Overview

The atomistic simulation of DNA

Charles A. Laughton[1]* and Sarah A. Harris[2]

We review the current state of the art relating to the atomistic simulation of the structure and dynamics of DNA. We begin with a brief historical overview to set the scene and introduce some of the key issues that had to be addressed to progress the field and then we divide our discussion of the current situation into two sections. First, we overview the role that simulation has played, closely intertwined with experimental studies, in increasing our understanding of the biomechanical properties of DNA, for example, the way in which its structure responds to perturbations such as stretching and over- and under-twisting. Second, we discuss how atomistic simulations are contributing to our deeper understanding of nucleic acid recognition—both by proteins and by small-molecule ligands. In both areas, we emphasize not only where simulation has been particularly successful but also where thorny problems remain to tax the ingenuity of computational scientists in close collaboration with their experimental colleagues.
 WIREs Comput Mol Sci 2011 1 590–600 DOI: 10.1002/wcms.46

INTRODUCTION

DNA is a dynamic molecule. Sometimes it is wrapped around proteins and highly compacted, other times it is expanded and exposed. It is transiently looped and bent, cut and resealed, damaged and repaired, and the strands pulled apart and then reannealed. From the early days, this behavior suggested a dynamic molecular structure, somewhat at odds with the elegant and regular models derived from fiber diffraction data.[1] So, perhaps one of the seminal moments in the development of our understanding of DNA came with the publication of the first crystal structure of an appreciable length of double-stranded DNA [the oligonucleotide d(CGCGAATTCGCG)$_2$) by Drew et al.[2]]. This revealed a far from canonical double helix (Figure 1), with bends, nonplanar base pairs, variations in groove width, and many other irregularities. This structure thus provided the first direct experimental evidence to inform the already wide-ranging discussion as to the flexibility of DNA, to what extent its structure and deformability were dependent on the local base sequence, and what the biological significance of this might be. Many of these issues were not amenable to direct experimental determination, but they were potentially addressable through the then-developing techniques of computational modeling.

*Correspondence to: Charles.laughton@nottingham.ac.uk

[1]School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham, UK

[2]School of Physics and Astronomy, University of Leeds, Leeds, UK

DOI: 10.1002/wcms.46

Here, we briefly discuss the history of the computational simulation of DNA before moving on to overview the current state of the art. For this, we have divided the discussion into two parts: one concentrating on what computational science has been able to reveal about the mechanical properties of DNA and the other concentrating on what has been revealed about the process of DNA recognition by both small molecules and proteins.

A Brief History of DNA Simulation

The first reported molecular dynamics (MD) simulation of DNA was by Levitt[3] in 1983 [90 picosecond simulations of the Drew and Dickerson dodecamer, and of d(A)$_{24}$·d(T)$_{24}$], closely followed by Tidor et al.[4] [a 60 picosecond simulation of the B-form hexamer d(CGCGCG)$_2$]. The same year saw the first RNA simulations reported by Prabhakaran et al.[5] (12 picosecond simulations of phenylalanine tRNA). These first reports highlight difficulties that were to dog the early years of DNA simulation: stable behavior was hard to achieve and initial structures quickly adopted highly distorted and unrealistic conformations unless artificial adjustments were made such as ignoring electrostatic interactions or enhancing hydrogen bonding. In many respects, this was not surprising: the force fields were very approximately parameterized, the high net charge on DNA was always going to make the correct treatment of electrostatic interactions critical, and early simulations were done without explicit consideration of the surrounding environment of water molecules

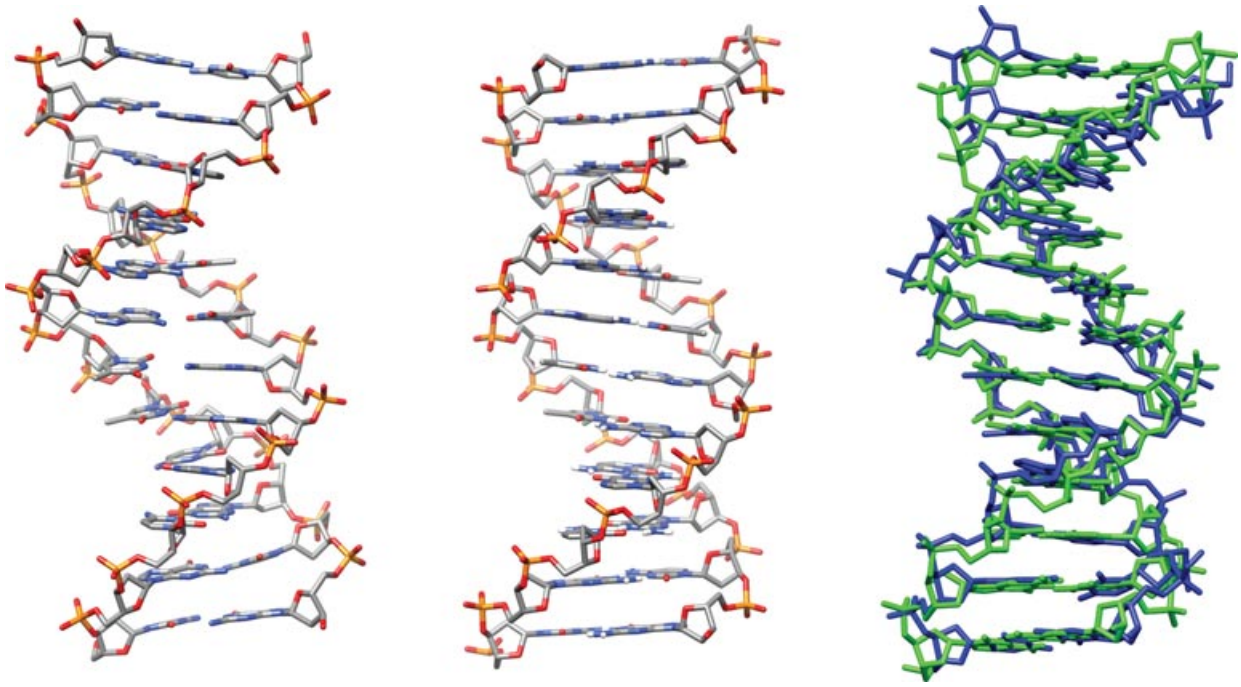

FIGURE 1 | Comparison of (left) the crystal structure of the DNA dodecamer d(CGCGAATTCGCG)$_2$ [PDB code 1BNA, (2)] and (center) a canonical B-form model of the same sequence. A variety of deviations from a fully regular structure are evident in the crystal structure, particularly if the two are overlaid (right; crystal structure in blue and regular B-form model in green).

and counterions, when it was already well known that DNA structure is very sensitive to hydration. So, in 1985, we have Seibel et al.[6] 106 picosecond simulation of the pentamer d(CGCGA)·d(TCGCG) in a droplet of 830 water molecules and in 1986 van Gunsteren et al.[7] 80 picosecond simulation of the octamer d(CGCAACGC)·d(GCGTTGCG) in a periodic box of 1231 water molecules. In the second case, the DNA contributed just 11% of the total number of the atoms in the simulation, so the limited computational resources were required to spend nearly 90% of their time considering the dynamics of the 'uninteresting' solvent. However, the improved dynamics of the DNA that resulted made it clear that this was probably a price that would have to be paid.

The period from the mid-1980s to the mid-1990s is therefore characterized by gradual improvements to force fields and longer and larger simulations as software was refined and hardware developed, coupled with a rapidly increasing database of experimental information from X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. However, it was clear that obtaining extended, reliable simulations of DNA seemed to remain as much an art as a science. The real breakthrough came in the mid-1990s with the development of computationally efficient methods to evaluate long-range electrostatic interactions, for example, those based on Ewald summations. Now simulations became far more robust, even when, as around this time, they reached the nanosecond timescale.[8–10]

Nevertheless, it remained the case that as longer and longer simulations became feasible, limitations in force fields would continue to be uncovered. For example, around 2005 it became clear that DNA simulations performed using one of the most popular force fields (AMBER parm99) often showed gradual and irreversible structural degradation when extended beyond 30 nanoseconds or so. This led to a very successful community-wide project to correct the problem,[11] and now stable simulations beyond the microsecond are achievable.[12]

In 25 years, atomistic MD simulations have moved from the regime of tens of picoseconds to the microsecond—an improvement of five orders of magnitude, comfortably outstripping Moore's law. But in common with protein simulation, timescales remain an issue. Many of the most biologically relevant dynamics of DNA are still beyond what can be simulated using 'vanilla' MD (Table 1). But the plethora of methodologies that has evolved to address this issue in protein simulation is equally applicable to the study of DNA dynamics as will be discussed in more detail later.

Simulations of DNA do however face one issue that is less prevalent in protein MD: we are seldom

TABLE 1 | Examples of DNA Dynamics Timescales; Currently, Unaccelerated Atomistic Molecular Dynamics Simulations are Generally Restricted to Timescales of a Microsecond or Less

Timescale	Types of Motion
Picosecond	Bond vibrations; local conformational transitions (e.g., in sugar–phosphate backbone); and residence times for water molecules at hydration sites.
Nanosecond	Formation of denatured regions in stressed DNA and large-scale conformational changes (e.g., writhing; A to B transitions).
Microsecond	Base breathing; ligand association; and association of complementary oligonucleotides.
Millisecond	DNA stretching experiments and repair protein and helicase procession rates.
Second	Topoisomerization; ligand disassociation; and nucleosome remodeling.

in a position to model the whole of the molecule of interest. The smallest human chromosome has about 5×10^7 base pairs (bp), whereas the vast majority of atomistic DNA simulations are on molecules of less than 20 bp. Fortunately, this is less problematic than might be expected. The flexible and dynamic nature of linear DNA means that in many regards the behavior of an individual short segment of genomic DNA is the same as if that segment was indeed an isolated oligonucleotide. In addition, simulations have focussed on improving our understanding the relationship between sequence, structure, flexibility, and recognition, and on uncovering rules that transcend length scales. An example of this would be the ongoing work of the international Ascona B-DNA Consortium, which has collaborated to produce a comprehensive study of the structure and flexibility of every unique tetranucleotide sequence.[13] It is also the case that most of the experimental data, for example, from X-ray and NMR structures, also relates to relatively short oligomers. Nevertheless, in recent years, there has been a move to the simulation of longer DNA sequences and DNA/histone complexes to complement newer structural methodologies such as cryoelectron microscopy, atomic force microscopy (AFM), and other single-molecule biophysical approaches. However, as these typically involve modeling the DNA in less than atomic detail, we will not be discussing them further here.

In the following sections, we provide an overview of the current state-of-the-art with respect to the atomistic simulation of DNA, and DNA-containing systems. For convenience we divide this discussion broadly into two areas: simulations concerned with investigating the mechanical properties of DNA, and then simulations concentrating on DNA recognition by other molecules. However, it will rapidly become apparent that in reality these two aspects of DNA are closely intertwined.

Simulating the Mechanical Properties of DNA

DNA must remain stable when not in use, to protect the genome. However, DNA processing motors, such as RNA and DNA polymerase, must also be able to open the duplex to read and to copy genes. These contrasting biological demands require a molecule with very special mechanical properties. DNA is carefully packaged to condense the enormous molecule into a conveniently small space. This is achieved both by twisting the DNA into a highly compacted supercoiled structure and also by wrapping the molecule neatly around nuclear proteins. Supercoiling occurs whenever the twist of the DNA is forced to differ from its equilibrium value ($\sim 34^\circ$), and can occur in several different guises: it is normally plectonemic in bacteria (like intertwined strands in a rope) but is toroidal in chromatin (like a solenoid).[14] Remarkably, DNA packing also enables the cell to fine-tune the mechanical properties of the duplex and ultimately genetic responses. The presence of an entire family of enzymes called topoisomerases, which control supercoiling by changing the number of times that the two DNA strands are wound around each other, illustrates the central importance of supercoiling. Topoisomerases can both introduce and relieve supercoils, depending on the biological details of the enzyme. One of the most important functions of the topoisomerases is to relieve the buildup of supercoiling (positive ahead and negative behind) due to the progression of the transcription or replication complex along the double helix,[14] which if left unchecked would cause sufficient mechanical stress to stall the process.

Although DNA is commonly depicted as a linear molecule *in vivo*, the DNA is both supercoiled and decorated by numerous binding proteins. The study of the mechanical properties of DNA is an important step towards a greater understanding of the 'storage solutions' that are available for packing the DNA as efficiently as possible and of DNA processing enzymes such as the polymerases and topoisomerases. Atomistic simulations have been used to study the response of duplex DNA to stress applied through a variety of mechanisms including stretching, bending, and over and under-twisting. These calculations complement a

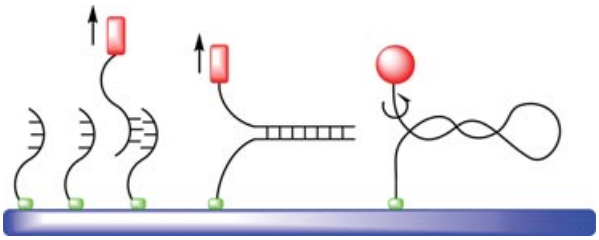

FIGURE 2 | A schematic representation of DNA stretching (left), unzipping (middle), and twisting (right) using atomic force microscopy (left and middle) and magnetic tweezers (right).

wealth of experimental investigations of the mechanical properties of DNA, which include experiments employing nanoscale force probes to manipulate individual DNA molecules,[15,16] the use of enzymatic techniques to measure the free energy of DNA bending into circular structures for progressively shorter DNA sequences,[17–20] and experiments that study the ability of topoisomerases to untwist very small DNA loops.[21]

Much of our current understanding of the mechanical and tensile properties of duplex DNA has come from experiments performed at the single molecule level. These nanomanipulation experiments chemically anchor one end of the DNA to a surface whereas the other is manipulated by a sensor that records the applied force. In optical traps and magnetic tweezer experiments, the force detector is a trapped microbead, whereas AFM uses a microscopic cantilever, as shown schematically in Figure 2.[16,22] The biomechanical properties of DNA are studied by measuring the force required to produce a particular extension up until the fracture point where the duplex disassociates into two single strands. Molecular modeling studies of DNA stretching and twisting have been very influential in the interpretation of the data from nanomanipulation experiments because it is only through computer simulation that it is possible to visualize the structural changes taking place in the biomolecule at the atomic level.

Simulations of DNA Stretching

DNA stretching experiments have been used to study very long DNA molecules (of order 50,000 bp),[23–26] as well as oligomers only 10 bp in length.[27,28] The force–extension curves are rather interesting: DNA initially extends elastically, then there is a plateau region in which it can be further extended plastically (i.e., at constant force), and then a second elastic regime before the molecule eventually ruptures. Pioneering simulations of DNA stretching performed by Konrad and Bolonick[29] and Lebrun and Lavery[30] on short oligomers revealed similar features in the calculated force–extension curves. The simulations showed that the plateau region in the force–extension curve corresponds to a conformational change in which the base stacking interactions in the DNA helix switched from being predominantly intrastrand (as in canonical B-DNA) to interstrand. This novel stretched DNA conformation was termed 'S-DNA' (see Figure 3 top left), and in line with the simulation, the plateau in the force–extension curves is known as the 'B-S plateau'. Further calculations by Lavery and Lebrun[32] performed in 1999, which compared the stretched DNA structures resulting from pulling on either the 5′5′ or the 3′3′ opposing strands showed that the DNA is also sensitive to the pulling direction. The simulations predicted that the structure resulting from pulling on the 5′5′ ends was far narrower than the flat ribbon structure obtained from stretching from the 3′3′ ends. Remarkably, a series of single molecule experiments in 2009 by Danilowicz et al.[33] using magnetic tweezers verified that the mechanical response of the DNA is indeed sensitive to the pulling direction. By performing the experiments at different ionic strengths and comparing the two sets of force–extension curves, Danilowicz et al.[33] concluded that the 3′3′ structure is well screened electrostatically but that the 5′5′ conformation is not, consistent with the computer simulations.

However, the interpretation of the simulations in terms of the experimental data remains controversial. On the basis of a series of experiments that probed the thermodynamics of the stretching process, Rouzina and coworkers[23,34,35] proposed that the B-S plateau signifies a melting transition rather than the conformational change suggested by simulation. In addition, the B-S transition is only observed in experiments that stretch very long DNA molecules, and not in experiments that pull the oligomeric sequences used in the calculations.[27,28] Support for this 'force-induced melting model' has been obtained by recent nanomanipulation experiments using biochemical probes that preferentially bind single-stranded DNA.[36,37]

Our limited computational resources make it impossible to literally reproduce these nanomanipulation experiments *in silico*. The laboratory measurements stretch the DNA in around a microsecond, whereas computer simulation studies have so far been limited to nanosecond timescales. Because the force measured experimentally is strongly dependent on the pulling rate,[26,28] it is difficult to make a quantitative comparison between nanomanipulation experiments and theory. In 2005, Harris et al.[38] and Piana[39] used complementary simulation techniques in

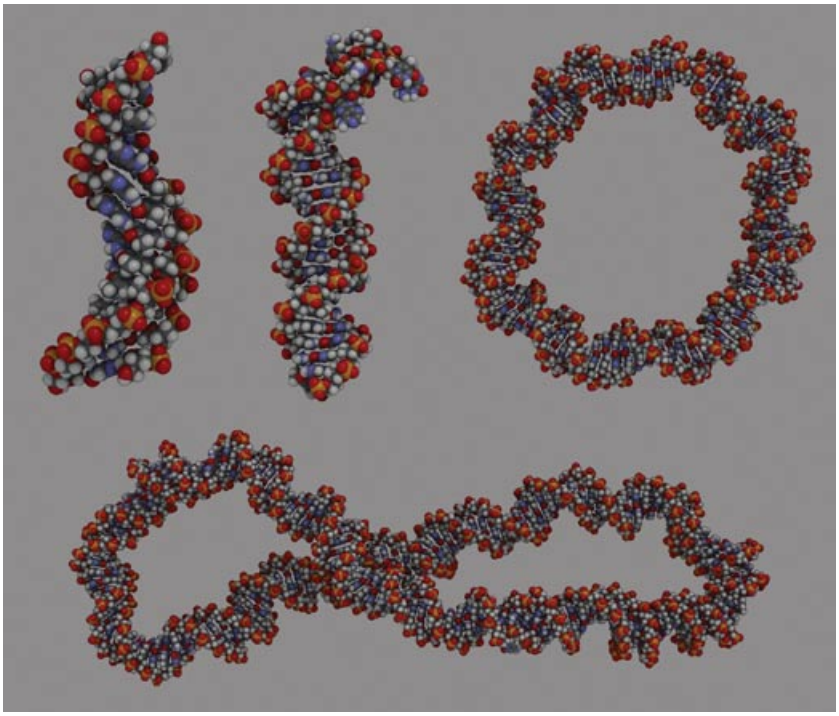

FIGURE 3 | DNA under mechanical stress. The S-DNA structure from molecular dynamics (top left) showing the unusual inclination of the DNA bases; an over-wound P-DNA structure[44] (top middle) showing a flipped out base at the top of the structure; a relaxed 90 bp DNA circle (top right); and an under-wound and writhed 178 bp circle (bottom), where the DNA has relieved torsional stress by adopting a 'figure of 8' type structure. This image was produced using QuteMol.[31]

an attempt to calculate the free energy associated with DNA stretching. Both of these studies showed that entropy and dynamics can be extremely important in determining the mechanical properties of biomolecules such as DNA. This is easily understood from a physical point of view. As the DNA is stretched, the hydrogen bond and base stacking interactions are weakened by the external force. Eventually, an extension is reached when thermal fluctuations are sufficiently large that the molecule simply shakes itself apart under the applied tension.[40,41] These calculations suggested that the S-DNA structure was in fact thermodynamically unstable and formed single-stranded bubble-like regions, given sufficient time. Most interestingly, the calculations performed by Piana[39] showed a Na^+ counterion localized within the denaturated bubble, suggesting that open conformations can be stabilized by interactions with the solvent. Relating the simulations directly to the experiments, probes our understanding of single molecule thermodynamics, kinetics, biomolecular mechanics, and the role of the environment.

Simulating Over- and Under-twisted DNA

It is generally believed that untwisting DNA (e.g., negative supercoiling) destabilizes the double helix and promotes processes that require strand separation (such as transcription). Denaturation due to negative supercoiling has indeed been observed by nanomanipulation experiments in which long DNA molecules were twisted with magnetic tweezers.[42,43] A series of MD calculations performed to twist the Dickerson and Drew dodecamer by Wereszczynski and Andricioaei[44] using the CHARMM force field also showed that under-twisting caused DNA denaturation. In contrast, they found that over-twisting produced a novel DNA structure in which the bases are located on the outside of the duplex structure, which they called P-DNA. The observation of P-DNA excited the community because it is the DNA structure originally proposed by Pauling, but later rejected in favor of the familiar Watson and Crick model. Wereszczynski and Andricioaei[44] then combined their calculations with existing experimental data to propose a phase-diagram for stretched and twisted DNA molecules, which showed the torsional/stretching forces at which the DNA would be expected to be found in each of the possible B-, S-, and P-DNA forms.

In 2009, Randall et al.[45] performed a detailed series of 19 simulations in which the DNA was held restrained for 10 nanoseconds in a series of increasingly under- and over-twisted states. The calculations showed that under-wound duplexes are prone to denaturation and base flipping at relatively low levels of supercoiling, whereas over-wound duplexes are significantly more robust and remain intact up until the threshold for the formation of P-DNA (see Figure 3, top and middle). The simulations showed 'waves of

conformational deformation' diffusing through the DNA during equilibration until certain sequences preferentially absorbed the twist deficit allowing others to relax back into a B-form structure. They also showed that over-winding increases the concentration of positively charged counterions in the grooves, which provides improved electrostatic screening in response to decreasing the distance between the negatively charged backbone strands. Fewer ions were observed close to regions containing denatured bases, in which the backbone is relatively open. Therefore, the simulations shows that the behavior of the environment in response to stress placed on the DNA can also be important in determining the mechanical response of the duplex.

DNA Circles: Bending and Twisting DNA

DNA is often bent and twisted. It is generally stored in tightly coiled structures in the cell, and a number of regulatory protein–DNA complexes (such as the Lac repressor[46]) involve the formation of highly bent DNA loops. DNA plasmids used for genetic engineering are circular and bacterial genomes are closed circular and negatively supercoiled. Recent advances in genetic manipulation techniques have enabled experimentalists to study DNA bending and twisting by constructing very small (as small as 64 bp) closed circular DNA structures.[19] As this is far shorter than the persistence length of duplex DNA (~150 bp[47]), bending into such a tight loop must be associated with a significant deformation energy. Twisting can also be studied as these circles can be supercoiled as well as bent.

The first long-timescale atomistic MD simulation of a DNA circle in explicit solvent was performed by Lankas et al.[48] They simulated a 90 bp nucleosome positioning sequence that was reported to be particularly flexible compared with theoretical expectations[17] (although these experiments were later disputed by other experimental groups[18,20]). The atomistic simulations showed the emergence of defects or 'kinks' over a timescale of ~50 nanoseconds at positions in which the stacking interactions between successive bases around the circle were weakest, namely at pyrimidine–purine steps, but especially at TpA steps, which are known to be particularly deformable.[49] However, more recent calculations performed by Curuksu et al.[50] to calculate the free energy of DNA bending have reported that kinks are less prevalent in simulations run using an updated version of the AMBER force field (PARMBSC0). Consequently, the precise atomic details of the structural transitions that take place due to bending and torsional stress in small DNA circles, as for stretched DNA conformations, are still under debate. Harris et al.[38] also ran a series of simulations of small DNA circles, but in these calculations the DNA was subjected to far larger torsional forces through higher levels of supercoiling. These simulations observed circular structures for low supercoiling, writhed DNA structures in which the DNA buckled into a 'figure of 8' structure to relieve torsional stress at higher supercoiling, and finally DNA denaturation at the highest levels of twisting/untwisting explored. Figure 3 shows examples of the circular (top right) and writhed structures (bottom). In a manner equivalent to Wereszczynski and Andricioaei,[44] Harris et al.[38] used these simulations to map out the phase diagram describing the topology of small closed DNA circles as a function of circle size, the supercoiling, and the salt concentration (which has a strong influence on the amount of electrostatic repulsion in a writhed structure in which the helices come into close contact at the crossing point).

The diversity of possible DNA structures and the rich behavior revealed by atomistic MD of DNA under force shows that a rather detailed model is necessary to perform a realistic simulation. These features would not be captured by simple polymer models that treat the duplex as a homogeneous elastic rod. However, these simulations of DNA sequences larger than the traditional dodecamer can emphasize small problems with the force fields that to date have not been so important. 'AMBER DNA' has a slightly lower twist than the B-form duplex (33° compared with 34°), and this must be taken into account when comparing the levels of supercoiling in the simulations to those imposed experimentally. In addition, simulations that distort DNA push the force fields beyond the regime in which they were originally parameterized, and so strongly test our understanding of DNA mechanics and the predictive capabilities of our simulation techniques.

DNA Recognition

DNA–Ligand Recognition

DNA is the target for a variety of naturally occurring and synthetic ligands. Broadly speaking, they fall into two categories: those that intercalate between base pairs, and those that are found in the grooves, particularly the minor groove, of the double helix. Binding may involve the formation of covalent links between reactive groups in the ligand and the DNA, or recognition may be entirely through noncovalent interactions (Figure 4). Examples include a number of

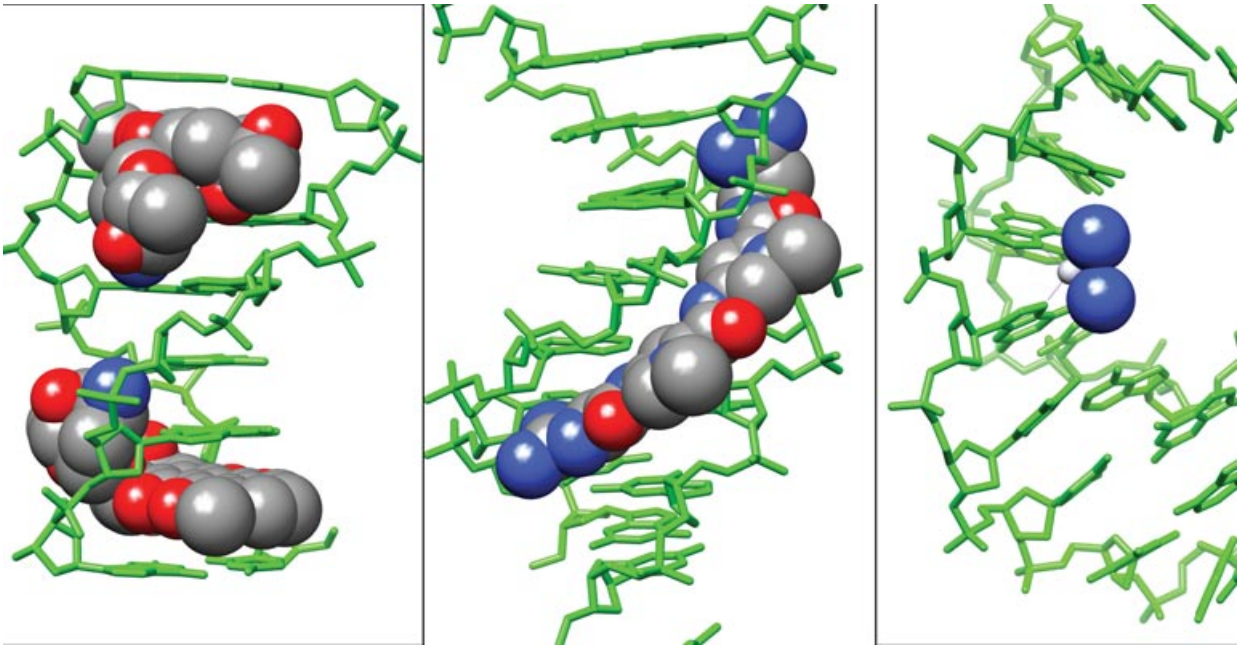

FIGURE 4 | Types of DNA–ligand interactions. (Left) Intercalation, exemplified here by daunomycin (PDB code 1DA0; Ref 51). (Center) Minor groove binding, exemplified here by netropsin (PDB code 261D; Ref 52). (Right) Covalent attachment, exemplified here by cisplatin (PDB code 1A84; Ref 53).

important cytotoxic drugs used in cancer chemotherapy, such as cisplatin, temozolomide, and doxorubicin. Some of these have a limited degree of sequence selectivity—for example, cisplatin and temozolomide preferentially alkylate polyG sequences, but this gives very little selectivity within a genomic context and is the key reason why these drugs are associated with significant host toxicity. It has been realized for a long time (pioneering work from the 1970s is reviewed by Nielsen[54]) that if the rules governing DNA–ligand recognition were better understood, it might be possible to design small molecule drugs with much better sequence-reading properties that could be selectively targeted to desired sites—for example, oncogenes—sparing the rest of the genome from damage. Such drugs could bring about a highly controlled modulation of gene expression, and so open up the possibility of therapies for many other diseases in addition to cancer. The main focus has been on small, modular molecules that bind in the minor groove of DNA. Taking inspiration from the structure and DNA-binding properties of netropsin, an antibacterial agent produced by *Streptomyces netropsis*, these are often referred to as 'lexitropsins'.[55] Molecular simulations have played a major part in increasing our understanding of the forces at work in DNA–lexitropsin interaction, but it is probably fair to say that they have been more successful in explaining why in practice, sequence-selective recognition is so hard to achieve than in providing, for example, reliable calculations of binding affinities. So, although general aspects of recognition such as the matching of electrostatic potential and shape complementarity[56] are well understood and can explain, for example, why netropsin binds better to AT-rich sequences than GC-rich ones, why this ligand prefers the sequence ATTTA over ATTAT[57] is much harder to explain. A major reason for this is that the binding of ligands to the minor groove of DNA appears to be mainly entropy-driven[58] and our ability to accurately predict entropic terms from simulation data remains limited.[59] The calculations require the accurate modeling of subtle sequence-dependent features such as DNA flexibility and hydration, and appropriate statistical mechanical treatment of the data—these remain serious challenges.

DNA–Protein Recognition

Life depends on the fact that proteins such as transcription factors have evolved to solve the problem of highly sequence-selective DNA recognition. On the contrary, other classes of proteins have had different problems to solve. Polymerases need to function with high affinity and fidelity irrespective of the sequence-dependent structural and chemical properties of their DNA substrates, and DNA repair proteins have to be able to recognize sometimes very subtle signals

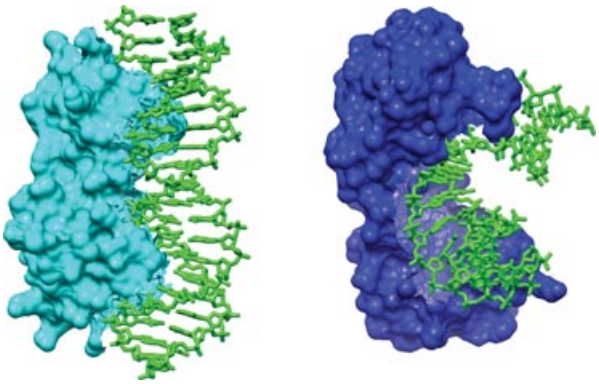

FIGURE 5 | Examples of 'direct' and 'indirect' sequence recognition in DNA–protein complexes. (Left) The complex of C2 repressor protein with DNA (PDB code 3JXB[60]) shows little deformation of the double helix. (Right) In contrast, the TATA-binding protein (PDB code 1CDW[61]) causes gross distortion of the DNA structure.

relating to alterations in DNA structure above the natural sequence-dependent noise. Crystal structures of DNA–protein complexes quickly revealed that nature has not produced a single, universal solution to these problems. In certain cases, recognition appears to be mainly direct—key amino acid side chains interrogate sequence-specific chemical features in the DNA major and/or minor grooves. But in other cases it seems that recognition has been provided mainly indirectly—by the ability of the protein to significantly distort the DNA away from its canonical structure in ways that are energetically allowable for the cognate sequence(s), but not for others (Figure 5).

Molecular modeling has proved to be a key approach to understand this in more detail. First, it provides a method to estimate the energetic contributions from different parts of the process to the overall recognition event (e.g., how much is from direct interactions and how much from induced fit) in a way that experimental methods (e.g., isothermal calorimetry) cannot. Second, to fully explain selectivity, we must address a combinatorial problem, for example, to explain why *Eco*RI preferentially binds to the sequence GAATTC, we should examine and analyze its binding to the other 4095 possible sequences as well—a tall order for experimental structure determination methods, but now within the scope of modeling approaches. For example, Lavery and coworkers,[50] using the power of grid computing, have studied the selectivity of recognition in the mixed human/*Xenopus* nucleosome core particle.[62] Theoretically, there are 3.2×10^{88} different 147 bp sequences whose binding to the nucleosome should be evaluated, but using a 'divide and conquer' approach and a linearization approximation, this was reduced to 284,672 independent simulations that were completed in 11 days using the EGEE (Enabling Grids for E-sciencE) grid computing resource. Resulting predictions for nucleosome positioning adjacent to transcription start sites were in good agreement with the experimental data.

Many of the most intriguing protein–DNA interactions are those in which dynamics is the key issue. For example, DNA repair proteins typically 'flip' the damaged base out of the DNA helical stack into a pocket in the enzyme. If the protein needs to flip a base into the pocket before it can tell if it needs repairing or not, and maybe only one out of every 10,000 bases is a lesion site, the process would appear to be very inefficient in terms of both time and energy.[63] It seems likely that the mechanism is more subtle than this—but in what way? Molecular simulations are an obvious way to investigate such issues, but require enhanced sampling or accelerated dynamics methods as the natural rate of the process lies beyond the timescales accessible by conventional atomistic simulation. For example, steered and targeted MD simulations of the recognition of 8-oxoguanine by MutM[64] have produced detailed free energy maps for the recognition and base extrusion process for both normal (G) and lesion sites and revealed that the protein is able to distinguish between the two at an early stage, as the protein distorts the local DNA structure into a still intrahelical, but 'extrudogenic', conformation.

Many important DNA–protein complexes are molecular motors, for example, topoisomerases and helicases. Molecular simulations are playing a key role in helping us understand how these machines actually function, identifying plausible pathways between the static structures that have been captured by, for example, X-ray crystallography. Work from the Schulten group[65] on the bacterial helicase PcrA is a prime example of this, combining a wide range of modeling methods from quantum mechanical calculations to elastic network analysis to provide insights into how ATP hydrolysis ultimately results in unidirectional travel of the protein along the DNA strand.

CONCLUSION

Computer models have proven invaluable in the interpretation of experiments involving mechanically distorted DNA structures, and have shown the importance of molecular flexibility in DNA recognition. Such simulations are often the only method available to study changes in DNA conformation at the atomistic level. They have inspired many of the experiments that have significantly improved our understanding of the mechanical properties of DNA,

such as the use of biochemical probes to look for denatured regions in stretched DNA. Simulations have been used to design potential new anticancer drugs and to show how proteins exploit DNA flexibility to perform their biological function. Though they are far from perfect, atomistic simulations are at present probably the best conceptual framework through which to test our understanding of biomolecular structure and dynamics, and we can often learn as much from cases where the simulations fail to accurately reproduce experimental data as from cases where they succeed.

For biomolecular simulation in general, the challenge for the future will be to model bigger systems for longer timescales. This requires continuing improvements in computer hardware and MD simulation software, and will probably reveal new issues with the MD force fields that will require adjustment. The possibility of accessing micro- or millisecond timescales for a large protein/DNA complex, for example, is particularly exciting because we will be in the realm of biological timescales, and so will be able to study these molecules without recourse to artificial acceleration methods. Calculations of free energy changes will be more accurate, improving our understanding of molecular recognition. As significant progress has recently been made in developing hardware specifically for the purposes of MD, promising a speed of a factor of approximately 1000, we hope that microsecond simulations will be routine within 5 years.[66]

For simulations of DNA in particular, we envisage that in the short term, the community will become increasingly interested in higher order DNA structures, such as quadruplex DNA, DNA hairpins, and DNA cruciforms. As simulation timescales improve into the microsecond and millisecond regime, it will become possible to model conformational switching between these different DNA states, such as single-stranded DNA folding into a quadruplex. Studies of protein–DNA interactions will also become more sophisticated. Although pioneering studies of single nucleosome–DNA complexes are now published,[67–69] within the next 10 years, we hope to see atomistic simulations of large assemblies of nucleosomes and exploration of chromatin condensation and chromatin opening at the fully atomistic level. Such calculations will also provide the opportunity to explore the influence of epigenetic factors (such as DNA methylation and histone acetylation) on the structure and dynamics of chromatin. Over this timescale, our computational facilities will be sufficient for us to 'watch' molecular motors such as helicases processing DNA. Given the dynamic nature of DNA and its interactions, it may be that the most remarkable abilities of this molecule are yet to be observed.

ACKNOWLEDGMENTS

We thank B.M. Pettitt for the data for the P-DNA structure.

REFERENCES

1. Watson JD, Crick FHC. Molecular structure of nucleic acids—a structure for deoxyribose nucleic acid. *Nature* 1953, 171:737–738.
2. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE. Structure of a B-DNA dodecamer—conformation and dynamics .1. *Proc Natl Acad Sci U S A* 1981, 78:2179–2183.
3. Levitt M. Computer simulation of DNA double-helix dynamics. *Cold Spring Harb Symp Quant Biol* 1983, 47 Pt 1:251–262.
4. Tidor B, Irikura KK, Brooks BR, Karplus M. Dynamics of DNA oligomers. *J Biomol Struct Dyn* 1983, 1:231–252.
5. Prabhakaran M, Harvey SC, Mao B, McCammon JA. Molecular-dynamics of phenylalanine transfer-RNA. *J Biomol Struct Dyn* 1983, 1:357–369.
6. Seibel GL, Singh UC, Kollman PA. A molecular-dynamics simulation of double-helical B-DNA including counterions and water. *Proc Natl Acad Sci U S A* 1985, 82:6537–6540.
7. van Gunsteren WF, Berendsen HJ, Geurtsen RG, Zwinderman HR. A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: comparison with experimental two-dimensional NMR data. *Ann N Y Acad Sci* 1986, 482:287–303.
8. Beveridge DL, Ravishanker G. Molecular-dynamics studies of DNA. *Curr Opin Struct Biol* 1994, 4:246–255.
9. Cheatham TE. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr Opin Struct Biol* 2004, 14:360–367.

10. Cheatham TE, Kollman PA. Molecular dynamics simulation of nucleic acids. *Annu Rev Phys Chem* 2000, 51:435–471.
11. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE III, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophy J* 2007, 92:3817–3829.
12. Perez A, Luque FJ, Orozco M. Dynamics of B-DNA on the microsecond time scale. *J Am Chem Soc* 2007, 129:14739–14745.
13. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T, Dixit S, Jayaram B, Lankas F, Laughton C, 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* 2010, 38:299–313.
14. Bates A, Maxwell A. *DNA Topology*. Oxford: Oxford University Press; 2005.
15. Lavery R, Leburn A, Allemand JF, Bensimon D, Croquette V. Structure and mechanics of single biomolecules: experiment and simulation. *J Phy : Condens Matter* 2002, 14:R383–R414.
16. Strick TR, Dessinnges MN, Charvin G, Dekker NH, Allemand JF, Bensimon D, Croquette V. Stretching of macromolecules and proteins. *Rep Prog Phys* 2003, 66:1–45.
17. Cloutier TEaW J. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes *Proc Natl Acad Sci U S A* 2005, 102:3645–3650.
18. Du Q, Smith C, Shiffeldrim N, Vologodskaia M, Vologodskii A. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc Natl Acad Sci U S A* 2005, 102:5397–5402.
19. Du Q, Kotlyar A, Vologodskii A. Kinking the double helix by bending deformation. *Nucleic Acids Res* 2008, 36:1120–1128.
20. Demurtas D, Amzallag A, Rawdon EJ, Maddocks JH, Dubochet J, Stasiak A. Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles. *Nucleic Acids Res* 2009, 37:2882–2893.
21. Bates AD, Maxwell A. DNA gyrase can supercoil DNA circles as small as 174 base-pairs. *Embo J* 1989, 8:1861–1866.
22. Lavery R, Lebrun A, Allemand JF, Bensimon D, Croquette V. Structure and mechanics of single biomolecules: experiment and simulation. *J Phys: Condens Matter* 2002, 14:R383–R414.
23. Williams MC, Rouzina I, Bloomfield VA. Thermodynamics of DNA interactions from single molecule stretching experiments. *Acc Chem Res* 2002, 35:159–166.
24. Rief M, Clausen-Schaumann H, Gaub HE. Sequence-dependent mechanics of single DNA molecules. *Nat Struct Biol* 1999, 6:346–349.
25. Clausen-Schaumann H, Rief M, Gaub HE. Sequence dependent mechanics of single DNA molecules. *Biophys J* 1999, 76:A151–A151.
26. Clausen-Schaumann H, Rief M, Tolksdorf C, Gaub HE. Mechanical stability of single DNA molecules. *Biophys J* 2000, 78:1997–2007.
27. Pope LH, Davies MC, Laughton CA, Roberts CJ, Tendler SJB, Williams PM. Force-induced melting of a short DNA double helix. *Eur Biophys J* 2001, 30:53–62.
28. Strunz T, Oroszlan K, Schumakovitch I, Guntherodt HJ, Hegner M. Dynamic force spectroscopy of single DNA molecules. *Proc Natl Acad Sci U S A* 1999, 96:11277–11282.
29. Konrad MW, Bolonick JI. Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *J Am Chem Soc* 1996, 118:10989–10994.
30. Lebrun A, Lavery R. Modelling extreme stretching of DNA. *Nucleic Acids Res* 1996, 24:2260–2267.
31. Tarini M, Cignoni P, Montani C. Ambient occlusion and edge cueing to enhance real time molecular visualization. *IEEE Trans Vis Comput Graph* 2006, 12:1237–1244.
32. Lavery R, Lebrun A. Modelling DNA stretching for physics and biology. *Genetica* 1999, 106:75–84.
33. Danilowicz C, Limouse C, Hatch K, Conover A, Coljee VW, Kleckner N, Prentiss M. The structure of DNA overstretched from the 5′ 5′ ends differs from the structure of DNA overstretched from the 3′ 3′ ends. *Proc Natl Acad Sci USA* 2009, 106:13196–13201.
34. Rouzina I, Bloomfield VA. Force-induced melting of the DNA double helix 1. Thermodynamic analysis. *Biophys J* 2001, 80:882–893.
35. Williams MC, Wenner JR, Rouzina I, Bloomfield VA. Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching. *Biophys J* 2001, 80:1932–1939.
36. Shokri L, McCauley MJ, Rouzina L, Williams MC. DNA overstretching in the presence of glyoxal: structural evidence of force-induced DNA melting. *Biophys J* 2008, 95:1248–1255.
37. van Mameren J, Gross P, Farge G, Hooijman P, Modesti M, Falkenberg M, Wuite GJL, Peterman EJG. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proc Natl Acad Sci U S A* 2009, 106:18231–18236.
38. Harris SA, Sands ZA, Laughton CA. Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA. *Biophys J* 2005, 88:1684–1691.
39. Piana S. Structure and energy of a DNA dodecamer under tensile load. *Nucleic Acids Res* 2005, 33:7029–7038.

40. Evans E. Probing the relation between force-lifetime- and chemistry in single molecular bonds. *Annu Rev Biophys Biomol Struct* 2001, 30:105–128.
41. Tinoco IJ, Bustamante C. The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys Chem* 2002, 101–102:513–533.
42. Allemand JF, Bensimon D, Lavery R, Croquette V. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc Natl Acad Sci U S A* 1998, 95:14152–14157.
43. Bryant Z, Stone MD, Gore J, Smith SB, Cozzarelli NR, Bustamante C. Structural transitions and elasticity from torque measurements on DNA. *Nature* 2003, 424:338–341.
44. Wereszczynski J, Andricioaei I. On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *Proc Natl Acad Sci U S A* 2006, 103:16200–16205.
45. Randall GL, Zechiedrich L, Pettitt BM. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res* 2009, 37:5568–5577.
46. Swigon D, Coleman BD, Olson WK. Modeling the Lac repressor–operator assembly: the influence of DNA looping on Lac repressor conformation. *Proc Natl Acad Sci U S A* 2006, 103:9879–9884.
47. Marko JF, Cocco S. The micromechanics of DNA. *Phys World* 2003, 16:37–41.
48. Lankas F, Lavery R, Maddocks JH. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* 2006, 14:1527–1534.
49. Mack DR, Chiu TK, Dickerson RE. Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. *J Mol Biol* 2001, 312:1037–1049.
50. Curuksu J, Zacharias M, Lavery R, Zakrzewska K. Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res*. 2009, 37:3766–3773.
51. Moore MH, Hunter WN, Destaintot BL, Kennard O. DNA–drug interactions. The crystal structure of d(CGATCG) complexed with daunomycin. *J Mol Biol* 1989, 206:693–705.
52. Nunn CM, Garman E, Neidle S. Crystal structure of the DNA decamer d(CGCAATTGCG) complexed with the minor groove binding drug netropsin. *Biochemistry* 1997, 36:4792–4799.
53. Gelasco A, Lippard SJ. NMR solution structure of a DNA dodecamer duplex containing a *cis*-diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. *Biochemistry* 1998, 37:9230–9239.
54. Nielsen PE. Sequence-selective DNA recognition by synthetic ligands. *Bioconjug Chem* 1991, 2:1–12.
55. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE. The molecular origin of DNA drug specificity in netropsin and distamycin. *Proc Natl Acad Sci U S A* 1985, 82:1376–1380.
56. Goodsell D, Dickerson RE. Isohelical analysis of DNA-groove binding drugs. *J Med Chem* 1986, 29:727–733.
57. Boger DL, Fink BE, Brunette SR, Tse WC, Hedrick MP. A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity. *J Am Chem Soc* 2001, 123:5878–5891.
58. Chaires JB. A thermodynamic signature for drug–DNA binding mode. *Arch Biochem Biophys* 2006, 453:26–31.
59. Wang H, Laughton CA. Evaluation of molecular modelling methods to predict the sequence-selectivity of DNA minor groove binding ligands. *Phys Chem Chem Phys* 2009, 11:10722–10728.
60. Watkins D, Mohan S, Koudelka GB, Williams LD. Sequence recognition of DNA by protein-induced conformational transitions. *J Mol Biol* 2010, 396:1145–1164.
61. Nikolov DB, Chen H, Halay ED, Hoffmann A, Roeder RG, Burley SK. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc Natl Acad Sci U S A* 1996, 93:4862–4867.
62. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* 2002, 319:1097–1113.
63. Verdine GL, Bruner SD. How do DNA repair proteins locate damaged bases in the genome? *Chem Biol* 1997, 4:329–334.
64. Qi Y, Spong MC, Nam K, Banerjee A, Jiralerspong S, Karplus M, Verdine GL. Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* 2009, 462:762–766.
65. Yu J, Ha T, Schulten K. How directional translocation is regulated in a DNA helicase motor. *Biophys J* 2007, 93:3783–3797.
66. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM* 2008, 51:91–97.
67. Bishop TC. Molecular dynamics simulations of a nucleosome and free DNA. *J Biomol Struct Dyn* 2005, 22:673–685.
68. Ruscio JZ, Onufriev A. A computational study of nucleosomal DNA flexibility. *Biophys J* 2006, 91:4121–4132.
69. Roccatano D, Barthel A, Zacharias M. Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers* 2007, 85:407–421.

