ACCOUNTS
of chemical research

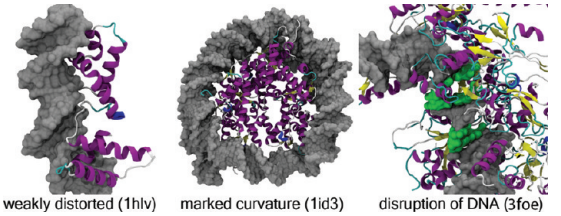
Frontiers in Molecular Dynamics Simulations of DNA

ALBERTO PÉREZ,[†] F. JAVIER LUQUE,[‡] AND MODESTO OROZCO[*,†,§]

[†]*Joint IRB-BSC Program in Computational Biology, Institute of Research in Biomedicine Barcelona, Baldiri i Reixac 10, Barcelona 08028, Spain,* [‡]*Department de Fisicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, Avgda Diagonal 643, Barcelona 08028, Spain, and* [§]*Departament de Bioquímica, Universitat de Barcelona, Avgda Diagonal 647, Barcelona 08028, Spain, and Instituto Nacional de Bioinformàtica, Parc Científic de Barcelona, Baldiri i Reixac 10, Barcelona 08028, Spain*

RECEIVED ON APRIL 24, 2011

CONSPECTUS



It has been known for decades that DNA is extremely flexible and polymorphic, but our knowledge of its accessible conformational space remains limited. Structural data, primarily from X-ray diffraction studies, is sparse in comparison to the manifold configurations possible, and direct experimental examinations of DNA's flexibility still suffer from many limitations.

In the face of these shortcomings, molecular dynamics (MD) is now an essential tool in the study of DNA. It affords detailed structural and dynamical insights, which explains its recent transition from a small number of highly specialized laboratories to a large variety of groups dealing with challenging biological problems. MD is now making an irreversible journey to the mainstream of research in biology, with the attendant opportunities and challenges. But given the speed with which MD studies of DNA have spread, the roots remain somewhat shallow: in many cases, there is a lack of deep knowledge about the foundations, strengths, and limits of the technique. In this Account, we discuss how MD has become the most important source of structural and flexibility data on DNA, focusing on advances since 2007 of atomistic MD in the description of DNA under near-physiological conditions and highlighting the possibilities and shortcomings of the technique.

The evolution in the field over the past four years is a prelude to the ongoing revolution. The technique has gained in robustness and predictive power, which when coupled with the spectacular improvements in software and hardware has enabled the tackling of systems of increasing complexity. Simulation times of microseconds have now been achieved, with even longer times when specialized hardware is used. As a result, we have seen the first real-time simulation of large conformational transitions, including folding and unfolding of short DNA duplexes.

Noteworthy advances have also been made in the study of DNA–ligand interactions, and we predict that a global thermodynamic and kinetic picture of the binding landscape of DNA will become available in a few years. MD will become a crucial tool in areas such as biomolecular engineering and synthetic biology. MD has also been shown to be an excellent source of parameters for mesoscopic models of DNA flexibility. Such models can be refined through atomistic MD simulations on small duplexes and then applied to the study of entire chromosomes. Recent evidence suggests that MD-derived elastic models can successfully predict the position of regulatory regions in DNA and can help advance our understanding of nucleosome positioning and chromatin plasticity. If these results are confirmed, MD simulations can become the ultimate tool to decipher a physical code that can contribute to gene regulation.

We are entering the golden age of MD simulations of DNA. Undoubtedly, the expectations are high, but the challenges are also enormous. These include the need for more accurate potential energy functionals and for longer and more complex simulations in more realistic systems. The joint research effort of several groups will be crucial for adapting the technique to the requirements of the coming decade.

Published on the Web 08/10/2011 www.pubs.acs.org/accounts
10.1021/ar2001217

Introduction

Five decades after the work by Watson and Crick (WC),[1] we know that DNA is extremely flexible and polymorphic,[2,3] but our knowledge of its accessible conformational space is still very limited. Currently the Protein Data Bank (PDB) contains around 1360 DNA structures and 1950 DNA–protein complexes, mainly solved by X-ray diffraction. They likely represent a small fraction of the DNA conformational space, since there is very limited information on unusual structures (e.g., mutated or stressed DNAs, hybrids, chimeras, or DNA in nonaqueous solvents or complexed with ligands). Furthermore, the structural coverage becomes even more compressed if it is limited to naked B-DNA (Figure 1). For example, the d(CG) dinucleotide step is the only base pair step represented more than 100 times in the PDB structures, and several dinucleotide steps have less than 30 instances, reducing the quality of the averages (Figure 1).[4] The scarcity of information is worse when one considers tetranucleotides, since none of the 136 unique tetrads is present in more than 50 experimental structures, only 26 tetrads are found in more than 5 structures, and no experimental information is available for 62 cases (Figure 1). Overall, the lack of experimental sequence-dependent structural information on DNA is dramatic, and there is no indication that it will be reduced in the near future.[5]

DNA is a very flexible polymer, and this property is crucial in the understanding of the wide range of conformations DNA can adopt in physiological conditions. Unfortunately, experimental determination of flexibility suffers from many limitations,[2,6] and most "direct" information comes from low-resolution techniques (e.g., circularization experiments, atomic force microscopy, optical tweezers, and permeation in nanopores),[7–10] which cannot provide an atomistic description of DNA flexibility. Thus, fluctuations in "indirect" descriptors derived from helical properties of the different steps in X-ray structures[4,11] are viewed as the gold standard for flexibility, but we cannot ignore that they present many practical problems, such as the lack of available experimental data and the required assumption of normality in their distributions (Figure 1).

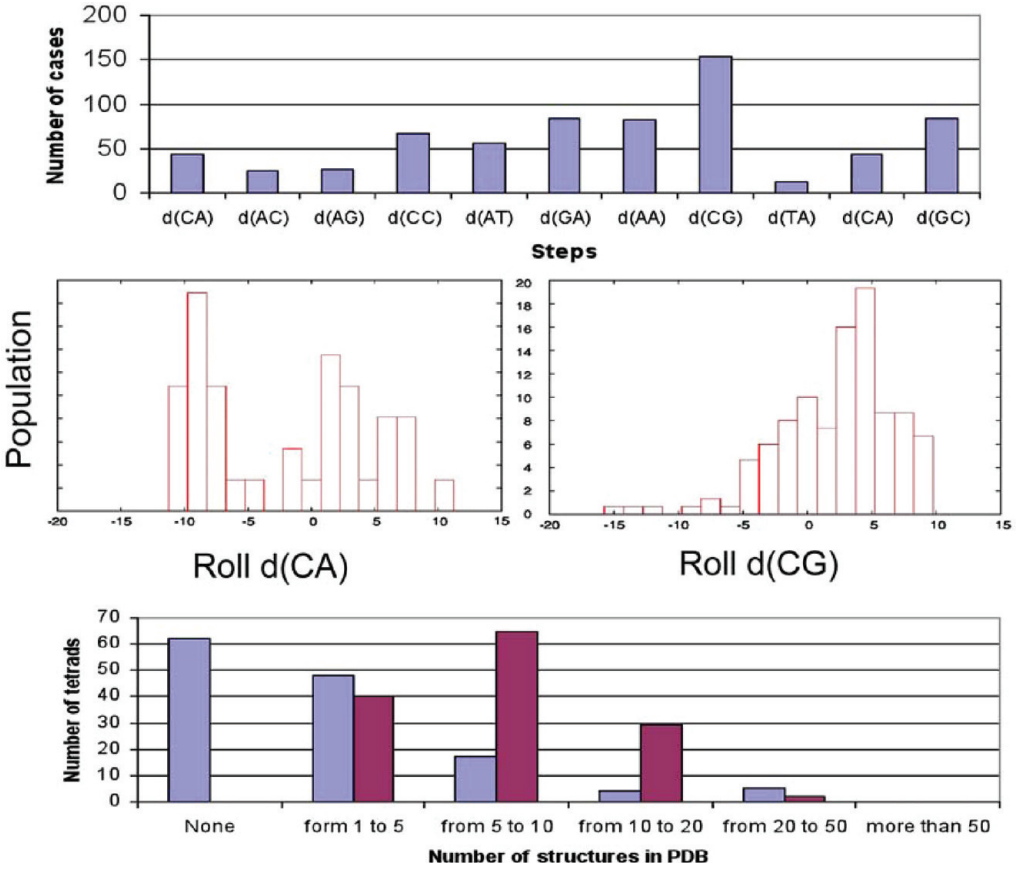


FIGURE 1. (top) Experimental coverage of the 10 unique dinucleotides of naked B-DNA, (middle) roll (in degrees) distribution for d(CA) and d(CG) steps found in X-ray structures, and (bottom) experimental coverage of the unique 136 tedrads when only naked B-DNA structures are considered (blue[4]) or when an extended database containing DNA–protein complexes is used (red).

The lack of reliable experimental data for both structure and flexibility of DNA hampers our ability to understand the relationship between physical properties of DNA and gene regulation mechanisms and explains the popularization of simulation techniques (>1000 papers with keyword "DNA simulation" in ISI-WOK year 2009). Particularly, molecular dynamics (MD)[12] has become in the last years the most important source of structural and flexibility data on DNA. Here we will focus on recent (2007−2010) advances of atomistic MD in the description of DNA in near-physiological conditions. The reader is addressed to other reviews for wider information on MD simulations of nucleic acids.[5,13−17]

Force Fields and Simulation Protocols

MD protocols have remained quite stable for the last years, with most simulations being performed using the isothermal−isobaric ensemble, explicit solvent, particle mesh Ewald, and 2 fs (or equivalent) integration step. Simulation protocols may change in the future in two directions: (i) longer integration steps might be adopted coupled to algorithms for removal of fast, irrelevant movements, and (ii) the use of GPU (graphical processing units) might cause the substitution of certain algorithms, such as the particle mesh Ewald in the treatment of long-range effects. Improvements in hardware and software have led to spectacular increases in the length of trajectories. The microsecond time scale barrier was broken 3 years ago for DNA,[18] and the development of specific purpose computers are putting the millisecond time scale within reach,[19] thus entering the biologically relevant time scale. Whether or not force fields will be robust in these time scales is still an unsolved question.

Two families of force fields, originated from the Karplus[20] and Kollman[21] groups, have dominated the simulation of DNA in the last decades. They have evolved under the impetus afforded by the increase in computer power that made longer simulations possible. For example, multipicosecond simulations revealed artifactual B → A transitions in MD simulations carried out with early versions of Charmm. Nanosecond trajectories helped to detect artifactual unfolding of DNA in all force fields due to the neglect of long-range effects, and unbalanced α/γ transitions were found in simulations (>30 ns) performed with Amber's parm99,[22] which caused severe DNA distortions.[23,24] The effort of four research groups was necessary to solve the latest problem in the newest Amber force field (parmbsc0),[22] which yields stable trajectories in the multi-microsecond scale.[18,25] The latest version of Charmm also performs well for DNA, at least in the 0.1 μs time scale, and in fact provides results not dramatically different from those of parmbsc0.[26] However, a critical evaluation of current force fields is still required, since we are aware of a variety of small problems, which might be the tip of the iceberg when moving to millisecond time scales or beyond, such as the tendency of Charmm27 to underestimate groove asymmetry and probably base-pairing stability in DNA (clearly this is the case in RNA)[27] or the underestimation of twist in parmbsc0 trajectories of B-DNA. Furthermore, both Charmm27 and parmbsc0 fail to reproduce special loop structures,[28] and no guarantee exists that they can capture some nonhelical conformations of DNA.

The reliability of force fields for describing DNA flexibility is a critical aspect due to the scarcity of experimental data (see above). In this context, convergence in the results derived from those force fields has been used as probe of the robustness of MD-derived flexibility descriptors. For instance, a similar flexibility pattern has been recently found for parmbsc0 and Charmm27,[26] and in fact, the nearest neighbor stiffness parameters derived from those force fields agree with "experimental" estimates by Olson's group.[11] Hartmann and co-workers[6] have argued that neither parmbsc0 nor Charmm27 can reproduce BI/BII equilibrium in different steps of DNA, thus raising a warning on their ability to describe local backbone dynamics. However, despite their criticisms Hartmann's data show good agreement between NMR and MD (parmbsc0 or Charmm27, even for parm94!) estimates for most of the steps (see Figure 2 in ref 6). It is clear that we need to be careful with MD-derived flexibility descriptors, but in the absence of accurate experimental measures they can be a reasonable tool to describe DNA flexibility.

Current MD simulations use explicit solvent and counterions (typically Na^+ or K^+) to achieve neutrality. Adding extra salt seems a reasonable choice to represent physiological conditions, but these conditions have given rise to the artifactual formation of salt crystals,[29,30] reducing the effective ion concentration around DNA, which might affect some conformational changes, like breathing or groove narrowing.[18,31] Fortunately, these artifacts do not appear for many ion force fields, even at high ionic strength,[32] and thus MD simulations are not dramatically dependent on the detailed ion model used.[32] Simulation of divalent ions,[33] like Mg^{2+}, Mn^{2+}, or Zn^{2+}, which are important in certain DNA structures,[33] is unfortunately very difficult using pairwise potentials.

Force-field formalisms developed in the 1970s have been prevalent in the field for more than 30 years, but future improvements in force fields will probably imply recalibration of nonbonded functionals, including explicit

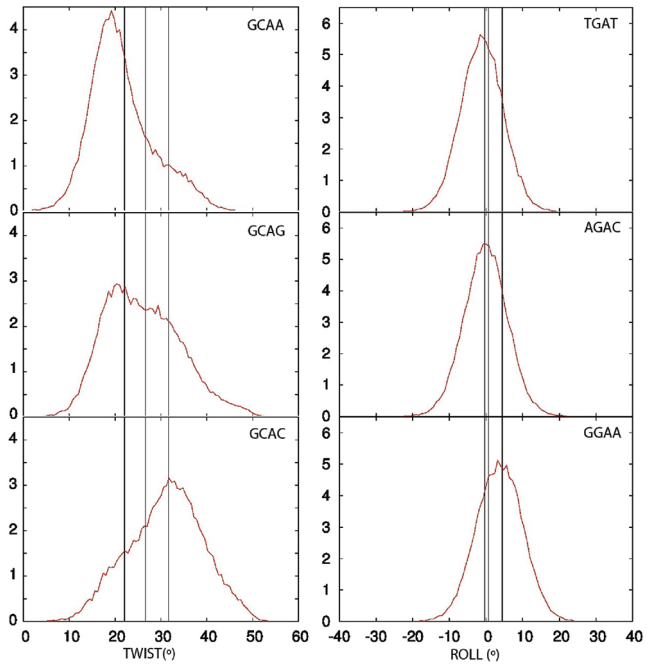


FIGURE 2. Distribution of roll and twist in ABC simulations[45] for the central steps d(CA) and d(GA) in different tetramer environments. Straight lines denote the average for the common central dinucleotide for each of the three stacked histograms. The effect of next nearest neighbors is large on the left side and negligible on the right side of the figure.

polarization.[33] In this area, the Charmm community[34,35] is taking the lead by developing new polarized force fields for DNA (at the present time only nucleobases are parametrized), which considering their computational efficiency might displace pairwise potentials in the future. New frontiers might be also opened from the development of coarse-grained force fields, which at the expense of a lost of resolution allow a very fast calculation of DNA potential energy, which might be useful for the study of very long DNA fragments and longer time scales.[36–40]

Elastic Properties of DNA

Mining MD trajectories to extract information about DNA flexibility is not trivial and can be approached in different ways. One of them is based on techniques such as PCA, which captures the deformation profile of the molecule (eigenvectors) and the relative importance of the eigenvectos (associated eigenvalues). Eigenvalues can easily be converted into force constants from which to derive an energetic measure of deformation cost.[4,13,14] A similar analysis can be performed on the DNA's natural helical space described by three translations (shift, slide, rise) and three rotations (tilt, roll, twist) relating the two base pairs of a base pair step. To preserve the nature of the movements, inversion of the covariance matrix is performed, yielding force constants (with their coupling terms), which are used to describe the energetic cost of deforming a base pair step along its natural 6 degrees of freedom.[11,41,42]

Different groups have derived stiffness parameters for all 10 unique base steps,[11,26,41] providing mesoscopic descriptors of DNA flexibility useful to describe the deformability in protein–DNA complexes.[43] However, these descriptors ignore sequence effects beyond the dinucleotide level, which can be an unreasonable simplification (Figure 2). Kono's group[44] was the first to explore next-nearest neighbor effects by analyzing trajectories of duplexes built by inserting the 136 unique tetramers in Dickerson's dodecamer. Unfortunately, their simulations were too short to guarantee convergence and were based on a force field with known caveats.[22] More recently, the Ascona B-DNA consortium (ABC)[23,45] has simulated 39 duplexes (parmbsc0, 50–100 ns trajectories) chosen to contain several copies of the 136 unique tetramers, which provides in our opinion the best description of sequence-dependent DNA properties. The results show that helical parameters do not follow normal behavior in some cases but present bimodal distributions, highlighting the shortcomings of the harmonic approximation implicit to elastic analysis (Figure 3). The generality and impact of bimodality on DNA properties has not been studied in full detail yet, but experimental data suggests that, at least in some cases, it must not be ignored (Figure 3).

More caveats of current elastic models have been recently discussed. For instance, Maddocks's group[46] has suggested the use of an alternative coordinate system consisting of rigid bases instead of rigid base pairs. Moreover, they have reported a new, elegant method to compute local stiffness parameters based on fitting the global (rather than local) flexibility. At present, it is necessary to evaluate whether this new algorithm and the associated rigid-base model features enough advantages to justify its larger formal complexity. Kinks also represent a relevant challenge to elastic models.[47] Potential of mean force calculations on model systems[47] have been recently used to describe the energetic profile for these distortions, and the existence of kinks in DNA duplex under different stress conditions has been examined by unrestrained MD in mini- or microcircle calculations.[48–50] There is not yet a consensus on the impact of kinks in DNA, but it is clear that nonharmonic distortions are common in DNA–protein complexes.[51]

Large Conformational Transitions

Classical transitions, such as A ↔ B and B ↔ Z, have been studied recently by MD simulations, often including

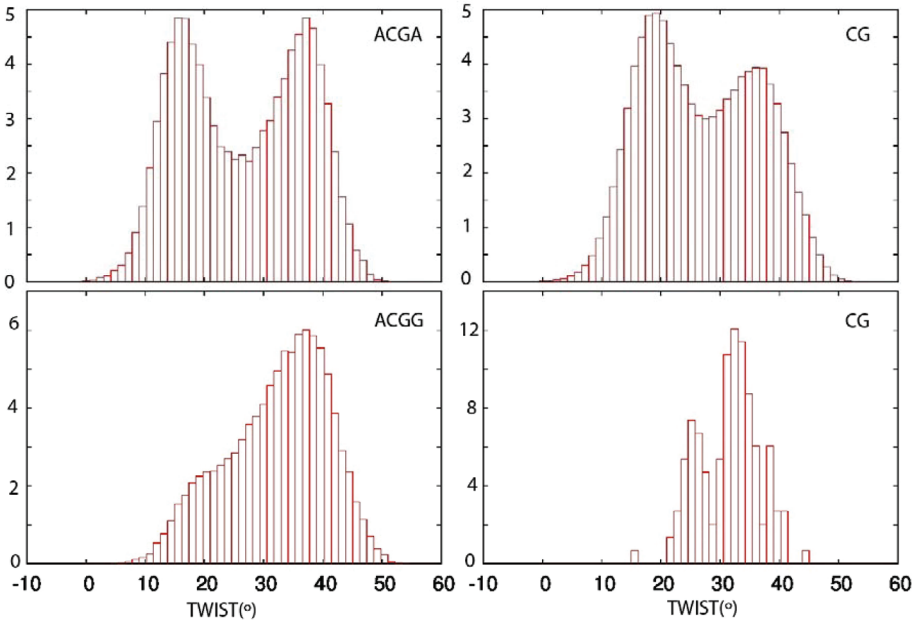


FIGURE 3. Distribution of d(CG) twist states found in (left) ABC-MD simulations of d(ACGA) and d(ACGG) tetrads, (right-top) all ABC-MD simulations of tetramers sharing a central d(CG) step, and (right-bottom) all d(CG) steps in experimental B-DNA structures.[4] The bimodality of twist distribution is clear from both MD simulations and X-ray structures.

experimental restraints.[52,53] Base opening, a crucial local distortion involved in DNA repairing and epigenetic imprinting, has been deeply studied by umbrella sampling MD,[16,51,54] finding connections between bending and base pair openings, which suggest interesting mechanisms to reduce the kinetic barrier of this process. The same methods, supplemented by accurate NMR data, have been used to describe the change between WC and Hoogsteen pairings,[2] a real landmark study illustrating the existence of minor pairings (<1%) coexisting with canonical ones in B-DNA in the millisecond regime, which confirms previous qualitative theoretical predictions.[55]

Ultrafast measuring techniques have shown that folding/unfolding of small DNAs can occur on the microsecond time scale,[56,57] stimulating the interest of the MD community in this issue. Replica exchange MD simulations (REXMD)[58−60] have been used to examine the folding/unfolding landscape and to predict "de novo" the folded state of small DNA oligomers. One of the main conclusions is that the unfolded state does not fit the concept of a disordered random coil but involves a myriad of very different compact non-native structures, in agreement with experimental suggestions by Zewail and co-workers.[56] Pande's group has also explored DNA folding using massive numbers of short MD simulations, suggesting that folding is a very rare but ultrafast process.[61,62] Zewail's group[63] has used short-to-medium parallel MD simulations (from 1 to 360 ns) at seven different temperatures (from 300 to 700 K) to study the folding of a short hairpin. The unfolding events agree with predictions made by an analytical (kinetic intermediate structure) model and with their own T-jump experiments,[56,63] reinforcing the idea that hairpin unfolding is not a two-state process and that the unfolded state is dominated by compact structures separated from native state by significant free energy barriers.

Our group published recently the first unbiased MD simulation of the chemo-thermal DNA duplex unfolding under realistic physical conditions.[64] While simulations under native conditions reported stable duplexes in the microsecond time scale, complete unfolding was detected in 6 of the 10 microsecond-long trajectories and local unfolding was evident in the remaining four. Contrary to common belief, unfolding does not follow a unique pathway for DNA, illustrating the complexity of the process (Figure 4). Thus, while 5 of the 6 unfolding trajectories follow a "fraying peeling mechanism",[65] with unfolding starting from terminal CG pairs, one trajectory follows a different route, where unfolding starts at the center of the duplex (d(AATT)·d(AATT) tetramer) and progresses to the ends in a "bubble" mechanism. More recently, REXMD simulations of the folding of a very short DNA hairpin[25] revealed a complex folding landscape dominated by a myriad of compact unfolded conformations, some of them extremely close in rmsd to the native form but having difficulties to evolve to the native state.

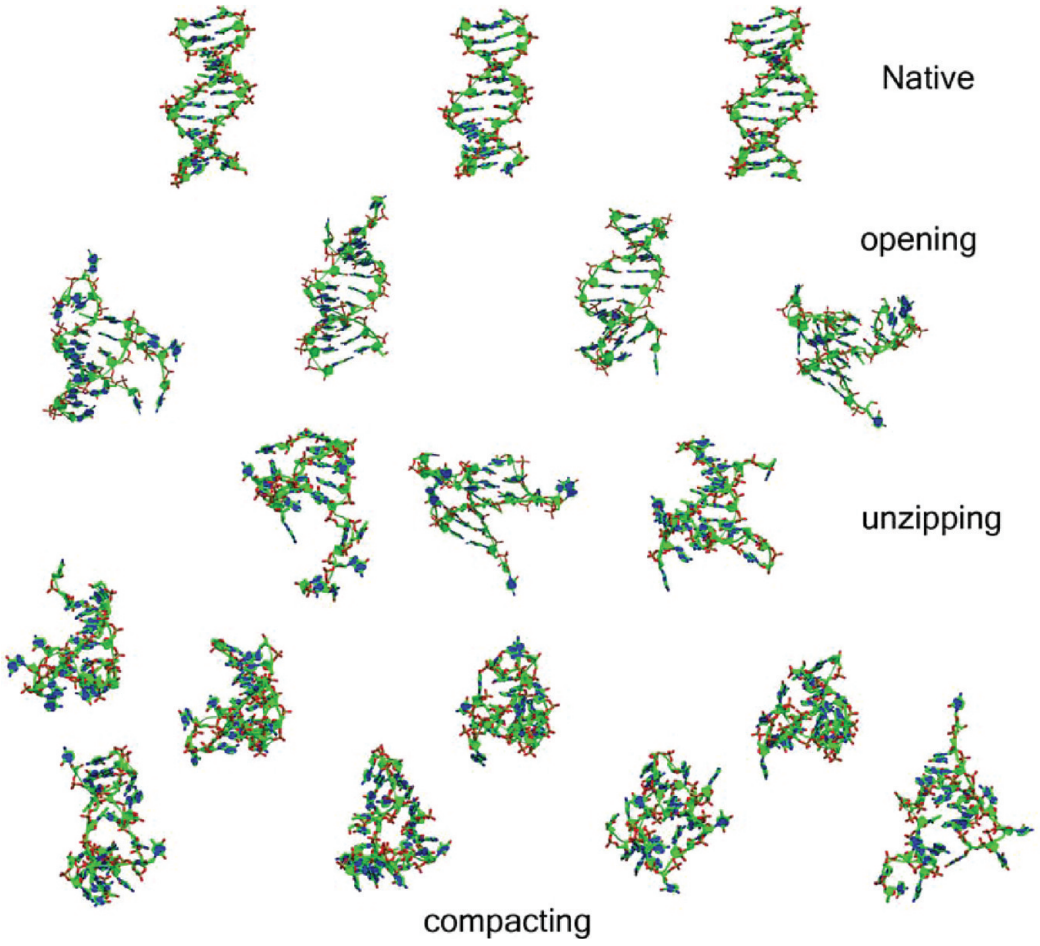


FIGURE 4. Ensembles obtained during different stages of the chemo-thermal unfolding of a DNA dodecamer.

Starting from a completely extended state, 14 of 20 trajectories converged to the native form in less than 4 μs. In fact, some folding routes converged into the native form in few nanoseconds, but others required more than 1 μs, suggesting that at least two major mechanisms coexist: (i) downhill folding,[66] which can drive extremely fast to the folded state, and (ii) detrapping, where the time-limiting step is the disruption of globular abortive conformations. MD simulations rule out simplistic views of hairpin folding, such as the popular two-state model, and support the idea of folding[67] as a competition between different folding pathways, some quickly attracted to native basin and others spending large periods in abortive basins. Overall, simulations offer a nice complement to experimental studies of DNA folding/unfolding.

Further advances in the field will probably arise from the improvement of computer resources, which will allow the study of slower transitions, from methods for improvement of sampling,[68] and certainly from the implementation of clever methods for biasing trajectories along transitions.[69,70]

Protein– and Drug–DNA Interactions

MD is widely used to describe DNA–protein and DNA–drug complexes.[5,16,17] Some groups are developing smart algorithms for the massive screening of protein and drug binding sites in DNA.[71–73] MD studies by Mukherjee et al.[74] have illustrated the complexity of the intercalation of small drugs to DNA, suggesting a mechanism that approaches the "detrapping" paradigm, where minor groove binding modes act as abortive complexes, while DNA opening in the presence of a properly oriented intercalator happens with a small free energy barrier, indicating that approach of the drug favors DNA unwinding. Metadynamics studies by Carloni's group[75] have illustrated the complexity of drug binding to DNA minor groove, showing how the drug navigates among many basins leading to nonspecific contacts before reaching the high-affinity spot. These pioneering studies illustrate the power of MD to clarify long-standing issues, such as the thermodynamics and kinetics of drug–DNA binding, and to rationalize aspects such as specificity and selectivity.

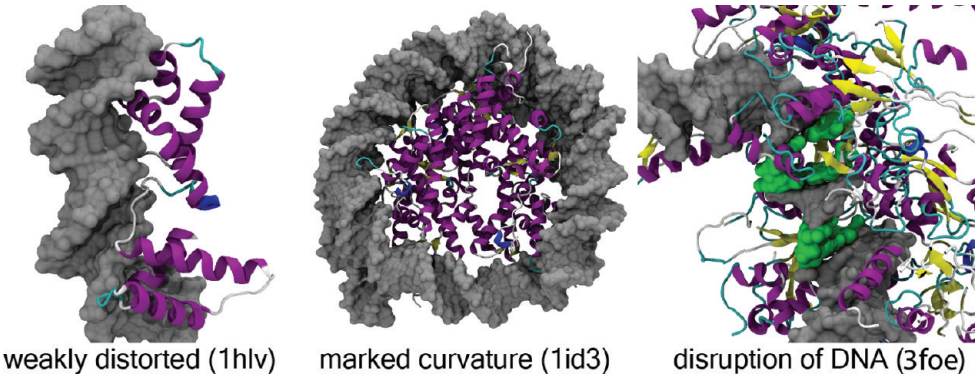


FIGURE 5. Structures of DNA–protein complexes introducing different degrees of stress into the DNA. In 3foe, region where duplex integrity is lost marked in green.

Based on the level of distortion of nucleic acids, we can distinguish three types of DNA–protein complexes (Figure 5):[72] (i) those where DNA structure is not altered, (ii) those where DNA is largely distorted but maintains the duplex integrity, and (iii) those where the duplex integrity is lost, either locally or globally. In this context, the specificity in DNA–protein interaction emerges from three sources: (a) specific hydrogen-bond interactions between protein and nucleobases, (b) electrostatic interactions mediated by grooves, whose geometry is sequence-dependent,[76] and (c) sequence-dependent DNA deformability. While factors a and b might dominate in complexes where DNA is slightly deformed, "indirect" c-type effects related to deformability might be crucial in other cases.[10,43,77–80]

The exact mechanism by which DNA deforms its geometry upon protein binding is still being debated, and simulation results support either the induced fit[81] or the conformational selection paradigms.[78] It is also unclear how proteins scan the genome to find the target sequence with a speed much greater than that expected by pure 3D diffusion.[51,82–84] MD calculations by Lavery's group[79,85] suggest that the DNA deformability might be used by the transcription factor SRY to probe the local base sequence. A similar mechanism might be exploited by other proteins to recognize their substrates.[54,84,86]

A particularly relevant protein–DNA complex is the nucleosome. Recent experiments[76,87,88] have shown that nucleosomes display preference for some sequences, whose selection is yet unclear. Different authors have suggested that indirect reading is the dominant effect in determining the position and phasing of nucleosomes,[43,77,89–92] something that agrees with the anticorrelation found between nucleosome positioning and the elastic deformation energy required to wrap a DNA around the protein core of the nucleosome (Figure 6). However, others have suggested that direct interactions play a non-negligible role in determining the best nucleosome positioning sequences.[47,72] Clearly, further work is necessary to identify the factors behind DNA sequence recognition by histones and to derive a physically based predictor of nucleosome positioning. MD is going to be crucial in this effort, which will shed light on the connection between physical properties of DNA and gene regulatory mechanisms,[93–95] which is still an unknown part of the genetic code.[96]

Final Remarks

The maturity of MD allows us to anticipate that high-quality atomistic details of major conformational changes, including DNA folding/unfolding, will be produced in the coming years, complementing ultrafast experimental techniques. MD is a crucial tool to understand the connection between physical properties and chromatin function, the ultimate objective of molecular biology in the postgenomic era. MD is also linked to X-ray crystallography in elucidating the mechanism of action of many drugs and proteins that modulate the DNA activity. MD will also have a tremendous scientific and technological impact in other areas not considered here, such as DNA conductivity, the analysis of DNA in nonaqueous solvents, and the characterization of unusual forms of DNA.[2] The past decade has demonstrated not only that MD is a technique useful to rationalize experimental behavior of nucleic acids but also that it can act as a central engine of entire experimental research lines. We predict that during this decade MD will definitively enter in the main stream of research in biology and make a transition from highly specialized "dry" labs with strong theoretical background to "misty" or even "wet" labs, often without such

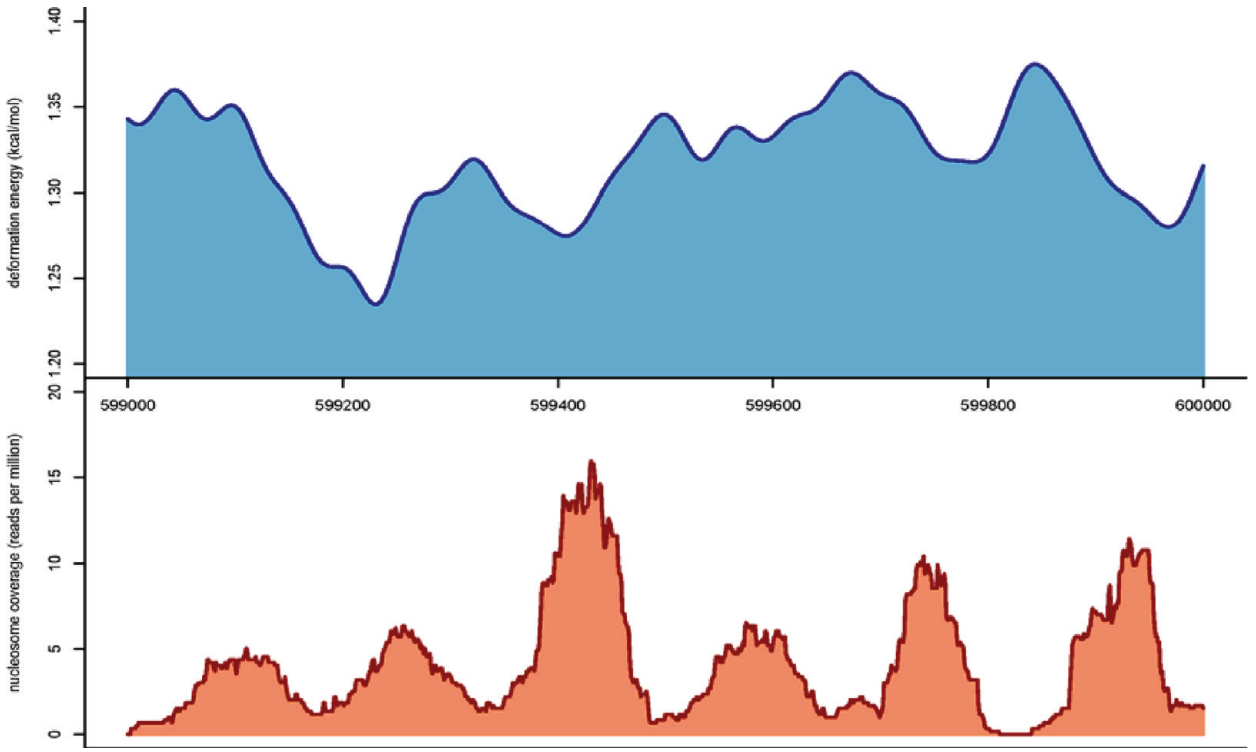


FIGURE 6. (top) Experimental nucleosome occupancy obtained from micrococal nuclease digestion and deep sequencing of yeast genome (Ozgen et al., submitted for publication) and (bottom) elastic deformation energy (eq 2) computed using elastic parameters derived from MD simulations. Region displayed corresponds to chromosome 2.

backgrounds, but working in crucial aspects of the biology of nucleic acids.

BIOGRAPHICAL INFORMATION

Alberto Pérez was born in Barcelona, Spain, in 1980. He received his Ph.D. in 2008 from the University of Barcelona under the supervision of Profs. Orozco and Luque. He is currently a post-doctoral associate working in folding of macromolecules at Stony Brook University under the supervision of Prof. Ken A. Dill thanks to an EMBO long-term fellowship.

F. Javier Luque was born in Barcelona, Spain, in 1962. He received his Ph.D. from the Universitat Autònoma de Barcelona in 1989. He joined the Departament de Física-Química of the Universitat de Barcelona in 1986, where he is now full professor of Physical Chemistry. He is group leader at the Institut de Biomedicina (IBUB). He has received several scientific awards and is a member of several editorial boards. He is the author of almost 260 papers in topics such as the representation of solvation effects, the study of polarization effects, drug design, and the analysis of macromolecular flexibility, including that of nucleic acids.

Modesto Orozco was born in Barcelona, Spain, in 1962. He received his Ph.D. from the University of Barcelona in 1990. He joined the Departament de Bioquímica i Biologia Molecular of the Universitat de Barcelona in 1986, where he is now full professor of Biochemistry at the same department. He is also group leader at the Institut de Recerca Biomèdica (IRB), director of the Life Science department of the Barcelona supercomputing center (BSC), and director of the joint IRB-BSC program on computational biology. He has received different national and international scientific awards and is member of several editorial boards. He has published more than 300 scientific papers in the area of computational biology and chemistry, with special emphasis in nucleic acids, and method development for the study of molecular interactions and solvent effects.

FOOTNOTES

*To whom correspondence should be addressed. E-mail: modesto.orozco@irbbarcelona.org.

REFERENCES

1 Watson, J. D.; Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738.

2 Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011**, *470*, 498–502.

3 Neidle, S. *Nucleic Acid Structure and Recognition*; Oxford University Press: Oxford, U.K., 2002.

4 Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. The relative flexibility of B-DNA and A-RNA duplexes: Database analysis. *Nucleic Acids Res.* **2004**, *32*, 6144–6151.

5 Rohs, R.; West, S. M.; Liu, P.; Honig, B. Nuance in the double-helix and its role in protein-DNA recognition. *Curr. Opin. Struct. Biol.* **2009**, *19*, 171–177.

6 Heddi, B.; Foloppe, N.; Oguey, C.; Hartmann, B. Importance of accurate DNA structures in solution: the Jun-Fos model. *J. Mol. Biol.* **2008**, *382*, 956–970.

7 Geggier, S.; Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15421–15426.

8 Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15431–15436.

9 Heng, J. B.; Aksimentiev, A.; Ho, C.; Marks, P.; Grinkova, Y. V.; Sligar, S.; Schulten, K.; Timp, G. The electromechanics of DNA in a synthetic nanopore. *Biophys. J.* **2006**, *90*, 1098–1106.

10 Strick, T. R.; Dessinges, M. N.; Charvin, G.; Dekker, N. H.; Allemand, J. F.; Bensimon, D.; Croquette, V. Stretching of macromolecules and proteins. *Rep. Prog. Phys.* **2003**, *66*, 1–45.

11 Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.

12 Mccammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585–590.

13 Orozco, M.; Noy, A.; Perez, A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185–193.

14 Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* **2003**, *32*, 350–364.

15 Cheatham, T. E. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360–367.

16 MacKerell, A. D.; Nilsson, L. Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194–199.

17 Laughton, C. A.; Harris, S. A. Atomistic simulation of DNA. *WIREs Comput. Mol. Sci.* **2011**, *1*, 590–600.

18 Perez, A.; Luque, F. J.; Orozco, M. Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.* **2007**, *129*, 14739–14745.

19 Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341–346.

20 MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257–265.

21 Cheatham, T. E.; Cieplak, P.; Kollman, P. A. 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.

22 Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

23 Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E.; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(C(p)G) steps. *Biophys. J.* **2004**, *87*, 3799–3813.

24 Varnai, P.; Zakrzewska, K. DNA and its counterions: A molecular dynamics study. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.

25 Portella, G.; Orozco, M. Multiple routes to characterize the folding of a small DNA hairpin. *Angew. Chem., Int. Ed.* **2010**, *49*, 7673–7676.

26 Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.* **2008**, *36*, 2379–2394.

27 Faustino, I.; Perez, A.; Orozco, M. Toward a consensus view of duplex RNA flexibility. *Biophys. J.* **2010**, *99*, 1876–1885.

28 Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. Molecular dynamics and quantum mechanics of RNA: Conformational and chemical change we can believe in. *Acc. Chem. Res.* **2010**, *43*, 40–47.

29 Savelyev, A.; Papoian, G. A. Electrostatic, steric, and hydration interactions favor Na(+) condensation around DNA compared with K(+). *J. Am. Chem. Soc.* **2006**, *128*, 14506–14518.

30 Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. Spontaneous formation of KCl aggregates in biomolecular simulations: A force field issue? *J. Chem. Theory Comput.* **2007**, *3*, 1851–1859.

31 Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. Exploring the counterion atmosphere around DNA: What can be learned from molecular dynamics simulations? *Biophys. J.* **2004**, *87*, 800–811.

32 Noy, A.; Soteras, I.; Luque, F. J.; Orozco, M. The impact of monovalent ion force field model in nucleic acids simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596–10607.

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

34 Baker, C. M.; Anisimov, V. M.; MacKerell, A. D. Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model. *J. Phys. Chem. B* **2011**, *115*, 580–596.

35 Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; Schulten, K.; Roux, B. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.* **2011**, *2*, 87–92.

36 Ouldridge, T. E.; Louis, A. A.; Doye, J. P. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys. Rev. Lett.* **2010**, *104*, No. 178101.

37 Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J. Chem. Phys.* **2010**, *132*, No. 035105.

38 Khalid, S.; Bond, P. J.; Holyoake, J.; Hawtin, R. W.; Sansom, M. S. DNA and lipid bilayers: self-assembly and insertion. *J. R. Soc., Interface* **2008**, *5* (Suppl 3), S241–S250.

39 Savelyev, A.; Papoian, G. A. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20340–20345.

40 Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711–1725.

41 Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **2003**, *85*, 2872–2883.

42 Lankas, F.; Spackova, N.; Moakher, M.; Enkhbayar, P.; Sponer, J. A measure of bending in nucleic acids structures applied to A-tract DNA. *Nucleic Acids Res.* **2010**, *38*, 3414–3422.

43 Sereda, Y. V.; Bishop, T. C. Evaluation of elastic rod models with long range interactions for predicting nucleosome stability. *J. Biomol. Struct. Dyn.* **2010**, *27*, 867–887.

44 Fujii, S.; Kono, H.; Takenaka, S.; Go, N.; Sarai, A. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.* **2007**, *35*, 6063–6074.

45 Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., 3rd; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. 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.

46 Lankas, F.; Gonzalez, O.; Heffler, L. M.; Stoll, G.; Moakher, M.; Maddocks, J. H. On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10565–10588.

47 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.

48 Demurtas, D.; Amzallag, A.; Rawdon, E. J.; Maddocks, J. H.; Dubochet, J.; Stasiak, A. Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles. *Nucleic Acids Res.* **2009**, *37*, 2882–2893.

49 Mitchell, J. S.; Laughton, C. A.; Harris, S. A. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.* **2011**, *39*, 3928–3938.

50 Randall, G. L.; Zechiedrich, L.; Pettitt, B. M. 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.

51 Huang, N.; MacKerell, A. D., Jr. Specificity in protein-DNA interactions: Energetic recognition by the (cytosine-C5)-methyltransferase from Hhal. *J. Mol. Biol.* **2005**, *345*, 265–274.

52 Noy, A.; Perez, A.; Laughton, C. A.; Orozco, M. Theoretical study of large conformational transitions in DNA: the B↔A conformational change in water and ethanol/water. *Nucleic Acids Res.* **2007**, *35*, 3330–3338.

53 Kastenholz, M. A.; Schwartz, T. U.; Hunenberger, P. H. The transition between the B and Z conformations of DNA investigated by targeted molecular dynamics simulations with explicit solvation. *Biophys. J.* **2006**, *91*, 2976–2990.

54 Qi, Y.; Spong, M. C.; Nam, K.; Banerjee, A.; Jiralerspong, S.; Karplus, M.; Verdine, G. L. Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **2009**, *462*, 762–766.

55 Cubero, E.; Luque, F. J.; Orozco, M. Theoretical study of the Hoogsteen-Watson-Crick junctions in DNA. *Biophys. J.* **2006**, *90*, 1000–1008.

56 Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. DNA folding and melting observed in real time redefine the energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 712–716.

57 Ansari, A.; Kuznetsov, S. V.; Shen, Y. Q. Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7771–7776.

58 Kannan, S.; Zacharias, M. Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10589–10595.

59 Kannan, S.; Zacharias, M. Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophys. J.* **2007**, *93*, 3218–3228.

60 Garcia, A. E.; Paschek, D. Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 815–817.

61 Sorin, E. J.; Rhee, Y. M.; Pande, V. S. Does water play a structural role in the folding of small nucleic acids? *Biophys. J.* **2005**, *88*, 2516–2524.

62 Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. Structural insight into RNA hairpin folding intermediates. *J. Am. Chem. Soc.* **2008**, *130*, 9676–9678.

63 Lin, M. M.; Meinhold, L.; Shorokhov, D.; Zewail, A. H. Unfolding and melting of DNA (RNA) hairpins: The concept of structure-specific 2D dynamic landscapes. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4227–4239.

64 Perez, A.; Orozco, M. Real-time atomistic description of DNA unfolding. *Angew. Chem., Int. Ed.* **2010**, *49*, 4805–4808.

65 Wong, K. Y.; Pettitt, B. M. The pathway of oligomeric DNA melting investigated by molecular dynamics simulations. *Biophys. J.* **2008**, *95*, 5618–5626.

66 Li, P.; Oliva, F. Y.; Naganathan, A. N.; Munoz, V. Dynamics of one-state downhill protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 103–108.

67 Thirumalai, D.; Woodson, S. A. Kinetics of folding of proteins and RNA. *Acc. Chem. Res.* **1996**, *29*, 433–439.

68 Gee, J.; Shell, M. S. Two-dimensional replica exchange approach for peptide-peptide interactions. *J. Chem. Phys.* **2011**, *134*, No. 064112.

69 de Marco, G.; Varnai, P. Molecular simulation of conformational transitions in biomolecules using a combination of structure-based potential and empirical valence bond theory. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10694–10700.

70 Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148–154.

71 Wang, H.; Laughton, C. A. Molecular modelling methods for prediction of sequence-selectivity in DNA recognition. *Methods* **2007**, *42*, 196–203.

72 Zakrzewska, K.; Bouvier, B.; Michon, A.; Blanchet, C.; Lavery, R. Protein-DNA binding specificity: A grid-enabled computational approach applied to single and multiple protein assemblies. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10712–10721.

73 Deremble, C.; Lavery, R.; Zakrzewska, K. Protein-DNA recognition: Breaking the combinatorial barrier. *Comput. Phys. Commun.* **2008**, *179*, 112–119.

74 Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T. On the molecular mechanism of drug intercalation into DNA: A simulation study of the intercalation pathway, free energy, and DNA structural changes. *J. Am. Chem. Soc.* **2008**, *130*, 9747–9755.

75 Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. Dissociation of minor groove binders from DNA: Insights from metadynamics simulations. *Nucleic Acids Res.* **2008**, *36*, 5910–5921.

76 Joshi, R.; Passner, J. M.; Rohs, R.; Jain, R.; Sosinsky, A.; Crickmore, M. A.; Jacob, V.; Aggarwal, A. K.; Honig, B.; Mann, R. S. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **2007**, *131*, 530–543.

77 Tolstorukov, M. Y.; Colasanti, A. V.; McCandlish, D. M.; Olson, W. K.; Zhurkin, V. B. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J. Mol. Biol.* **2007**, *371*, 725–738.

78 Mura, C.; McCammon, J. A. Molecular dynamics of a kappaB DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucleic Acids Res.* **2008**, *36*, 4941–4955.

79 Bouvier, B.; Lavery, R. A free energy pathway for the interaction of the SRY protein with its binding site on DNA from atomistic simulations. *J. Am. Chem. Soc.* **2009**, *131*, 9864–9865.

80 Balasubramanian, S.; Xu, F.; Olson, W. K. DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. *Biophys. J.* **2009**, *96*, 2245–2260.

81 Qin, F.; Jiang, Y.; Chen, Y.; Wu, M.; Yan, G.; Ye, W.; Li, Y.; Zhang, J.; Chen, H. F. Conformational selection or induced fit for Brinker and DNA recognition. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1407–1412.

82 von Hippel, P. H. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 79–105.

83 Verdine, G. L.; Bruner, S. D. How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* **1997**, *4*, 329–334.

84 Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **2005**, *434*, 612–618.

85 Bouvier, B.; Zakrzewska, K.; Lavery, R. Protein-DNA recognition triggered by a DNA conformational switch. *Angew. Chem., Int. Ed.* **2011**, *50*, 6516–6518.

86 Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. Dynamics of a benzo[a]pyrene-derived guanine DNA lesion in TGT and CGC sequence contexts: enhanced mobility in TGT explains conformational heterogeneity, flexible bending, and greater susceptibility to nucleotide excision repair. *J. Mol. Biol.* **2007**, *374*, 292–305.

87 Jiang, C.; Pugh, B. F. Nucleosome positioning and gene regulation: Advances through genomics. *Nat. Rev. Genet.* **2009**, *10*, 161–172.

88 Choi, J. K.; Kim, Y. J. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat. Genet.* **2009**, *41*, 498–503.

89 Wang, X.; Bryant, G. O.; Floer, M.; Spagna, D.; Ptashne, M. An effect of DNA sequence on nucleosome occupancy and removal. *Nat. Struct. Mol. Biol.* **2011**, *18*, 507–509.

90 Bishop, T. C. Geometry of the nucleosomal DNA superhelix. *Biophys. J.* **2008**, *95*, 1007–1017.

91 Miele, V.; Vaillant, C.; d'Aubenton-Carafa, Y.; Thermes, C.; Grange, T. DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic Acids Res.* **2008**, *36*, 3746–3756.

92 Ponomarev, S. Y.; Putkaradze, V.; Bishop, T. C. Relaxation dynamics of nucleosomal DNA. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10633–10643.

93 Goñi, J. R.; Fenollosa, C.; Perez, A.; Torrents, D.; Orozco, M. DNAlive: A tool for the physical analysis of DNA at the genomic scale. *Bioinformatics* **2008**, *24*, 1731–1732.

94 Goñi, J. R.; Perez, A.; Torrents, D.; Orozco, M. Determining promoter location based on DNA structure first-principles calculations. *Genome Biol.* **2007**, *8*, No. R263.

95 Singhal, P.; Jayaram, B.; Dixit, S. B.; Beveridge, D. L. Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations. *Biophys. J.* **2008**, *94*, 4173–4183.

96 Heddi, B.; Abi-Ghanem, J.; Lavigne, M.; Hartmann, B. Sequence-dependent DNA flexibility mediates DNase I cleavage. *J. Mol. Biol.* **2010**, *395*, 123–133.