Dynamics of B−DNA on the Microsecond Time Scale

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · NOVEMBER 2007

Impact Factor: 12.11 · DOI: 10.1021/ja0753546 · Source: PubMed

3 AUTHORS, INCLUDING:

Alberto Pérez
Stony Brook University
43 PUBLICATIONS **2,301** CITATIONS

SEE PROFILE

J|A|C|S
ARTICLES
Published on Web 11/07/2007

Dynamics of B-DNA on the Microsecond Time Scale

Alberto Pérez,[†,‡] F. Javier Luque,[§] and Modesto Orozco*[,†,‡,||]

Contribution from the Institut de Recerca Biomèdica and Instituto Nacional de Bioinformática, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Computational Biology Program, Barcelona Supercomputer Center, Jordi Girona 31, Edifici Torre Girona, Barcelona 08028, Spain, Departament de Fisicoquímica, Facultat de Farmàcia, Avgda Diagonal sn, Barcelona 08028, Spain, and Departament de Bioquímica, Facultat de Biología, Avgda Diagonal 647, Barcelona 08028, Spain

Received July 18, 2007; E-mail: modesto@mmb.pcb.ub.es

Abstract: We present the first microsecond MD simulation of B-DNA. Trajectory shows good agreement with available data and clarifies the μs dynamics of DNA. The duplex is sampling the B-conformation, but many relevant local transitions are found, including S → N repuckers (up to 7 N-sugars are found simultaneously), local B_{II} transitions (15% of the dinucleotides are in B_{II}-form; some of these forms are stable for up to 7 ns), and sequence-dependent α/γ transitions (happening in the 7−50 ns time scale, and being stable for up to 80 ns). Partial and total openings are often detected, but no base flipping is found. A·T openings happen after amplification of propeller twist movements, while G·C pairs (which can be opened for up to 1 ns) are opened by a complex mechanism which is often catalyzed by cations. A high affinity Na^+ binding site is found in the center of the minor groove. Access to this site by cations is difficult (average entry time 400 ns), but once inside, the ion remains for long periods of time (10−15 ns), producing a sizable narrowing of the minor groove. The essential dynamics of DNA fits well with the pattern of deformation needed to (i) sample uncommon right-handed forms and (ii) sample conformations adopted by DNA when bound to proteins. Clearly, DNA has evolved to be not only a stable structure able to maintain and transmit the genetic information but also a flexible entity whose intrinsic pattern of deformability matches its functional needs.

Introduction

The history of molecular dynamics (MD) simulations of DNA is marked by some landmark simulations which pushed the field at that time (see discussion on refs 1−8). To our knowledge, the first DNA simulations should be credited to Levitt[9] and Karplus's groups[10] in the middle 1980s, several years after the first MD simulation of proteins.[11] Those simulations were performed introducing artificial restraints since otherwise the DNA unfolded. These restraints were used until the middle 1990s, when Kollman's group performed nanosecond-long simulations of DNA in water by using an improved force field and a more accurate treatment of long-range electrostatic effects.[12−14] Further major improvements in both force fields and simulation protocols appeared at the end of the century, coming mostly from CHARMM and AMBER teams.[15−17] These improvements explained the widespread use of MD techniques to the analysis of nucleic acids (see 1−7 for review).

Most of the MD simulations published in the past 10 years were done in the 1−10 ns time scale for 10−15-mer fully hydrated oligonucleotides (1−7), though a few simulations such as those of the Ascona B-DNA consortium covered 10−20 ns,[18−20] very rarely approaching the 50−100 ns time scale.[21,22]

† IRBB Parc Cientific de Barcelona.
‡ Barcelona Supercomputer Center.
§ Facultat de Farmàcia.
|| Facultat de Biología.

(1) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10* (2), 182−196.
(2) Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−71.
(3) Cheatham, T. E. *Curr. Opin. Struct. Biol.* **2004**, *14* (3), 360−367.
(4) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35* (6), 350−357.
(5) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32* (6), 350−364.
(6) Cheatham, T. E.; Young, M. A. *Biopolymers* **2000**, *56* (4), 232−256.
(7) Sponer, J.; Lankas, F. *Computational Studies of RNA and DNA*; Springer: Dordrecht, 2006; Vol. 2, p 636.
(8) Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E.; Sponer, J. *Biophys. J.* **2004**, *87* (1), 227−242.
(9) Levitt, M. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *47* Pt 1, 251−62.
(10) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983**, *1* (1), 231−52.
(11) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267* (5612), 585−90.
(12) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (14), 4193−4194.
(13) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(14) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197.
(15) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117* (48), 11946−11975.
(16) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21* (2), 86−104.
(17) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16* (4), 845−862.
(18) Beveridge, D. L., et al. *Biophys. J.* **2004**, *87* (6), 3799−3813.
(19) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. *Biophys. J.* **2005**, *89* (6), 3721−3740.
(20) Dixit, S. B.; Beveridge, D. L. *Bioinformatics* **2006**, *22* (8), 1007−1009.
(21) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (41), 14771−14775.
(22) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32* (14), 4269−4280.

10.1021/ja0753546 CCC: $37.00

These later simulations were extremely useful as they revealed problems in force fields not detected in shorter simulations,[22,23] which have stimulated force-field refinements, such as the recently reported *parmbsc0*[23] refinement of the parm98 force field,[14,17] thus opening the door to perform very extended trajectories. In this paper we present the first microsecond-long (1.2 μs) MD simulation of the fully hydrated Dickerson−Drew Dodecamer (DD)[23] using explicit solvent and "state-of-the-art" simulation protocols. The use of the AMBER *parmbsc0* force field and of massive supercomputer resources (CPU cost equivalent to 15 years) allowed us to obtain a high quality picture of the structural and flexibility pattern of the DNA duplex in solution during a biologically relevant time scale. Relevant aspects such as the ability of DNA to adopt highly distorted structures, the dynamic properties of DNA, and its solvent environment are finally clarified. The results should represent a milestone in the atomistic simulations of DNA, since for the first time we demonstrate that a pure physical description of DNA can reliably represent with a femtosecond resolution a biologically relevant piece of DNA in water.

Methods

Simulation Protocol: As is common in the field, DD (d(CGCGAATTGCGC)$_2$) was chosen as a model system, since DD[24,25] has been largely studied and up to a dozen of NMR and X-ray structures have been deposited in the Protein Data Bank. DD is a very stable B-duplex containing a full helical turn, which combines a variety of different steps and local sequence-dependent structural elements. Using a standard protocol, the crystal structure of DD[23] was neutralized by Na^+ counterions and hydrated by 4998 water molecules. The model system was then optimized, thermalized (T = 298 K), and pre-equilibrated using our well-established multistep protocol[25,26] doubling the lengths of the windows, followed by a final re-equilibration for 10 ns. The production run extended for 1.2 μs using periodic boundary conditions and Particle Mesh Ewald.[13] SHAKE[27] was used for hydrogen atoms in conjunction with an integration time step of 2 fs. The AMBER parm98 with *parmbsc0* corrections[14,17,23] and TIP3P[28] water molecules were used to represent molecular interactions. Snapshots were collected for analysis every picosecond, leading to a total of 1.2×10^6 structures. The dry compressed (to a 95% variance coverage) trajectory[29] is available from http://mmb.pcb.ub.es/microsecond. Simulations were done using AMBER8.0.[30]

Analysis of Trajectories: The structural analysis was performed using *in house* software and standard codes like 3DNA[31] and PTRAJ module of AMBER8.0. The solvent and ionic atmospheres were analyzed by computing solvent/ions densities in regular grids and by following waters and ions with long residence times in internal coordinates. The flexibility was examined using a variety of techniques (see Supporting Information), including (a) measurement of entropies as determined by Andrioaiaei's method with Harris extrapolation,[5,32,33] (b) essential dynamics,[5,34,35] and (c) harmonic stiffness analysis as suggested by Lankas and Olson.[5,36,37] The ability of the essential deformation modes of DNA to explain biologically relevant conformational transitions was analyzed by computing the capability of essential deformation movements to trace spontaneous transitions (B → C, D, T or A conformations) and to simulate the deformation needed to adapt the DNA structure to that adopted in 50 nonredundant protein−DNA complexes taken from the NDB (see Supporting Information). The similarity between the unbound → bound movement and the essential deformation modes of free DNA was measured by using eq 1,[5] and the energy cost associated with the biological deformation was determined by assuming harmonic displacements along normal modes;[37,38] see eq 2 and Supporting Information for further details).

$$\gamma = \sum_{i=1}^{m} |r \bullet \nu_i|^2 \tag{1}$$

where m stands for the "important space" of essential deformation modes, r is the transition vector, and ν stands for unit eigenvectors (the statistical quality of the similarity index is obtained from the associated Z-scores as described elsewhere[39]).

$$E = \sum_{i}^{n} \frac{kT}{\lambda_i} \Delta X_i^2 \tag{2}$$

where k is the Boltzman's constant, T is the temperature, λ stands for eigenvalues (in distance units), and ΔX denotes the optimum displacement along a given eigenvector (see Supporting Information).

Results and Discussion

General Structure: After 6×10^8 integrations of Newton equations the DNA maintains its canonical structure, keeping backbone angles, groove dimensions, sugar puckering, and helical parameters within the values expected for a B-DNA duplex (see Figure 1 and Figures S1−S3 in Supporting Information). Interestingly, the molecular dynamics (MD) simulation captures fine structural details of the duplex, like the sequence difference in twist or sequence-dependent groove widths (see Figure S4), including a correct ratio between minor and major groove widths (see Figure 1 and S4). The distribution of average helical parameters along the trajectory (see Figure S1) is Gaussian except for the average twist which has a small asymmetry toward low values, reflecting the existence of minor conformers with reduced twist (see below). The distributions of average α, g and χ angles are Gaussians. On the contrary, the δ-distribution displays asymmetry related to the large population of East puckerings (see below). Finally, β and especially ϵ distributions are bimodal due to the very common BI/BII transitions (see below). As experimentally suggested, groove oscillations (especially those at the minor groove) are quite sequence-dependent (Figures S4 and S5) and display asymmetry due to the steric problems in extreme narrowings of the grooves. Finally, our simulations suggest that the grooves

(23) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92* (11), 3817−29.
(24) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151* (3), 535−556.
(25) Dickerson, R. E.; Ng, H. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (13), 6986−6988.
(26) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119* (32), 7463−7469.
(27) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(29) Meyer, T.; Ferrer-Costa, C.; Perez, A.; Rueda, M.; Bidon-Chanal, A.; Luque, F. J.; Laughton, C. A.; Orozco, M. *J. Chem. Theory Comput.* **2006**, *2* (2), 251−258.
(30) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26* (16), 1668−1688.
(31) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31* (17), 5108−5121.
(32) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115* (14), 6289−6292.
(33) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123* (50), 12658−12663.
(34) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17* (4), 412−425.
(35) Perez, A.; Blas, J. R.; Rueda, M.; Lopez-Bes, J. M.; de la Cruz, X.; Orozco, M. *J. Chem. Theory Comput.* **2005**, *1* (5), 790−800.
(36) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. *Biophys. J.* **2003**, *85* (5), 2872−2883.
(37) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (19), 11163−11168.
(38) Noy, A.; Perez, A.; Laughton, C. A.; Orozco, M. *Nucleic Acids Res.* **2007**.
(39) Rueda, M.; Chacon, P.; Orozco, M. *Structure* **2007**, *15* (5), 565−575.

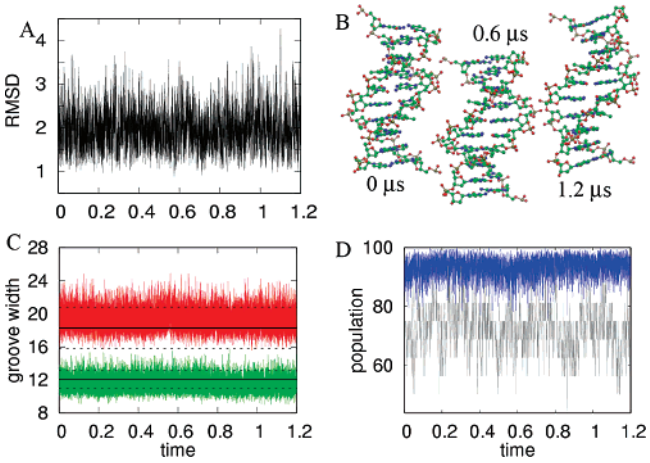

Figure 1. General characteristics of the trajectory. (A) RMSd (in Å) from crystal structure. (B) Structures at the beginning, middle, and end of the trajectory. (C) Groove widths (in Å; major, red; minor, green); experimental NMR values are shown in black solid lines with associated standard deviations (dashed lines; all values in Å). (D) Population (in %) of nucleotides with South puckering (blue) and full canonical (α,γ,ϵ,ζ and south puckering) backbones (in gray). Time scales are in microseconds in all the cases.

are very flexible, specially the major one (standard deviations up to 2.2 Å; Figure 1), in agreement with NMR data and supporting the idea that groove plasticity is a key requirement to achieve good complementary in the binding of DNA to small ligands or proteins. In summary, we can conclude that the DNA duplex shows a stable but very flexible structure, which behaves in general quite harmonically, but where some conformational parameters display a more complex behavior.

Backbone Geometry. The conformational freedom of the DNA backbone is dominated by three major elements: (i) sugar puckering, (ii) rotations around ζ/ϵ torsions, and (iiii) rotations around α/γ torsions. The high flexibility of DNA makes that very often few torsions sample noncanonical regions and around 25%, of nucleotides have at least one backbone torsion out of the canonical region (see Figure 1). As expected for B-DNA, 2′-deoxyriboses are mostly in the South region, but North puckerings are present around 5% of the time (see Figure 1). As experimentally found,[40] North conformers are more popular in pyrimidines (especially for C) than in purines (see Table S1). The S → N transition produces a moderate increase in the width of the minor groove, which becomes quite evident when several North puckers occur simultaneously (see Figure S6). Sugar transitions in the S ↔ E region are very fast and frequent (on average every 40 ps), while S → N transitions are less common, with typical transition times ranging from 100 (T and C) to 600 (G) ps and very short residence time (15−20 ps) in the North region. According to our results, S → N transitions occur in a stochastic, uncooperative way. The accumulation of North puckerings in the duplex is disfavored, but a few snapshots contain 6 (0.02% of trajectory) and up to 7 (0.002%) North puckerings. Then, it can be concluded that individual sugar repuckering is fast and easy and does not have a dramatic structural impact. On the contrary, global repuckering, which should lead to dramatic conformational transitions, is very rare event for B-philic duplexes like DD in aqueous solution.

(40) Gonzalez, C.; Stec, W.; Reynolds, M. A.; James, T. L. *Biochemistry* **1995**, *34* (15), 4969−4982.

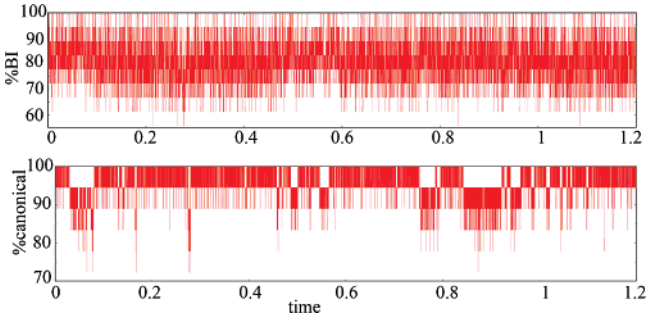

Figure 2. Evolution along the time of the number of nucleotides with (top) canonical BI (ϵ/ζ) conformation and (bottom) canonical α,γ conformation. Time scale in microseconds.

The concerted rotation around ζ/ϵ torsions generates two major confomers: B_I and B_{II}.[41] In agreement with experimental data[41,42] (see Figure 2 and S7), a dinucleotide spends around 85% of the time in the canonical B_I conformer and 15% in the B_{II} form. The population of B_{II} conformers leads to a small increase in local twist (see Figure S8), in agreement with experimental suggestions,[42] but does not produce major distortions in the duplex. Transitions to the B_{II} conformation are more common and fast in G·C segments (21% of B_{II} with a transition $B_I \rightarrow B_{II}$ every 0.4 ns) than in the central AATT tetramer (3% of B_{II} conformer with average time for $B_I \rightarrow B_{II}$ around 6 ns) (see Figure S7). The average life time of B_{II} conformers ranges between 0.1 (A·T) and 0.2 (G·C) ns, with some transitions to B_{II} being stable for up to 7 ns. Taken together, we can conclude that the dodecamer has most of the time at least one pair of ζ/ϵ torsions in the B_{II} conformation and that on average every 40 ps a $B_I \leftrightarrow B_{II}$ transition occurs somewhere in the structure (see Figure S7), supporting the idea that even in a small oligonucleotide there is sizable structural diversity and a high rate of conformational transitions. In addition, our results indicate that B_I/B_{II} transitions occur stochastically, without cooperative effects. Accumulation of B_{II} conformers is disfavored, but up to eight torsions are found simultaneously in the B_{II} region (0.007% of the trajectory) and structures with 6 ζ/ϵ torsions in the B_{II} region are not uncommon (4%).

Though conformational changes around α/γ torsions are less prevalent (see Figure 2), the population of noncanonical α/γ torsions is not negligible, since a dinucleotide spends around 1.5% of the time in these regions in agreement with experimental evidence.[43] The population of the noncanonical α/γ torsions is sequence-dependent, with lower probability on the central AATT tetramer (see Figure S9). This means that the dodecamer has at least one noncanonical α/γ pair around 30% of the time, while there are simultaneously two such unusual conformations in the dodecamer around 6% of the time (structures with more than three transitions are very rare, it being less than 0.05%). The presence of isolated noncanonical α/γ conformers leads to a local decrease in the twist (see Figure S10) but has little impact in other helical parameters. Transitions from canonical to noncanonical α/γ conformers occur on average every 7 (G·C section) or 50 (central AATT) ns and have average life times in the range 0.03−5 ns (longer values always found for G·C

(41) Hartmann, B.; Piazzola, D.; Lavery, R. *Nucleic Acids Res.* **1993**, *21* (3), 561−568.
(42) Bertrand, H. O.; Ha-Duong, T.; Fermandjian, S.; Hartmann, B. *Nucleic Acids Res.* **1998**, *26* (5), 1261−1267.
(43) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. *Nucleic Acids Res.* **2002**, *30* (24), 5398−5406.

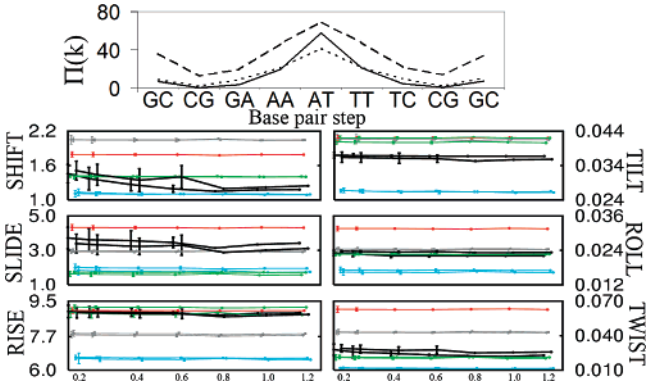

Figure 3. (Top) Global stifness index (Π in (kcal/mol Å deg)^6 in continuous line) along the sequence (central 10-mer) and associated rotational (in kcal^3/mol^3 deg^6 in dotted line) and translational (in kcal^3/mol^3 Å^6 in dashed line) indexes. Total, rotational, and translational indexes were scaled differently to display three profiles in the same plot. (Bottom) Stiffness parameters for helical deformations (left: translations, in kcal/molÅ^2 and right: rotations in kcal/moldeg^2) obtained for the different steps using different samplings. Standard deviations are shown in all cases. Color code: blue, CpG; green, d(GpA); black, d(GpC); gray, d(ApA); red, d(ApT) steps. Lines with the same color correspond to steps related by symmetry.

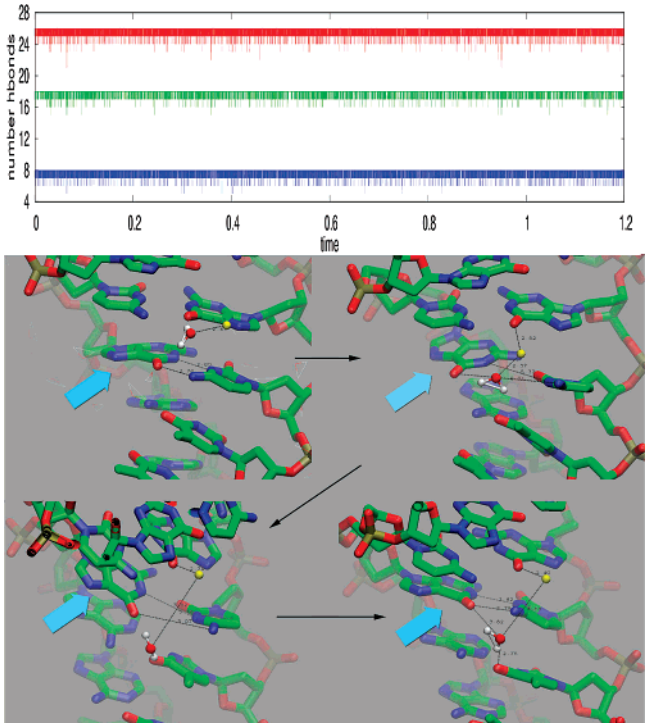

Figure 4. (Top) Time evolution (in microseconds) of the total number of hydrogen bonds (red), those due to G·C (green), and those due to A·T (blue) pairs. (Bottom) Four snapshots representative of a typical mechanistic process of G·C opening. Key water and sodium and selected distances are displayed.

pairs). Some of the transitions in the G·C region have quite long residence times (around 80 ns), indicating that noncanonical α/γ transitions can lead to metastable conformers with potential biological importance.

Base Pair Step Dynamics. Taking advantage of the general Gaussian distribution of helical parameters, stiffness analysis can be used to determine the rigidity of the different steps to distortions along helical coordinates (see Supporting Information). The global flexibility index indicates that the dodecamer is not behaving like an ideal rod but has a strongly sequence-dependent stiffness profile (see Figure 3), with the most rigid segment at the central d(AT).d(AT) step and strongly marked hinge points at the d(CG)·d(CG) steps, in agreement with X-ray derived data (see Figure S11). The large rigidity of the central d(AT)·d(AT) step is due to both translational (mostly slide) and rotational (mainly twist and roll) components, while the very soft nature of the two d(CG)·d(CG) steps is general for all translational and rotational movements. As a final remark, present results support the general quality of helical stiffness parameters obtained in shorter simulations (100 ns time scale; see Figure 3), but we should remark that convergence in some specific stiffness parameters can be slow, thus requiring longer calculations.

Local Nucleobase Dynamics. The internal geometries of base pairs oscillate around equilibrium values during the entire trajectory, leading to structures which often do not fit "ideal Watson−Crick" pairs (Figure 4). For example, base pairs at the central AATT track displays an average propeller twist of 17°−18°, and a non-negligible buckle is found for both A·T and G·C pairs (see Table S2). As expected, G·C pairs are in general less deformable by internal translations and rotations than the A·T ones (Table S2 and Figure S12). Many partial opening events leading to disruption of one hydrogen bond (H-bond) occur during the trajectory, especially at A·T pairs (one of the 4 A·T pairs has lost one H-bond in 8% of the trajectory; see Figure S12), but the canonical geometry is recovered in the ps time scale. The simultaneous loss of more than two H-bonds is rare (<1%); even a few snapshots are detected where up to five hydrogen bonds are lost (they represent less than 0.002% of trajectory). In average, one of the 10 central base pairs is "opened" (two H-bond distances greater than 4.5) every 7.3 ns (indicating an average time for individual base pair opening of around 70 ns). The average lifetime of an "open" base pair is typically 10−20 ps, but a few opening events last for 100−700 ps, leading to a strong reduction in local twist and to a unique solvation pattern where water molecules bridge the opened pairs (as anticipated by Lavery and co-workers from umbrella sampling calculations[44]). Surprisingly, these long-living events are found more often in G·C pairs, and openings always occur toward the major groove. Finally, the opening, which is often preceeded by the binding of Na^+ to the major groove, involves different H-bond breaking mechanisms in A·T and G·C pairs,[44] since, in the former case, it is mediated by an amplification of "propeller twist" and, in the latter, by an "opening" rotation in the base-pair plane (see a typical example in Figure 4; more exhaustive data are available at http://mmb.pcb.ub.es/microsecond). The different nature of the A·T and G·C opening can be also shown by analyzing the distribution of "opening" angles in the structures displaying disruption of one or two hydrogen bonds in A·T and G·C steps (see Supporting Information Figure S13).

Overall, we should conclude that partial and total openings are quite frequent and in most cases have a very short life time. However, in a few cases the open structures have life times that allow proton interchange with solvent without local unfolding of the DNA (see distribution of solvent accessible surfaces in Figure S14). The opening of base pairs mainly reflects competition with solvent molecules at the major groove

(44) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31* (5), 1434−43.

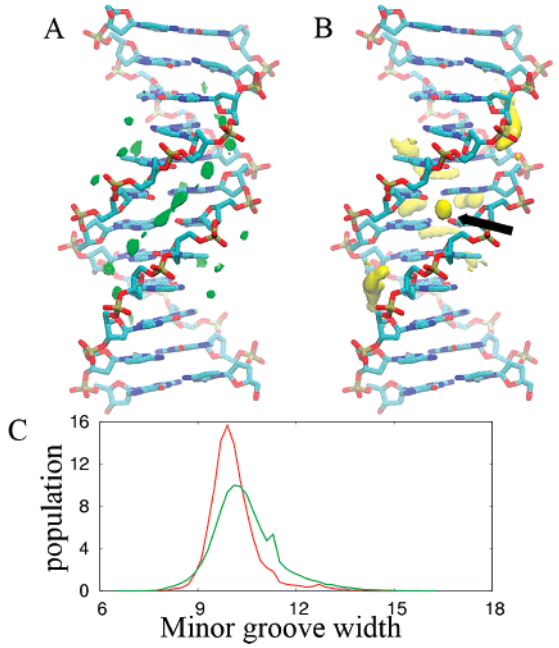

Figure 5. (Top) Density contours for (A) water in green and (B) Na^+ in yellow. Contours represented correspond to 2.5 (water) and 2.0 (Na^+) times the background densities. The position of a long residence Na^+ is marked with an arrow. (Bottom) Population (in %) of structures with AATT minor grooves of different widths (in Å) considering all the snapshots (green) or only those where the central position is occupied by a Na^+ (red).

rather than intrinsic strength/weakness of base pairs. Furthermore, the opening mechanism depends on the type of base pair (G·C and A·T) and of the flanking sequence which can modulate the water-induced stabilization of opened bases.

Solvent Atmosphere: The hydration free energy of each nucleotide pair (see ref 45 and Supporting Information) increases as we move to the center of the dodecamer (see Supporting Information Figure S15). However, taking the central 6-mer $d(GAATTC)_2$ there is little difference in the solvation free energy of G·C and A·T pairs, in disagreement with generally accepted ideas that A·T pairs are better solvated than the G·C ones in B-DNA. On average, we found around 39 waters in the vicinities (less than 5 Å) of polar atoms of each nucleotide pair in the central hexamer, again with little difference between steps. There is a strong region of high water density at the minor groove of A·T steps (see Figure 5), which reproduces Dickerson's spine of hydration.[24] Less defined regions of preferential hydration are located at the minor groove of the central G·C step and in major grooves, where these regions appear in tandem (see Figure 5). Water molecules are very mobile, and the longest residence times are around 200−500 ps. However, two events were found during the trajectory where waters were trapped for 1 ns in the same position. All long-living waters are found at the bottom of the central portion of the minor groove, a narrow region with a large electronegative potential and strong hydrogen-bonding potential.

The Na^+ ions are also very mobile visiting around 99% of the accessible space during the trajectory, suggesting that the ion atmosphere around DNA can be well sampled on the μs time scale. The strongest Na^+−nucleotide interactions happen in the central hexamer, with a preference for d(A·T) pairs in front of the d(G·C) ones (see Figure 5). A strong, very localized region of high Na^+ density (up to 9 times above background) is found in the center of the minor groove, with Na^+ being mainly coordinated to O_2 of T_7 and T_{19} (Figure 5). More diffuse binding regions are located symmetrically in the minor groove at $d(CG\boldsymbol{CGA}ATT\boldsymbol{GCG}C)_2$ sites and in the major groove of the central $d(AATT)_2$ segment parallel to the interbase pair plane (see Figure 5). Ions located at diffuse regions in the two grooves have residence times of 0.1−0.2 ns. The high affinity position at the center of the duplex is occupied during 4% of the trajectory (residence times of 10−15 ns), thus reflecting the Na^+ binding site previously suggested by shorter MD simulations[46] and accurate NMR or X-ray experiments.[47,48] It is worth noting, however, that the entrance of Na^+ to the high affinity site is very rare (it was detected only 3 times during the trajectory) and might be undetectable in sub-microsecond trajectories.

It is very clear from the trajectory that there is competition between water and Na^+ ions for the minor groove, especially at the central AATT track. The simultaneous occupancy of the groove by several ions is uncommon, but the presence of one Na^+ is not rare. Our data indicate 250 times more water than Na^+ at the minor groove (even at the high affinity binding site for Na^+, water is around 25 times more abundant). As previously noticed,[21,22,46] the presence of Na^+ in unspecific positions in the groove does not lead to relevant changes in the helix. However, occupancy of the high affinity binding site significantly reduces the width of the minor groove (see Figure 5), supporting the suggestions by Williams and co-workers.[47,48] It is worth noting that the total amount of time in which the high affinity position is occupied by Na^+ is rather small and, accordingly, the narrow amplitude of the minor groove in the AATT track found experimentally should be mostly due to the intrinsic characteristics of this sequence, and not to the presence of cations inside the groove.

Global Deformations: Entropy analysis shows that DNA is quite flexible and large simulation times are needed to explore the configurational space. For example, taking the entropy extrapolated at infinitive simulation time ($S_\infty = 2.21$ kcal mol^{-1} K^{-1}; see Supporting Information) as reference, 97% of the entropy space is visited in 1.2 μs, which decreases to 92% or 88% if the first 100 or 10 ns of trajectory are used (see Figure S16). The deformability of the duplex is dominated by 10 movements, which represent around 70% of the total variance. The three first modes (with harmonic deformation constants between 3 and 6 cal mol^{-1} Å^{-2}) explain around 40% of the variance and are annotated as (i) global twisting/untwisting related to correlated changes in the minor and major grooves and global changes in twist; (ii) bending/twisting around hinge points at d(GA)·d(TC) junctions and linked to anticorrelated changes of twist roll and groove widths at the ends of the duplex; and (iii) global tilting with strong and correlated changes in the major groove width and twist (see movies at http://mmb.pcb.ub.es/microsecond). The measures of self-similarities (see Supporting Information and Figure S17) between different portions of the trajectory show that there is good convergence when short trajectories are considered (self-similarity indexes

(45) Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Proteins: Struct., Funct., Bioinf.* **2004**, *57* (3), 458−467.
(46) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004**, *87* (2), 800−811.
(47) Denisov, V. P.; Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (2), 629−633.
(48) Williams, L. D. *DNA Binders and Related Subjects* **2005**, *253*, 77−88.

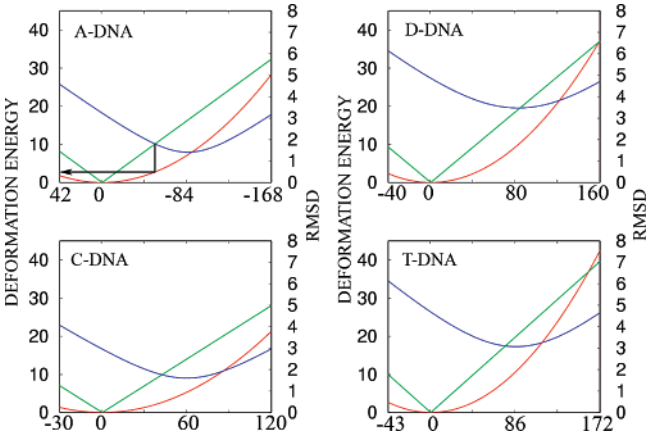

Figure 6. Analysis of the ability of the first relaxed mode of DD in water to guide conformational transitions: B → A (top, left), B → D (top, right), B → C (bottom, left), and B → T (bottom, right) transitions. The *x*-axis stands always for deformation along the first mode (in Å). The *y*-axis reads for the RMSd (Å) from the relaxed structure (green), the RMSd from the target structure (in blue), and the deformation energy (red) in kcal/mol. The way in which the energy for a transition from pure B to a B/A structure is determined can be seen in the first panel.

of 0.85−0.95 for 50 ns trajectories), which improves for 100-ns long trajectories (self-similarity indexes of 0.92−0.99), and that any (small) memory effect is completely lost after 100 ns. In summary, even though very long simulations are needed to completely sample the conformational space, shorter simulations can provide a qualitatively correct picture of the DNA flexibility, supporting many of the already published simulations.

Functionally Important Deformations: The large and well-defined flexibility of DNA is not fortuitous, but it is carefully designed by evolution to facilitate the biological role of the molecule. Thus, the first deformation mode overlaps very well (see Table S3) with the transition vector moving from B-DNA to other biologically relevant conformations (C, D, T, and A). This excellent ability of the B-DNA deformation space to favor relevant transition explains the low energy associated with these (between 2 (C) and 9 (T) kcal/mol to reach the point where rms(B) = rms(X), where X denotes a different conformation; see Figure 6), indicating that the B-DNA can adopt unusual conformations (of biological importance) quite easily.

Further insight into the relation between DNA flexibility and function comes from the analysis of 50 nonredundant DNA−protein complexes. Distortions in 33 of these complexes were small (less than 2 Å), but in other cases distortions were large (up to 14 Å; see Supporting Information Figure S18), which points out the need for high flexibility in the DNA. The deformation required to adapt the equilibrium geometry of the DNA to that in the protein−DNA complex overlaps quite well with the first and, to a lesser extent, with the third eigenvectors of the relaxed DNA (see Figure 7). In fact, when the first 10 eigenvectors are considered, the similarity index (γ; eq 1) between the unbound → bound transition vector and the essential dynamics of the DNA is quite large (γ in the range 0.3−0.8, and Z-scores in the range 30−350 indicating that this cannot be expected for a random deformation pattern). The best overlap between transition and essential space is found in those cases where the DNA structural distortion is more severe, that is, where structural flexibility is most demanding. The ability of

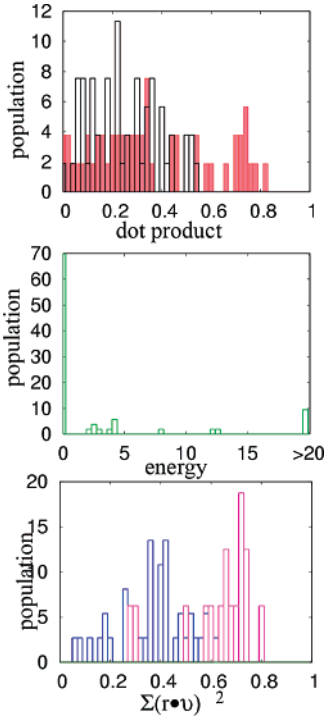

Figure 7. Analysis of the ability of essential deformation modes to sample transitions in B-DNA from unbound in solution to DNA−protein complexes. (Top) Dot product between transition vectors and the first (pink) and third (white) eigenvectors. (Middle) Histogram distribution (in %) of deformation energies (in kcal/mol) associated with the deformation of relaxed DNA to conformation adopted in a set of 50 different DNA−protein structures (the central 12-mer showing the most deformed conformation is taken for each complex). (Bottom) Similarity indexes (γ; see Methods) between the essential dynamics of DNA (considering first 10 eigenvectors) and the unbound → bound transition vectors; values in blue correspond to small unbound → bound transitions (less than 2 Å), and those in magenta, to large transitions.

DNA to spontaneously sample biologically relevant transitions also justifies the small energy cost (see eq 2, Supporting Information and Figure 7) required to drive the DNA to the bound state. Clearly, evolution has designed not only the sequence but also the flexibility pattern of DNA to make it suitable to perform its biological function.

Conclusions

More than a decade from the first μs simulation of a protein,[49] we demonstrate here for the first time that using state-of-the-art force fields, simulation protocols, and supercomputer resources that B-DNA in aqueous solution can be simulated in the μs time scale providing samplings that agree well with experimental data. The DNA is a stable structure close to the canonical B form, but at the same time it is able to support major reversible conformational transitions in time scales ranging from pico- to microseconds. Simulations suggest that most of the time the DNA is locally sampling noncanonical conformations. The solvent environment and the dynamics of transitions are clearly outlined showing the existence of subtle connections linking the presence of ions or waters in the vicinities of the duplex to local changes in the structure of the nucleotide pairs. Finally, our results strongly suggest that the intrinsic flexibility of DNA has been carefully designed by

(49) Duan, Y.; Kollman, P. A. *Science* **1998**, *282* (5389), 740−744.

evolution to exploit its biological function, which often requires major conformational changes. Overall, it can be envisaged that atomistic MD simulations are able to explore processes in a time scale of biological importance, such as those conformational changes involved in chromatin packing/repacking or DNA binding to proteins.

Acknowledgment. We are indebted to many members of the ABC team for helpful discussions. This work was supported by the Spanish Ministry of Education and Science (BIO2006-01602) and the National Institute of Bioinformatics (Structural Bioinformatics Node). Calculations were performed using the *MareNostrum* supercomputer at the Barcelona Supercomputer Center.

Supporting Information Available: Supporting figures, tables, and methods that complement the article are available. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0753546