HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2015 April 29.

Published in final edited form as:

Nat Commun. ; 5: 5152. doi:10.1038/ncomms6152.

On the absence of intra-helical DNA dynamics on the μs to ms timescale

Rodrigo Galindo-Murillo, **Daniel R. Roe**, and **Thomas E. Cheatham III**[*]

Department of Medicinal Chemistry, L.S. Skaggs Pharmacy Institute, 2000 East 30 South Skaggs 307, University of Utah, Salt Lake City, UT 84112

Abstract

DNA helices display a rich tapestry of motion on both short (< 100 ns) and long (> 1 ms) timescales. However, with the exception of mismatched or damaged DNA, experimental measures indicate that motions in the 1 μs to 1 ms range are effectively absent, which is often attributed to difficulties in measuring motions in this time range. We hypothesized that these motions have not been measured because there is effectively no motion on this timescale, as this provides a means to distinguish faithful Watson-Crick base paired DNA from damaged DNA. The absence of motion on this timescale would present a "static" DNA sequence-specific structure that matches the encounter timescales of proteins, thereby facilitating recognition. Here we report long timescale (~10-44 μs) molecular dynamics simulations of a B-DNA duplex structure that addresses this hypothesis using both an "Anton" machine and large ensembles of AMBER GPU simulations.

Introduction

Beyond the familiar and characteristic structure of the DNA Watson-Crick double helix[1], the differential dynamics and deformability of DNA are very important for its biological functions. At the macro-scale, torsional stress and supercoiling play key roles in many of the functions of DNA ranging from transcription and packing in the chromosome to DNA replication and its regulation[2,3]. Likewise, at the micro-scale and on the level of individual base pairs, DNA breathing, bending, twisting, groove fluctuations, and base pair opening are critical for function and important for recognition processes[4,5]. Many different experimental approaches have been applied to give insight into DNA dynamics, which can be separated into measurements on very fast timescales (less than 100 nanoseconds) to those probing slower events — such as internal base pair opening — on timescales of milliseconds and longer.

[*]To whom correspondence should be addressed (tec3@utah.edu).

Author contributions. DRR wrote the CPPTRAJ code and performed the principal component analysis, RG-M performed data processing, analysis, and figure generation, TEC3 performed simulation, analysis, analyzed the literature, and wrote the initial drafts with the help of RG-M. Subsequent editing and improvement is the work of all the authors.

Competing financial interests: The authors declare no competing financial interests.

On the fast timescales, many different experiments probe specific indicators of motion, such as a particular pair interactions or site interactions that can be uncovered by the applied instrument. Characteristic vibrations and/or interactions can be probed and exposed on femtosecond to nanosecond timescales with Fourier transform IR difference spectroscopy[6], 2D IR spectroscopy[7], nonlinear ultrafast vibrational spectroscopy[8], triplet anisotropy decay[9,10], field cycling NMR methods[11], as well as electron paramagnetic resonance and pulsed electron-electron double resonance measurements of active nitroxide or other spin labels[12–15], All of these methods expose identifiable signals of motion at particular and specific timescales in the femtosecond to nanosecond regime. For example, time resolved electron emission spectroscopy monitoring the dye Hoechst 33258 bound to DNA shows components of DNA relaxation at 40 ps and 12.3 ns[16], while solution NMR studies of poly-adenine tracts showed enhanced sugar puckering and backbone transitions at the junctions on the ps-ns timescale[17]. In contrast, there are other experimental measures that provide a more general picture of the motions on the fast timescale — dynamics that are collectively averaged over all the motions of the DNA, solvent, and ions. For example, time-resolved Stokes shifts of dyes bound to duplex DNA display a rich power law behavior in the fast timescale motions, interpreted from femtoseconds to ~40 nanoseconds, with motions that cannot be easily decomposed into subsets of motions on particular timescales[18–20]. Essentially, all of the different experimental approaches paint a consistent picture that suggests a rich and dynamic environment of DNA motions on the fast timescale across the entire fs-ns timescale, a picture that is supported by molecular dynamics simulations on nanosecond timescales[21,22]. Considering timescales longer than 1 millisecond, the most accurate probes of DNA dynamics and flexibility are likely from NMR spectroscopy[5]. Internal base pair opening of Watson-Crick paired bases, as inferred from measured imino proton exchange in NMR experiments, is estimated to be on the order of ~5 ms or longer, with open-state lifetimes in the ~100 ns range[23–25]. Exceptions to this are slightly faster internal base pair opening rates with isolated A-T base pairs (not in A-tracts)[26] and also in $d(CG)_n$ repeats,[27] however still with opening rates greater than ~1 ms.

Taken together, the experimental investigations suggest rich dynamics within DNA duplexes on timescales faster than a few hundred nanoseconds in terms of bending, twisting, backbone dynamics, and sugar puckering, followed on longer timescales by significant dynamics due to internal base pair opening at milliseconds and beyond. Interpretation of the experimental data suggests that there is a gap in the dynamics of Watson-Crick base paired DNA on timescales from ~1-5 microseconds to ~1 millisecond. Is this gap real or does it appear as a result of the difficulty in measuring dynamics on this timescale? There are three indirect pieces of evidence that suggest that the gap in dynamics is real and also likely to be important for recognition: specialized NMR measurements[28,29], differential scanning calorimetric (DSC) thermodynamic measurements[30], and molecular dynamics (MD) simulations.

MD simulations can accurately probe nucleic acid structure and dynamics[21]. However, unlike proteins which have been simulated to the millisecond timescale,[31] the longest published simulations of DNA duplexes are on the 1-4 μs timescale[32–34]. Although many of the properties of the internal helix are converged on the 250-300 ns timescale, the data are

not *fully* converged and longer simulations are necessary to completely relax or sample the backbone conformational sub-states (such as B_I/B_{II}, α/γ, sugar puckers), ion distributions, and bimodal twist distributions at particular base pair steps. Additionally, as terminal base pair opening occurs on a microsecond timescale and can occur in multiple ways, for example with opening into the major or minor groove, clearly the "end-effects" will not be converged. In order to understand the dynamics of DNA and convergence in the structural properties on the microsecond timescale, a relatively long ~44 μs MD simulation (ANTON) was run using the special purpose "Anton" MD engine. The DNA is an 18-mer with the sequence d(GCACGAACGAACGAACGC), hereafter referred to as GAAC, which is one of the sequences used by the ABC consortium to study sequence-dependent DNA structure and dynamics in all tetra-nucleotide repeats[35–37]. Since the initial results observed in the Anton runs were unexpected, with an effective absence of motion beyond the 1-5 μs timescales, an additional ensemble (ENS) set of simulations were performed using the GPU MD engine in AMBER. Specifically, 100 independent MD simulations were run using the standard AMBER force field for DNA at the time (ff99SB with parmbsc0 or equivalently ff12)[38–40] for on the order of ~900 ns, each with different initial ion placements, leading to an ~80+ μs aggregate MD trajectory. Moreover, in order to determine if convergence in the dynamics of the internal portion of the helix is also observed in simulations with a completely different nucleic acid force field, an additional set of ensemble simulations was performed with the CHARMM all36 (C36) force field[41]. The results suggest that MD simulations can reproducibly converge the structure and dynamics of the internal portion of a DNA helix. Combined with interpretations from experiment, the MD simulations support the assertion that the gap in Watson-Crick base paired DNA helix dynamics from 1 μs to 1 ms is real.

Results

Evidence for a gap in the dynamics on the μs-ms timescale

There are three indirect pieces of evidence that suggest that the gap in dynamics from the microsecond to the millisecond range in duplex DNA is real and also likely to be important for recognition: Specialized NMR measurements, DSC thermodynamic measurements, and MD simulations. From NMR experiments, measures of the exchange kinetics of imino protons from resonance line widths and selective longitudinal relaxation times indicate that when a G-T mismatch is present in a duplex, base pair opening times occurred faster than 1 ms[29]. More recently, by applying selective off-resonance carbon $R_{1\rho}$ NMR relaxation dispersion spectroscopy, Al-Hashimi and coworkers were able to more clearly resolve exchange processes of a 1,N6-ethenoadenine (eA) lesion/mismatch in a DNA duplex with timescales on the order of 26 ±8 μs[28]. Similar NMR experiments on native Watson-Crick base paired duplexes without mismatches or lesions are not able to resolve any exchange processes faster than 1 millisecond — with the exception of terminal base pair fraying which occurs on the microsecond timescale[42–44]. Similarly, thermodynamic signatures from DNA damage are clearly evident in DSC experiments that are likely to impact the dynamics and which are not seen in undamaged DNA[30].

MD simulations of the GAAC 18-mer duplex

To better understand the convergence of the structure and dynamics of a B-DNA duplex in MD simulations and reproducibility of independent simulations, a variety of MD simulations were performed on the "Anton" MD engine and also with AMBER. Figure 1 displays the root mean square deviation (RMSD), or overall similarity of each trajectory to a reasonably converged reference structure as a function of time, with lower RMSD indicating greater similarity. For the AMBER Anton and ENS simulations the reference is the average structure over 10 μs from an additional independent Anton run. Since the C36 runs had larger fluctuations compared to the other two, its reference is the average structure over the first 10 μs of the aggregate C36 runs. Overall, the structures sampled in the MD simulations remained close to their respective reference structures on the 10-40 μs timescale. The RMSD values fluctuate between 1-6 Å (or up to ~8 Å with C36), noting that the transient larger deviations (seen as peaks or bumps in the running average plot) are due to terminal base pair opening events (Figure 1a-c). Such events can occur on either one end of the duplex or on both sides simultaneously and are complex molecular processes. For example, the base pairs can partially open, open completely into the minor or major groove, and may open multiple base pairs. In each of the simulations multiple short-lived terminal base pair opening events are present, with significantly faster terminal base pair opening and larger excursions from the reference structure evident with the C36 force field. As the opening events are complex, it is clear that insufficient events are observed to demonstrate convergence on the timescale of these simulations. The RMSD plot for the longer Anton simulation (Figure 1a) shows a long-lasting deviation that represents base flipping into the minor groove followed by an opening of the second base pair as well. The event starts at ~21.5 μs and by the end of the simulation the two opened base pairs completely reform their Watson-Crick (WC) base pairing. The RMSD plot for the AMBER ENS simulations (Figure 1b) does not show such large deviations mainly due to the shorter duration of the individual MD trajectories, which have less of a chance to find and sustain a terminal base pair opening event.

Although the traditional RMSD plots suggest that the MD sampled structures remain near the reference structure, particularly when the terminal base pairs are omitted (Figure 1d-f), they do not directly imply stability or provide insight into "convergence" of the MD simulations. Convergence, in this context, means effectively a complete sampling of conformational events observed over particular timescales. A simple way to visualize this is by comparison of time-averaged structures calculated over different regions of the MD trajectory. When this is done for MD structures on 1-5 microsecond timescales from the initial Anton MD trajectories, we were very surprised to find that average structures compared over different microsecond time intervals were nearly identical. Given that DNA is known to be dynamic, with characteristic bending, twisting, and breathing motions, the absence of motion on the microsecond timescale — despite a dearth of experimental measures of dynamics on this timescale with the exception of terminal base pair fraying — initially caused us concern. Given that the Anton machine is known to make approximations for speed, we initially hypothesized that the effective lack of motion may be due to either Anton or alternatively artifacts of the applied force field. This caused us to reproduce MD trajectories independently using AMBER on GPUs and CPUs, and also — thanks to access to large-scale GPU resources on XSEDE KIDS/Keeneland and the Blue Waters Petascale

Resource — to perform ensembles of independent MD runs with both the AMBER ff12 and CHARMM C36 force fields.

Convergence of the DNA structure over time

Shown in Figure 2 are molecular graphics representations of the representative structures (those closest to their cluster centroid) from ten clusters generated by clustering the MD trajectories based on structural similarity after smoothing over different time intervals (specifically 50 ns, 325 ns, 2 μs and 8 μs) with CPPTRAJ[45]. The terminal base pairs on each end were not used during the clustering. The ten representative 50 ns averaged structures show that even at this short timescale, the motions of the bases on the internal portions of the helix are fairly well converged with structural deviations, or differences, most evident on the backbone and in the termini. By 325 ns of time averaging, the internal helix is even more converged or overlapping, with the most significant structural differences occurring on the termini. By 2 μs of time-averaging, the ten overlapping structures are extremely close to each other, with only small differences observed at the termini. By 8 μs of time averaging there is a near-complete overlap of the structures with only very small deviations at the termini.

As an additional metric to explore convergence, we also applied our previously developed method for calculating the RMSD average correlation (RAC) which shows how the similarity of structures within a trajectory changes when averaged over increasing time windows. The faster the RAC decays, the more rapidly the behavior of the system approaches the overall average. This is shown in Figure 3 for each of the simulations using only the heavy atoms of the 10 central base pairs d(GAACGAACGA) using the same reference structure (the 10 μs average structure from the initial Anton simulation). The decay of the RAC values is extremely fast, such that on the timescale of ~80-130 ns the average RMSD values are already less than 0.1 Å for the AMBER simulations. The decay of the RAC with the ff12 force field continues as the simulation progresses and the average RMSD values plateau. The fast decay suggests a rapid convergence of the dynamics and fluctuations of the central base pairs such that when the averaging occurs over the 1-5 μs timescale, essentially the structure is converged and all the dynamics are absent. This agrees with the previously discussed experimental evidence suggesting a gap in DNA dynamics on the 1 μs to 1 ms timescale. Considering the RAC analysis, the slope is essentially flat by 2.5-5 μs for all of the systems with deviations of less than 0.05 Å by 2 μs for the simulations with the AMBER ff12 force field. In terms of atomic positional fluctuations of the atoms around the average reference structure, by 1 μs of time averaging, the base atom fluctuations are already less than 0.1 Å and the phosphate and sugar atom fluctuations are less than 0.15 Å. With the CHARMM C36 force field, the convergence is slightly slower due to structural distortions induced by the frequent base pair opening which traverses deeper into the helix. This is clearly evident in Figure 1 in the bottom panel for the C36 simulation which shows large RMSD fluctuations away from the reference even when the terminal four base pairs on each end are omitted. Despite this, average RMSD values are still less than 0.075 Å by 2.5 μs. These are *very* small deviations in the structure. Together, the MD simulations suggest that no significant structural changes are to be expected in the central regions of the DNA beyond ~1 μs unless an internal base pair opening event occurs, which as previously

mentioned is expected at time ranges between 5-100 milliseconds. The inset plot shows the RAC decay at shorter timescales and suggests, interestingly, that the dynamics decay initially faster with the CHARMM C36 force field and then slow; again, this is likely due to the disruption in the internal helix structure resulting from the larger and more frequent terminal base pair opening events. Despite subtle differences, it is clear with either force field that the decay in the RAC is very rapid, with deviations of less than 1 Å already within 10 ns of MD simulation. Given the computational cost of these calculations and expiration of our allocation of computer time on Anton, we do not yet have the results of MD simulations with mismatches. However, further ensemble simulations are underway on the Blue Waters Petascale Resource and various XSEDE resources with faster GPU nodes to explore mismatches and ideally dynamic events from internal base pair opening on the low μs timescale. Additionally, although the observed converged average structures with both force fields show general agreement with sequence specific structural trends expected from experiment,[35–37,46] there is no direct way to compare this structure, chosen initially because of its bimodal twist distribution at the central CpG step observed in the original ABC experiments,[35–37] to experiment as an NMR solution structure has not been published. Details on the structure and observed differences with different force fields are discussed in greater detail in subsequent work.

Convergence and reproducibility of the MD simulations

Principal component (PC) analysis can be used to reveal the dominant modes of motion in a trajectory, and comparing PC projections between different MD simulations can provide a metric of similarity between the dominant modes of motion sampled by the different MD trajectories[47,48]. To measure how well the motions (i.e. the dynamic properties) of DNA in the Anton and ENS simulations converge with respect to each other, we performed PC analysis using both trajectories. Figure 4 shows the overlap of histograms of the PC projections for the five most dominant models of motion from the Anton and ENS simulations (shown as solid and dashed lines respectively). Figure 4a shows the results of the PCA performed on all base pairs. While there is some overlap between the 1st, 2nd, and 5th PCs (corresponding to global DNA bending/twisting motions), there is significantly less overlap for the 3rd and 4th PCs which happen to correspond to terminal base-pair opening events. The Anton simulations have wider distributions for these PCs, indicating these motions are less prevalent in the ENS simulations, which is consistent with the observations of the RMSD plots (Figure 1) and the fact that the ENS MD simulations are each individually shorter in duration. If the first and last base pairs are removed from the analysis, we notice an overall better overlap between the PCs of both simulations (Figure 4b). If only the central ten base pairs are included in the analysis (Figure 4c), the histograms are essentially identical. These results suggest that overall the same types of motion are sampled in the two independent simulations, but also that terminal base pair opening events are enhanced in the longer timescale Anton simulation.

Rapid fluctuations in the groove width

Another measurement of interest is groove width, as the space in the grooves is critically important for drug and protein binding. Figure 5 shows the average and standard deviations of the major and minor groove widths for the Anton and ENS simulations. In general, the

difference in groove widths between the Anton and ENS simulations are very small, on the order of 0.05 Å (noting that the two terminal base pairs have been omitted from the calculation), and only moderately higher at A15 and C16 with values of ~0.10 Å and ~0.20 Å for the major and minor groove, respectively, which is due to the Anton long-lived base pair opening event at that end of the helix. Although the distribution of the groove widths are well converged between the two simulations, the standard deviations in the groove widths themselves across each simulation are quite large, on the order of ~1.5-1.9 Å. The observed fluctuations in groove width are considerably larger than the sequence specific structural differences. The MD simulations show that on very rapid timescales a wide variety of groove widths are sampled, which then effectively converge to the average groove widths on timescales on the order of μs. This corresponds to the timescale for protein or ligand association with the DNA and likely is important for recognition. With the dynamics effectively averaged out in the internal DNA helix on the 1 μs to 1 ms timescale, potential ligands effectively see a consistent structure and can bind by conformational selection (as opposed to induced fit). This contrasts with DNA mismatches or lesions, as discussed previously, which alters DNA dynamics to signal an unusual “beacon” of motions on the μs timescale that may help DNA repair protein recognition. Moreover, as internal base pair opening events occur on much longer timescales (>1 ms) with rapid base pair closing, induced fit by the protein or ligand binding are likely the operational mechanisms for recognition of open states and/or to induce kinking and/or bending. These results, taken together with the current experimental understanding of DNA dynamics, suggest that this gap in DNA dynamics from 1 μs to 1 ms may provide a means for fidelity in recognition.

Discussion

The results of this work suggest that MD simulations can converge the structure and dynamics of the internal portion of a DNA helix and demonstrate reproducibility from independent simulations. The results clearly show fast decay in the dynamics of the DNA helix with both the latest AMBER and CHARMM nucleic acid force fields, and that a large ensemble of independent simulations compared to a single long Anton MD simulation give equivalent results in terms of structure and dynamics, at least for internal base pairs. Motions faster than μs in Watson Crick base paired B-DNA include sampling of different backbone conformational states, rapid fluctuations in the groove widths, sugar repuckering, bending, twisting and sampling of a wide variety of helicoidal parameter distributions, and terminal base pair opening. On the μs timescale, with the exception of terminal base pair fraying, all these motions are effectively averaged out. In the absence of mismatches or lesions in the DNA helix (and with the exception of terminal base pair fraying) other motions are not expected until the ms timescale and this motion will largely be internal base pair opening. The results from MD simulation combined with interpretations from experiment support the assertion that the gap in Watson-Crick base paired DNA helix dynamics from 1 μs to 1 ms is real.

Methods

The B-DNA sequence d(GCACGAACGAACGAACGC) was generated following the original ABC protocol[35–37]. Simulations were run using the Amber ff99SB[38,39] force field

using the parmbsc0 modifications[40], the SPC/E water model[49], neutralizing charge with K^+ and an excess of K^+ and Cl^- to reach a ~150mM concentration (i.e. an additional 60 ions of each type) with the Dang ion parameters[50]. The initial ion positions were randomized by swapping with a random water molecule such that no ion was closer than 4 Å to any other like ion, and such that all ions were at least 6 Å from the DNA structure. The simulation protocol applied is equivalent to the earlier ABC simulations. Production runs with AMBER were performed at 300 K with weak coupling algorithm for temperature and pressure control (5 ps relaxation time). Long-range interactions were calculated using PME using default settings, SHAKE on bonds to hydrogen, and a 2 fs integration time step. For the Anton simulation, the coordinates were recorded every 50 ps using the specialized MD engine Anton, built by DE Shaw Research, Inc. that has been loaned to the Pittsburgh Supercomputing Center. Allocations of computer time on Anton were made available through a competitive proposal and allocation process to researchers in the United States. This process was administered by the Pittsburgh Supercomputing Center and a National Academies Review Committee. The Anton simulations were performed through two allocations of computer time on Anton, each for 1 week on the 512 computing element machine in 2011 and 2012. The total Anton simulation time for the longer simulation was of 44.06 μs. In addition, an earlier initial simulation on Anton was run for over 10 μs. Multiple different versions of the Anton software and microcode (initially 2.4.1 and then 2.4.5) were applied and the available “amber_topNrst2cms.py” script on anton.psc.edu was used with Desmond to convert AMBER parameter and topology inputs into a format appropriate for Anton. During error checking, we noticed that with the parmbsc0 force field C5’ atoms were erroneously specified with a zero mass which was fixed. The Anton supplied “guess_chem”, “refinesigma”, and “subboxer” programs set up the system and then production runs were initiated at constant 300 K temperature and 1 bar pressure with weak coupling using a coupling time “tau” of 10.0, a maximum and minimum velocity scaling of 1.2 and 0.85, and a maximum and minimum expansion per step of 1.05 and 0.97 and kappa of $4.5x10^{-5}$ were imposed. The “max_strain” was set to 0.08 and a 2 fs integration time step with RESPA on the long-range non-bonded interactions every third step was performed. The resulting Anton trajectories were converted to DCD trajectory format using VMD.

The ENS (ensemble) simulations were generated using 100 independent simulations starting from the same structure but with randomized ion positions. Each AMBER ff12 simulation was run for no less than 300 ns using either the AMBER12 or AMBER14 PMEMD.cuda[51–55] version on either the XSEDE Keeneland or KIDS GPU cluster from Georgia Tech / NICS. The initial 100 ns of MD simulation from each trajectory were deleted to remove any artifacts from equilibration and the remaining frames were concatenated together, resulting in one large 22+ μs trajectory. Given that the dynamics converge so rapidly, the arbitrary choice of 100 ns for omission appears justified. Given queue waits, this required approximately 7 months on nVIDIA M2090 GPUs. While this article was in review, the initial 300 ns trajectories were extended to over ~900 ns per independent run on the Blue Waters Petascale Resource leading to an aggregate trajectory of 80+ trajectory; the extension of these trajectories did not alter any of the results significantly, except for better, albeit not converged, sampling of the terminal base pair opening events. To determine if the gap in the dynamics of the internal portion of the helix are also observed in simulations with

a completely different nucleic acid force field, an additional set of ensemble simulations was performed with the CHARMM all36 (C36) force field.[41] The 100 independent runs on at least the 900 ns timescale were performed with the CHARMM C36 force field, built in an equivalent manner using the CHARMM C36 force field, but with CHARMM TIP3P and ion parameters instead. The files were built using CHARMM version c37b2 after the force fields were converted over to AMBER/PDB/standard DNA naming conventions, and then converted into AMBER compatible input coordinate and parameter/topology formats with the chamber utility of AmberTools. The C36 runs were performed on the NCSA Blue Waters resource. Thanks to access to faster nVIDIA K20 GPUs on Blue Waters, improvements in the performance of the AMBER GPU code, and shorter queue waits, this longer set of trajectories only required two months to generate. All the analysis was performed using a development version of CPPTRAJ45 and this code is now freely available in the released version of Amber 14. To create average structures as references, the coordinates from the MD trajectories at 50 ps intervals were RMS-fit to the first frame over all DNA atoms and a straight coordinate average over the specified interval was performed. Clustering by RMSD was done using the average-linkage algorithm on the MD trajectories at 50 ps intervals using a sieve of 250 frames. DNA structural analysis was made using PTRAJ, CPPTRAJ and Curves+[56]. All the analysis commands, inputs, raw and derived simulation data are available for download at http://www.amber.utah.edu/DNA-dynamics/GAAC.

Acknowledgments

This work was supported by the NIH R01-GM081411 and the NSF CHE-1266307, with extensive computational support from NSF XSEDE XRAC MCA01S027 (including both friendly user and allocated time on the Georgia Tech/NICS KIDS and Keeneland GPU resources), NSF/NCSA/U Illinois Blue Waters Petascale Resource (PRAC OCI-1036208 and OCI 07-25070), and also the Center for High Performance Computing at the University of Utah.

References

1. Watson JD, Crick FHC. A Structure for Deoxyribose Nucleic Acid. Nature. 1953; 171:737–738. [PubMed: 13054692]
2. Baranello L, Levens D, Gupta A, Kouzine F. The importance of being supercoiled: how DNA mechanics regulate dynamic processes. Biochim. Biophys. Acta. 2012; 1819:632–638. [PubMed: 22233557]
3. Lavelle C. Pack, unpack, bend, twist, pull, push: the physical side of gene expression. Curr. Opin. Genet. Dev. 2014; 25C:74–84. [PubMed: 24576847]
4. Von Hippel PH, Johnson NP, Marcus AH. Fifty years of DNA “Breathing”: Reflections on old and new approaches. Biopolymers. 2013; 99:923–954. [PubMed: 23840028]
5. Frank-Kamenetskii MD, Prakash S. Fluctuations in the DNA double helix: A critical review. Phys. Lfe Rev. 2014; 11:153–170.
6. Rüdisser S, Hallbrucker A, Mayer E. B-DNA’s Conformational Substates Revealed by Fourier Transform Infrared Difference Spectroscopy. J. Am. Chem. Soc. 1997; 119:12251–12256.
7. Greve C, Elsaesser T. Ultrafast two-dimensional infrared spectroscopy of guanine-cytosine base pairs in DNA oligomers. J. Phys. Chem. B. 2013; 117:14009–14017. [PubMed: 24127664]
8. Szyc Ł, Yang M, Nibbering ETJ, Elsaesser T. Ultrafast vibrational dynamics and local interactions of hydrated DNA. Angew. Chem. Int. Ed. Engl. 2010; 49:3598–3610. [PubMed: 20446279]
9. Hogan M, Wang J, Austin RH, Monitto CL, Hershkowitz S. Molecular motion of DNA as measured by triplet anisotropy decay. Proc. Natl. Acad. Sci. U.S.A. 1982; 79:3518–3522. [PubMed: 6954497]

10. Georghiou S, Bradrick TD, Philippetis A, Beechem JM. Large-amplitude picosecond anisotropy decay of the intrinsic fluorescence of double-stranded DNA. Biophys. J. 1996; 70:1909–1922. [PubMed: 8785350]
11. Roberts MF, Cui Q, Turner CJ, Case DA, Redfield AG. High-resolution field-cycling NMR studies of a DNA octamer as a probe of phosphodiester dynamics and comparison with computer simulation. Biochemistry. 2004; 43:3637–3650. [PubMed: 15035634]
12. Hustedt EJ, Spaltenstein A, Kirchner JJ, Hopkins PB, Robinson BH. Motions of short DNA duplexes: An analysis of DNA dynamics using an EPR-active probe. Biochemistry. 1993; 32:1774–1787. [PubMed: 8382521]
13. Keyes RS, Bobst EV, Cao YY, Bobst AM. Overall and internal dynamics of DNA as monitored by five-atom-tethered spin labels. Biophys. J. 1997; 72:282–290. [PubMed: 8994613]
14. Marko A, et al. Conformational flexibility of DNA. J. Am. Chem. Soc. 2011; 133:13375–13379. [PubMed: 21702503]
15. Okonogi TM, Reese AW, Alley SC, Hopkins PB, Robinson BH. Flexibility of duplex DNA on the submicrosecond timescale. Biophys. J. 1999; 77:3256–3276. [PubMed: 10585948]
16. Banerjee D, Pal SK. Direct observation of essential DNA dynamics: melting and reformation of the DNA minor groove. J. Phys. Chem. B. 2007; 111:10833–10838. [PubMed: 17676797]
17. Nikolova EN, Bascom GD, Andricioaei I, Al-Hashimi HM. Probing sequence-specific DNA flexibility in A-tracts and pyrimidine-purine steps by nuclear magnetic resonance ^{13}C relaxation and molecular dynamics simulations. Biochemistry. 2012; 51:8654–8664. [PubMed: 23035755]
18. Andreatta D, et al. Power-law solvation dynamics in DNA over six decades in time. J. Am. Chem. Soc. 2005; 127:7270–7271. [PubMed: 15898749]
19. Brauns EB, Madaras ML, Coleman RS, Murphy CJ, Berg MA. Measurement of Local DNA Reorganization on the Picosecond and Nanosecond Time Scales. J. Am. Chem. Soc. 1999; 121:11644–11649.
20. Brauns E, Madaras M, Coleman R, Murphy C, Berg M. Complex Local Dynamics in DNA on the Picosecond and Nanosecond Time Scales. Phys. Rev. Lett. 2002; 88:158101. [PubMed: 11955218]
21. Beveridge DL, Cheatham TE 3rd, Mezei M. The ABCs of molecular dynamics simulations on B-DNA, circa 2012. J. Biosci. 2012; 37:379–397. [PubMed: 22750978]
22. Sen S, Andreatta D, Ponomarev SY, Beveridge DL, Berg MA. Dynamics of water and ions near DNA: comparison of simulation to time-resolved stokes-shift experiments. J. Am. Chem. Soc. 2009; 131:1724–1735. [PubMed: 19191698]
23. Guéron M, Kochoyan M, Leroy JL. A single mode of DNA base-pair opening drives imino proton exchange. Nature. 1987; 328:89–92. [PubMed: 3037381]
24. Leijon M, Gräslund A. Effects of sequence and length on imino proton exchange and base pair opening kinetics in DNA oligonucleotide duplexes. Nucleic Acids Res. 1992; 20:5339–5343. [PubMed: 1331987]
25. Moe JG, Russu IM. Proton exchange and base-pair opening kinetics in 5'-d(CGCGAATTCGCG)-3' and related dodecamers. Nucleic Acids Res. 1990; 18:821–827. [PubMed: 2156233]
26. Leroy JL, Charretier E, Kochoyan M, Guéron M. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. Biochemistry. 1988; 27:8894–8898. [PubMed: 3233210]
27. Leijon M, Leroy JL. Internal motions of nucleic acid structures and the determination of base-pair lifetimes. Biochimie. 1997; 79:775–779. [PubMed: 9523020]
28. Hansen AL, Nikolova EN, Casiano-Negroni A, Al-Hashimi HM. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon $R_{1\rho}$ NMR spectroscopy. J. Am. Chem. Soc. 2009; 131:3818–3819. [PubMed: 19243182]
29. Moe JG, Russu IM. Kinetics and energetics of base-pair opening in 5'-d(CGCGAATTCGCG)-3' and a substituted dodecamer containing G T mismatches. Biochemistry. 1992; 31:8421–8428. [PubMed: 1327102]
30. Gold B, Stone MP, Marky LA. Looking for waldo: a potential thermodynamic signature to DNA damage. Acc. Chem. Res. 2014; 47:1446–1454. [PubMed: 24702131]

31. Shaw DE, et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010; 330:341–346. [PubMed: 20947758]
32. Pérez A, Luque FJ, Orozco M. Dynamics of B-DNA on the microsecond time scale. J. Am. Chem. Soc. 2007; 129:14739–14745. [PubMed: 17985896]
33. Dans PD, Pérez A, Faustino I, Lavery R, Orozco M. Exploring polymorphisms in B-DNA helical conformations. Nucleic Acids Res. 2012; 40:10668–10678. [PubMed: 23012264]
34. Dršata T, et al. Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer. J. Chem. Theory Comput. 2012; 9:707–721. [PubMed: 23976886]
35. Beveridge DL, et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. Biophys. J. 2004; 87:3799–3813. [PubMed: 15326025]
36. Dixit SB, et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. Biophys. J. 2004; 89:3721–3740. [PubMed: 16169978]
37. Lavery R, et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. Nucleic Acids Res. 2009; 38:299–313. [PubMed: 19850719]
38. Cheatham TE 3rd, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J. Biomol. Struct. Dyn. 1999; 16:845–862. [PubMed: 10217454]
39. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 2000; 21:1049–1074.
40. Pérez A, et al. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J. 2007; 92:3817–3829. [PubMed: 17351000]
41. Hart K, et al. 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. [PubMed: 22368531]
42. Leroy JL, Kochoyan M, Huynh-Dinh T, Guéron M. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. J. Mol. Biol. 1988; 200:223–238. [PubMed: 2836594]
43. Nonin S, Leroy J-L, Gueron M. Terminal Base Pairs of Oligodeoxynucleotides: Imino Proton Exchange and Fraying. Biochemistry. 1995; 34:10652–10659. [PubMed: 7654719]
44. Andreatta D, et al. Ultrafast dynamics in DNA: “fraying” at the end of the helix. J. Am. Chem. Soc. 2006; 128:6885–6892. [PubMed: 16719468]
45. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput. 2013; 9:3084–3095. [PubMed: 26583988]
46. Schneider B, Neidle S, Berman HM. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. Biopolymers. 1997; 42:113–124. [PubMed: 19350745]
47. Bergonzo C, et al. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. J. Chem. Theory Comput. 2014; 10:492–499. [PubMed: 24453949]
48. Roe DR, Bergonzo C, Cheatham TE 3rd. Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. J. Phys. Chem. B. 2014; 118:3543–3552. [PubMed: 24625009]
49. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. J. Phys. Chem. 1987; 91:6269–6271.
50. Dang LX. Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study. J. Am. Chem. Soc. 1995; 117:6954–6960.
51. Case DA, et al. The Amber biomolecular simulation programs. J. Comput. Chem. 2005; 26:1668–1688. [PubMed: 16200636]

52. Pearlman DA, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput. Phys. Commun. 1995; 91:1–41.

53. Le Grand S, Götz AW, Walker RC. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. Comput. Phys. Commun. 2013; 184:374–380.

54. Götz AW, et al. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput. 2012; 8:1542–1555. [PubMed: 22582031]

55. Salomon-Ferrer R, Götz AW, Poole D, Grand S, Le, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. 2013; 9:3878–3888. [PubMed: 26592383]

56. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+ Nucleic Acids Res. 2009; 37:5917–5929. [PubMed: 19625494]

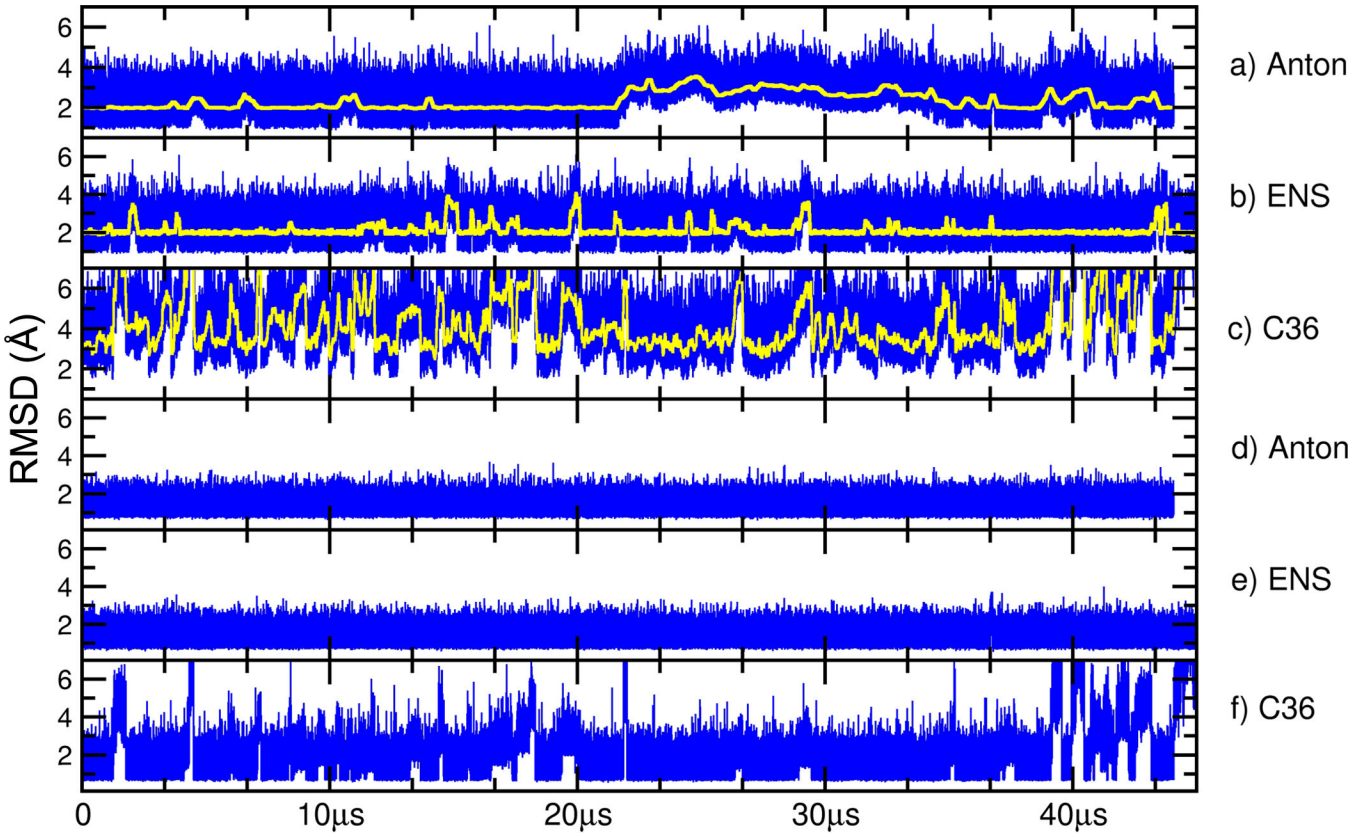

Figure 1.

Root mean squared deviation (RMSD) of the frames sampled in MD simulations to time averaged reference structures. Shown in the top three panels are RMSD values over all atoms (y-axis in Å) in blue and running averages over 5000 frames in yellow for the Anton simulation (a), the ENS simulations (b), and the CHARMM C36 simulations (c). In the bottom three panels are the RMSD values omitting the four terminal base pairs on each end for Anton (d), ENS (e), and CHARMM C36 (f). Whereas the AMBER ff12 data is flat due to no significant base pair opening beyond the first 2-3 base pairs, the C36 data still shows excursions away from lower RMSD values due to more significant perturbation from base pair fraying on the ends of the helix. The reference structure for the Anton and ENS simulations is an average structure calculated over 10 μs from an initial independent Anton simulation with ff12. The reference structure for the C36 simulations is an average structure calculated over the first 10 μs of the C36 ENS runs. When calculating the average structure, frames were RMS-fit to the first frame.

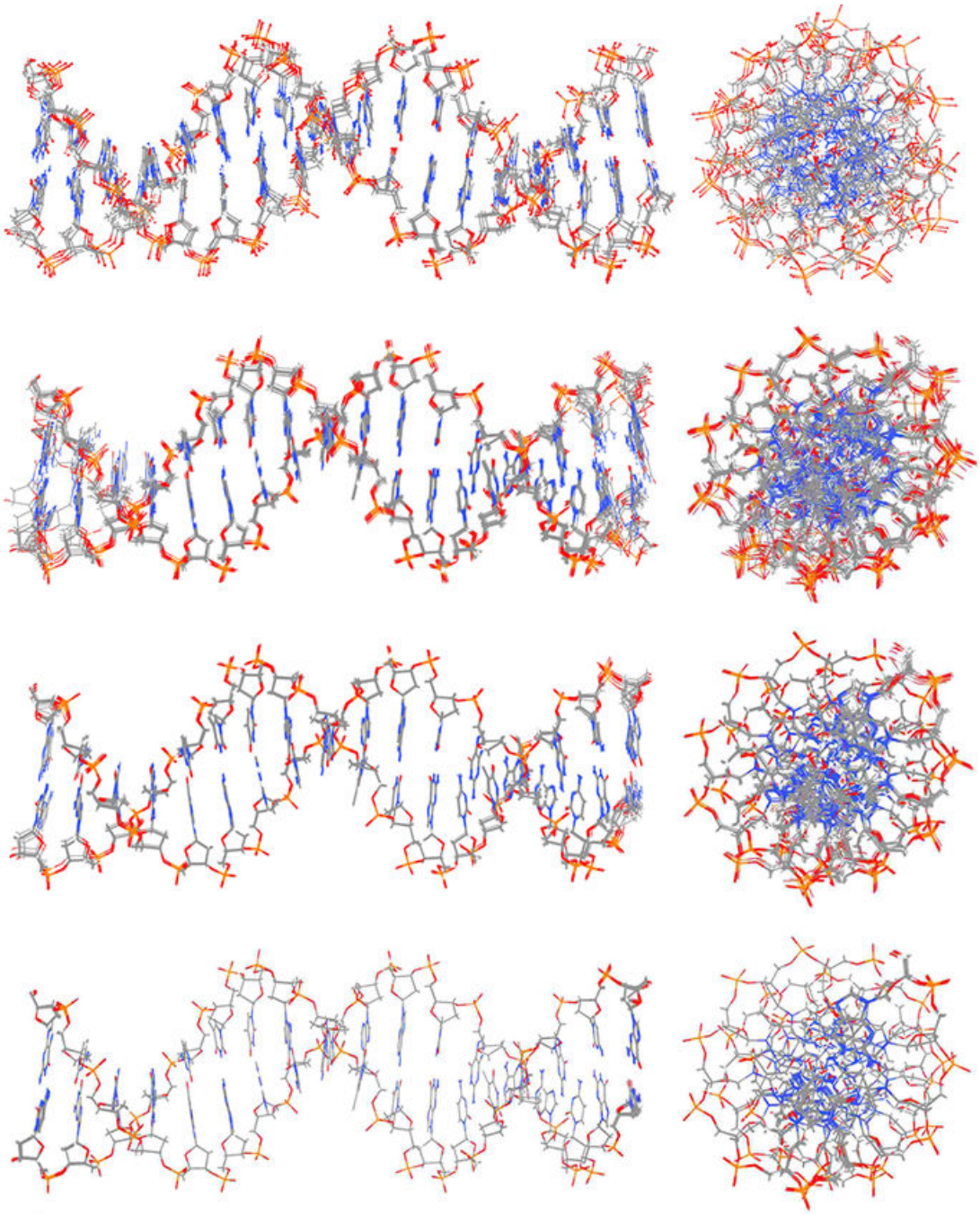

Figure 2.

Structures from time-averaged or smoothed MD trajectories. Straight coordinate running averages over different time windows (50 ns, 325 ns, 2 μs, and 8μs) of RMS fit snapshots at 50 ps intervals were calculated for the AMBER ENS MD simulations and the resulting trajectories were clustered into ten clusters (using pairwise RMSD calculations over all atoms omitting the terminal base pairs from the clustering). The 10 most representative structures (i.e. those closest to the centroid of each cluster) are displayed as molecular graphics.

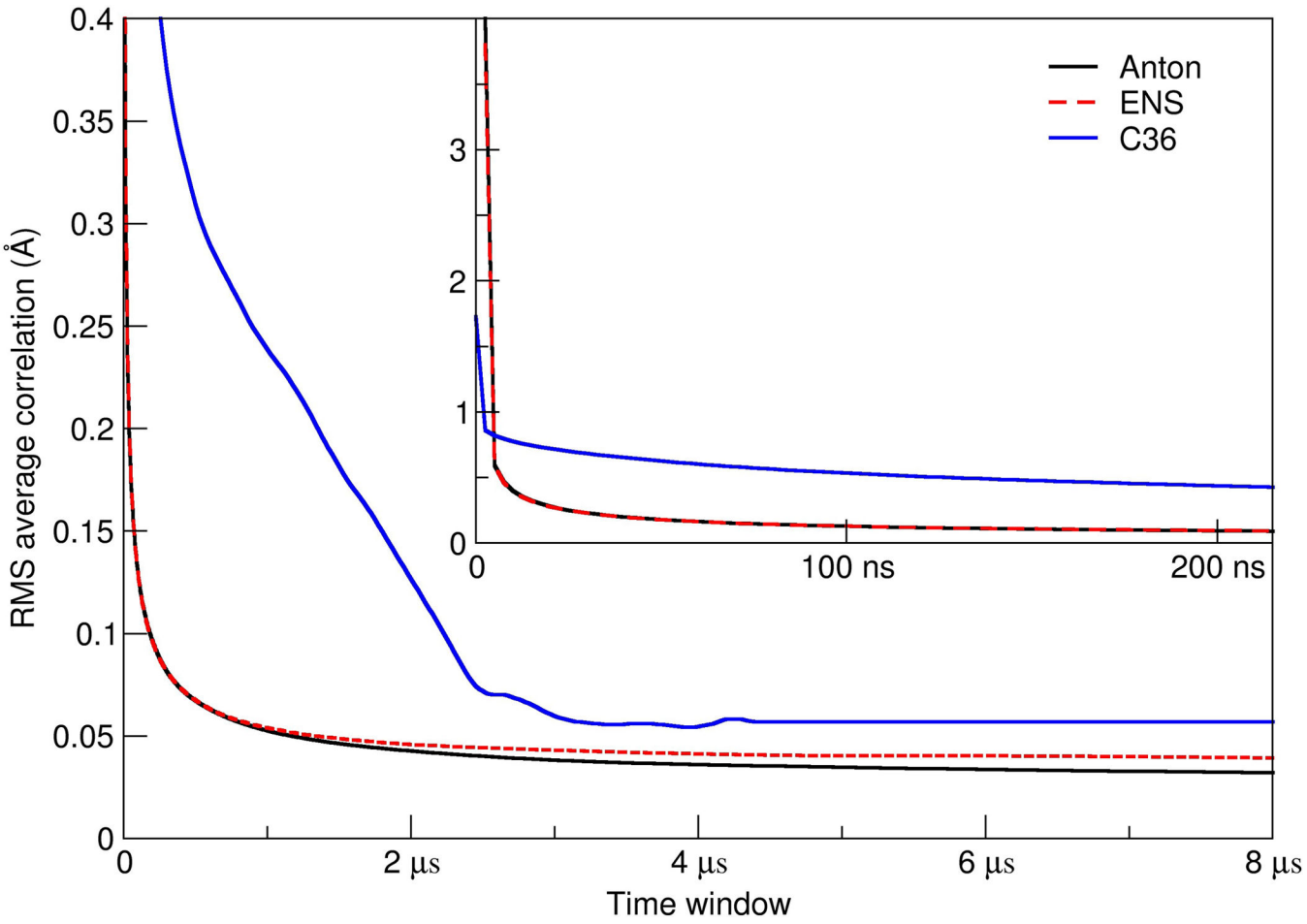

Figure 3.

Decay in the average RMSD values as the window for time averaging increases. Shown is the root mean squared deviation average correlation (RAC) for the Anton (solid black line), AMBER ENS (dashed red line), and CHARMM C36 ENS (solid blue line) simulations based on the heavy atoms of the DNA omitting the four terminal base pairs on each end of the helix (i.e. using only the 10 central base pairs). Given that the two force fields do not converge to the same average structure, to approach low values in the RAC analysis, the reference structure for the ff12 runs was the 0-10 μs average structure from the initial Anton simulation while that for the C36 ensemble was the 0-10 μs average structure from the C36 ENS simulations. The inset plot shows the same information, focusing on the initial decay of the average RMSD values over shorter time windows.

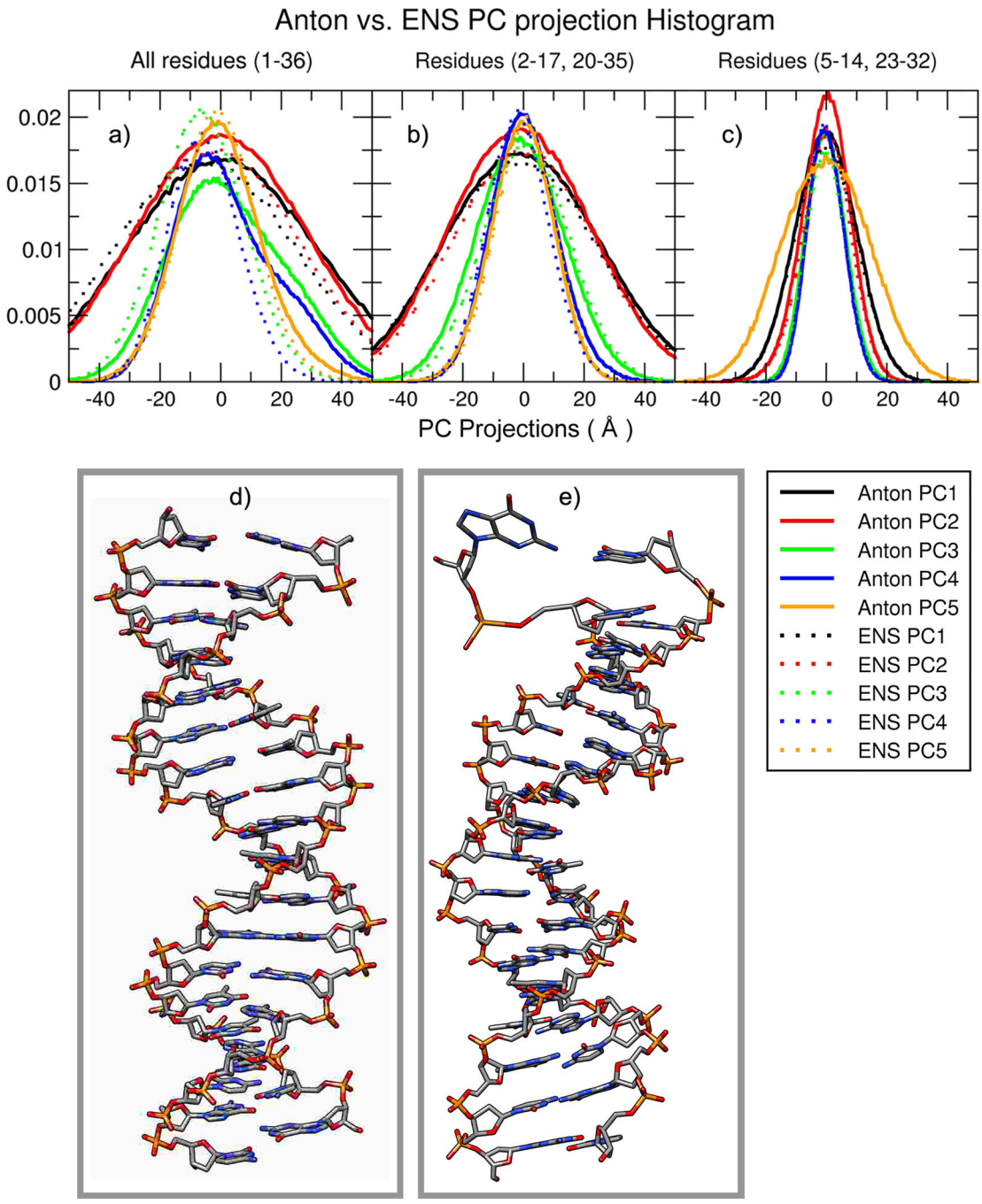

Figure 4.

Plots (a), (b) and (c) shows the overlap of the independent principal component (PC) projection histograms from each of the MD simulations from PCs calculated in Cartesian space from the combined Anton and ENS MD trajectories over all DNA atoms. Visual inspection shows significant divergence in the histogram overlap of the third (green) and fourth PC (blue). As the terminal base pairs are removed from the analysis (b) and (c), the PC histograms more clearly overlap and sample the same space. The molecular graphics in (d) and (e) are representative structures obtained from the projected frames of the third and

fourth PC mode respectively, which show conformational difference due to terminal base pair opening events.

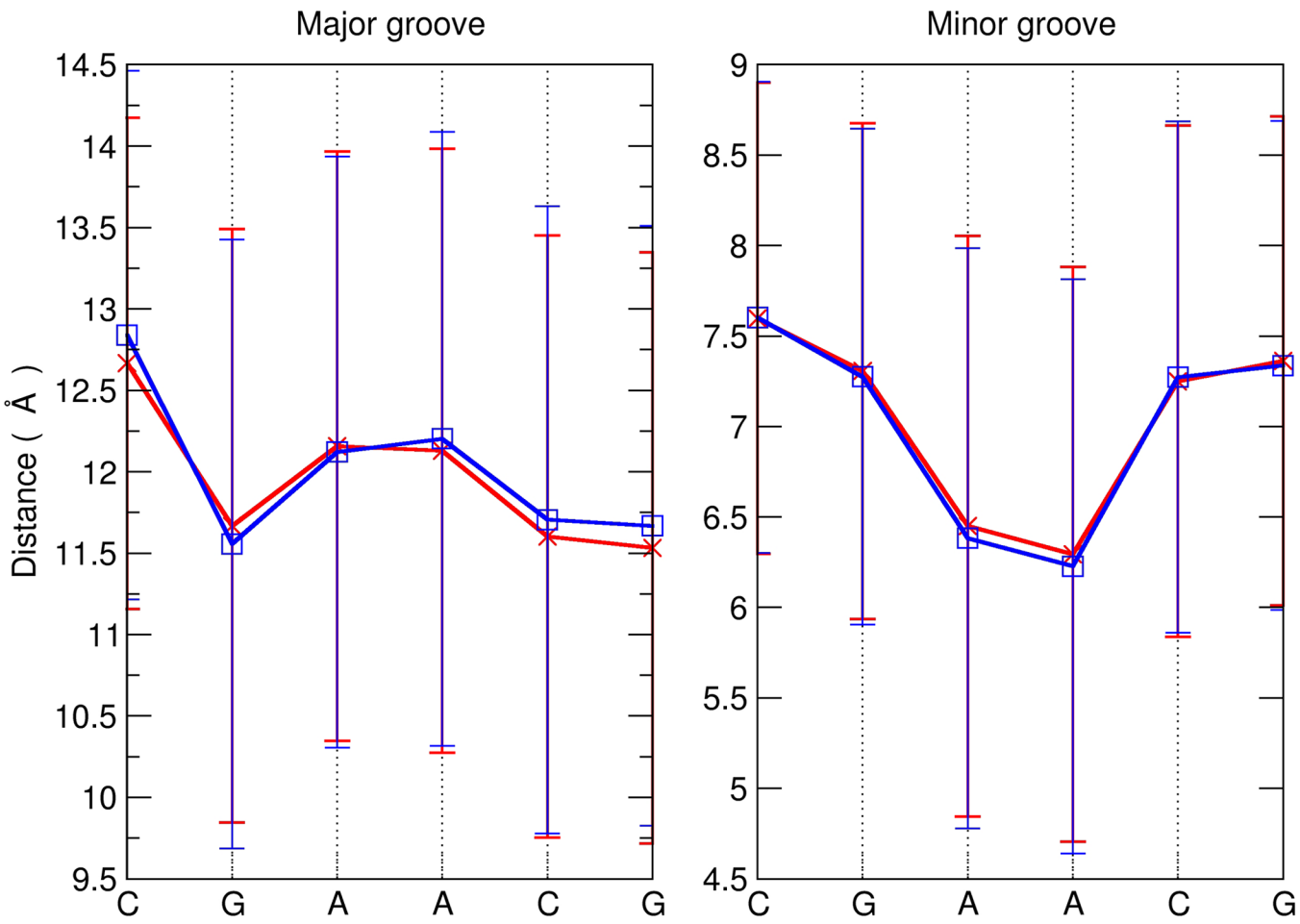

Figure 5.

Average and standard deviation of major and minor groove widths for the ENS (blue lines) and Anton (red lines) simulations. The widths for the central CGAACG portion of the helix (residues 8-13, 24-29) are shown. Data was calculated for each frame from the MD trajectory at 50 ps intervals using the Curves+ program.