An in-depth look at DNA crystals through the prism of molecular dynamics simulations

Antonija Kuzmanic[1,2], Pablo D. Dans[1,2], and Modesto Orozco[1,2,3,*]

[1]Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology. Baldiri Reixac 10-12, 08028 Barcelona, Spain.
[2]Joint BSC-IRB Research Program in Computational Biology. Baldiri Reixac 10-12, 08028 Barcelona, Spain.
[3]Department of Biochemistry and Molecular Biology. University of Barcelona, 08028 Barcelona, Spain.

[*]Correspondence should be addressed to: Prof. Modesto Orozco, Tel: +34 934037155, Email: modesto.orozco@irbbarcelona.org.

ABSTRACT

X-ray crystallography has been traditionally considered as the primary tool for the determination of biomolecular structures and its derived models are taken as the gold standard in structural biology. However, contacts formed through the crystal lattice are known to affect the structures, especially in the case of small and flexible molecules, like DNA oligos, introducing drastic changes in the structure with respect to the solution phase. Furthermore, it is still unknown why molecules crystallize in certain symmetry groups and how the associated lattice impacts their structure. The role of crystallization additives and whether they are just innocuous and unspecific catalyzers of the crystallization process also remains unclear. On account of a massive computational effort and the use of the latest generation force field, we were able to describe with unprecedented level of detail the nature of intermolecular forces that participate in the stabilization of B-DNA crystals in various symmetry groups and in different solvent environments. We showed that the stability of the crystal lattice and the type of crystallization additives are tightly coupled, and certain symmetry groups are only stable in the presence of a specific crystallization additive (i.e., spermine). Additives and crystal contacts induce small but non-negligible changes in the physical properties of DNA.

INTRODUCTION

Ever since Linus Pauling laid the foundations of structural biology, X-ray crystallography has been the cornerstone method for solving biomolecular 3D structures. Hundreds of thousands of models derived from X-ray data and deposited in the Protein Data Bank (PDB) are the best witnesses of the power of this technique, which has also become the gold standard for validating other structural methods. However, just as any other method, X-ray crystallography has its own shortcomings and obtaining diffractable crystals is not always an easy task.[1–3] Through robotic equipment, numerous variants of crystallization buffers are tested until a suitable one is found.[4] However, why a given buffer promotes crystallization, what its influence is on the symmetry of the unit cell, as well as the overall structure and physical properties of the biomolecule, is typically unknown.

Thousands of X-ray derived DNA structures have been crucial in understanding the fine details of isolated and protein-bound DNAs.[5–7] Unfortunately, artifacts in DNA crystals can lead to conformations which are otherwise undetected in solution (like the A-form of DNA). Even in ideal cases, such as the Drew-Dickerson dodecamer (DDD[8,9]), where crystal structures resemble the solution ones reasonably well, it is unclear how crystallization conditions modulate the symmetry space in which the molecule crystalizes, or the fine details of the structure and the physical properties of the DNA.[1,2] Knowing the molecular interactions that stabilize the crystals would help to understand the actual properties of DNA in such systems and the mechanisms through which highly charged molecules are packed in severely crowded conditions, e.g. the cellular nucleus.

Here we use a new generation of supercomputers and a recently developed DNA force field[10–12] to analyze the nature of DDD in three known crystal lattices and to explore the effects of crystallization additives on the stability of the crystals and the properties of the DNA.[13–15] Through extensive unbiased MD simulations on the multi-microsecond timescale, we demonstrate how the stability of DNA crystals depends on subtle interactions between the packed DNA molecules and the components of the crystallization buffer. We obtained, for the first time, stable atomistic simulations of DNA crystals in biologically relevant timescales,[16–21] in various symmetry groups (in which DDD has been crystallized), and in different solvent environments, which allowed us to understand with unprecedented level of detail the nature of intermolecular interactions that stabilize the crystals. Through the analysis of numerous simulations, we characterized the physical properties of DNA in a highly crowded environment resembling chromatin, and how the presence of additives affects them.

METHODS

Systems setup. X-ray structures of the Drew-Dickerson dodecamer used as the starting conformations in the unbiased MD simulations are those deposited under PDB codes 1EHV,[14] 1FQ2,[13] and 463D.[15] Both 1EHV and 1FQ2 entries contain the full DNA duplex, while the 5' cytosines are missing in 463D and were modelled in with PyMOL v1.8.2[22] by using 1EHV as the template. The dodecahedral simulation boxes for solution systems (labelled sol) were created by solvating a single DNA duplex with TIP3P water molecules,[23,24] leaving a 10-Å solvent layer around the duplex (~20,000 atoms). K^+, Na^+, and Cl^- ions (Smith and Dang parameters)[25,26] were added to reach neutrality and the final concentration of 400 mM (with 4:3 potassium to sodium ratio). In the case of 1FQ2 1:6 sol system, 6 spermine molecules were randomly added prior to solvation with TIP3P water and neutrality was reached with Cl^- ions. For the crystal systems, the simulation boxes were built following the dimensions of the unit cells reported in the PDB using PyMOL v1.8. The Na/KCl and $MgCl_2$ systems were completed by first adding TIP3P water and then the ions to reach neutrality and the final concentrations listed in Table 1 (using Smith and Dang[25,26] parameters for Na/KCl systems and Allnér-Nilsson-Villa[27] parameters for $MgCl_2$). The $SpmCl_4$ and $MgCl_2{\cdot}6H_2O$ crystal systems were completed by first randomly adding spermine or hexacoordinated Mg^{2+} ions (using Yoo and Aksimentiev[28,29] parameters, quantities in Table 1), then TIP3P waters followed by monovalent counter ions. System sizes ranged from 44,000-60,000 atoms for the smaller crystals, while the 1EHV systems with 81 duplexes had ~170,000 atoms.

Relaxation and equilibration of the systems. Energy of the simulated systems (listed in Table 1) was initially minimized through steepest-descent energy minimization. The initial velocities for the atoms were taken from Maxwell distribution at 100 K and the systems were subsequently heated to the final temperature (corresponding to the temperatures at which the crystals were grown; Table 1) in five steps of ~50 K simulated for either 100 or 1000 ps at constant volume and temperature using velocity rescale thermostat.[30] In parallel, atomic position restraints for the DNA molecules were uniformly relaxed. To achieve internal pressure close to 1 bar in crystal simulations, additional TIP3P water molecules were randomly placed within each system, which was first minimized and equilibrated, and then 50-ns production simulations were run, in a trial and error process until a suitable pressure was obtained (see Supp. Methods for further details).

Production trajectories. Once simulation conditions matched those expected in the crystal or in solution, production runs (1-4 μs long) were generated using GROMACS 5.0.4 biomolecular simulation package[31] with a 2-fs integration step and coordinate output of 10 ps. Constant volume conditions were employed for the crystal simulations, while the constant pressure (1 bar) was used in solution

simulations. All simulations were done at constant temperature (284-295 K depending on the reported crystallization conditions). Parrinello-Rahman barostat[32] and velocity rescale thermostat were used.[30] Periodic boundary conditions and particle mesh Ewald[33] were used in conjunction with state-of-the-art simulation conditions. DNA was described using the recently developed parmbsc1 force field.[10] See Methods for additional details of the simulations.

Analysis. Center of mass (COM) displacements were calculated after the optimal roto-translational fitting of the whole system to the ideal crystal. RMSD values were calculated for each duplex after the optimal roto-translational fitting to the original X-ray structures (see Suppl. Methods for details). Diffusion coefficients were calculated from a linear fit of mean square displacements over time using Einstein's equation. The effective temperature of water molecules in crystals was estimated from the corresponding diffusion coefficients which were first scaled by 1/2.56 (to account for the overestimation of the water diffusion by the TIP3P water model) as described in our previous work,[34] while effective DNA temperature was inferred at the base pair step level from the analysis of helical covariance matrices, which provided the stiffness matrices associated to harmonic deformations in the helical space.[35] Essential dynamics was determined by the diagonalization of the Cartesian covariance matrices. Similarity in essential deformation modes were computed using Hess[36] and Pérez *et al*. metrics,[35] without considering the capping base pairs. Additional details on the analysis and the associated packages are shown in Suppl. Methods.

Data availability. All trajectories reported here are deposited in the MuG-BigNASim[37] database and can be freely downloaded.
http://www.multiscalegenomics.eu/MuGVRE/modules/BigNASimMuG/

RESULTS AND DISCUSSION

Are the simulations of DNA crystals stable in the μs regime? The Drew-Dickerson dodecamer is a prototypical B-DNA molecule, whose structure has been determined by X-ray crystallography in various space groups and under diverse conditions,[13–15] as well as through high-quality NMR data in solution. We chose to perform our simulations on structures crystallized in 3 different space groups: $P3_212$, $P2_12_12_1$, and H3 (PDB codes: 1EHV, 1FQ2, and 463D, respectively). We followed the multistep protocol developed by the Case group,[17] which includes a careful stabilization of the internal pressure by calibrating the number of water molecules in the system (see the Supp. Methods section for details and Supp. Table S1). Once the internal pressure was adjusted, we ran multi-microsecond long simulations of each crystal, which contained respectively 27, 24, and 36 copies of DDD (Table 1). To our surprise, two of our systems showed a systematic degradation of the crystal lattice, as shown by the center-of-mass (COM)

displacements of each duplex with respect to the ideal positions in the crystal (Figure 1). The distortion of the lattice was particularly pronounced along the z-axis and got worse with simulation time for 1EHV and 1FQ2, while the 463D system was completely stable thorough the trajectory. Not only was the lattice intact in 463D, but the MD conformations sampled in the helical space on the base pair step (bps) level were in complete agreement with the average X-ray values (Figure 2), thus establishing that our force field is fully capable of representing local and global features of crystal structures.

We thus centered our efforts on stabilizing the other two crystals starting first with the smallest system: 1EHV. We simulated a larger crystal (81 instead of 27 DDD molecules) with 5 different Na/KCl concentrations (Table 1), ranging from 15 mM to 800 mM (i.e., up to 5 times the assumed physiological salt concentration of the cellular nucleus), without being able to stabilize it (Supp. Figure S1). Considering that the increase in crystal's size had little effect on its stability (Supp. Figure S2), we decided to proceed with the smaller system (with 27 DDD copies). The addition of 400 mM $MgCl_2$ produced some improvement in the lattice integrity, but its degradation was still evident during the 4-μs-long simulations (Supp. Figure S3, and Supp. Methods for more details). We thus checked if the global degradation of the crystals was due to internal distortions of the double helices by analyzing the 2D distributions of the bps helical parameters from MD simulations. We obtained a good agreement between the values calculated from MD simulations and the X-ray structures (Supp. Figure S4), which discards the notion that the lattice distortions were caused by the deterioration of duplex geometries. Even the flexible ends (residue pairs 1-24 and 12-13), which exhibit fast but moderate fraying in solution,[11] had in most cases an average RMSD with respect to the X-ray structures under 4 Å, oscillating between 3D conformations compatible with the experimental electron density (Supp. Figure S5). We obtained similar results when trying to stabilize, without success, the $P2_12_12_1$ 1FQ2 crystal (data not shown), which allowed us to discard the possibility of a force field artifact that is space-group specific.

How important are the buffer components in the stabilization of DNA crystals? As we had discarded any obvious explanation for the degradation of the crystal lattice, we focused our attention on the chemical additives used to obtain the DDD crystals in the studied space groups. We noticed that 463D system was crystallized without using spermine (SPM),[15] while both 1EHV[14] and 1FQ2[13] crystals were formed in the presence of this polyamine (+4 charge at pH 7). Spermine is normally found in all eukaryotic cells and has been used as an "inert molecular glue" for obtaining thousands of DNA crystals in all the major forms (A, B, Z, etc.).[19,20,38] However, the molecular basis of its mechanism of action is mostly unknown, and in fact, spermine electron densities are often absent from the crystal. To check whether the lack of spermine in our simulations was the reason of the degradation of $P3_212$ and $P2_12_12_1$ simulations, we repeated the simulations

for both crystals at three different spermine concentrations ranging from 1:3 (3 spermine molecules per duplex) up to 1:12 (Table 1). At last, the lattice integrity for both space groups was now preserved in a consistent concentration dependent manner (Figure 3). At the medium spermine concentration tested, 1:9 for 1EHV and 1:6 for 1FQ2, the observed stability is already comparable to the one obtained for 463D with no perceptible drift of the crystal lattice along the y/z-axes (Supp. Figures S6 and S7). Clearly the "inert molecular glue" has a major role in preserving the integrity of some of the crystal lattices.

Similarly to 463D, once properly stabilized, we obtained an impressive agreement between the X-ray structures (1EHV/1FQ2) and the MD crystals. This is visible when comparing the rotational helical space (Figure 4), or the groove dimensions, a property which has always been difficult to reproduce accurately by modern force fields (Supp. Figure S8). The dynamics of the end residues also seems to be stable according to their RMSD values, which are typically below 2 Å with respect to the crystal position (Supp. Figures S9 and S10). In the end, to illustrate the observed structural agreement, we also compared the average structures obtained from crystal MD simulations with the deposited X-ray structure for the three systems stabilized in the μs regime is shown in Supp. Figure S11.

To follow each ion along the trajectories of each duplex, we used the curvilinear helicoidal coordinate (CHC)[39] method as implemented in Canion/Curves+, which allowed us to calculate ion populations and concentrations in the inner and outer areas of major and minor grooves of DNA duplexes (see Supp. Methods).[40,41] We found that the sequence-dependent binding sites observed for K^+ and the amino groups of spermine were the same in solution and in crystals, although the sequence dependence was weaker in crystals and the cation distributions seem to be fuzzier (Supp. Figure S12).[11] The comparison of aqueous simulations with and without spermine (left column in Supp. Figure S12) showed that spermine mainly localizes outside the grooves, which was also previously observed in crystal simulations.[19,20,38,42] Interestingly, the interiors of the grooves are depleted of cations in all crystal simulations compared to the solution phase.

We also analyzed the cation environment in all the individual duplexes present in each crystal. First, a clear similarity was found in the cation distribution in all the duplexes of a given crystal, confirming the robustness of the average results discussed here (see Suppl. Figures S13-15). Second, spermine distribution was similar around the duplexes in simulations of two different space groups (1EHV and 1FQ2), indicating that the ion distribution is independent of the crystal packing (see Supp. Figures S14 and S15). Moreover, spermine mostly preferred the exterior of the central part of the major groove, however, upon entering the groove, spermine was located at non-central sites, which we also observed for solution simulations (see Suppl. Figure S12). Spermine could concentrate up to 12 M around the phosphate groups in the central portion of the duplexes where no

DNA intermolecular contacts are present. In contrast, these contacts are particularly pronounced in terminal bases (Supp. Figure S16). In summary, spermine has a tendency to move around the exterior of the duplex without any perceptible sequence dependence (as observed in NMR[43] and RAMAN[44] experiments), most likely screening intermolecular phosphate-phosphate repulsion and facilitating crystal packing. Combined with its rare long-term presence within the insides of the grooves, it is not surprising that spermine electron density is seldom captured.

While the temperature for both experiments and MD simulations was in the 284-293 K range for 1EHV, 1FQ2, and 463D (see Supp. Methods), the intermolecular DNA contacts in the crystal froze the duplex structure. DNA's internal effective temperature was lowered by 200 K (with respect to the thermostat temperature) at the end of the duplexes, which are mostly rigidified, while we observed only ~50 K decrease in the central portion of the duplex (Supp. Figure S17). Water molecules and SPM are also affected by the crowded and highly negatively charged environment produced by the packing of the backbone phosphate groups of neighboring duplexes. The rescaled diffusion coefficient[34] of water molecules (Supp. Figure S18 and Supp. Methods), was reduced by one order of magnitude - from $1.70 \cdot 10^{-5}$ cm^2 s^{-1} in solution DNA simulations to $1.67 \cdot 10^{-6}$ cm^2 s^{-1} in crystal simulations. The reduction in water mobility is visible from a 30-K decrease in its effective temperature. Furthermore, water mobility in crystals is anisotropic, with the preference for the z-axis in 463D and 1EHV, and the y-axis in 1FQ2 following the water channels that are created in each crystal due to the orientation of the duplexes (Supp. Figure S19). The effect was even more pronounced for SPM when comparing the 1FQ2 1:6 solution simulations with the 1FQ2 crystal with the same SPM concentration: the diffusion coefficient in the crystal was reduced by three orders of magnitude compared to the solution: from $1.90 \cdot 10^{-6}$ cm^2 s^{-1} to $2.62 \cdot 10^{-9}$ cm^2 s^{-1}. Although the mobility of SPM is greatly reduced in the crystal, the lack of specificity in its binding to DNA (extensive and non-specific binding outside the major groove) makes its detection in the experimental electron densities very difficult (see Suppl. Figure S13). As expected, the diffusion of water and SPM in the crystal was correlated and exhibited an SPM-concentration dependent effect as shown in Supp. Figure S20.

How do the structural and dynamical properties of DNA crystals compare with solution? We computed the global RMSD of MD simulations (crystal and solution) with respect to the experimental crystal configuration (Table 2), confirming that the average MD structure in solution deviates much more from the crystal configuration (with larger deviations in the backbone) than the average structure from the crystal MD simulations,[17] confirming that MD simulations are able to capture the effect of the crystal environment on the DNA. Irrespectively of the original conformation present in the crystal, when we transferred the duplexes to solution, all the simulations converged to the same ensemble (Supp. Figure S21

and S22), showing a strong consensus and the lack of memory in the simulations. Moreover, as confirmed by WAXS experiments,[38] SPM has a negligible effect on the structure of DNA in aqueous solution (Supp. Figure S21 and S22). As shown in Figure 5, the solution simulations were generally able to sample the X-ray values within the first standard deviation, indicating that the overall structures in solution and in the crystal are very similar. However, a close inspection shows small local discrepancies between the average MD solution simulations and the X-ray values at several bps, although these differences (Supp. Figure S23 and S24) tend to disappear when average properties of the entire duplex are computed.[11]

We also analyzed how crystal modifies the flexibility and dynamical behavior of the duplex. As expected, DNA in crystals was much more rigid than in solutions, which can be seen from the force constants associated with helical deformations. As expected from helical parameters, the differences are especially pronounced for the terminal residues which are much more mobile in water (Supp. Figure S25). Interestingly, same sequence-dependent pattern of flexibility is found for the crystal and solution simulations - for all degrees of freedom and for all studied crystals. This suggests that the strong reduction of terminal movements is a consequence of the specific DNA packing required for crystallization. The analysis of the essential dynamics (see Methods) of DNA in solution and in the crystal indicates a high degree of similarity: 93-98% for solution *vs* solution simulations, while the similarity decreases to 78-95% for crystal *vs* crystal MD comparisons, and to 75-84% when comparing crystal to solution (Supp. Figure S26). This finding suggests that the essential dynamics of the duplexes in the crystals slightly differs from the one in solution, and that the variability in the type of movements between different crystal lattices is larger than in solution. These results indicate that the small structural discrepancies in the average helical parameters, discussed above, arose from dynamic, rather than static, properties of the DNA. However, they also suggest a reasonable maintenance of the intrinsic deformation pattern of DNA in crystals, which would explain how processes which require deformation of the DNA can be reproduced (on longer timescales) in crystal environments, e.g., binding of small molecules in soaking experiments.[45,46]

We finally focused our analysis on two paradigmatic structural polymorphisms that seem to be causally connected in solution, namely the backbone BI/BII polymorphism (related to ε and ζ torsions), and the bimodal behavior of the twist helical parameter observed for d(CpG) steps.[11,40,47–49] BII populations are detected in both crystal and solution simulations, but the base pair steps for which such minor populations were observed are different (Supp. Figure 27). In agreement with the experiments (X-ray and 1H/31P-NMR),[50] BII state was observed less frequently for d(CpG) and d(GpA) steps in the crystal than in the solution simulations, while the opposite was observed for d(GpC) steps. Quite similar results were observed for d(ApA) and d(ApT) steps. This would suggest that the differences in BII propensities between X-ray and NMR structures are not spurious

effects related to the uncertainties of the refinement process, but a consequence of crystal packing.

Polymorphisms at the base level (i.e. the existence of two stable conformational substates) were much more difficult to detect experimentally.[51] On the one hand, NOEs often lack the required resolution to introduce dual restraints in the simulation procedure, and on the other hand, "substate-washing-effect" produced by the averaging of structures present in the crystal annihilate minor substates in the final refined structures.[51,52] As we have individual data for each duplex in the crystal, we can analyze minor polymorphism with high accuracy. Thus, we explored the twist polymorphism at the d(CpG) step for the three simulated crystals and compared the results to the solution trajectories. We found that the bimodal behavior was mitigated in the crystals due to both averaging effects and a lower propensity for the bimodal twist (Supp. Figures 28-30).[40,49] The two and three substates observed in the 2D density plots of roll vs twist and twist vs tilt, respectively, were all found among the duplexes present in the crystal units, but not simultaneously in a unique duplex as in the solution simulations. This correlated perfectly with the reduced amount of cations found inside the minor groove of d(CpG) steps in the crystals simulations,[40] and with the smaller mobility of crystal DNA that reduces the kinetics of substate inter-conversion. These results suggest that the solvent environment created around the duplexes in the crowded crystal affects the DNA dynamics and limits its conformational space compared to the free DNA in solution. Therefore, taking extra caution is necessary before using X-ray data to evaluate small-scale polymorphisms which can be crucial for DNA recognition by ligands.

CONCLUSIONS

Even with recent advances in cryoEM, X-ray crystallography remains a dominant technique in structural biology that shows no signs of slowing down. However, one of its main challenges still lies in obtaining diffractable crystals. Conditions that lead to such crystals are a priori unknown and the knowledge gained in the field has been mainly empirical. Typically, several types of solvent buffers are usually tried out but since the electron density of ions and chemical additives is rarely observed, it has been impossible to know which buffer components finally end up in the crystal and at which concentration. For example, spermine has been used to obtain thousands of DNA crystals, yet the molecular basis of its mechanism of action was mostly unknown until now. Furthermore, atomistic MD simulations were not able to provide relevant information on the intermolecular forces that dominate within the crystal structure, due to issues with obtaining stable crystal simulations on relevant timescales. In this work, we presented for the first time stable multi-microsecond-long MD simulations of crystals of the Drew-Dickerson

dodecamer in all the symmetry groups it has been experimentally crystallized in. We achieved this by using the latest generation classical force field and by modeling correctly the solvent environment around DNA duplexes which allowed us to understand the importance of chemical additives in the crystallization buffer. Once stable, an impressive global and local structural agreement was obtained between the modeled and experimental DNA crystals. This allowed us to unravel the interactions and physical properties of water molecules, ions and spermine at the atomic level, giving their quantifiable overview. Water molecules and spermine diffused 10 and 1,000 times more slowly in the crowded crystal environment than in solution, respectively. Most of the spermine molecules, which could concentrate up to 12 M around DNA, were bound to phosphate groups outside of the major groove without a specific sequence dependence. Moreover, we showed that previous discrepancies in local structure between solution simulations and X-ray structures, or the ones between solution NMR and X-ray structures in the description of the backbone polymorphism, were mainly due to packing effects that modify the solvent-solute interactions, affecting the dynamics of the DNA and limiting its conformational space compared to its free behavior in solution. Our work expands the limits of the field and opens the door to further systematic studies of the effect of crystallization conditions on the obtained structures. In the near future, MD simulations of crystals could be used to anticipate and predict the effects of a given additive, reducing the number of trial and error iterations usually involved in the experimental obtainment of stable DNA crystals.

SUPPLEMENTARY DATA

Further details on the setup and analysis of the trajectories, and additional results are provided in the Supplementary Methods, Figures and Tables respectively.

ACKNOWLEDGEMENTS

M.O. is an ICREA (Institució Catalana de Recerca i Estudis Avançats) academia researcher. P.D.D. is a PEDECIBA (Programa de Desarrollo de las Ciencias Básicas) and SNI (Sistema Nacional de Investigadores, Agencia Nacional de Investigación e Innovación, Uruguay) researcher.

FUNDING

Spanish Ministry of Science [BFU2014-61670-EXP, BFU2014-52864-R], Catalan SGR, the Instituto Nacional de Bioinformática; European Research Council (ERC SimDNA), European Union's Horizon 2020 research and innovation program

[676556], Biomolecular and Bioinformatics Resources Platform (ISCIII PT 13/0001/0030) co-funded by the Fondo Europeo de Desarrollo Regional (FEDER) (all awarded to M.O.); MINECO Severo Ochoa Award of Excellence (Government of Spain) (awarded to IRB Barcelona). The project was awarded computer resources from the PRACE 12th call (awarded to A.K., P.D.D. and M.O.).

REFERENCES

1. Zheng, J. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461,** 74–77 (2009).
2. Zhang, W., Szostak, J. W. & Huang, Z. Nucleic acid crystallization and X-ray crystallography facilitated by single selenium atom. *Front. Chem. Sci. Eng.* **10,** 196–202 (2016).
3. Saenger, W. in 29–50 (1984). doi:10.1007/978-1-4612-5190-3_3
4. Yeung, H. *et al.* Radiation Damage and Racemic Protein Crystallography Reveal the Unique Structure of the GASA/Snakin Protein Superfamily. *Angew. Chemie Int. Ed.* **55,** 7930–7933 (2016).
5. Mooers, B. H. M. Crystallographic studies of DNA and RNA. *Methods* **47,** 168–176 (2009).
6. Westhof, E. in 3–8 (Humana Press, New York, NY, 2016). doi:10.1007/978-1-4939-2763-0_1
7. Egli, M. Nucleic acid crystallography: current progress. *Curr. Opin. Chem. Biol.* **8,** 580–591 (2004).
8. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2179–83 (1981).
9. Drew, H. ., Samson, S. & Dickerson, R. E. Structure of a B-DNA dodecamer at 16 K. *Proc.Natl.Acad.Sci.USA* **79,** 4040–4044 (1982).
10. Ivani, I. *et al.* Parmbsc1: A refined force field for DNA simulations. *Nat. Methods* **13,** 55–58 (2015).
11. Dans, P. D. *et al.* Long-timescale dynamics of the Drew-Dickerson dodecamer. *Nucleic Acids Res.* **44,** 4052–4066 (2016).
12. Dans, P. D. *et al.* How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* **45,** 4217–4230 (2017).
13. Chad C. Sines, Lori McFail-Isom, Shelley B. Howerton, Don VanDerveer, and & Williams*, L. D. Cations Mediate B-DNA Conformational Heterogeneity†. (2000). doi:10.1021/JA002244P
14. Johansson, E., Parkinson, G. & Neidle, S. A new crystal form for the dodecamer C-G-C-G-A-A-T-T-C-G-C-G: symmetry effects on sequence-dependent DNA structure. *J. Mol. Biol.* **300,** 551–561 (2000).
15. Liu, J. & Subirana, J. A. Structure of d(CGCGAATTCGCG) in the presence of Ca(2+) ions. *J. Biol. Chem.* **274,** 24749–52 (1999).
16. 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.* **117,** 5001–5002 (1995).

17. Liu, C., Janowski, P. A. & Case, D. A. All-atom crystal simulations of DNA and RNA duplexes. *Biochim. Biophys. Acta* **1850,** 1059–71 (2015).
18. Babin, V., Baucom, J., Darden, T. A. & Sagui, C. Molecular dynamics simulations of polarizable DNA in crystal environment. *Int. J. Quantum Chem.* **106,** 3260–3269 (2006).
19. Korolev, N., Lyubartsev, A. P., Nordenskiöld, L. & Laaksonen, A. Spermine: an "invisible" component in the crystals of B-DNA. A grand canonical Monte Carlo and molecular dynamics simulation study. *J. Mol. Biol.* **308,** 907–917 (2001).
20. Korolev, N., Lyubartsev, A. P., Laaksonen, A. & Nordenskiöld, L. On the competition between water, sodium ions, and spermine in binding to DNA: a molecular dynamics computer simulation study. *Biophys. J.* **82,** 2860–75 (2002).
21. Korolev, N., Lyubartsev, A. P., Laaksonen, A. & Nordenskiöld, L. A molecular dynamics simulation study of oriented DNA with polyamine and sodium counterions: diffusion and averaged binding of water and cations. *Nucleic Acids Res.* **31,** 5971–81 (2003).
22. Schrodinger, L. L. C. The PyMOL Molecular Graphics System. (2015).
23. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
24. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112,** 8910 (2000).
25. Smith, D. E. & Dang, L. X. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.* **100,** 3757–3766 (1994).
26. Dang, L. X. Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study. *J. Am. Chem. Soc.* **117,** 6954–6960 (1995).
27. Allnér, O., Nilsson, L. & Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **8,** 1493–1502 (2012).
28. Yoo, J. & Aksimentiev, A. Improved Parametrization of Li^{+}, Na^{+}, K^{+}, and Mg^{2+} Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* **3,** 45–50 (2012).
29. Yoo, J. & Aksimentiev, A. Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J. Chem. Theory Comput.* **12,** 430–443 (2016).
30. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).

31. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2,** 19–25 (2015).
32. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
33. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).
34. Hospital, A., Candotti, M., Gelpí, J. L. & Orozco, M. The Multiple Roles of Waters in Protein Solvation. *J. Phys. Chem. B* **121,** 3636–3643 (2017).
35. Pérez, A. *et al.* Exploring the Essential Dynamics of B-DNA. *J. Chem. Theory Comput.* **1,** 790–800 (2005).
36. Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E. Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics* **62,** 8438–48 (2000).
37. Hospital, A. *et al.* BIGNASim: A NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.* **44,** D272–D278 (2016).
38. Katz, A. M. *et al.* Spermine Condenses DNA, but Not RNA Duplexes. *Biophys. J.* **112,** 22–30 (2017).
39. Lavery, R., Maddocks, J. H., Pasi, M. & Zakrzewska, K. Analyzing ion distributions around DNA. *Nucleic Acids Res.* **42,** gku504- (2014).
40. Dans, P. D. *et al.* Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.* **42,** 11304–11320 (2014).
41. Pasi, M., Maddocks, J. H. & Lavery, R. Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res.* **43,** 2412–23 (2015).
42. Yoo, J. & Aksimentiev, A. The structure and intermolecular forces of DNA condensates. *Nucleic Acids Res.* **44,** 2036–2046 (2016).
43. Wemmer, D. E., Srivenugopal, K. S., Reid, B. R. & Morris, D. R. Nuclear magnetic resonance studies of polyamine binding to a defined DNA sequence. *J. Mol. Biol.* **185,** 457–9 (1985).
44. Deng, H., Bloomfield, V. A., Benevides, J. M., Thomas, G. J. & Jr. Structural basis of polyamine-DNA recognition: spermidine and spermine interactions with genomic B-DNAs of different GC content probed by Raman spectroscopy. *Nucleic Acids Res.* **28,** 3379–85 (2000).
45. Das, K., Martinez, S. E., Bauman, J. D. & Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat. Struct. Mol. Biol.* **19,** 253–9 (2012).
46. Burmeister, W. P. *et al.* Crystal Structure of the Vaccinia Virus Uracil-DNA Glycosylase in Complex with DNA. *J. Biol. Chem.* **290,** 17923–34 (2015).
47. Hartmann, B., Piazzola, D. & Lavery, R. BI-BII transitions in B-DNA. *Nucleic Acids Res.* **21,** 561–8 (1993).
48. Balaceanu, A. *et al.* The Role of Unconventional Hydrogen Bonds in

Determining BII Propensities in B-DNA. *J. Phys. Chem. Lett.* **8,** 21–28 (2017).
49. Pasi, M. *et al.* μABC: A systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* **42,** 12272–12283 (2014).
50. Heddi, B., Oguey, C., Lavelle, C., Foloppe, N. & Hartmann, B. Intrinsic flexibility of B-DNA: The experimental TRX scale. *Nucleic Acids Res.* **38,** 1034–1047 (2009).
51. Maehigashi, T. *et al.* B-DNA structure is intrinsically polymorphic: even at the level of base pair positions. *Nucleic Acids Res.* **40,** 3714–3722 (2012).
52. Kuzmanic, A., Pannu, N. S. & Zagrovic, B. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nat. Commun.* **5,** 3220 (2014).

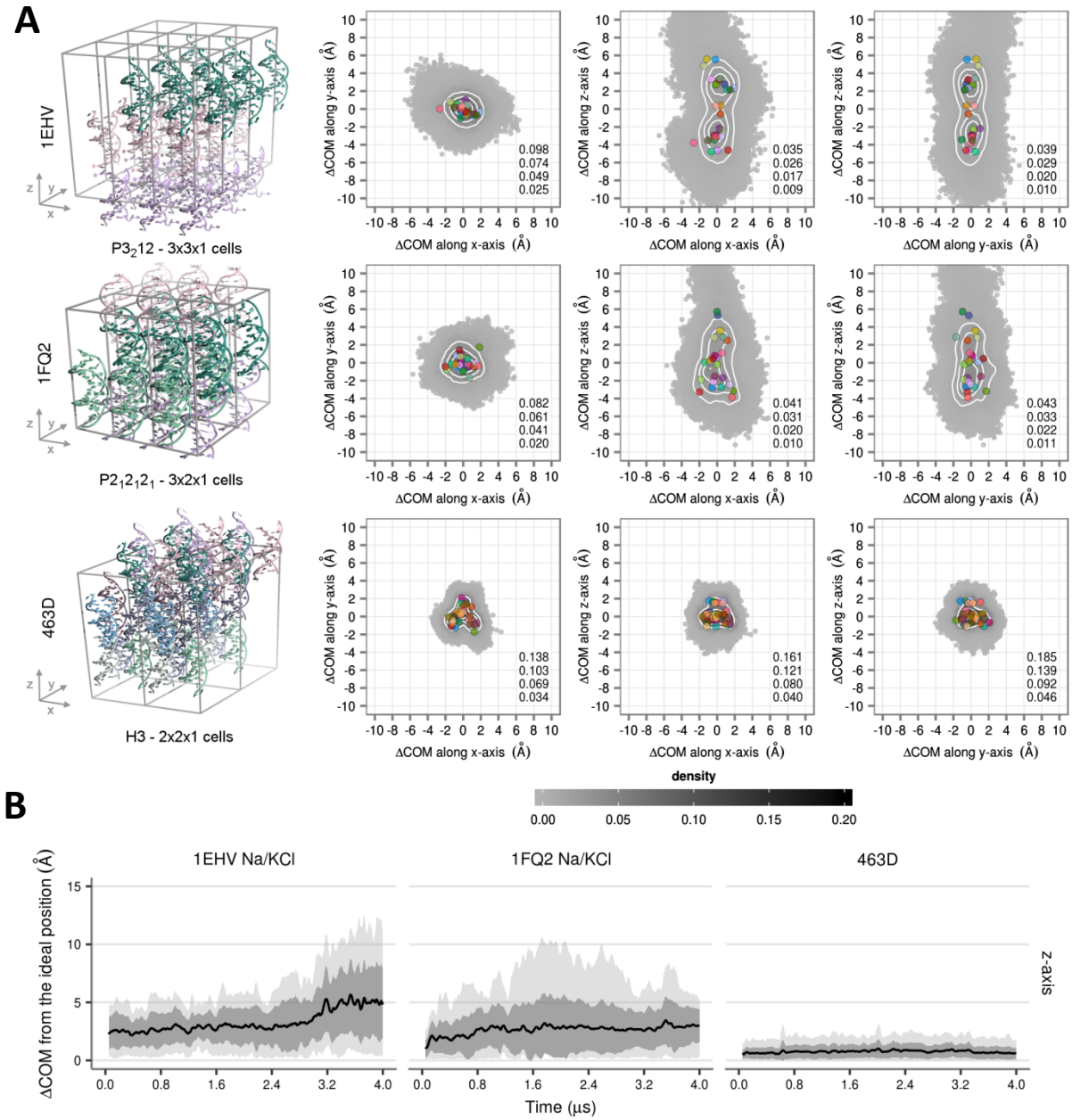

Figure 1. A) COM displacements from the ideal crystallographic positions shown for each duplex in the simulated supercell systems (1EHV, 1FQ2, 463D) along principal axes. Points are colored in greyscale according to the smooth 2D densities, estimated by fitting the observed distributions to a bivariate normal kernel (evaluated on a square grid of 90 × 90 bins). Four isodensity curves are shown in white and are quantified on the bottom right side of each plot. The average positions of the duplexes are given as colored circles. B) COM displacements in time along the z-axis shown for all three systems. Mean values are shown as black lines, standard deviations are shown as dark grey ribbons, while the extreme values are shown with light grey ribbons (lower panels).

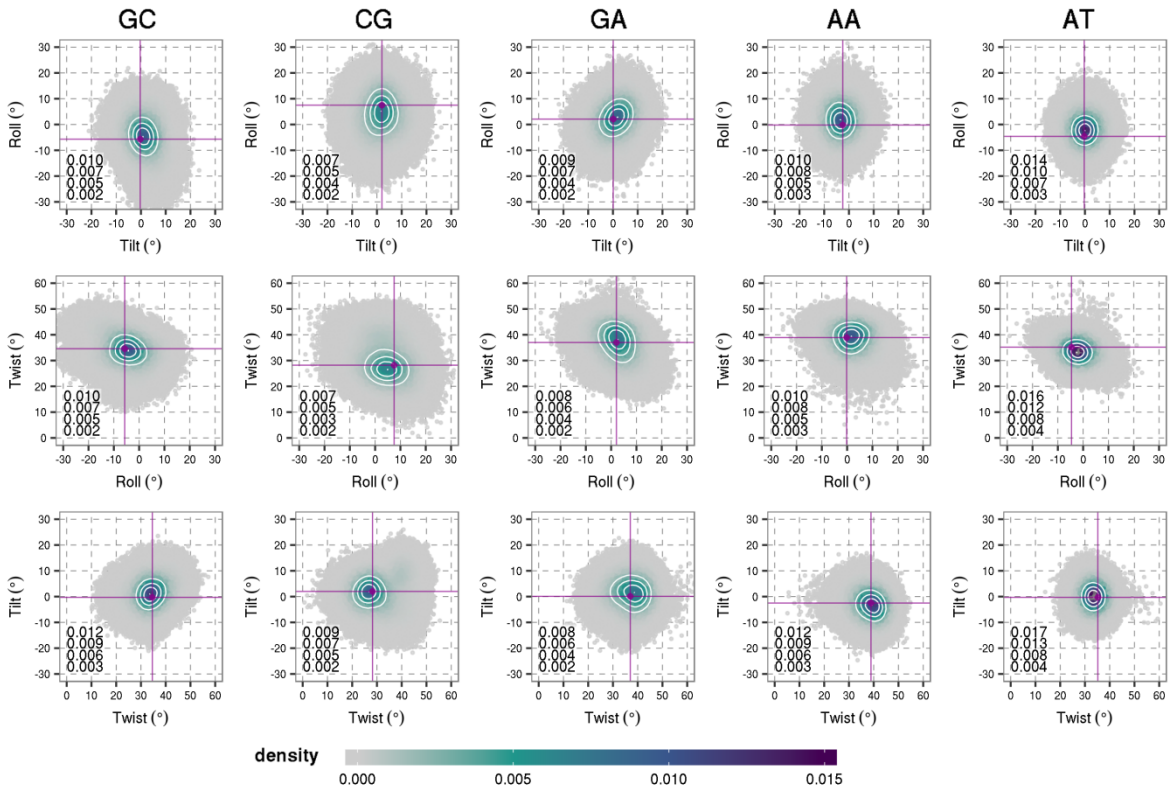

Figure 2. Rotational helical parameters (roll, twist, tilt) of all the duplexes in the 463D system (given for distinct base pair steps in DDD after removing the ends). The values calculated from the X-ray structure are shown with purple lines. Points are colored according to the smooth 2D densities, estimated by fitting the observed distributions to a bivariate normal kernel (evaluated on a square grid of 90 × 90 bins). Four isodensity curves are shown in white and are quantified on the bottom right side of each plot.

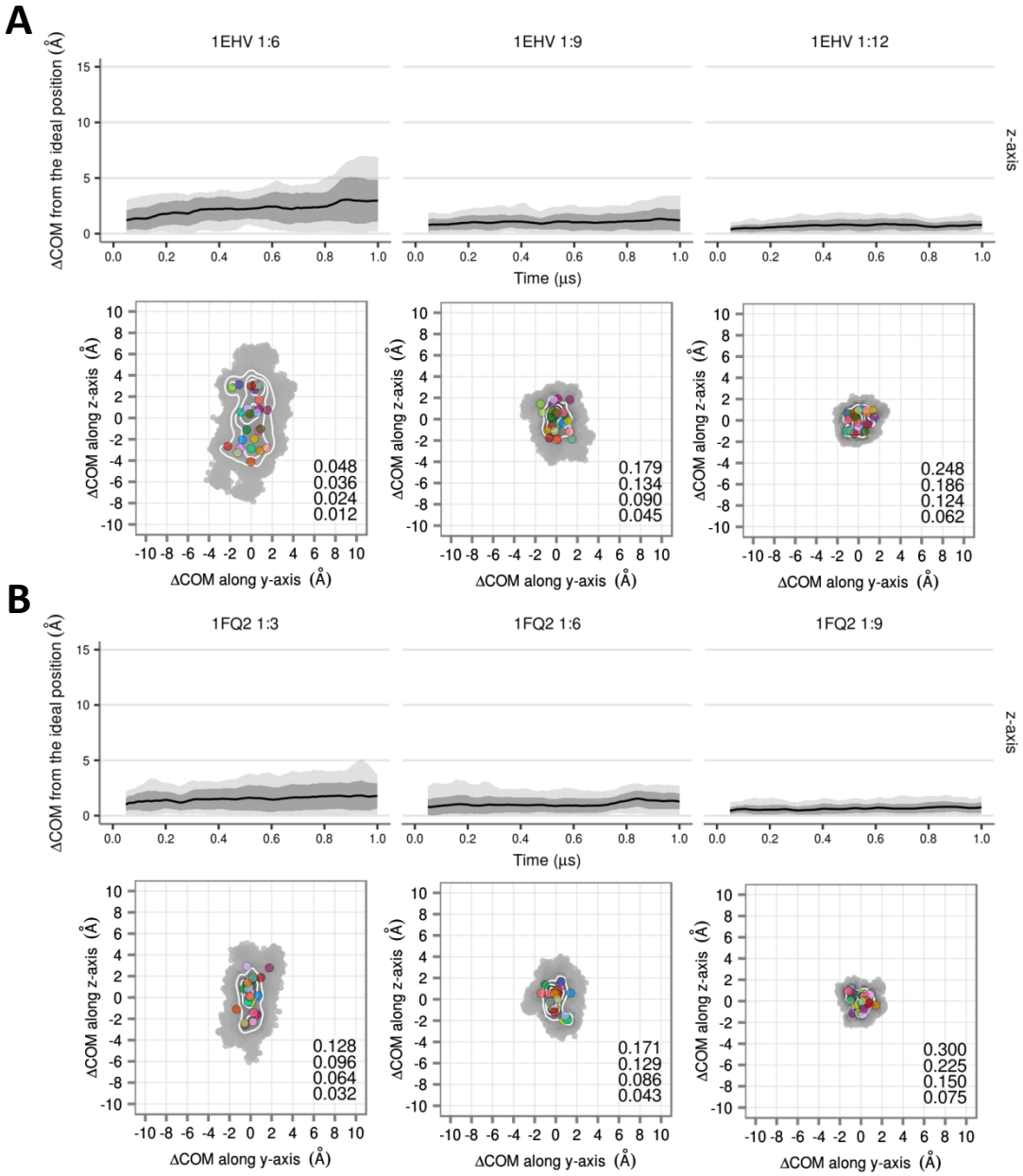

Figure 3. COM displacements from the ideal crystallographic positions for different spermine concentrations in A) 1EHV and B) 1FQ2 systems. The upper panels show the COM displacements in time along the z-axis where mean values are shown as black lines, standard deviations are shown as dark grey ribbons, while the extreme values are shown with light grey ribbons. The lower panels show COM displacements shown for each duplex along y and z axes with the average positions of the duplexes given as colored circles. Points are colored in greyscale according to the smooth 2D densities, estimated by fitting the observed distributions to a bivariate normal kernel (evaluated on a square grid of 90 × 90 bins). Four isodensity curves are shown in white and are quantified on the bottom right side of each plot.

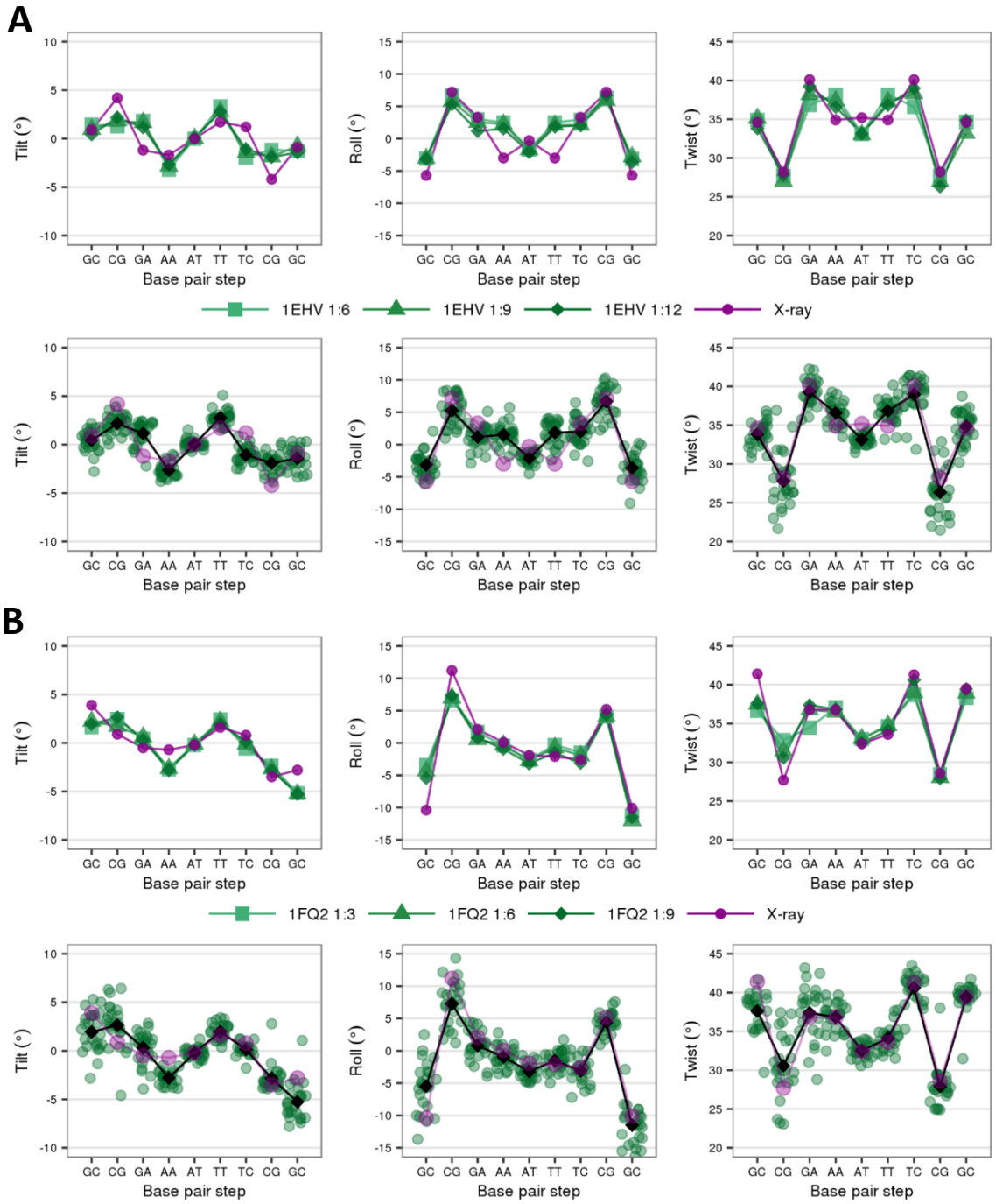

Figure 4. Rotational helical parameters (roll, twist, tilt) (given for all base pair steps in DDD after removing the ends) for different spermine concentrations in A) 1EHV and B) 1FQ2 systems. The upper panels show average values across all duplexes for varying spermine concentrations (colored green), with the purple line giving values calculated for the X-ray structure. The lower panels show the average values for each duplex in the systems with the highest spermine concentration (shown as green circles), while the average over all duplexes is shown as a black line and the X-ray values in purple.

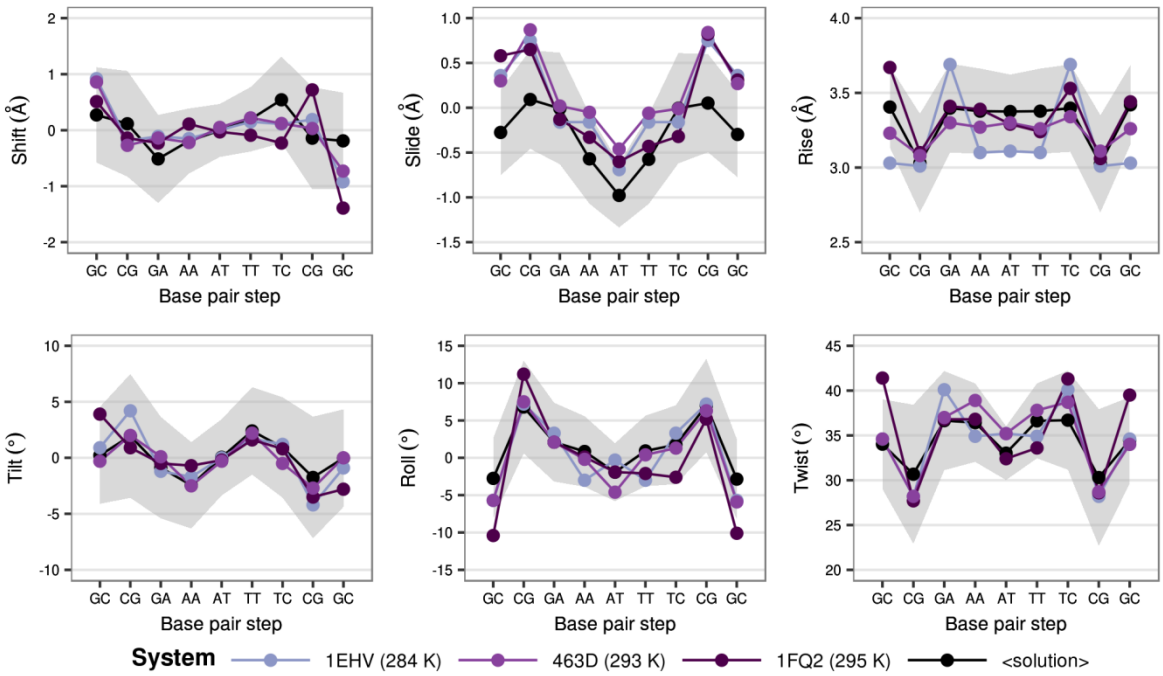

Figure 5. Translational (shift, slide, rise) and rotational (roll, twist, tilt) helical parameters (given for all base pair steps in DDD after removing the ends) for the solution simulations of single duplexes starting from different X-ray structures. The average values across the simulations are shown in black, while the average standard deviations are shown as grey ribbons. Values calculated from X-ray models are shown as colored lines.

TABLES

Table 1. Studied systems and simulated conditions.

System name	PDB	Environment	Space group	Nº of unit cells	Nº of DNA duplexes	Salt type	[Salt] (mM)	dsDNA to Spm ratio	Temp. (K)	Equilibration time (ps)	Simulation time (μs)
1EHV sol	1EHV	solution	---	---	1	Na/KCl	400	---	284	600	1
15 mM	1EHV	crystal	$P3_212$	3x3x3	81	Na/KCl	15	---	284	600	1
200 mM	1EHV	crystal	$P3_212$	3x3x3	81	Na/KCl	200	---	284	600	1
400 mM	1EHV	crystal	$P3_212$	3x3x3	81	Na/KCl	400	---	284	600	4
600 mM	1EHV	crystal	$P3_212$	3x3x3	81	Na/KCl	600	---	284	600	1
800 mM	1EHV	crystal	$P3_212$	3x3x3	81	Na/KCl	800	---	284	600	1
1EHV Na/KCl	1EHV	crystal	$P3_212$	3x3x1	27	Na/KCl	400	---	284	600	4
1EHV Mg	1EHV	crystal	$P3_212$	3x3x1	27	$MgCl_2$	400	---	284	600	4
1EHV Mg hex	1EHV	crystal	$P3_212$	3x3x1	27	$MgCl_2{\cdot}6H_2O$	400	---	284	600	4
1EHV 1:6	1EHV	crystal	$P3_212$	3x3x1	27	$SpmCl_4$	---	1:6	284	6000	1
1EHV 1:9	1EHV	crystal	$P3_212$	3x3x1	27	$SpmCl_4$	---	1:9	284	6000	1
1EHV 1:12	1EHV	crystal	$P3_212$	3x3x1	27	$SpmCl_4$	---	1:12	284	6000	1
1FQ2 sol	1FQ2	solution	---	---	1	Na/KCl	400	---	295	600	1
1FQ2 1:6 sol	1FQ2	solution	---	---	1	$SpmCl_4$	---	1:6	295	600	1
1FQ2 Na/KCl	1FQ2	crystal	$P2_12_12_1$	3x2x1	24	Na/KCl	400	---	295	600	4
1FQ2 Na/KCl equ	1FQ2	crystal	$P2_12_12_1$	3x2x1	24	Na/KCl	400	---	295	6000	1
1FQ2 1:3	1FQ2	crystal	$P2_12_12_1$	3x2x1	24	$SpmCl_4$	---	1:3	295	6000	1
1FQ2 1:6	1FQ2	crystal	$P2_12_12_1$	3x2x1	24	$SpmCl_4$	---	1:6	295	6000	1
1FQ2 1:9	1FQ2	crystal	$P2_12_12_1$	3x2x1	24	$SpmCl_4$	---	1:9	295	6000	1
463D sol	463D	solution	---	---	1	Na/KCl	400	---	293	600	1
463D	463D	crystal	H3	2x2x1	36	Na/KCl	400	---	293	6000	4

Table 2. Root mean square deviations (RMSD, Å) from the deposited crystal structures for 463D, 1EHV, and 1FQ2. The statistics in each box are heavy atoms, backbone atoms, and base atoms RMSD of average MD structure with respect to the crystal structure. The values in parentheses exclude the terminal residues.

	All heavy atoms	Backbone	Base	All heavy atoms (sol)	All heavy atoms (sol + SPM)
463D	1.72 (0.86)	1.39 (0.99)	2.03 (0.65)	3.20 (1.20)	-
1EHV 1:12	0.96 (0.83)	1.00 (0.95)	0.90 (0.64)	3.20 (1.15)	-
1FQ2 1:9	0.57 (0.52)	0.64 (0.60)	0.47 (0.41)	1.45 (1.19)	1.49 (1.23)