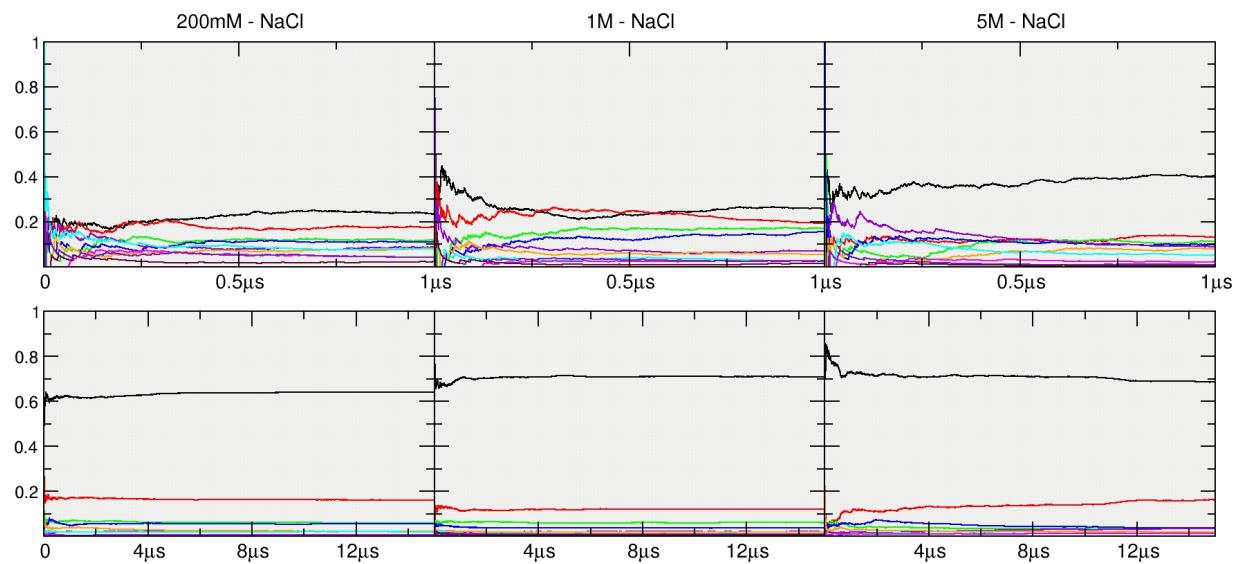
- [1] Sharp KA, Honig B. Salt effects on nucleic acids. Current opinion in structural biology. 1995; 5(3):323–8. doi: [10.1016/0959-440X\(95\)80093-X](https://doi.org/10.1016/0959-440X(95)80093-X).
- [2] Saenger W. Principles of Nucleic Acid Structure. Springer-Verlag; 1988.
- [3] McFail-Isom L, Sines CC, Williams LD. DNA structure: cations in charge? Current Opinion in Structural Biology. 1999; 9(3):298–304. doi: [10.1016/S0959-440X\(99\)80040-2](https://doi.org/10.1016/S0959-440X(99)80040-2).
- [4] Kulkarni M, Mukherjee A. Understanding B-DNA to A-DNA transition in the right-handed DNA helix: Perspective from a local to global transition. Progress in Biophysics and Molecular Biology. 2017; 128:63–73. doi: [10.1016/j.pbiomolbio.2017.05.009](https://doi.org/10.1016/j.pbiomolbio.2017.05.009).
- [5] Waters JT, Lu XJ, Galindo-Murillo R, Gumbart JC, Kim HD, Cheatham TE, Harvey SC. Transitions of Double-Stranded DNA Between the A- and B-Forms. The Journal of Physical Chemistry B. 2016; 120(33):8449–8456. doi: [10.1021/acs.jpcb.6b02155](https://doi.org/10.1021/acs.jpcb.6b02155).
- [6] Saenger W, Hunter WN, Kennard O. DNA conformation is determined by economics in the hydration of phosphate groups. Nature. 1986; 324(6095):385–388. doi: [10.1038/324385a0](https://doi.org/10.1038/324385a0).
- [7] Cheatham 3rd TE, Crowley MF, Fox T, Kollman PA. A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. Proceedings of the National Academy of Sciences of the United States of America. 1997; 94(18):9626–30. doi: [10.1073/pnas.94.18.9626](https://doi.org/10.1073/pnas.94.18.9626).
- [8] Bhattacharyya D, Mirihana Arachchilage G, Basu S. Metal Cations in G-Quadruplex Folding and Stability. Frontiers in chemistry. 2016; 4:38. doi: [10.3389/fchem.2016.00038](https://doi.org/10.3389/fchem.2016.00038).
- [9] Langridge R, Wilson HR, Hooper CW. The molecular configuration of deoxyribonucleic acid:: I. X-ray diffraction study of a crystalline form of the lithium salt. Journal of Molecular Biology. 1960; 2(1):19–37. doi: [10.1016/S0022-2836\(60\)80004-6](https://doi.org/10.1016/S0022-2836(60)80004-6).

- [10] Marvin DA, Spencer M, Wilkins MHF, Hamilton LD. The molecular configuration of deoxyribonucleic acid III. X-ray diffraction study of the C form of the lithium salt. *Journal of Molecular Biology*. 1961; 3(5):547–IN14. doi: [10.1016/S0022-2836\(61\)80021-1](https://doi.org/10.1016/S0022-2836(61)80021-1).
- [11] Wedeking JE, Dutta D, Belashov IA, Jenkins JL. Metalloriboswitches: RNA-based inorganic ion sensors that regulate genes. *J Biol Chem*. 2017; 292(23):9441–9450. doi: [10.1074/jbc.R117.787713](https://doi.org/10.1074/jbc.R117.787713).
- [12] Leamy KA, Assmann SM, Mathews DH, C BP. Bridging the gap between in vitro and in vivo RNA folding. *Q Rev Biophys*. 2016; e10(49).
- [13] Auffinger P, Westhof E. Water and ion binding around RNA and DNA (C,G) oligomers. *Journal of Molecular Biology*. 2000; (5):1113–1131. doi: [10.1006/jmbi.2000.3894](https://doi.org/10.1006/jmbi.2000.3894).
- [14] Auffinger P, Westhof E. RNA solvation: A molecular dynamics simulation perspective. *Biopolymers*. 2001; (56):266–274. doi: [10.1002/1097-0282](https://doi.org/10.1002/1097-0282).
- [15] Cheatham 3rd TE, Case DA. Twenty-five years of nucleic acid simulations. *Biopolymers*. 2013; 12:969–977. doi: [10.1002/bip.22331](https://doi.org/10.1002/bip.22331).
- [16] Feig BM, Pettit M. Sodium and Chlorine ions as part of the DNA solvation shell. *Biophys J*. 1999; 77(4):1769–81. doi: [10.1016/S0006-3495\(99\)77023-2](https://doi.org/10.1016/S0006-3495(99)77023-2).
- [17] Hamelberg D, McFail-Isom L, Williams DL, Wilson WD. Flexible Structure of DNA: Ion dependence of minor-groove structure and dynamics. *J Am Chem Soc*. 2000; 122(43):10513–20. doi: [10.1021/ja0007071](https://doi.org/10.1021/ja0007071).
- [18] Panteva MT, Giambasu MG, York MD. Force Field for Mg<sup>2+</sup>, Mn<sup>2+</sup>, Zn<sup>2+</sup> and Cd<sup>2+</sup> Ions that have balanced interactions with nucleic acids. *J Phys Chem B*. 2015; 119(50):15460–70. doi: [10.1021/acs.jpcb.5b10423](https://doi.org/10.1021/acs.jpcb.5b10423).
- [19] Cheatham 3rd TE, Kollman PA. Observation of the A-DNA to B-DNA Transition During Unrestrained Molecular Dynamics in Aqueous Solution. *Journal of Molecular Biology*. 1996; 259(3):434–444. doi: [10.1006/JMBI.1996.0330](https://doi.org/10.1006/JMBI.1996.0330).
- [20] Mazur AK. Titration in silico of reversible B<->A transitions in DNA. *J Am Chem Soc*. 2003; 125(26):7849–59. doi: [10.1021/ja034550j](https://doi.org/10.1021/ja034550j).
- [21] Pastor N. The B- to A-DNA transition and the reorganization of solvent at the DNA surface. *Biophys J*. 2005; 88(5):3262–75. doi: [10.1529/biophysj.104.058339](https://doi.org/10.1529/biophysj.104.058339).
- [22] Mamatkulov S, Schwierz N. Force field for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties. *J Chem Phys*. 2018; 148(7):074505. doi: [10.1063/1.5017694](https://doi.org/10.1063/1.5017694).
- [23] Li J, Wang F. Pariwise-additive force fields for selected aqueous monovalent ions from adaptive force matching. *J Chem Phys*. 2015; 143(19):194505. doi: [10.1063/1.4935599](https://doi.org/10.1063/1.4935599).
- [24] Yoo J, Aksimentiev A. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys Chem Chem Phys*. 2018; (20):8432. doi: [10.1039/c7cp08185e](https://doi.org/10.1039/c7cp08185e).
- [25] Aqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. *Journal of Physical Chemistry*. 1990; 94(21):8021–8024. doi: [10.1021/j100384a009](https://doi.org/10.1021/j100384a009).
- [26] Smith DE, Dang LX. Computer simulations of NaCl association in polarizable water. *The Journal of Chemical Physics*. 1994; 100(5):3757. doi: [10.1063/1.466363](https://doi.org/10.1063/1.466363).
- [27] Zichi DA. Molecular Dynamics of RNA with the OPLS Force Field. Aqueous Simulation of a Hairpin Containing a Tetranucleotide Loop. *Journal of the American Chemical Society*. 1995; 117(11):2957–2969. doi: [10.1021/ja00116a001](https://doi.org/10.1021/ja00116a001).
- [28] Li P, Song LF, Merz KMJ. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J Chem Theory Comput*. 2015; 11(4):1654–1657. doi: [10.1021/ct500918t](https://doi.org/10.1021/ct500918t).
- [29] Joung IS, Cheatham 3rd TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B*. 2008; 112:9020–9041. doi: [10.1021/jp8001614](https://doi.org/10.1021/jp8001614).
- [30] Joung IS, Cheatham 3rd TE. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *The Journal of Physical Chemistry B*. 2008; 113(40):13279–13290. doi: [10.1021/jp902584c](https://doi.org/10.1021/jp902584c).
- [31] Noy A, Soteras I, Luque FJ, Orozco M. The impact of monovalent ion force field model in nucleic acids simulations. *Physical chemistry chemical physics*. 2009; 11(45):10596–607. doi: [10.1039/b912067j](https://doi.org/10.1039/b912067j).
- [32] Rueda M, Cubero E, Laughton CA, Orozco M. Exploring the counterion atmosphere around DNA: what can be learned from molecular dynamics simulations? *Biophysical journal*. 2004; 87(2):800–11. doi: [10.1529/biophysj.104.040451](https://doi.org/10.1529/biophysj.104.040451).
- [33] Song C, Xia Y, Zhao M, Liu X, Li F, Ji Y, Huang B, Yin Y. The effect of salt concentration on DNA conformation transition: a molecular-dynamics study. *Journal of Molecular Modeling*. 2006; 12(3):249–254. doi: [10.1007/s00894-005-0023-9](https://doi.org/10.1007/s00894-005-0023-9).
- [34] Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE. Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*. 1981; 78(4):2179–83.
- [35] Pérez A, Marchán I, Svozil D, Šponer J, Cheatham 3rd TE, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophysical Journal*. 2007; 92(11):3817–3829. doi: [10.1529/biophysj.106.097782](https://doi.org/10.1529/biophysj.106.097782).
- [36] Galindo-Murillo R, Roe DR, Cheatham 3rd TE. On the absence of intra-helical DNA dynamics on the micro to ms timescale. *Nature Communications*. 2014; 5:5152. doi: [10.1038/ncomms6152](https://doi.org/10.1038/ncomms6152).
- [37] Galindo-Murillo R, Roe DR, Cheatham 3rd TE. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAAACGAACGC). *Biochimica et biophysica acta*. 2014; 1850(5):1041–1058. doi: [10.1016/j.bbagen.2014.09.007](https://doi.org/10.1016/j.bbagen.2014.09.007).

- [38] Cheatham 3rd TE, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *Journal of Biomolecular Structure & Dynamics*. 1999; 16(4):845–862. doi: [10.1080/07391102.1999.10508297](https://doi.org/10.1080/07391102.1999.10508297).
- [39] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*. 1995; 117(19):5179–5197. doi: [10.1021/ja00124a002](https://doi.org/10.1021/ja00124a002).
- [40] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*. 1983; 79(2):926. doi: [10.1063/1.445869](https://doi.org/10.1063/1.445869).
- [41] Roe DR, Cheatham 3rd TE. PTraj and CPPTRAj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*. 2013; 9(7):3084–3095. doi: [10.1021/ct400341p](https://doi.org/10.1021/ct400341p).
- [42] Roe DR, Cheatham 3rd TE. Parallelization of CPPTRAj enables large scale analysis of molecular dynamics trajectory data. *Journal of Computational Chemistry*. 2018; doi: [10.1002/jcc.25382](https://doi.org/10.1002/jcc.25382).
- [43] Pastor RW, Brooks BR, Szabo A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics*. 1988; 65(6):1409–1419. doi: [10.1080/00268978800101881](https://doi.org/10.1080/00268978800101881).
- [44] Darden TA, York D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*. 1993; 98(12):10089. doi: [10.1063/1.464397](https://doi.org/10.1063/1.464397).
- [45] Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of chemical theory and computation*. 2012; 8(5):1542–1555. doi: [10.1021/ct200909j](https://doi.org/10.1021/ct200909j).
- [46] Izadi S, Anandakrishnan R, Onufriev AV. Building Water Models: A Different Approach. *The journal of physical chemistry letters*. 2014; 5(21):3863–3871. doi: [10.1021/jz501780a](https://doi.org/10.1021/jz501780a).
- [47] Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Juřečka P, Cheatham 3rd TE. Assessing the Current State of Amber Force Field Modifications for DNA. *Journal of Chemical Theory and Computation*. 2016; 12(8):4114–4127. doi: [10.1021/acs.jctc.6b00186](https://doi.org/10.1021/acs.jctc.6b00186).
- [48] Case DA, Cheatham 3rd TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *Journal Of Computational Chemistry*. 2005; 26(16):1668–1688. doi: [10.1002/jcc.20290](https://doi.org/10.1002/jcc.20290).
- [49] Shao J, Tanner WS, Thompson N, Cheatham 3rd TE. Clustering Molecular dynamics Trajectories: 1. Characterizing the performance of Different Clustering Algorithms. *J Che Theory Comput*. 2007; 3(6):2312–34. doi: [10.1021/ct700119m](https://doi.org/10.1021/ct700119m).
- [50] Lavery R, Maddocks JH, Pasi M, Zakrzewska K. Analyzing ion distributions around DNA. *Nucleic Acids Research*. 2014; 42(12):8138–8149. doi: [10.1093/nar/gku504](https://doi.org/10.1093/nar/gku504).
- [51] Lavery R, Moakher M, Maddocks JH, Petkeviute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research*. 2009; 37(17):5917–5929. doi: [10.1093/nar/gkp608](https://doi.org/10.1093/nar/gkp608).
- [52] Hawkins GD, Cramer CJ, Truhlar DG. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem*. 1996; (100):19824–19839. doi: [10.1021/jp961710n](https://doi.org/10.1021/jp961710n).
- [53] Salomon-Ferrer R, Götz AW, Poole D, Grand SL, Walker RC, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*. 2013; 9(9):3878–3888. doi: [10.1021/ct400314y](https://doi.org/10.1021/ct400314y).
- [54] Zgarbová M, Otyepka M, Šponer J, Lankaš F, Jurečka P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *Journal of Chemical Theory and Computation*. 2014; 10(8):3177–3189. doi: [10.1021/ct500120v](https://doi.org/10.1021/ct500120v).
- [55] Banas. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of Chemical Theory and Computation*. 2010; 6(12):3836–3849. doi: [10.1021/ct100481h](https://doi.org/10.1021/ct100481h).
- [56] Zgarbová M, Šponer J, Otyepka M, Cheatham 3rd TE, Galindo-Murillo R, Jurečka P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *Journal of chemical theory and computation*. 2015; 11(12):5723–5736. doi: [10.1021/acs.jctc.5b00716](https://doi.org/10.1021/acs.jctc.5b00716).
- [57] Bergonzo C, Cheatham 3rd TE. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *Journal of chemical theory and computation*. 2015; 11(9):3969–72. doi: [10.1021/acs.jctc.5b00444](https://doi.org/10.1021/acs.jctc.5b00444).
- [58] Nikolova E, Hashim Al-Hashimi M. Preparation, resonance assignment, and preliminary dynamics characterization of residue specific <sup>13</sup>C/<sup>15</sup>N-labeled elongated DNA for the study of sequence-directed dynamics by NMR. *Journal of Biomolecular NMR*. 2009; 1-2(45):9–16. doi: [10.1007/s10858-009-9350-y](https://doi.org/10.1007/s10858-009-9350-y).
- [59] Nikolova E, Gavin B, Loan A, Hashim Al-Hashimi M. Probing sequence-specific DNA flexibility in A-Tracts and pyrimidine-purine steps by NMR. *Biochemistry*. 2012; 43(51):8654–8664. doi: [10.1021/bi3009517](https://doi.org/10.1021/bi3009517).
- [60] Li H, Chowdhary J, Huang L, He X, MacKerell AD, Roux B, Roux B. Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids. *Journal of chemical theory and computation*. 2017; 13(9):4535–4552. doi: [10.1021/acs.jctc.7b00262](https://doi.org/10.1021/acs.jctc.7b00262).

**Appendix Table 4.** Selected helicoidal parameters calculated from a 10  $\mu$ s average structure for each system. Standard deviation calculated from the 10  $\mu$ s trajectory data. NMR reference data obtained from the PDB code 1NAJ. All the data is calculated using the central base-pairs only (3-10 and 15-22)

	Major	Std. Dev.	Minor	Std. Dev.	H-rise	Std. Dev.	H-twist	Std. Dev.	Stretch	Std. Dev.	Stagger	Std. Dev.	Buckle	Std. Dev.	Propeller	Std. Dev.	Slide	Std. Dev.	Rise	Std. Dev.	Roll	Std. Dev.	Twist	Std. Dev.	Inclination	Std. Dev.	Bend			
NMR ref	11.52	4.18	3.20	36.2	5.5	-0.30	0.11	0.08	0.38	-0.0	10.5	-12.5	10.0	-0.06	0.66	3.32	0.28	1.6	6.3	35.9	5.6	1.8	5.9	1.0	0.9					
LiCl	net	11.82	1.38	3.78	1.33	3.33	0.30	36.0	5.5	-0.03	0.11	0.08	0.38	-0.0	10.5	-12.5	9.9	-0.03	0.67	3.32	0.28	1.8	6.3	35.8	5.7	2.0	5.9	1.0	0.9	
	200mM	11.81	1.38	3.78	1.35	3.33	0.30	35.9	5.6	-0.04	0.11	0.07	0.38	-0.0	10.5	-12.5	10.0	-0.04	0.66	3.32	0.28	1.8	6.2	36.0	5.6	2.0	5.8	0.9	0.9	
	1M	11.71	1.37	3.76	1.33	3.33	0.30	36.1	5.6	-0.03	0.11	0.08	0.38	0.1	10.5	-12.6	10.0	-0.04	0.66	3.32	0.28	1.8	6.2	36.0	5.6	2.0	5.8	0.9	0.9	
	5M	11.43	1.37	3.88	1.33	3.31	0.30	35.7	5.6	-0.03	0.11	0.06	0.38	-0.4	10.5	-12.6	10.0	-0.02	0.6	3.31	0.28	2.4	6.2	35.6	5.6	3.2	5.8	0.6	0.9	
	net	11.84	1.38	3.78	1.35	3.34	0.30	36.1	5.6	-0.03	0.11	0.08	0.38	-0.1	10.5	-12.5	9.9	-0.06	0.67	3.32	0.28	1.6	6.3	36.0	5.6	1.7	5.9	0.9	0.9	
	200mM	11.75	1.37	3.76	1.35	3.33	0.30	35.9	5.6	-0.03	0.11	0.07	0.38	0.0	10.6	-12.5	10.1	-0.04	0.67	3.32	0.29	1.6	6.3	35.9	5.7	2.1	5.9	0.9	0.9	
	1M	11.75	1.37	3.81	1.36	3.33	0.30	35.8	5.6	-0.03	0.11	0.06	0.38	-0.1	10.5	-12.5	9.9	-0.02	0.68	3.31	0.29	1.9	6.3	35.7	5.7	2.3	5.9	1.0	0.9	
	5M	11.53	1.36	3.96	1.36	3.31	0.30	35.5	5.6	-0.03	0.11	0.05	0.38	-0.1	10.5	-12.5	9.8	-0.01	0.68	3.30	0.29	2.4	6.3	35.4	5.7	3.3	5.9	0.6	0.9	
	net	11.88	1.40	4.10	1.37	3.33	0.30	35.6	5.8	-0.03	0.11	0.05	0.38	0.2	10.6	-11.8	10.0	-0.10	0.67	3.32	0.29	2.3	6.2	35.5	5.8	3.0	5.8	0.9	0.9	
	200mM	11.82	1.39	4.04	1.37	3.33	0.30	35.8	5.7	-0.04	0.11	0.06	0.38	0.0	10.6	-12.3	9.9	-0.08	0.67	3.32	0.29	2.3	6.2	35.7	5.8	2.8	5.8	0.9	0.9	
	1M	11.73	1.38	3.91	1.35	3.33	0.29	36.0	5.7	-0.03	0.11	0.08	0.38	0.0	10.6	-12.7	9.9	-0.06	0.66	3.32	0.28	2.0	6.2	35.9	5.7	2.2	5.9	0.9	0.9	
	5M	11.71	1.37	3.52	1.34	3.29	0.30	35.3	5.7	-0.03	0.11	0.11	0.38	0.6	10.6	-12.7	9.0	-0.01	0.66	3.28	0.28	1.5	6.2	35.2	5.7	1.5	5.9	1.1	0.9	
	NaBr	net	11.92	1.40	4.10	1.38	3.33	0.30	35.7	5.7	-0.04	0.12	0.06	0.38	-0.1	10.6	-12.1	9.8	-0.10	0.67	3.33	0.29	2.3	6.2	35.6	5.8	2.9	5.8	0.9	0.9
	200mM	11.82	1.38	4.05	1.37	3.33	0.30	35.8	5.7	-0.03	0.11	0.07	0.38	-0.1	10.6	-12.3	9.8	-0.09	0.67	3.32	0.29	2.2	6.2	35.7	5.7	2.7	5.8	0.9	0.9	
	1M	11.68	1.39	3.91	1.38	3.31	0.30	35.5	5.9	-0.03	0.12	0.05	0.38	0.3	10.7	-12.0	10.2	-0.04	0.67	3.31	0.29	2.0	6.2	35.4	6.0	2.7	5.9	0.9	0.9	
	5M	11.91	1.39	3.80	1.38	3.30	0.30	35.4	5.9	-0.04	0.12	0.10	0.38	-0.1	10.7	-14.0	10.2	-0.05	0.67	3.29	0.29	2.3	6.2	35.2	6.0	2.7	5.9	1.0	0.9	
	KCl	net	11.92	1.39	3.91	1.31	3.33	0.30	35.8	5.8	-0.03	0.11	0.07	0.38	0.1	10.6	-12.1	10.2	-0.11	0.67	3.32	0.29	2.0	6.2	35.6	5.9	2.3	5.8	1.0	0.9
	200mM	11.80	1.39	3.88	1.32	3.32	0.30	35.6	5.7	-0.03	0.12	0.07	0.38	-0.2	10.6	-12.0	10.1	-0.08	0.67	3.31	0.29	2.0	6.2	35.5	5.9	2.3	5.9	1.0	0.9	
	1M	11.67	1.36	3.75	1.36	3.32	0.31	35.6	5.9	-0.03	0.11	0.09	0.38	-0.1	10.8	-12.5	10.0	-0.04	0.68	3.31	0.30	1.5	6.3	35.4	6.1	2.2	6.0	1.1	0.9	
	5M	11.46	1.36	3.40	1.36	3.30	0.31	35.7	5.9	-0.04	0.12	0.13	0.38	-0.1	10.1	-13.0	10.0	-0.01	0.68	3.29	0.30	1.3	6.3	35.6	6.1	0.9	6.0	1.2	0.9	
	KBr	net	11.95	1.40	3.93	1.33	3.33	0.30	35.7	5.7	-0.04	0.11	0.07	0.38	0.1	10.5	-12.1	9.9	-0.10	0.68	3.32	0.29	2.1	6.3	35.6	5.8	2.4	5.8	1.1	0.9
	200mM	11.85	1.38	3.87	1.31	3.33	0.30	35.9	5.6	-0.04	0.11	0.08	0.38	0.0	10.5	-12.3	9.9	-0.09	0.67	3.31	0.29	1.9	6.2	35.8	5.7	2.0	5.8	1.0	0.9	
	1M	11.70	1.37	3.73	1.33	3.33	0.31	35.8	5.7	-0.03	0.11	0.09	0.38	-0.1	10.7	-12.5	10.0	-0.04	0.67	3.31	0.29	1.7	6.3	35.7	5.8	1.8	5.9	1.0	0.9	
	5M	11.40	1.37	3.72	1.33	3.27	0.31	34.3	5.7	-0.02	0.11	0.09	0.38	-0.6	10.7	-13.0	10.0	-0.11	0.67	3.27	0.29	2.4	6.3	34.2	5.8	3.6	5.9	1.1	0.9	



**Appendix Figure 8.** Normalized cluster population analysis over time for the  $1\ \mu\text{s}$  and  $15\ \mu\text{s}$  simulations. Clustering was performed using the hierarchical agglomerative (bottom-up) approach focusing on the inner 8 residues (3-10 and 15-22). Each line represent one of the 10 calculated clusters for each simulation.