Article pubs.acs.org/JPCB

# Framework for Conducting and Analyzing Crystal Simulations of Nucleic Acids to Aid in Modern Force Field Evaluation

Sölen Ekesan<sup>®</sup> and Darrin M. York<sup>\*®</sup>

Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States

Supporting Information

ABSTRACT: Crystal simulations provide useful tools, along with solution simulations, to test nucleic acid force fields, but should be interpreted with care owing to the difficulty of establishing the environmental conditions needed to reproduce experimental crystal packing. These challenges underscore the need to construct proper protocols for carrying out crystal simulations and analyzing results to identify the origin of deviations from crystallographic data. Toward this end, we introduce a novel framework for B-factor decomposition into additive intramolecular, rotational, and translational atomic



fluctuation components and partitioning of each of these components into individual asymmetric unit and lattice contributions. We apply the framework to a benchmark set of A-DNA, Z-DNA, and B-DNA double helix systems of various chain lengths. Overall, the intramolecular deviations from the crystal were quite small ( $\leq 1.0$  Å), suggesting high accuracy of the force field, whereas crystal packing was not well reproduced. The present work establishes a framework to conduct and analyze crystal simulations that ultimately take on issues of crystal packing and can provide insight into nucleic acid force fields.

# ■ INTRODUCTION

The function of nucleic acids is intimately tied to their structure and dynamics.<sup>1</sup> Structural biology experiments such as X-ray crystallography,<sup>2</sup> NMR,<sup>3</sup> and cryo-EM<sup>4</sup> are powerful tools that provide data about nucleic acid structure and dynamics but require at least some level of modeling to interpret. Molecular simulations with realistic, physics based force field models are poised to aid in the interpretation of a wide range of experimental structural biology data, and provide detailed conformational ensembles of the system that can be used to predict not only structure but thermodynamics and kinetics.<sup>5–10</sup>

Currently, X-ray crystallography has been one of the primary sources of structural data for nucleic acids.<sup>11</sup> Crystal structures, therefore, often serve as the departure points for molecular dynamics (MD) simulation studies that further probe mechanistic details about the structure and dynamics in solution that are not easily inferred from the experimental data.<sup>12</sup> Crystal simulations play an important validation role to help deconstruct possible artifacts of the models from features of the simulations that reflect true relaxation of the structure from a crystalline to a solution environment and afford a mechanism to compare results directly with crystallographic data.<sup>13–19</sup> In the case of nucleic acids, crystal simulations of a B-DNA dodecamer<sup>20</sup> were first used to demonstrate that with rigorous treatment of long-range electrostatics, nucleic acids could be stably simulated without artificial restraints or charge scaling. More recently, crystal simulations of the hammer-<sup>1</sup> hairpin,<sup>22</sup> twister,<sup>23</sup> and TS<sup>24</sup> ribozymes have been head,<sup>2</sup>

used to validate the force field and molecular simulation protocol and helped lead to the identification of the most plausible active dynamical state in solution.

The value of molecular simulation results, however, relies heavily on the accuracy of the underlying force field. Due to the high charge of nucleic acid biopolymers, which requires special consideration of electrostatics, solvation, and ionic interactions, the development of molecular simulation force fields for DNA and RNA has faced significant challenges.<sup>25–29</sup> Traditionally, the accuracy of these force fields has lagged behind those for proteins. Recently, there has been a resurgence of effort to develop improved, next-generation force fields for nucleic acids, enabled largely by the concurrent evolution of specialized high-performance computing resources such as graphics processing units (GPUs) and the Anton supercomputer.<sup>9</sup> As these resources and algorithms become more generally accessible, it becomes feasible to integrate more detailed sets of experimental data into the force field development pipeline such that nucleic acid models can more rapidly advance.

Molecular simulations of nucleic acid crystals provide rich benchmark data from which modern force fields can be evaluated and refined.<sup>8,9,18,30,31</sup> In this regard, crystal simulations have the advantages that (1) statistical sampling is highly efficient as many copies of the nucleic acid can be

Received: December 11, 2018 Revised: March 8, 2019 Published: April 19, 2019

simulated without need for large solvent boxes, (2) information is gained for intramolecular and intermolecular interactions, both of which are important for higher order nucleic acid structures, (3) simulation results can be directly compared with X-ray crystallographic data. In this way, crystal simulations of nucleic acids will play an increasingly important role in the evolution of next-generation force fields for nucleic acids. However, it is critical to note that crystal simulations alone are far from sufficient to fully benchmark performance of force fields or assess their limitations. Caution should be made in interpreting data from crystal simulations, the crystal packing environment for which are especially sensitive to conditions such as temperature and crystallization agents<sup>32</sup> (the concentrations of which might not be known in the crystal itself), and may be further limited by unbalanced solutesolvent interactions and incomplete sampling.<sup>33</sup> These factors could produce crystal packing results that lead to potentially false positive or false negative conclusions.<sup>34</sup> Recently, a critical study of results from crystal simulations of a DNA dodecamer has been examined in the context of intermolecular forces that arise from crystal packing and crystallization agents.<sup>32</sup>

Toward that end, it is important to establish best practices and provide a framework from which consistent benchmark crystal simulations of nucleic acids can be performed and analyzed. In the present work, we describe methods for performing and analyzing crystal simulations of nucleic acids. We begin by describing the construction and preparation of the fundamental unit cell (UC) and its replication into an appropriately sized supercell (SC) for production simulation, collection of statistics, and comparison with crystallographic data. Next, we develop a novel B-factor decomposition scheme that allows contributions to atomic fluctuations to be separated into additive intramolecular, rotational, and translational components as well as partitioning of each of these components into individual asymmetric unit (e.g., duplex "molecule") and lattice (supercell) contributions. We then apply this framework to examine DNA duplex structure and fluctuations for benchmark A-, B-, and Z-form DNA systems using the AMBER OL15<sup>35</sup> DNA force field. We first examine helical parameters, base-pair (bp) interactions, torsion angle, and sugar pucker distributions. Following, we examine intermolecular crystal packing interactions by analyzing the preservation of native crystal contacts and using B-factor decomposition analysis. In this way, we can characterize the degree to which disruption of native packing leads to artificially large B-factor estimates. It should be pointed out that this analysis is not meant to be a critical assessment of the OL15 force field, as the many potential caveats described above for crystal simulations<sup>32</sup> have not been addressed. For these reasons, the simulation results presented here are expected to reasonably represent intramolecular structure and dynamics but likely do not well represent the intermolecular interactions that affect the crystal packing environment. Rather, the consistent framework for conducting and analyzing crystal simulations that we develop here is meant to capture and quantify these deficiencies and hence facilitate more in-depth studies of a broad range of crystals that could ultimately be used to benchmark nucleic acid force fields and assess their limitations.

# METHODS

All simulations were carried out using the AMBER16<sup>36</sup> package, employing the OL15<sup>35</sup> DNA force field, with

TIP4P/Ew<sup>37</sup> water model and corresponding Joung and Cheatham monovalent<sup>38</sup> and Li/Merz divalent<sup>39</sup> ion parameters, under periodic boundary conditions at 300 K. An 8 Å nonbond cut-off is applied, and the electrostatics beyond cut-off are accounted for by particle mesh Ewald.<sup>40</sup> Langevin thermostat with  $\gamma = 5 \text{ ps}^{-1}$  and Berendsen isotropic barostat with  $\tau = 1$  ps are used when necessary. Heavy-atom time step of 1 fs (as opposed to the more standard 2 fs, for consistency with other simulations performed in the lab such as quantum mechanics/molecular mechanics and certain free energy simulations) is used along with the SHAKE algorithm<sup>41</sup> for hydrogens.

The number of water molecules needed for each system is determined through multiple unconstrained NPT simulations until the system (or interpolation between two simulations) yields within 0.5% of experimental unit cell volume (Tables S2 and S4), which is comparable to the tolerance used in other crystal simulations reported in the literature.<sup>17</sup> It should be pointed out that NPT simulations of crystals are difficult to set up for the reason that packing solvent is a tedious process of trial and error, causing most crystal simulations to be performed in the NVT ensemble. Each system is solvent equilibrated in a multistage fashion through a cycle of minimization, NVT heating, and NPT equilibration where the constraint weight on the solute is approximately halved at each turn  $[50, 25, 10, 5, 2 \text{ kcal}/(\text{mol } \text{Å}^2)]$ , yielding in a total of ~100 ns solvent equilibration with fixed solute. Supercells are then constructed by propagating the solvent equilibrated unit cells and are further solvent equilibrated through the same cycle.

For each system, production simulations of supercells (SCs) with 12 unit cells (UCs) are run for 1.2  $\mu$ s with frames collected every 10 ps. The first 200 ns are discarded, and analyses are performed on the remaining 1  $\mu$ s of simulations. The structural indexes and distributions monitored in the present study are reasonably well converged on this time scale, and the structures are observed to be equilibrated and stably fluctuate after 200 ns, as discussed in the Results and Discussion Section. Drifting of the crystal is removed by reimaging the system around the solute center of mass, such that the center of mass of the system is fixed. For analyses at the molecular level (i.e., all except crystal contacts), each symmetry related molecule is mapped onto the asymmetric unit using reverse symmetry operations.

For each discussed property *P*, either an overall average,  $\langle P \rangle$ , or a molecular average,  $\langle P \rangle_{mol}$ , is reported.  $\langle P \rangle$  is calculated for the full ensemble of structures obtained from combining reverse symmetry applied to the time series for all DNA duplex trajectories, whereas  $\langle P \rangle_{mol}$  is average of P calculated separately for all individual DNA duplex trajectories. Root mean square deviations (RMSD), B-factor, native contacts, helical parameters, hydrogen bond distances, and torsional angle distributions are all calculated using corresponding cpptraj modules, with the exception of sugar puckers which were calculated using  $\nu_1$  and  $\nu_3$  angles.<sup>42</sup> Crystal contacts are calculated every 10 ps.

# RESULTS AND DISCUSSION

**Fundamentals of Crystal Simulations.** Constructing the Unit Cell (UC). Crystals have long-ranged order that involves replication of a unit cell, the fundamental repeating unit of the crystal. The unit cell may have internal symmetry dictated by the space group. The asymmetric unit is the

	1LJX <sup>49</sup>	137D <sup>48</sup>	5DN	5DNB <sup>43</sup>		119D <sup>51</sup>	1BNA <sup>52</sup>
conformation	Z	А	I	В		В	В
bp <sup>a</sup> in duplex	6	10	1	0	10	12	12
duplexes in UC	4	4	2	2		4	4
Xtal							
ion type	Mg <sup>2+</sup>		Mg <sup>2+</sup>		Mg <sup>2+</sup>	Mg <sup>2+</sup>	
ion count	8		1	4	8	2	
water count	232	404	30	08	614	545	320
Extra							
ion type	Mg <sup>2+</sup>	Na <sup>+</sup>	Na <sup>+</sup>	Mg <sup>2+</sup>	Mg <sup>2+</sup>	Mg <sup>2+</sup>	Mg <sup>2+</sup>
ion count	12	72	8	4	28	42	44
water count	250	697	224	228	430	570	1092
Total							
ions	20	72	22 <sup>b</sup>	18	36	44	44
water	482	1101	532	536	1044	1115	1412
atoms	3460	5887	2882	2890	5696	6429	7312
Unit Cell							
space group	$P2_{1}2_{1}2_{1}$	$P2_{1}2_{1}2_{1}$	C	22	$P2_{1}2_{1}2_{1}$	C2	$P2_{1}2_{1}2_{1}$
a (Å)	21.18	24.9	32.	.25	38.93	64.84	24.87
b (Å)	28.36	44.84	25.	.53	39.63	35.36	39.63
c (Å)	44.44	47.97	34	.38	33.3	25.35	33.3
$\alpha$ (deg)	90	90	9	0	90	90	90
$\beta$ (deg)	90	90	11	3.4	90	92.24	90
γ (deg)	90	90	9	0	90	90	90
volume (Å <sup>3</sup> )	26 700	53 560	25	980	51 375	58 080	66 500
volume/bp (Å <sup>3</sup> )	1113	1339	12	99	1284	1210	1385
resolution (Å)	1.64	1.7	1.	.4	1.5	2.25	1.9
collection $T(K)$	293	290	27	73	273	288	290 <sup>c</sup>

"The "bp" abbreviation stands for base pair. <sup>b</sup>Total of Na<sup>+</sup> and Mg<sup>2+</sup> ions. <sup>c</sup>Crystallization temperature, as collection temperature is not reported for this crystal.

symmetry-independent part of the unit cell that can be used to generate the complete unit cell from the space group symmetry operations and is typically the only coordinates that are deposited in crystallographic databases. In the present case, the complete unit cell contains molecules of DNA closely packed, with solvent and ions (typically between 30 and 70%) filling the interstitial spaces. It should be emphasized that the symmetry of the crystal, and indeed its periodicity, is a result of an ensemble average over all unit cells and time, i.e., perfect symmetry and periodicity will not be apparent from an instantaneous configuration taken from a snapshot in time. Hence, simulations that introduce instantaneous symmetry and/or periodicity with "periodic boundary conditions" alone are artificial even for a perfect crystal. These artifacts can be systematically reduced by considering larger supercells (SCs) created by replicating copies of the unit cell along with its lattice vector translations and considering all atoms in the supercell as independent degrees of freedom in solving the equations of motion. Although this procedure is theoretically more rigorous, by not constraining simulations to preserve either the internal space group symmetry or exact periodicity of the fundamental unit cell, freedom is provided for the packing and fluctuations in the crystal to break symmetry and diverge from their experimental values. This then provides a sensitive test of the simulation models to reproduce intermolecular interactions in the crystal that affect packing, which may also have relevance as models for tertiary interactions in higher order nucleic acid structures and complexes.

Constructing a model supercell is not without its challenges, as while the asymmetric unit contains the DNA molecule(s), typically only a small number of ordered solvent molecules are able to be resolved from crystallographic data. Hence, one must develop robust procedures to pack the interstitial space with water and ions. The only information available to guide this process is the experimental volume of the unit cell (and in rare cases the density of the crystal) and the identities of solution conditions used in crystallization. The precise composition of crystallization agents in the crystal itself is often unknown if these components are not resolved from the X-ray diffraction, as the conditions in the crystal will differ from those of the crystallization buffer. These conditions are nonetheless important for forming correct crystal packing and thus present a significant challenge to set up in simulations, as discussed in a recent in-depth study of a DNA dodecamer.<sup>3</sup> We do not explore these issues in depth for the broad range of DNA crystals under study, but rather confine our attention to overall packing of the crystal with solvent and counter ions so as to closely reproduce the unit cell volume. In the present study, we constructed UCs by keeping all of the ion and solvent atomic positions from the X-ray data, added counter ions to neutralize the system, and packed water molecules so as to preserve the experimental lattice vectors and unit cell volume. Simulated unit cell contents in the present study are summarized in Table 1.

Determining Optimal Size of the Supercell (SC). As mentioned previously, the use of SCs affords a mechanism to systematically eliminate artifacts of enforcing periodic boundary conditions to mimic an infinite periodic crystal of exact



**Figure 1.** Effect of supercell size on RMSDs and B-factors. Top and bottom panels show overall and molecular average properties, respectively. Molecular averages are obtained by averaging values of each symmetry related molecules. Each SC size is shown with a different color and is consistent in all panels. Squares and circles denote NVT and NPT, respectively. For NPT simulations with multiple SC propagations leading to the same SC size triangles are used for differentiation. RMSDs (left, Å) are with respect to the experimental structure. Standard deviations in  $(RMSD)_{mol}$  correspond to the variation among molecules. B-factors (right, Å<sup>2</sup>) are obtained from fluctuations in the simulation and are shown as residual averages alongside experimental data (black.) The bracket notation is used to maintain consistency with other sections.

cell replicas at every instant. In this sense, the larger the SC, the less pronounced are periodic boundary artifacts. Computational effort increases roughly linearly with the size of the SC, however, these requirements are at least partially offset by the increased number of independent DNA molecules that are sampled and contribute to statistics. Thus, it is important to find an optimum SC size that is sufficient for proper modeling of the crystal while being practical to compute.

To explore the effects of the SC size, we chose a test system of a 10 base-pair B-DNA double helix (PDB: 5DNB<sup>43</sup>) which has been well-studied previously in crystal environment.<sup>44-4</sup> We simulated this system with various SC sizes in the NPT and NVT ensembles for 250 ns (see the Supporting Information Table S2 for details about NPT volume fluctuations). We examined B-factors and root mean square deviations (RMSD) (see Figure 1) of all replicas of the asymmetric unit, which for this system is a single DNA strand. B-factors are calculated using the atomic fluctuations from the simulation and give a measure of dynamics, whereas RMSD values provide a measure of the structural variation of the simulation average from the experimental structure. For each property, we present results for the overall average (average over all symmetry-related molecules in the SC and time) and the molecular average (average of results obtained for individual molecules in the SC).

We have seen that the overall B-factors and both sets of RMSD values all initially increase with the SC size (Figure 1). This is an expected result as the system is given more degrees of freedom and an indicator of improved sampling with increasing SC size. As the SC size continues to increase, the cell-average RMSD values converge to roughly 0.7 Å with time-averages for individual molecules consistently higher around 0.8 Å. B-factor root mean square error values from the cell-average values increase monotonically with SC size (Table S3), whereas the time-average values stabilize more quickly. As will be explained in more detail below, the former is due to improper crystal packing with large SC sizes, whereas the latter indicates that intramolecular structure and fluctuations are relatively stable. Overall, the best balance occurs with a SC size of  $2 \times 2 \times 3$  (12 UC replicas) that affords sufficient freedom to allow molecules to repack and is the minimum propagation length ensuring that contacts in all directions are with independent molecules in the crystal (i.e., there are no self or neighbor image "intermolecular" interactions).

There are some variations observed between NVT and NPT ensembles. In general, the NVT ensemble gives results comparable to NPT for a single UC but shows more substantial deviations when applied to a larger  $2 \times 2 \times 1$  SC. Since the goal of this work is to provide benchmark simulation tests that are sensitive to errors in force fields, we will henceforth use the NPT ensemble which will allow volume fluctuations that can be compared with the experiment.

Decomposition of Crystal Fluctuations into Intramolecular and Intermolecular (Translational and Rotational) Components. Consider the atomic fluctuations arising from a crystal simulation of a system of identical DNA molecules arranged in a supercell, as described previously. Each molecule in the supercell contains independent atomic degrees of freedom (i.e., there are no explicit symmetry constraints). During dynamics, individual molecules will undergo internal (intramolecular) fluctuations, in addition to translational and rotational fluctuations about their equilibrium positions within the lattice. Further, since there are no explicit symmetry constraints, the supercell is free to break perfect symmetry of the original crystallographic lattice. This will cause the average structures of individual molecules, as well as their positions and orientations within the lattice, to deviate from one another. The result is that these deviations will yield larger positional fluctuations derived from the overall ensemble of lattice structures at each time point than the average fluctuations derived from the time series of individual molecules in the lattice.

Herein, we propose a simple but general framework for analysis of atomic fluctuations from crystal simulations. This framework enables decomposition of fluctuations into additive translational, rotational, and intramolecular components as well as the separation of each of these components into molecular (asymmetric unit) and lattice (supercell) contributions.

We first introduce the following notation. The overall fluctuations in atomic positions  $\mathbf{u}_i$  for atom *i* are derived from consideration of all replicated molecules in the supercell and all time points for each molecule. This ensemble can be constructed by applying reverse supercell translations and inverse symmetry operations so as to transform the atomic coordinates of each molecule to a common reference frame. Unless otherwise indicated, this common origin will be the center of mass origin of the fundamental asymmetric unit (i.e.,

the asymmetric unit from which all other asymmetric units in the unit cell are generated using forward symmetry operations). We designate the overall (total) atomic fluctuations derived from all molecules in the lattice as  $\langle \sigma_i^2 \rangle$ , where the use of brackets  $\langle \cdots \rangle$  indicates an average over the lattice ensemble that contains the transformed coordinates for all molecules at every sampled time point. Alternatively, we can consider the atomic fluctuations of individual molecules in the lattice and then average the individual time ensemble fluctuations for each molecule. We designate this fluctuation average as  $\langle \sigma_i^2 \rangle_{mol}$ , where we use the short-hand notation  $\langle \cdots \rangle_{mol}$  to indicate the sequential time ensemble average for each molecule, averaged over all molecules. We generically designate the difference between the atomic fluctuations derived from the lattice average  $\langle \cdots \rangle$  and molecular timeaverage  $\langle \cdots \rangle_{mol}$  atomic fluctuations as

$$\Delta \sigma_i^2 = \langle \sigma_i^2 \rangle - \langle \sigma_i^2 \rangle_{\text{mol}} \tag{1}$$

We now consider sequential transformations of the ensemble such that translational and rotational fluctuations are removed. Translational fluctuations can be removed from an ensemble by recentering each molecule of the ensemble to a common center of mass origin. Rotational fluctuations about a common center of mass origin can be removed from an ensemble by reorienting each molecule to a common reference frame. This can be achieved by finding the set of rotations that minimize the mass-weighted variance of atomic positions within the ensemble. Considering that the total fluctuations include intramolecular (I), rotational (R), and translational (T) components (i.e.,  $\langle \sigma_i^2 \rangle = \langle \sigma_i^2 \rangle^{\text{IRT}}$ ), we designate that the atomic fluctuations computed in the transformed ensembles with translational fluctuations removed as  $\langle \sigma_i^2 \rangle^{\text{IR}}$  and both translational and rotational fluctuations removed as  $\langle \sigma_i^2 \rangle^{\text{I}}$ .

Next, we consider a cluster expansion of the total lattice fluctuations as

$$\begin{split} \langle \sigma_i^2 \rangle &= \langle \sigma_i^2 \rangle^{\mathrm{I}} + \left( \langle \sigma_i^2 \rangle^{\mathrm{IR}} - \langle \sigma_i^2 \rangle^{\mathrm{I}} \right) + \left( \langle \sigma_i^2 \rangle^{\mathrm{IRT}} - \langle \sigma_i^2 \rangle^{\mathrm{IR}} \right) \\ &= \langle \sigma_{i,\mathrm{Intra}}^2 \rangle + \langle \sigma_{i,\mathrm{Rot}}^2 \rangle + \langle \sigma_{i,\mathrm{Trans}}^2 \rangle \end{split}$$
(2)

where

$$\langle \sigma_{i,\text{Intra}}^2 \rangle = \langle \sigma_i^2 \rangle^{\text{I}} \tag{3}$$

$$\langle \sigma_{i,\text{Rot}}^2 \rangle = (\langle \sigma_i^2 \rangle^{\text{IR}} - \langle \sigma_i^2 \rangle^{\text{I}})$$
(4)

$$\langle \sigma_{i,\mathrm{Trans}}^2 \rangle = \left( \langle \sigma_i^2 \rangle^{\mathrm{IRT}} - \langle \sigma_i^2 \rangle^{\mathrm{IR}} \right) \tag{5}$$

Note that  $\langle \sigma_{i,\text{Intra}}^2 \rangle$  corresponds to purely intramolecular atomic fluctuations from the ensemble having translational and rotational fluctuations removed, and  $\langle \sigma_{i,\text{Rot}}^2 \rangle$  and  $\langle \sigma_{i,\text{Trans}}^2 \rangle$  correspond to additional contributions that arise from pure rotations and translations. The decomposition above can also be applied to the molecular average fluctuations from the time ensembles  $\langle \sigma_i^2 \rangle_{\text{mol}}$  as

$$\langle \sigma_i^2 \rangle_{\rm mol} = \langle \sigma_{i,\rm Intra}^2 \rangle_{\rm mol} + \langle \sigma_{i,\rm Rot}^2 \rangle_{\rm mol} + \langle \sigma_{i,\rm Trans}^2 \rangle_{\rm mol} \tag{6}$$

with analogous definitions for the intramolecular, rotational, and translational components.

Similarly to eq 1, we define the differences between the atomic fluctuation components derived from the lattice average and molecular time-average atomic fluctuations as

$$\Delta \sigma_{i,\text{Intra}}^2 = \langle \sigma_{i,\text{Intra}}^2 \rangle - \langle \sigma_{i,\text{Intra}}^2 \rangle_{\text{mol}}$$
(7)

$$\Delta \sigma_{i,\text{Rot}}^{2} = \langle \sigma_{i,\text{Rot}}^{2} \rangle - \langle \sigma_{i,\text{Rot}}^{2} \rangle_{\text{mol}}$$
(8)

$$\Delta \sigma_{i,\text{Trans}}^2 = \langle \sigma_{i,\text{Trans}}^2 \rangle - \langle \sigma_{i,\text{Trans}}^2 \rangle_{\text{mol}}$$
(9)

The  $\Delta \sigma_{i,X}^2$  values directly report on the differences between the average molecule time ensembles and the overall lattice ensemble and thus provide insight into the degree to which the intramolecular structure, orientation, and position of the molecules deviate from one another and break the symmetry of the ideal crystal. The general framework developed here, thus, enables the following additive decomposition relations, illustrated in Scheme 1.

#### Scheme 1. Atomic Fluctuation Decomposition

$\langle \sigma_{i,Intra}^2 \rangle_{mol}$	+	$\langle \sigma_{i,Rot}^2 \rangle_{mol}$	+	$\langle \sigma_{i,Trans}^2 \rangle_{mol}$	=	$\langle \sigma_i^2 \rangle_{mol}$
+		+		+		+
$\Delta\sigma^2_{i,Intra}$	+	$\Delta\sigma_{i,Rot}^2$	+	$\Delta \sigma^2_{i,Trans}$	=	$\Delta\sigma_i^2$
				II		
$\langle \sigma_{i,Intra}^2 \rangle$	+	$\langle \sigma_{i,Rot}^2 \rangle$	+	$\langle \sigma_{i,Trans}^2 \rangle$	=	$\langle \sigma_i^2 \rangle$

Transformed Atomic Fluctuations of the Test System. To decompose fluctuations to all of the individual components mentioned above, the first step is to calculate the fluctuation ensembles  $\langle \sigma_i^2 \rangle^{\text{I}}$ ,  $\langle \sigma_i^2 \rangle^{\text{IR}}$ , and  $\langle \sigma_i^2 \rangle^{\text{IRT}}$  as well as their corresponding molecular counterparts ( $\langle \sigma_i^2 \rangle_{\text{mol}}^{\text{I}}$ ,  $\langle \sigma_i^2 \rangle_{\text{mol}}^{\text{IR}}$ , and  $\langle \sigma_i^2 \rangle_{\rm mol}^{\rm IRT}$ ). We simulated the same B-DNA system with 12 UC for 1.2  $\mu$ s, and monitored the RMSD time series for each asymmetric unit (molecule), as well as the RMSD time series for the instantaneous cell-average structure (i.e., the structure resulting from averaging the conformation of each of the molecules in the unit cell at the time point). Figure 2a shows the RMSD time series and indicates that the structures are well converged after 200 ns and fluctuate stably with average molecule RMSD values below 1.2 Å and instantaneous cellaverage RMSD values below 0.8 Å. In the remainder of the manuscript, sampling of fluctuations used for the B-factor decomposition and other statistical analyses are performed over the last 1  $\mu$  of simulation. The average atomic B-factors are calculated from the atomic fluctuations through the relation

$$\langle B_i \rangle = \frac{8\pi^2}{3} \langle \sigma_i^2 \rangle \tag{10}$$

where  $\langle B_i \rangle$  is the B-factor for atom *i* and  $\langle \sigma_i^2 \rangle$  is the corresponding atomic fluctuation. Figure 2b summarizes the B-factor results in two panels, with the top focusing on supercell level ( $\langle \cdots \rangle$ ) and the bottom on molecular level ( $\langle \cdots \rangle_{mol}$ ) fluctuations. For the overall system fluctuations, the values obtained from 1  $\mu$ s trajectory are almost same as the previous 250 ns. On the other hand, we see ~10 point increase in the  $\langle \sigma_i^2 \rangle_{mol}^{21}$  (or  $\langle B\text{-factors} \rangle_{mol}$ ) values. This can be explained by recognizing that although individual DNA duplexes sampled more conformations during the longer simulation, the differences between the duplexes did not change significantly.

Comparing the calculated B-factor decomposition results, one observes that the  $\langle \sigma_i^2 \rangle_{mol}^I$  (bottom, red) values correlate and overlap the best with the experimental values. Another is that the specific B-factor component does not affect the general



**Figure 2.** RMSD and B-factor analysis of 5DNB simulated in 12-unit SC. (a) Molecular average (black) RMSD and standard deviation (gray) shown alongside RMSD of the average trajectory (blue) over 1.2  $\mu$ s of simulation. The average trajectory is created by obtaining a structural average of all replicas at each frame. (b) B-factor decomposition for the last 1  $\mu$ s of simulation. Top (overall lattice,  $\langle \cdots \rangle$ ) and bottom (molecular averages,  $\langle \cdots \rangle_{mol}$ ) panels show experimental values (black, Å<sup>2</sup>) in comparison with intramolecular (I, red) center of mass corrected (IR, blue) and uncorrected (IRT, green) B-factors.

trends but rather produces an overall shift/amplification of the curves. These two results point out the strength of the force field in modeling intramolecular interactions but also bring to the surface the deficiencies of the current simulations in packing. These are discussed in greater detail for the full set of benchmark systems studied in the next sections.

**Benchmark DNA Crystal Systems.** In addition to the 5DNB system discussed in the previous section, we consider five additional crystal systems comprising A-DNA (137D<sup>48</sup>), Z-DNA (1LJX<sup>49</sup>), and three other B-DNAs (1D23,<sup>50</sup> 119D,<sup>51</sup> and 1BNA<sup>52</sup>), as summarized in Table 1. These systems were chosen for study based on their resolution, size, structural diversity, and near room-temperature data collection. For each system, we constructed supercells consisting of 12 UCs as described in the previous section and ran microsecond MD simulations for analysis (see Methods for details and Table S4 in the Supporting Information for the volume information).

In the remainder of the paper, we apply the protocols and analysis methods to characterize the structure, dynamics, and crystal packing of the benchmark set of crystal systems described above using the AMBER OL15<sup>35</sup> DNA force field. As stated previously, the goal is not to assess the force field itself, particularly with respect to intermolecular interactions and crystal packing that are sensitive to precise crystallization conditions. Rather the purpose is to develop and apply the framework for conducting and analyzing the simulations to set the stage for further study aimed at aiding in force field validation. We organize the analysis and discussion that follow in sections that examine separately intramolecular DNA duplex structure and fluctuations and intermolecular crystal packing interactions. We take a hierarchical approach for data reduction that uses average structures for overall metrics like RMSD. We report only average helical values and standard deviations for helical parameters that, as discussed below, almost uniformly show small standard deviations. We report full angular distributions (also indicating averages and standard deviations) for torsion degrees of freedom and sugar pucker, since these distributions can be multimodal.

DNA Duplex Structure and Fluctuations. In this section, we examine structure, fluctuations, and interactions within each DNA duplex. We start first by using the RMSD as an overall measure of the structural similarity between the simulation average structures and the crystal structures. Second, we compare helical parameters and their fluctuations to examine variations between different DNA duplex forms. Third, we examine duplex base-pair hydrogen bonding and compare with crystallographic values. Fourth, we examine the distributions of DNA backbone  $(\alpha-\zeta)$  and glycosidic bond  $(\chi)$  torsion angles, and sugar puckering phase (P) angles derived from the simulation, and compare residue averages with crystallographic values.

*RMSD of Average Structures.* We obtained average structures from combined 1  $\mu$ s trajectories and, for each system, examined root mean square deviation (RMSD) from experimental structures. Figure 3 shows simulation averages (colored) aligned and overlaid with experimental structures (gray), along with RMSD (Å) values. All of the simulation average structures were very close to the crystal structures and had RMSD values below 1.0 Å (shown in parentheses) except for A-DNA (137D), which shows some deviations at the ends



Figure 3. Overlay of experimental (gray) and simulation average (colored) structures for the six DNA systems. Heavy-atom RMSD (Å) and RMSD per base pair (Å/bp) are shown in parentheses and brackets, respectively.

Table 2. Helical Parameters for all DNA Systems Obtained from Molecular Dynamics Simulation (Sim) Versus Experiment (Expt)<sup>a</sup>

		Z	А		1	В	
		1LJX	137D	5DNB	1D23	119D	1BNA
rise (Å)	Expt	$3.6 \pm 0.2$	$3.3 \pm 0.2$	$3.3 \pm 0.2$	$3.4 \pm 0.2$	$3.3 \pm 0.3$	$3.4 \pm 0.2$
	Sim	$3.6 \pm 0.3$	$3.3 \pm 0.1$	$3.4 \pm 0.1$	$3.3 \pm 0.1$	$3.3 \pm 0.1$	$3.3 \pm 0.1$
tilt (deg)	Expt	$-0.7 \pm 1.0$	$-0.2 \pm 1.4$	$0.0 \pm 2.5$	$-0.2 \pm 1.9$	$1.6 \pm 4.0$	$-0.2 \pm 2.8$
-	Sim	$-0.2 \pm 2.2$	$-0.2 \pm 1.6$	$-0.1 \pm 1.6$	$0.0 \pm 1.3$	$0.5 \pm 2.7$	$0.0 \pm 1.9$
roll (deg)	Expt	$-2.4 \pm 2.6$	$6.3 \pm 4.9$	$2.2 \pm 5.8$	$0.0 \pm 4.4$	$2.2 \pm 3.3$	$0.0 \pm 5.4$
	Sim	$-2.8 \pm 2.9$	$10.4 \pm 4.1$	$2.0 \pm 3.0$	$1.1 \pm 3.8$	$2.3 \pm 3.3$	$0.2 \pm 3.4$
twist (deg)	Expt	26.7 ± 19.7	$30.7 \pm 4.3$	$35.2 \pm 8.7$	$37.1 \pm 5.0$	$35.5 \pm 5.1$	35.6 ± 4.9
	Sim	$27.5 \pm 14.6$	$29.3 \pm 1.8$	$34.4 \pm 6.7$	$36.2 \pm 5.6$	$35.3 \pm 6.3$	$35.5 \pm 3.6$
x-disp (Å)	Expt	$12.5 \pm 12.8$	$-4.5 \pm 1.5$	$0.5 \pm 1.8$	$0.2 \pm 1.0$	$0.3 \pm 1.1$	$0.1 \pm 0.8$
	Sim	9.6 ± 9.1	$-3.2 \pm 0.9$	$0.1 \pm 1.2$	$0.0 \pm 0.7$	$-0.1 \pm 1.1$	$-0.2 \pm 0.6$
y-disp (Å)	Expt	$1.2 \pm 1.8$	$-0.1 \pm 1.4$	$0.0 \pm 1.0$	$0.0 \pm 0.6$	$-0.2 \pm 1.0$	$0.0 \pm 0.9$
	Sim	$-0.3 \pm 1.6$	$0.0 \pm 0.5$	$0.0 \pm 0.4$	$-0.1 \pm 0.7$	$-0.2 \pm 0.8$	$0.0 \pm 0.7$
minor (Å)	Expt	11.7	$17.4 \pm 0.6$	$11.6 \pm 1.0$	$11.2 \pm 1.5$	$12.1 \pm 1.9$	$10.9 \pm 1.5$
	(Sim)	11.6	$17.1 \pm 0.7$	$11.2 \pm 0.8$	$11.0 \pm 1.5$	$11.4 \pm 1.8$	$10.6 \pm 1.0$
major (Å)	Expt	21.4	$18.3 \pm 1.2$	$17.9 \pm 2.0$	$17.4 \pm 0.6$	$16.9 \pm 1.5$	$17.5 \pm 0.5$
, , , ,	(Sim)	21.2	19.1 ± 1.1	$18.5 \pm 0.4$	$17.9 \pm 0.4$	$17.2 \pm 1.4$	$18.1 \pm 0.5$

<sup>a</sup>Simulation helical parameters are averages over every frame of the trajectory, except for minor and major grooves which are calculated on the simulation average structures (Sim). Both simulation and experimental values are reported as single average values for each system with standard deviations showing the magnitude of variation among residues.

that likely arise from the disruption of crystal packing (described in the Crystal Packing section below). Hence, overall, the average structure of each system was very close to the experimental crystal structure.

We see that the best result (lowest RMSD) is obtained with Z-DNA (1LJX, 0.57 Å) and one of the dodecamers (119D, 0.64 Å) and the worst with A-DNA (137D, 1.07 Å). However, since RMSD usually increases with the size of the molecule (due to the global fitting procedure), we felt it valuable to also report the average RMSD per base pair (values in brackets) to help normalize values for duplexes of different lengths. With the base-pair normalized values, we see that the dodecamers (119D and 1BNA) are closest to experiment with values of 0.05 and 0.06 Å/base pair and are very close with each other as well. Following the dodecamers are the two decamers (1D23 and 5DNB), where 1D23 is better than 5DNB. The worst two values are from the A-DNA 12-mer (137D, 0.11 Å/base pair) and Z-DNA 6-mer (1LJX, 0.10 Å/base pair).

Helical Parameters. Next, we expand our measure of overall structural features to helical parameters. We calculated helical parameters using the 3DNA software<sup>53</sup> for each system at each frame of the trajectories. Table 2 compares the experimental and simulation average values over all base pairs for each system, along with standard deviations. The average helical parameter values calculated in this way from the trajectories are essentially identical to values calculated from the corresponding simulation average structure (see Table S6 in the Supporting Information). For major and minor groove widths, we report the value calculated from the average structure alone.

Helical parameters calculated from simulation agree reasonably well with experimental values, especially for the rise distance where both the average values and the standard deviations are in close agreement. Similarly, both major and minor grooves are well modeled by the simulation average structures.

For twist angle, the average values agree well (less than  $1^{\circ}$ ) with experiment with only the A-DNA (137D) system being a minor exception (1.4° twist deviation). The residue standard deviations from the simulations are generally less than those of the crystal structures, suggesting that the simulations are producing slightly more uniform helices than the crystal structures. We see this reduced variation in the simulation results for other parameters as well.

For tilt angle, although most systems give good agreement with experiment, the values for 1LJX (Z-DNA 6-mer) and 119D (B-DNA dodecamer) are both predicted to be slightly closer to zero (less tilt) in the simulation. In the case of Z-DNA, the standard deviation is also higher than the experiment, which is opposite of the general trend seen in simulations. For roll angles, 137D (A-DNA) and 1D23 (B-DNA decamer) are overestimated by more than  $1^{\circ}$ , and for the A-DNA system the difference is about  $4^{\circ}$ .

In x- and y-displacements, the B-DNA systems agree well with the experimental values, but for the A-DNA and Z-DNA systems, there are clear discrepancies. The average value for xdisplacement is underestimated for A-DNA, as well as the standard deviation, but the major problems lie with the Z-DNA system. For both x- and y-displacements, Z-DNA simulations have issues, such that the x-displacement averages

	1LJX	Expt	Sim		5DNB	Expt	Sim		1D23	Expt	Sim
1	Т — А	2.91	2.92	1	C - G	2.87	2.85	1	C - G	2.82	2.88
2	G - C	2.88	2.87	2	C — G	2.91	2.88	2	G - C	2.92	2.89
3	C — G	2.93	2.87	3	А — Т	2.98	2.91	3	А — Т	2.96	2.92
4	G — C	2.91	2.87	4	А — Т	2.91	2.88	4	Т — А	2.86	2.89
5	C — G	2.92	2.91	5	C — G	2.84	2.88	5	C — G	2.82	2.88
6	А — Т	2.92	2.94	6	G — C	2.91	2.89	6	G - C	2.87	2.89
				7	Т — А	2.99	2.91	7	А — Т	2.93	2.95
				8	Т — А	2.89	2.89	8	Т — А	2.89	2.91
				9	G - C	2.89	2.86	9	C - G	2.82	2.89
				10	G - C	2.87	2.86	10	G - C	2.78	2.88
	137D	Expt	Sim		119D	Expt	Sim		1BNA	Expt	Sim
1	G - C	2.71	2.90	1	С — G	2.81	2.88	1	С — G	2.74	2.87
2	C — G	2.75	2.88	2	G — C	2.68	2.91	2	G — C	2.78	2.86
3	G — C	2.76	2.86	3	Т — А	2.67	2.96	3	C — G	2.72	2.88
4	G — C	2.82	2.88	4	А — Т	2.69	2.94	4	G — C	2.75	2.89
5	G — C	2.83	2.87	5	G — C	2.58	2.89	5	А — Т	2.95	2.95
6	C — G	2.75	2.85	6	А — Т	2.63	2.97	6	А — Т	3.11	2.98
7	C — G	2.74	2.86	7	Т — А	2.65	2.92	7	Т — А	2.97	2.96
8	C — G	2.76	2.86	8	C — G	2.75	2.91	8	Т — А	2.83	2.90
9	G — C	2.82	2.88	9	Т — А	2.62	2.94	9	C — G	2.77	2.90
10	C — G	2.74	2.89	10	А — Т	2.46	2.92	10	G — C	2.77	2.88
				11	C — G	2.71	2.88	11	C — G	2.85	2.89
				12	G - C	2.76	2.91	12	G - C	3.01	3.00
		a	verages		Expt			Sim			
		A	А — Т		$2.84 \pm 0.16$			$2.93 \pm 0.03$			
		C	C — G		$2.81 \pm 0.08$			$2.88 \pm 0.02$			

### Table 3. Donor-Acceptor Hydrogen Bond Distances (Å) for Simulated DNA Duplexes<sup>a</sup>

<sup>a</sup>Reported values are averages of base-pair hydrogen bonding heavy-atom distances. Experimental distances (Expt) are calculated from the PDB structure, whereas simulation distances (Sim) are from the overall simulation average structures.

are underestimated by almost 3 Å, whereas the *y*-displacement average is off by 1.5 Å and has the wrong sign.

For the 137D (A-DNA decamer) and particularly the 1LJX (Z-DNA 6-mer), there are minor to moderate deviations in helical parameters with respect to the crystallographic values. Nonetheless, overall, the experimental and simulation B-DNA structures have very similar helices.

*Hydrogen Bond Distances.* Hydrogen bonding between base pairs (base pairing) of nucleic acids is fundamental to their structure and biological function.<sup>54</sup> It is, therefore, essential for force fields to model these interactions accurately. Here, we calculated and compared hydrogen bond distances in the experimental and simulation average structures (Table 3).

Overall, simulation values are in close agreement with experiment, especially for the Z-DNA (1LJX) and two decamer B-DNA (5DNB and 1D23) systems, which happen to have co-axial base-stacking only. In the case of the two dodecamers, the differences are higher. In 1BNA, although differences are still less than 0.25 Å, simulation gives slightly longer distances than the experiment. This is more amplified in 119D where all simulation distances are longer than the experiment, and 7 of which (6 A–T and 1 C–G) are by more than 0.25 Å. This is mostly due to the fact that the experimental values for this system are significantly below the average hydrogen bond distances, the simulation values are very close to simulation values of other systems (and therefore the average simulation values.) A-DNA (137D), only C–G pairs, similarly produces longer distances than the experiment. Even though all of the

differences are less than 0.25 Å, hydrogen bonds are systematically  $\sim$ 0.1 Å longer.

We also calculated base-pair-specific average H-bond distances from all of the experimental and simulation average structures. The experimental structures exhibited moderate standard deviation (0.08 Å) value for C-G pairs and fairly large (0.16 Å) value for A-T pairs. The simulation average structures exhibit much smaller standard deviation values (0.2–0.3 Å). The experimental and simulation average distances for C-G and A-T pairs are quite close with a difference less than 0.1 Å, which is only marginally statistically significant provided the standard deviations and number of data points.

Torsional Angle Distributions. We calculated all backbone and  $\chi$  torsions of DNA along with the sugar pucker (i.e., phase angle, *P*) for each system. We show the distribution density (Figure 4) since some torsions have bimodal distributions and reporting only averages can be less informative.

Looking at all of the distributions, one torsional at a time (i.e., looking down a column), we can see clear patterns among different systems. For B-DNA, there are slight variations in the narrowness (seen as light gray smudges) but overall the distributions for all torsional angles are very similar, more so than their corresponding experimental values. For example, for  $\alpha$  angle, although SDNB, 1D23, and 1BNA all have very narrow distribution for the experimental values (as seen by the red arc denoting the standard deviation among residues), 119D has a much wider arc and blue ticks are more spread out. Yet, the simulation distribution for 119D is very similar to that



**Figure 4.** Simulated torsional angle distributions for the six DNA systems. Simulation distributions calculated over all frames are shown in shades of gray according to the population density. Experimental average and the standard deviation among residues are shown with red tick marks and arcs, respectively, individual values for residues are shown with blue tick marks. In every dial, the origin (zero) of angular axis is the horizontal right ("east") reference direction and increases moving counterclockwise.



Figure 5. Crystal packing of benchmark DNA systems. Each system is labeled by its PDB ID and space group. Molecules making up the unit cell are colored, and the repeating units are shown in gray. Each system is represented at a view angle portraying packing and stacking most clearly, layers of nonunit cell molecules are hidden to further enhance clarity.

4619

of 5DNB, 1D23, and 1BNA. More specifically, 119D never populates  $\alpha = -4^{\circ}$ , even though it is the experimental value for one of its residues. We call this situation "no representation", when the distribution from the simulation does not overlap so as to populate an experimental value. We see a few more no representations among two B-DNAs 119D and 1D23: in 119D,  $\gamma = -106^{\circ}$  and  $P = 95-125^{\circ}$  (3 residues); in 1D23  $\gamma = -173^{\circ}$ ,  $\delta = 63^{\circ}$  and  $P = 35^{\circ}$ . It needs to be noted that since these distributions combine torsions of all residues under one plot, residue-specific nuances are not distinguished.

Other than the few no representations mentioned above, torsional angle distributions of B-DNA are generally well replicated in the simulations. The same is not true for A- and Z-DNA. For these systems, there are a few instances where the experimental values are not populated. For the Z-DNA, these include  $\gamma = -60^{\circ}$ ,  $\delta = 48-52^{\circ}$  (3 residues),  $\epsilon = -172^{\circ}$  (2 residues), and  $\zeta = 44^{\circ}$  (2 residues). For A-DNA,  $\alpha = 143^{\circ}$  and  $\gamma = -177^{\circ}$  are values with no representation. Unlike B-DNA, the experimental distributions for these torsional angles in A- and Z-DNA are very narrow, and the no representation values correspond to singled out values.

The main issue with A- and Z-DNA, however, is not with the no representation set but a more problematic set of misrepresentation values, where the simulation significantly populates a region that does not have an experimental counterpart. We see this in Z-DNA for 80 <  $\delta$  < 100° along with a light shadow seen around  $\zeta = -80^{\circ}$ . In A-DNA, it is seen at 140 <  $\delta$  < 150°, -100 <  $\delta$  < -120°, and 140 < P < 160°. One thing to note about the misrepresentations seen in each of the three angles in A-DNA and  $\zeta$  of Z-DNA is that the misrepresentation region corresponds to highly populated regions of B-DNA. From the current simulations, one cannot tell whether these are due to temperature, misbalance of solvent, ions and crystallization agents, insufficient sampling, and/or force field deficiencies. Further studies of more diverse sets of A- and Z-form helices are required to draw general conclusions about the source of these deviations. Nonetheless, next-generation molecular simulation force fields for nucleic acids may need to pay close attention to parameters that distinguish A-DNA and Z-DNA from B-DNA helices.

Intermolecular Crystal Packing Interactions. The nucleic acid crystals studied here exhibit a diverse array of crystal packing (Figure 5). All studied systems except for the A-DNA  $(137D^{48})$  have some form of co-axial stacking of the helices, which is common in nucleic acid crystals. Z-DNA (1LJX<sup>49</sup>) and the two decamer B-DNAs (5DNB<sup>43</sup> and 1D23<sup>50</sup>) form uniform columns and behave like "infinite" DNA helices, where the only co-axial interaction between molecules is base stacking. The dodecamer B-DNA (119D<sup>51</sup> and 1BNA<sup>52</sup>) have zig-zag co-axial stacking induced by the special intermolecular interactions denoted as Dickerson Interactions (DI)<sup>55</sup> where sets of three base pairs at both ends hydrogen bond with the consecutive molecules. The A-DNA decamer  $(137D^{48})$  is an orthorhombic crystal with helices that, unlike the other crystals, are not co-axially stacked, but rather form a more complex packing arrangement with the reference duplex interacting with two molecules that are asymmetrically tilted away from the center of the duplex. It has been pointed out that the packing arrangement of the crystalline environment can strongly influence local DNA conformation and helix parameters.

These crystals, thus, present a wide range of crystal packing arrangements, in addition to sharing some common features, that enable us to make a critical assessment of the ability of our current simulations to model intermolecular interactions between DNA helices. Interactions similar to these are important for modeling tertiary or quaternary interactions in more complex nucleic acid systems and macromolecular assemblies.<sup>56–59</sup> Here, we analyze results and characterize the degree to which our current simulations are able to model this diverse array of packing interactions by comparing the preservation of native contacts and examining the origins of discrepancies with crystallographic data through B-factor decomposition analysis.

*Crystal Contacts during Simulation.* To analyze the conservation of crystal contacts during simulations, we monitored the number of native intermolecular contacts for each residue. Specifically, we defined a contact as any pairwise atomic distance shorter than 3.5 Å between an atom of the residue and any atom of a neighboring duplex. The contact distance cut-off was chosen to represent roughly the vdW contact distance between carbons (1.7 Å). We report the calculated native crystal contacts for each residue where they exist alongside the simulation averages for all systems in Table 4.

Table 4. Conservation of Native Contacts in Simulation<sup>a</sup>

1LJX (5.2)					5DNB (3.0)				1D23 (3.4)				
R	es	Sim	Native	R	es	Sim	Native	]	Res	Sim	Native		
Α	1	$2.1\pm1.0$	6	A	1	$1.8 \pm 0.6$	8	A	1	$2.3\pm1.1$	12		
Α	3	$0.2 \pm 0.2$	3	Α	4	$0.1\pm0.1$	2	Α	3	$0.2\pm0.2$	2		
Α	4	$2.8 \pm 0.7$	6	Α	10	$2.8 \pm 0.7$	12	Α	4	$0.0 \pm 0.1$	1		
Α	6	$5.3 \pm 1.3$	18	B	11	$2.6 \pm 1.0$	14	Α	6	$1.0\pm0.5$	4		
• B	7	$2.9 \pm 0.9$	-15	В	13	$0.2\pm0.2$	2	Α	8	$0.2 \pm 0.2$	1		
в	8	$0.9\pm0.4$	2	В	14	$0.3 \pm 0.2$	2	Α	10	$5.1\pm1.3$	14		
В	10	$1.0 \pm 0.4$	3	В	20	$3.8 \pm 1.0$	19	B	11	$2.9 \pm 1.0$			
в	12	$3.3 \pm 1.1$	9					В	12	$0.2 \pm 0.2$	3		
								В	13	$0.1 \pm 0.1$	1		
137D (3.2)				119D (4.6)		В	16	$1.7\pm0.3$	6				
R	es	Sim	Native	R	es	Sim	Native	В	17	$0.2 \pm 0.2$	1		
Α	1	$1.1 \pm 0.9$	7	Α	1	$4.6 \pm 1.4$	13	В	20	$5.5 \pm 1.2$	14		
Α	2	$0.3 \pm 0.4$	1	Α	2	$4.4\pm0.7$	9						
Α	3	$0.6\pm0.5$	1	Α	3	$4.6\pm1.1$	12	1BNA (2.2)					
Α	4	$0.6\pm0.8$	5	Α	4	$0.4 \pm 0.2$	1	]	Res	Sim	Native		
Α	5	$0.0 \pm 0.1$	3	Α	6	$0.3 \pm 0.3$	1	A	1	$0.7 \pm 0.4$	3		
Α	7	$0.4 \pm 0.5$	3	Α	10	$1.5 \pm 0.3$	4	Α	2	$2.2\pm0.5$	3		
Α	8	$0.5 \pm 0.3$	6	Α	11	$0.6 \pm 0.3$	1	Α	3	$2.0\pm0.7$	8		
Α	10	$1.5 \pm 0.6$	5	Α	12	$8.3\pm1.1$	15	Α	8	$0.5\pm0.4$	2		
B	11	$1.4 \pm 1.3$		B	13	$\overline{4.6} \pm \overline{1.9}$	14	Α	12	$5.2 \pm 1.1$	8		
В	12	$0.2 \pm 0.3$	3	В	14	$4.7 \pm 0.7$	7	B	14	$\overline{2.3} \pm \overline{0.3}$			
В	13	$0.5 \pm 0.3$	4	В	15	$4.6\pm1.0$	7	В	15	$2.1\pm0.5$	5		
В	14	$0.2 \pm 0.2$	2	в	19	$0.3 \pm 0.2$	1	В	16	$2.2 \pm 0.8$	8		
В	15	$0.4 \pm 0.6$	3	В	20	$0.3 \pm 0.3$	4	В	17	$0.4 \pm 0.2$	1		
В	16	$0.3 \pm 0.3$	2	В	22	$0.8 \pm 0.6$	1	В	22	$0.9\pm0.2$	1		
в	18	$0.5 \pm 0.4$	2	В	23	$1.8 \pm 0.8$	6	В	24	$6.9\pm1.6$	10		
В	19	$1.3\pm0.6$	2	В	24	$8.8\pm1.5$	14						

<sup>a</sup>Number of contacts each residue makes in the native crystal and fraction of contacts observed over the simulations are presented. Cutoff distance for defining a contact is taken as 3.5 Å. For each system, the number of native contacts per residue are shown in parentheses next to PDB ID. Simulation contacts are reported as fraction averages over symmetry related duplexes along with standard deviations. Dashed lines are used to separate the two strands within a helix.

Looking at native contacts alone gives a sense of both the compactness and the packing motifs of the crystals. The number of total native contacts and the corresponding per residue values (shown in parentheses next to PDB IDs) are good indicators of how densely the crystals pack and correlate well with the volume per base-pair (bp) values (Table 1), except for a switch in the order between 5DNB and 137D. As for the crystal packing motifs, the axial base-stacking systems Z-DNA (1LJX) and two decamer B-DNAs (5DNB and 1D23) have the highest number of contacts at the ends of the duplex

helices with much fewer contacts in between. The dodecamer B-DNAs (119D and 1BNA) also have a high number of contact at the ends of the helices, in addition to residues flanking the ends. On the other hand, A-DNA (137D) has a much broader contact distribution, which reflects its packing arrangement that does not involve co-axial stacking.

Simulation averages portray a drastic loss of native crystal contacts through the simulations due to deviations from the native crystal packing. There is no clear pattern of contact loss among different regions of the molecules, but this can be explained by considering that a slight shift of the molecular orientation that disrupts a contact will propagate to residues involved in other contacts (since only native contacts are monitored). Overall, the best-conserved system (1BNA) has only 48% of its contacts preserved in the simulation, and the worst conserved system is for A-DNA (137D) with a mere 17% preserved native contacts. With the exception of 5DNB (20%), we see similar percentage conservation for systems that have similar packing motifs. The two dodecamers with DI interactions (119D and 1BNA) are at the 45% range, and the remaining two of the co-axial stacking systems (1LJX and 1D23) are at 30%. These values suggest that despite having an overall good description of intramolecular structure and fluctuations with the AMBER OL15<sup>35</sup> DNA force field, our simulations fail to maintain the overall intermolecular contacts and crystal packing environment.

*B-factor Decomposition Analysis.* We performed B-factor decomposition analysis following the framework introduced in Scheme 1 for all systems (Table 5). At the single molecule

Table 5. Decomposition of B-factors  $(Å^2)$  for all DNA Systems Following the Set up Shown in Scheme 1

	Intra	$\operatorname{Rot}$	Trans	Total	1LJX		Intra	$\operatorname{Rot}$	Trans	Total	5DNB
$\langle \cdots \rangle_{mol}$	9.9	4.6	6.3	20.8		$\langle \cdots \rangle_{mol}$	16.6	8.6	11.0	36.2	
	32%	15%	20%	67%			29%	15%	19%	63%	
$\Delta$	3.5	3.5	3.0	10.0		$\Delta$	4.4	9.0	8.2	21.5	
+	11%	11%	10%	33%	Expt	+	8%	16%	14%	37%	$\operatorname{Expt}$
$\langle \cdots \rangle$	13.3	8.1	9.3	30.8	15.9	$\langle \cdots \rangle$	21.0	17.5	19.2	57.7	13.9
	43%	26%	30%	100%			36%	30%	33%	100%	
	Intra	$\operatorname{Rot}$	Trans	Total	137D		Intra	$\operatorname{Rot}$	Trans	Total	1D23
$\langle \cdots \rangle_{mol}$	32.4	15.9	27.3	75.6		$\langle \cdots \rangle_{mol}$	16.1	6.6	13.3	36.1	
	25%	12%	21%	58%			30%	12%	24%	66%	
$\Delta$	18.4	12.8	23.2	54.4		$\Delta$	3.2	3.2	12.1	18.5	
+	14%	10%	18%	42%	Expt	+	6%	6%	22%	34%	Expt
$\langle \cdots \rangle$	50.8	28.7	50.5	129.9	23.1	$\langle \cdots \rangle$	19.3	9.8	25.4	54.5	12.9
	39%	22%	39%	100%			35%	18%	47%	100%	
	Intra	$\operatorname{Rot}$	Trans	Total	119D		Intra	$\operatorname{Rot}$	Trans	Total	1BNA
$\langle \cdots \rangle_{mol}$	11.1	2.4	4.0	17.6		$\langle \cdots \rangle_{mol}$	21.9	13.0	29.4	64.4	
	25%	5%	9%	39%			25%	15%	33%	72%	
$\Delta$	8.4	7.2	11.5	27.1		$\Delta$	6.5	6.6	11.6	24.7	
+	19%	16%	26%	61%	Expt	+	7%	7%	13%	28%	Expt
$\langle \cdots \rangle$	19.5	9.6	15.5	44.6	23.1	$\langle \cdots \rangle$	28.4	19.6	41.0	89.1	37.3
	44%	22%	35%	100%			32%	22%	46%	100%	

level, intramolecular, rotational, and translational fluctuation components give a sense for average fluctuations of individual duplexes around their lattice positions, whereas at the (super)cell level, these components incorporate the additional fluctuations introduced by considering the structural ensemble arising from all independent duplexes. In an ideal crystal simulation where average symmetry and crystal packing were preserved, and each duplex sampled the same conformational space, the  $\Delta \sigma_{i,X}^2$  values would be zero. The  $\Delta \sigma_{i,X}^2$  track the fluctuation differences between the average molecule (duplex) time ensembles and the overall lattice ensemble. However, we see that this is far from the case for the current simulations. For most systems, the  $\Delta \sigma_i^2$  component makes up 30–40% of total

simulation fluctuations, reaching as high as 61% for 119D, and alone is higher than the experimental B-factor in few cases.

Looking at percent contributions to total fluctuations, we see some trends among different systems. For all systems,  $\langle \sigma_{i,\mathrm{Intra}}^2 
angle_{\mathrm{mol}}$  values are within 25–32%, and except for the dodecamers (119D and 1BNA), have the highest contribution to the total fluctuations. In general,  $\langle \sigma_{i,\text{Rot}}^2 \rangle_{\text{mol}} < \langle \sigma_{i,\text{Trans}}^2 \rangle_{\text{mol}}$ with  $\langle \sigma_{i,\text{Rot}}^2 \rangle_{\text{mol}}$  10–15% and  $\langle \sigma_{i,\text{Trans}}^2 \rangle_{\text{mol}}$  19–33%, where 119D is again an exception with its 5%  $\langle \sigma_{i,\text{Rot}}^2 \rangle_{\text{mol}}$  and 9%  $\langle \sigma_{i,\text{Trans}}^2 \rangle_{\text{mol}}$ contributions. In terms of  $\Delta$  values, we see that  $\Delta \sigma_{i,\text{Intra}}^2$  is relatively small with 6–19% contribution, and  $\Delta\sigma^2_{i,\mathrm{Trans}}$  is the largest of the set in systems except for Z-DNA (1LJX) and SDNB. Finally,  $\langle \sigma_{i,X}^2 \rangle$  results yield in  $\langle \sigma_{i,\text{Intra}}^2 \rangle$  having a range of 32–44%, with  $\langle \sigma_{i,\text{Rot}}^2 \rangle$  18–30% and  $\langle \sigma_{i,\text{Trans}}^2 \rangle$  30–46%. Overall, the B-values calculated from the intramolecular  $\langle \sigma_{i,\text{Intra}}^2 \rangle_{\text{mol}}$ fluctuations have comparable magnitude to the experimentally derived values, but these fluctuations represent less than 1/3 of the full simulated fluctuations. The B-value components arising from translational fluctuations around a lattice site  $(\langle \sigma_{i,\text{Trans}}^2 \rangle_{\text{mol}})$  and between lattice sites  $(\langle \sigma_{i,\text{Trans}}^2 \rangle)$  have the next largest contributions, which reflects defects of packing.

Comparing simulation B-factors with experimental values, we see that even the intramolecular fluctuations alone, either at molecular or supercell levels, are enough for half of the systems to overshoot the experimental B-factors. Only the Z-DNA (1LJX) and the two dodecamer B-DNAs (119D and 1BNA) have intramolecular B-factor fluctuations ( $\langle \sigma_{i,\text{Intra}}^2 \rangle$ ) lower than the experimental values. From the molecular level fluctuations ( $\langle \sigma_i^2 \rangle_{\text{mol}}$ ), only 119D has B-factor values  $\langle \sigma_i^2 \rangle_{\text{mol}}$  less than experimental B-factors. The disruption of crystal packing in the simulations, thus, leads to greatly exaggerated fluctuations when considering the ensemble of conformations arising from the supercell, including rotational and translational components.

One interesting result is that similar systems did not necessarily yield similar trends in B-values, the most striking example being the dodecamers where 119D gives B-factors much closer to experiment relative to 1BNA. These differences can be traced, in part, to differences in the overall tightness of the crystal packing environment. A quantitative measure of packing tightness is the volume per base pair. We saw that 119D, the better behaving system, had 10% lower base-pair volume compared to the other dodecamer 1BNA and overall is the second smallest after the Z-DNA (1LJX). The two decamers had similar midrange volume, and A-DNA had the highest (see Table 1). This order agrees with the trend we see in the overall system B-factors, suggesting that the less tightly packed, the more solvent there is in the unit cell, and thus the more freedom there is to lose crystal contacts and break symmetry causing large fluctuations.

#### CONCLUSIONS

Crystal simulations are valuable tools to aid in assessing the reliability of nucleic acid force fields used to gain predictive insight into complex biological problems. Crystal simulations have the advantages that they provide dense statistical sampling (many copies of the asymmetric unit) at relatively low cost and afford a mechanism to compare directly to observables from X-ray crystallographic data. Further, crystal simulations not only predict intramolecular structure and dynamics but also serve as a sensitive probe of intermolecular interactions that may be important for modeling nucleic acid tertiary interactions. It should be emphasized that crystal

simulations alone are not sufficient for force field assessment, and caution must be taken in interpreting data from these simulations which are sensitive to environmental conditions, including temperature and crystallization agents that might not be known in the crystal, in addition to sampling requirements. In the present work, we establish a framework from which consistent, benchmark crystal simulations of nucleic acids can be performed and analyzed. We examine a benchmark set of four B-DNA, one A-DNA, and one Z-DNA crystals using the AMBER OL15 DNA force field with TIP4P/Ew water model and balanced ion parameters. As supercell size was increased, observed fluctuations from the full lattice ensemble initially increased but reasonably converged with 12 replicas of the fundamental unit cell. Intramolecular structure and fluctuations, as depicted by monitoring helical parameters, base pairing, and torsion angle/sugar puckering profiles, were overall quite good, with the best agreement with crystallographic data obtained for the B-DNA systems and considerably worse agreement obtained for the A- and Z-DNA systems. A novel B-factor decomposition scheme was introduced that enables contributions to atomic fluctuations to be separated into additive intramolecular, rotational, and translational components as well as partitioning of each of these components into individual duplex molecule (asymmetric unit) and lattice (supercell) contributions. This framework was applied to study fluctuations arising from the structural ensemble in the supercell and helped to pinpoint artifacts that arise as a consequence of improper crystal packing. Overall, the intramolecular deviations from the crystal were quite small (typically less than 1.0 Å), suggesting relative high accuracy of the force field, whereas crystal packing was not well reproduced in these simulations. These simulation results, however, do not represent a critical assessment of the force field, as the challenges of addressing environmental conditions and factors that impact crystal packing were not explored, and hence one cannot make conclusions about the origin of observed artifacts. Rather, the framework developed in the current work enables a systematic mechanism to conduct and analyze crystal simulations that do take on these issues. Analysis of such simulations would provide insight into the force field balance between solute, solvent, salt, and crystallization agents that are expected to much more significantly impact interactions between different molecules than within individual molecules.

# ASSOCIATED CONTENT

#### **S** Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b11923.

Tabulations of the dependence of RMSD on supercell size; volume fluctuations of each supercell simulated; root mean square error and linear correlation coefficients for B-factors; volume fluctuations for all systems; and average structure RMSD for all systems (PDF)

### AUTHOR INFORMATION

## **Corresponding Author**

\*E-mail: Darrin.York@rutgers.edu.

# ORCID 🔍

Şölen Ekesan: 0000-0002-5598-5754 Darrin M. York: 0000-0002-9193-7055

#### Notes

The authors declare no competing financial interest.

# ACKNOWLEDGMENTS

The authors are grateful for financial support provided by the National Institutes of Health (GM107485 and GM62248). Computational resources were provided by the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey, the National Institutes of Health under Grant No. S10OD012346, and by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation (ACI-1548562 and OCI-1053575). The authors thank Trich E. Chidae (Crystal River, FL) for engaging interactions and motivation for this work.

#### REFERENCES

(1) Saenger, W. Principles of Nucleic Acid Structure; Springer: NY, 1984.

(2) Egli, M. Nucleic acid crystallography: current progress. Curr. Opin. Chem. Biol. 2004, 8, 580–591.

(3) Mollova, E. T.; Pardi, A. NMR solution structure determination of RNAs. *Curr. Opin. Struct. Biol.* **2000**, *10*, 298–302.

(4) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **2008**, *16*, 673–683.

(5) Chen, A. A.; García, A. E. High-resolution reversible foling of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 16820–16825.

(6) Lemkul, J. A.; MacKerell, A. D., Jr. Polarizable force field for DNA based on the classical Drude oscillator: II. Microsecond molecular dynamics simulations of duplex DNA. *J. Chem. Theory Comput.* **2017**, *13*, 2072–2085.

(7) Zgarbová, M.; Jurečka, P.; Lankaš, F.; Cheatham, T. E., III; Šponer, J.; Otyepka, M. Influence of BII backbone substates on DNA twist: a unified view and comparison of simulation and experiment for all 136 distinct tetranucleotide sequences. *J. Chem. Inf. Model.* **2017**, *57*, 275–287.

(8) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J. P.; Ponder, J. W.; Ren, P. AMOEBA polarizable atomic multipole force field for nucleic acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.

(9) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 1346–1355.

(10) Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; et al. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **2018**, *118*, 4177–4338.

(11) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(12) Panteva, M. T.; Dissanayake, T.; Chen, H.; Radak, B. K.; Kuechler, E. R.; Giambaşu, G. M.; Lee, T.-S.; York, D. M. In *Multiscale Methods for Computational RNA Enzymology*; Chen, S.-J., Burke-Aguero, D. H., Eds.; Elsevier, 2015; Chapter 14.

(13) York, D. M.; Darden, T. A.; Pedersen, L. G.; Anderson, M. W. Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. *Biochemistry* **1993**, *32*, 1443–1453.

(14) York, D. M.; Wlodawer, A.; Pedersen, L. G.; Darden, T. Atomic level accuracy in simulations of protein crystals. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8715–8718.

(15) York, D. M.; Darden, T.; Pedersen, L. G. The Effect of Hydrostatic Pressure on Protein Crystals Investigated by Molecular Simulation. In *Modelling of Biomolecular Structures and Mechanisms*; Springer: Dordrecht, 1995.

(16) Cerutti, D. S.; Freddolino, P. L.; Duke, R. E., Jr.; Case, D. A. Simulations of a protein crystal with a high resolution X-ray structure:

evaluation of force fields and water models. J. Phys. Chem. B 2010, 114, 12811-12824.

(17) Janowski, P. A.; Cerutti, D. S.; Holton, J.; Case, D. A. Peptide crystal simulations reveal hidden dynamics. J. Am. Chem. Soc. 2013, 135, 7938–7948.

(18) Liu, C.; Janowski, P. A.; Case, D. A. All-atom crystal simulations of DNA and RNA duplexes. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1059–1071.

(19) Janowski, P. A.; Liu, C.; Deckman, J.; Case, D. A. Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. *Protein Sci.* **2016**, *25*, 87–102.

(20) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. Toward the accurate modeling of DNA: the importance of long-range electrostatics. *J. Am. Chem. Soc.* **1995**, *117*, 5001–5002.

(21) Martick, M.; Lee, T.-S.; York, D. M.; Scott, W. G. Solvent structure and hammerhead ribozyme catalysis. *Chem. Biol.* **2008**, *15*, 332–342.

(22) Heldenbrand, H.; Janowski, P. A.; Giambaşu, G.; Giese, T. J.; Wedekind, J. E.; York, D. M. Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. J. Am. Chem. Soc. **2014**, *136*, 7789–7792.

(23) Gaines, C. S.; York, D. M. Ribozyme catalysis with a twist: active state of the twister ribozyme in solution predicted from molecular simulation. *J. Am. Chem. Soc.* **2016**, *138*, 3058–3065.

(24) Gaines, C. S.; York, D. M. Model for the functional active state of the TS ribozyme from molecular simulation. *Angew. Chem., Int. Ed.* **2017**, *129*, 13577–13580.

(25) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771–1782.

(26) Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T. E., III; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; et al.  $\mu$ ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272–12283.

(27) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., III Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578–1590.

(28) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E., III Assessing the current state of AMBER force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127.

(29) Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; Gonzalez, C.; Orozco, M. How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* **2017**, *45*, 4217–4230.

(30) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Jr. Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348–362.

(31) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., Jr. An empirical polarizable force field based on the classical Drude oscillator model: development history and recent applications. *Chem. Rev.* **2016**, *116*, 4983–5013.

(32) Kuzmanic, A.; Dans, P. D.; Orozco, M. An in-depth look at DNA crystals through the prism of molecular dynamics simulations. *Chem* **2019**, *5*, 649–663.

(33) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1041–1058.

(34) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.

(35) Zgarbová, M.; Sponer, J.; Otyepka, M.; Cheatham, T. E., III; Galindo-Murillo, R.; Jurecka, P. Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. J. Chem. Theory Comput. 2015, 11, 5723–5736.

(36) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N. et al. *AMBER 16*; University of California: San Francisco, CA, 2016.

(37) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. **2004**, *120*, 9665–9678.

(38) Joung, I. S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(39) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M., Jr. Rational design of particle mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748.

(40) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Hsing, L.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(41) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327–341.

(42) Huang, M.; Giese, T. J.; Lee, T.-S.; York, D. M. Improvement of DNA and RNA sugar pucker profiles from semiempirical quantum methods. *J. Chem. Theory Comput.* **2014**, *10*, 1538–1545.

(43) Privé, G. G.; Yanagi, K.; Dickerson, R. E. Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. J. Mol. Biol. 1991, 217, 177–199.

(44) Bevan, D. R.; Li, L.; Pedersen, L. G.; Darden, T. A. Molecular dynamics simulations of the  $d(CCAACGTTGG)_2$  decamer: influence of the crystal environment. *Biophys. J.* **2000**, *78*, 668–682.

(45) Baucom, J.; Thomas, T.; Fuentes-Cabrera, M.; Krahn, J. M.; Darden, T. A.; Sagui, C. Molecular dynamics simulations of the  $d(CCAACGTTGG)_2$  decamer in crystal environment: comparison of atomic point-charge, extra-point, and polarizable force fields. *J. Chem. Phys.* **2004**, *121*, 6998–7008.

(46) Babin, V.; Baucom, J.; Darden, T. A.; Sagui, C. Molecular dynamics simulations of DNA with polarizable force fields: convergence of an ideal B-DNA structure to the crystallographic structure. *J. Phys. Chem. B* **2006**, *110*, 11571–11581.

(47) Babin, V.; Baucom, J.; Darden, T. A.; Sagui, C. Molecular dynamics simulations of polarizable DNA in crystal environment. *Int. J. Quantum Chem.* **2006**, *106*, 3260–3269.

(48) Ramakrishnan, B.; Sundaralingam, M. Evidence for crystal environment dominating base sequence effects on DNA conformation: crystal structures of the orthorhombic and hexagonal polymorphs of the A-DNA decamer d(GCGGGCCCGC) and comparison with their isomorphous crystal structures. *Biochemistry* **1993**, *32*, 11458–11468.

(49) Thiyagarajan, S.; Kumar, P. S.; Rajan, S. S.; Gautham, N. Structure of  $d(TGCGCA)_2$  at 293 K: comparison of the effects of sequence and temperature. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, 58, 1381–1384.

(50) Grzeskowiak, K.; Yanagi, K.; Prive, G. G.; Dickerson, R. E. The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. J. Biol. Chem. **1991**, 266, 8861–8883.

(51) Leonard, G. A.; Hunter, W. N. Crystal and molecular structure of d(CGTAGATCTACG) at 2 · 25 Å resolution. *J. Mol. Biol.* **1993**, 234, 198–208.

(52) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179–2183.

(53) Zheng, G.; Lu, X.-J.; Olson, W. K. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **2009**, *37*, 240–246.

(54) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. In *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.

(55) Tereshko, V.; Subirana, J. A. Influence of packing interactions on the average conformation of B-DNA in crystalline structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 810–819.

(56) Butcher, S. E.; Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* 2011, 44, 1302–1311.

(57) Todolli, S.; Perez, P. J.; Clauvelin, N.; Olson, W. K. Contributions of sequence to the higher-order structures of DNA. *Biophys. J.* **2017**, *112*, 416–426.

(58) Salmon, L.; Yang, S.; Al-Hashimi, H. M. Advances in the determination of nucleic acid conformational ensembles. *Annu. Rev. Phys. Chem.* **2014**, *65*, 293–316.

(59) Jones, C. P.; Ferré-D'Amaré, A. R. RNA quaternary structure and global symmetry. *Trends Biochem. Sci.* 2015, 40, 211–220.