

PARMBSC1: A REFINED FORCE-FIELD FOR DNA SIMULATIONS

Ivan Ivani^{1,2}, Pablo D. Dans^{1,2}, Agnes Noy³, Alberto Pérez⁴, Ignacio Faustino^{1,2}, Adam Hospital^{1,2}, Jürgen Walther^{1,2}, Pau Andrio^{2,5}, Ramon Goñi^{2,5}, Alexandra Balaceanu^{1,2}, Guillem Portella^{1,2,6}, Federica Battistini^{1,2}, Josep Lluís Gelpí^{2,7}, Carlos González⁸, Michele Vendruscolo⁶, Charles A. Laughton⁹, Sarah A. Harris³, David A. Case¹⁰ and Modesto Orozco^{1,2,7*}

¹Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, 08028 Barcelona, Spain.

²Joint BSC-IRB Research Program in Computational Biology, Baldiri Reixac 10-12, 08028 Barcelona, Spain.

³School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, UK.

⁴Laufer Center for Physical and Quantitative Biology, Stony Brook University, USA.

⁵Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona, Spain.

⁶Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

⁷Department of Biochemistry and Molecular Biology, University of Barcelona, 08028 Barcelona, Spain.

⁸Instituto de Química Física "Rocasolano", CSIC, Serrano 119, 28006 Madrid, Spain.

⁹School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK.

¹⁰Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08540, USA.

We present here parmbsc1, a new force-field for DNA atomistic simulation, which has been parameterized from high-level quantum mechanical data and tested for nearly 100 systems (~140 μ s) covering most of the DNA structural space. Parmbsc1 provides results of high quality in diverse systems, solving problems of previous force-fields. Parmbsc1 aims to be a reference force-field for the study of DNA in the next decade. Parameters and validation trajectories are available at <http://mmb.irbbarcelona.org/ParmBSC1/>.

Force-field development is tightly connected to the extension of simulation time scales. As molecular dynamics (MD) trajectories are extended, errors previously undetected in short simulations emerge, creating the need to improve the force-fields¹. For example, AMBER parm94/99 was the most used force-field in DNA simulations until multi-nanosecond simulations revealed severe artifacts^{2,3}, thus fueling the development of parmbsc0^{2,4}, which, in turn, started to show deviations from experimental data in the μ sec regime (for example an underestimation of the twist, biases in sugar puckering and ϵ/ζ torsions, excessive terminal fraying^{2,5}, and severe problems in representing certain non-canonical DNAs^{1,6}). Various modifications have been proposed to address these problems, such as the OL-ones^{5,6} designed to reproduce specific forms of DNA. While these and other tailor-made modifications are useful, there is an urgent need for a new general-purpose AMBER force-field for DNA simulations to complement recent advances in the CHARMM family of force-fields (see Methods). The parmbsc1 force-field presented here is designed to solve these needs, and aims to become a widely-used general-purpose (see simulated systems in Table S1) force-field for DNA simulations.

As described in the Methods, parmbsc1 shows a great ability to fit QM data, improving parmbsc0 results. The QM-derived parameters were first tested on the *Drew-Dickerson dodecamer* (DDD), the most studied DNA structure^{2,7}, typically used as benchmark in force-field developments. The trajectories sampled a stable B-type duplex that remained close to the experimental structures (Figure 1 and Table S2),

preserving hydrogen bonds and helical characteristics, even at the terminal base pairs, where fraying artifacts are common using other force-fields^{2,8} (see Methods, Table S2 for a comparison of parmbsc1 with other force-fields, and the Supplementary Material for an experimental study of fraying). The average, and sequence-dependent helical parameters (Figure 1 and Figures S1-S2) and BI/BII conformational preferences (Table S2 and Figure S3) match experimental values (for the comparison with estimates obtained with other force-fields see Methods and Supplementary Methods). Furthermore, parmbsc1 reproduce residual dipolar couplings (Q-factor=0.3) and NOEs (only 2 violations), giving success metrics similar to those obtained in the NMR-refined structures (Table S3).

We evaluated the ability of parmbsc1 to represent *sequence-dependent structural features* from simulations on 28 B-DNA duplexes (Table S4). The agreement between simulation and experiment is excellent (RMSd/base pair of 0.1- 0.2 Å). Almost no artifacts arising from terminal fraying were present, and the average helical parameters (twist and roll from simulations: 33.9° and 2.5° respectively), matched values from the analysis of the PDB (33.6° and 2.9°)⁹. Moreover, parmbsc1 was able to reproduce the unique properties of A-tracts¹⁰ (Figures S4-S6), and capture sequence-dependent structural variability (see Figure S7). We also studied longer duplexes (up to 50 bp) to ensure that a possible accumulation of small errors given by the force-field did not compromise the description of the DNA, finding excellent results (Table S5). The expected spontaneous curvature was clearly visible in both static and dynamical descriptors, demonstrating that parmbsc1 trajectories are able to capture complex polymeric effects (Table S5).

We also explored the ability of parmbsc1 to represent *unusual DNAs*, such as: a Holliday junction, a complex duplex/quadruplex structure which was fully preserved in μ sec-long trajectories (Figures S8-S9), or the Z-DNA a *levo* duplex containing nucleotides in *syn*, for which parmbsc1 not only provided stable trajectories (Figure 2a), but also reproduced the experimentally known salt

dependence, confirming that the conformation is stable only at high (4 M) salt concentration¹¹. For Hoogsteen-DNA (H-DNA), simulations with parmbsc1 showed a stable duplex for more than 150 ns (Figure 2b), and severe distortions in longer simulation periods (Figure S10), as expected from its metastable nature¹². Equivalent results were obtained for another metastable structure: the parallel poly-d(AT) DNA (Figure S11)¹³. Parmbsc1 simulations not only reproduced the known structure of parallel d(T-A·T) and d(G-G·C) triplexes (Figure 2c,d), but also showed correctly that the equivalent antiparallel structures are unstable in normal conditions (Figure 2e)¹⁴. Finally, parmbsc1 was able to reproduce experimental structures of both parallel and antiparallel DNA quadruplexes (RMSd < 2 Å, Figure 2f,g).

We explored also the ability of parmbsc1 to reproduce the complex conformation of hairpins and loops, exceptionally challenging structures for force-fields¹⁵. We performed μ s simulations of the d(GCGAAGC) hairpin (PDB: 1PQT), the 4T-tetraloop in Oxytricha Nova quadruplex d(G₄T₄G₄)₂ (OxyQ; PDB: 1JRN), and the junction loops in the human telomeric quadruplex (HTQ; PDB: 1KF1). Parmbsc1 provided excellent representations (RMSd around 1 Å) of the d(GCGAAGC) hairpin (Figure 2h), and of the OxyQ quadruplex (Figure 2i). For the very challenging HTQ structure, parmbsc1 maintained the stem structure 20 times longer than in previous simulations¹⁵, and recognized the large flexibility of the loops in the absence of the lattice-contacts (Figure S12), showing that, as predicted¹⁶, not only the crystal, but also other loop conformations were sampled (Figure 2j).

As an additional critical test of the new force-field we predicted *NMR observables* from parmbsc1 trajectories (see Methods). We obtained equivalent NOE violation statistics to those determined from NMR-derived ensembles (Tables S6 and S7, and Figure S13). This agreement was maintained in *de novo* predictions, i.e. in those cases where NMR observables were collected in one of our laboratories after parmbsc1 development (Table S8). Finally, it is worth noting that parmbsc1

trajectories reproduced the structure of DNA in *crystal environments*, giving a RMSd between the simulated and crystal structures of only 0.7 Å, and average twist differences below 1°, improving any previous calculations (see Methods and Figures S14-S15).

Our final structural test was to explore the ability of parmbsc1 to reproduce the conformation of DNA in complex with other molecules. We studied four diverse *protein DNA complexes* (PDB: 1TRO, 2DGC, 3JXC and 1KX5), and two prototypical drug DNA complexes. In all cases excellent agreement (RMSd for DNA around 2-3 Å for DNA in protein-DNA and 1-2 Å in drug-DNA complexes) with experiments was found (Figures 3 and S16-S17).

A force-field should not only reproduce the structure of DNA, but also its mechanical properties¹. To evaluate the performance of parmbsc1 we firstly evaluated the μ -scale dynamics of the central 10-mer of the DDD. The agreement between parmbsc0 and parmbsc1 *normal modes and entropy estimates* (Methods and Table S9) demonstrated that parmbsc1 does not “freeze” the DNA structure (a risk for a force-field reproducing well average properties), this was also confirmed by the ability of parmbsc1 to reproduce the DNA dielectric constant (8.0 ± 0.3 for DDD *versus* the experimental estimate of 8.5 ± 1.4 ; see Figure S18 and Methods), and the cooperative binding (around 0.7 kcal/mol) of Hoechst 33258 to DNA. We then computed the *helical stiffness matrices* for the 10 unique base pair steps^{17,18}. Parmbsc1 values were intermediate between parmbsc0 and CHARMM27 stiffness parameters¹⁸, and significantly smaller than those suggested by Olson and coworkers¹⁷ (Table S10 and Figure S19); the dependence of the stiffness parameters on sequence were similar for parmbsc1 and parmbsc0¹⁷.

The persistence length and, torsional and stretching moduli were obtained from simulations of long (up to 56 bp) duplexes (see Methods). Parmbsc1 predicted persistence lengths in the range of 40-57 nm (Table S11), close to the generally

accepted value of 50 nm. The computed static persistence length, stretch and twist torsion modules were around 500 nm, 1100-1500 pN, and 50-100 nm respectively, also in agreement with experimental values (Table S11). Finally, we explored the ability of parmbsc1 to describe relaxed and stressed *DNA minicircles*. We performed 3x 100 ns simulations of a 106-bp minicircle with 10 turns (106t10), which should have zero superhelical density ($\sigma=0$) and therefore no denatured regions^{19,20} (Figure S20). A kink was observed only in a single replica for one of the register angles, while in the remaining simulations the DNA remained intact (Figure S20). On the contrary, negatively supercoiled 100-bp (100t9; $\sigma=-0.05$) and 106-bp (106t9, $\sigma=-0.10$) minicircles formed distortions due to the superhelical stress, as previously reported experimentally using enzymes that digest single stranded DNA^{19,20}.

After demonstrating the ability of parmbsc1 to describe stable and metastable DNA structures and DNA flexibility, we finally studied *conformational transitions*. Parmbsc1 reproduced the spontaneous A→B transition in water, and the A form was found, as expected, to be stable in 200 ns control simulations in a 85% ethanol/15% water mixture (Figure S21). Parmbsc1 also reproduced the unfolding of DNA d(GGCGGC)₂ in a 4 Molar pyridine (Figure S21), and the effective folding of d(GCGAAGC) in water (Figure S22), suggesting the ability to capture long-scale conformational changes in DNA.

Based on the wide series of tests reported here we conclude that parmbsc1 provides good representations of the static and dynamic properties of DNA, and therefore anticipates that parmbsc1 will be a valuable reference force-field for atomistic DNA simulations under a diverse range of conditions.

METHODS

Methods and associated references are available in the online version of the paper. Additional details are shown in the Supplementary Material.

ACKNOWLEDGEMENTS

MO thanks Spanish Ministry of Science (BIO2012-32868), the Catalan SGR, the *Instituto Nacional de Bioinformática*, and the European Research Council (ERC_SimDNA) for support. MO is an ICREA academia researcher. MO thanks CPU/GPU time on MareNostrum/MinoTauro (BSC). CAL, SAH and AN thanks the UK HECBioSim Consortium for HPC time on ARCHER (Grant EP/L000253/1). AN was supported by the Biotechnology and Biological Sciences Research Council (BBSRC, grant number BB/I019294/1), and thanks ARC Leeds for computational resources. PDD is a PEDECIBA and SNI (ANII, Uruguay) researcher. DAC thanks Chunmei Liu for assistance with the crystal simulation analysis.

AUTHOR CONTRIBUTION

II derived the parmbsc1 force-field parameter set. II, PDD, AN, AP, IF, AH, JW, AB, GP, FB, CAL and SAH performed validation simulations. CG, MV, and GP validate results from NMR. CG did *de novo* NMR measures. DAC performed crystal MD simulations. RM, PA, AH, and JLG created the database infrastructure and web application. All authors contributed to the analysis of data. MO had the idea, directed the project, and wrote the manuscript which was improved by the rest of the authors.

REFERENCES

1. Pérez, A., Luque, F. J. & Orozco, M. *Acc. Chem. Res.***45**, 196–205 (2011).
2. Pérez, A., Luque, F. J. & Orozco, M. *J. Am. Chem. Soc.***129**, 14739–14745 (2007).
3. Varnai, P. & Zakrzewska, K. *Nuc. Acids Res.***32**, 4269–80 (2004).
4. Pérez, A. *et al. Biophys. J.***92**, 3817–3829 (2007).
5. Zgarbová, M. *et al. J. Chem. Theory Comput.***9**, 2339–2354 (2013).
6. Krepl, M. *et al. J. Chem. Theory Comput.***8**, 2506–2520 (2012).
7. Wing, R. *et al. Nature***287**, 755–758 (1980).
8. Lavery, R. *et al. Nucleic Acids Res.***38**, 299–313 (2010).
9. Dans, P. D., Pérez, A., Faustino, I., Lavery, R. & Orozco, M. *Nucleic Acids Res.***40**, 10668–10678 (2012).
10. Lankaš, F., Špačková, N., Moakher, M., Enkhbayar, P. & Šponer, J. *Nucleic Acids Res.***38**, 3414–3422 (2010).

11. Thamann, T. J., Lord, R. C., Wang, A. H. J. & Rich, A. *Nucleic Acids Res.***9**, 5443–5458 (1981).
12. Abrescia, N. G. A., González, C., Gouyette, C. & Subirana, J. A. *Biochemistry***43**, 4092–4100 (2004).
13. Cubero, E., Luque, F. J. & Orozco, M. *J. Am. Chem. Soc.***123**, 12018–12025 (2001).
14. Soyfer, V. N. & Potaman, V. N.: *Triple-helical nucleic acids*. (Springer - Verlag, New York, USA, 1996).
15. Fadrná, E. *et al. J. Chem. Theory Comput.***5**, 2514–2530 (2009).
16. Martín-Pintado, N. *et al. J. Am. Chem. Soc.***135**, 5344–5347 (2013).
17. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. *Proc. Natl. Acad. Sci.***95**, 11163–11168 (1998).
18. Pérez, A., Lankas, F., Luque, F. J. & Orozco, M. *Nucleic Acids Res.***36**, 2379–2394 (2008).
19. Moroz, J. D. & Nelson, P. *Proc. Natl. Acad. Sci.***94**, 14418–14422 (1997).
20. Du, Q., Kotlyar, A. & Vologodskii, A. *Nucleic Acids Res.***36**, 1120–1128 (2008).

ON-LINE METHODS

Parameterization procedure. All backbone torsion angles, except the recently corrected α/γ^4 were parameterized using representative model compounds (Figure S23), for which torsional profiles were obtained at the MP2/aug-cc-pVDZ level using B3LYP/6-31++G(d,p)-optimized geometries²¹⁻²³. Single-point calculations at crucial points of the conformational space were performed at the CCSD(T)/complete basis set (CBS) level²⁴⁻²⁶. Solvent effects were introduced using our MST^{28,29} method as implemented in Gaussian²⁸ (see Supplementary Methods for additional details on QM calculations).

Parameters were fitted using a flexible Monte-Carlo procedure⁴, which minimizes the error between QM reference profiles (in solution) and classical potentials of mean force-calculations in aqueous solution obtained from umbrella sampling calculations³⁰. By default we used gas phase-fitted values as first guess, and always limited the torsional representation to a 3 Fourier expansion terms, while reinforcing in the fitting the weight of the points described at the highest level of theory and those geometrical regions that are specially populated in experimental structures. Around 5-10 acceptable solutions of the Monte Carlo refinement were tested on short MD simulations (around 50-100 ns) for one small duplex d(CGATCG)₂ rejecting those leading to distorted structures. The optimum parameter set, without additional refinement was extensively tested against experimental results. Additional details (and references) on the parametrization procedure are given in the Supplementary Methods. Note that the way in which the parameters were derived does not guarantee their validity for RNA simulations. The use of others, already validated, RNA force-fields is recommended.

As shown in the on-line Methods Table 1, refined parmbsc1 parameters fit very well high-level QM data. The syn-anti equilibrium, which was non-optimal in parmbsc0, is now well reproduced (Fig. S24). The fitting to sugar puckering profile was improved by increasing the East barrier, and by displacing the North and South minima to more realistic regions (on-line Methods Table 1 and Fig. S25). Additionally, parmbsc1 provides ϵ/ζ conformational map almost indistinguishable from the CCSD(T)/CBS results in solution (Fig. S26), with errors in the estimates of relative BI/BII stability and transition barrier equal to 0.2 and 0.0 kcal/mol respectively.

Validation MD simulations. We performed an extensive set of simulations of a wide variety of DNA systems (Table S1) with a total simulation time of $\sim 140 \mu\text{s}$, which represents the most comprehensive analysis of DNA dynamics published to date. MD simulations were performed using AMBER³¹ or GROMACS³² (for the

impact of using one or the other simulation engine in the calculations see Methods and Fig. S27). Unless otherwise stated, calculations were done using TIP3P³³ solvated systems under NPT (P=1 atm; T=300 K) conditions. For discussion on the impact of ionic strength in the trajectories see Supplementary Methods and Figure S28. Simulations mimicking crystal environments were carried out as described elsewhere³⁴ for d(CGATCGATCG)₂ (PDB: 1D23) using 2- μ sec simulation with 12 unit cells (or 32 duplexes) in the simulation periodic box (Figure S14); for a total of 64 μ sec of duplex simulation.

A variety of analysis was performed to characterize the mechanical properties of DNA based on MD simulations. These include pseudo-harmonic analysis as described elsewhere³⁵⁻³⁷ (see Supplementary Methods for additional details). The calculation of polymer deformation parameters (persistence length, stretch and twist torsion modules) was done following different approaches to reduce errors associated to the use of a single method to move from atomistic simulations to macroscopic descriptors: i) extrapolation of base step translations and rotations^{17,37}, ii) analysis of the correlations in the conformations and fluctuations of the DNA at different lengths³⁸, and iii) an implementation of Olson's hybrid approach, which requires additional Monte Carlo simulations using MD-derived stiffness matrices³⁹. Dielectric constants of DNA were computed using Pettit's procedure^{40,41}.

The trajectories were analyzed using AMBER³², GROMACS³³, NAFlex³⁵, and Curves+tools⁴², as well as with in-house scripts.

Geometry annotation. We followed standard default geometrical definitions for defining the conformational regions of the different torsion angles and for annotating hydrogen bonds (see Supplementary Methods). Reference A-DNA and B-DNA fiber conformations were taken from Arnott's values⁴³.

Availability of force-field parameters and porting to different MD codes. The refined parameters are incorporated in amber-format libraries accessible from <http://mmb.irbbarcelona.org/ParmBSC1/>. Porting to GROMACS format was done from amber topology files using external utilities (amb2gmx⁴⁴ and acpype⁴⁵ tools accessible at <https://simtk.org/home/mmttools> and <https://github.com/choderalab/mmttools>). As shown in a test case in Figure S27, the use of GROMACS or AMBER (CPU or GPU versions of the code) does not introduce any significant change in the trajectories. Porting to NAMD is not required since direct reading of AMBER topology files is possible.

NMR analysis. We used MD ensembles to compute theoretical estimates of NMR observables using standard methodologies (see Supplementary Methods). Such estimates were compared with those available in BioMagResBank⁴⁶. When NMR data was determined *de novo*, the spectra were collected using default strategies summarized in Supplementary Methods.

Comparison with previous force-fields. It is out of the scope of this work to compare the performance of parmbsc1 with respect to other force-fields for all possible families of DNA. However we performed some tests for DDD, the most known B-DNA duplex, using in addition to parmbsc1, the default parmbsc0^{4,4748}, parmbsc0-OL1⁵, parmbsc0-OL4⁶, parmbsc0-OL1+OL4^{5,6}, Charmm36⁴⁹ and a modified parmbsc0 developed by mixing corrected χ values and scaled-down van der Waals interactions⁵⁰. In all cases simulations were extended for at least 1 μ s under identical simulation conditions. As shown in Table S2 parmbsc1 provides clearly the best description of the duplex, while some of the existing force-field showed non negligible artifacts (a more detailed discussion is provided in Supplementary Methods and Figures S28-S30).

The effect of ionic strength and the nature of counterion. To evaluate potential differences in simulations arising from the ionic strength we performed additionally

2 μ s simulations of DDD with extra salt: Na⁺Cl⁻ 150 mM and 500 mM. These additional calculations were performed using the same conditions outlined previously, showing results that are quite independent on the exact choice (in the 0-500 mM range) of the added extra salt (Supplementary Methods and Figure S31).

A more detailed description of the methods used for parametrization, trajectory collection and analysis, with additional references are included in Supplementary Methods.

Data Management. Trajectories and the analysis performed were placed in a novel dual database framework for nucleic acid simulations using Apache's Cassandra to manage trajectory data, and MongoDB to manage trajectory metadata and analysis. Results are available at <http://mmb.irbbarcelona.org/ParmBSC1/>. Details on the Barcelona's nucleic acids database will be presented elsewhere.

On-line Methods Table 1: Differences between QM and force-field estimates for the parameterized systems. Values refer to calculations performed in water.

Torsion	Adenosine	Guanosine	Cytosine	Thymidine
Glycosidic torsion (χ)				
<i>Geometries (degrees)</i> ^a				
Anti	14 / 40	9 / 40	2.5 / 1	2.5 / 1
Barrier	1.5 / 11	2.5 / 15	13 / 10	11 / 11
Syn	7 / 32	2.5 / 30	12 / 30	12 / 30
<i>Energies (kcal/mol)</i> ^b				
Anti/Syn	0.0 / -0.3	-0.4 / -0.6	-1.1 / 1.3	-0.8 / 1.7
Barrier ^c	0.3 / -2.0	0.0 / -2.1	-0.6 / -0.7	-0.9 / -1.2
Profile	0.3 / 2.5	1.2 / 2.8	0.9 / 4.0	0.9 / 3.9
Phase angle (P)				
<i>Geometries (degrees)</i> ^a				
North	10 / 30	10 / 10	10 / 40	0 / 10
East	0 / 10	0 / 0	10 / 10	0 / 10
South	0 / 0	10 / 10	0 / 0	0 / 0
<i>Energies (kcal/mol)</i> ^b				
North/South	-0.1 / -1.5	0.0 / -1.0	-0.6 / -1.6	0.5 / -0.5
East Barrier	-0.2 / 0.4	-0.5 / 0.7	-0.1 / 1.2	-0.8 / 0.0
Profile	0.4 / 0.6	0.5 / 0.4	0.4 / 0.7	0.2 / 0.5

^a Errors in the position of the minima and transition state when parmbc1 (first number in the cell) or parmbc0 (second number in the cell) values are compared with MP2 geometries. ^bErrors in the estimates of the relative stability and transition barrier when parmbc1 (first number in the cell) or parmbc0 (second number in the cell) values are compared with single-point CCSD(T)/CBS results. ^c Energy values refer to barrier at χ around 120 degrees, note that the large barrier located at χ around 0 is very well reproduced at the parmbc1 level, but very poorly at the parmbc0 one (Figure S24).

On-line Methods references

- [21] Krishnan, R., Binkley, J. S., Seeger, R., & Pople, J. A. *J. Chem. Phys.***72**, 650–654 (1980).
- [22] Woon, D. E. & Dunning Jr, T. H. *J. Chem. Phys.***98**, 1358–1371 (1993).

- [23] Becke, A. D. *J. Chem. Phys.***98**, 5648–5652 (1993).
- [24] Head-Gordon, M., Pople, J. A., Frisch, M. J. *Chemical Physics Letters***153**, 503–506 (1988).
- [25] Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W. & Olsen, J. *Chem. Phys. Lett.*, **302**, 437–446 (1999).
- [26] Řezáč, J., Hobza, P. *J. Chem. Theor. Comput.***9**, 2151–2155 (2013).
- [27] Miertuš, S., Scrocco, E. & Tomasi, J. *Chem. Phys.* **55**, 117–129 (1981).
- [28] Soteras, I., Curutchet, C., Bidon-Chanal, A., Orozco, M. & Luque, F. J. *J. Mol. Struct. Theochem***727**, 29–40 (2005).
- [29] Frisch, M. J. *et al.*, *Gaussian 09*, Gaussian, Inc., Wallingford, CT, USA, 2009.
- [30] Torrie, G.M., Valleau, J.P., *J. Comput. Phys.***23**, 187–199 (1977).
- [31] Case, D. A. *et al.* *Amber 14*. (2014).
- [32] Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. *J. Chem. Theory Comput.***4**, 435–447 (2008).
- [33] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. *J. Chem. Phys.***79**, 926–935 (1983).
- [34] Liu, C., Janowski, P. A. & Case, D. *Biochim. Biophys. Acta (BBA)-General Subj.* **1850**, 1059–1071 (2014).
- [35] Hospital, A. *et al.* *Nucleic Acids Res.***41**, W47–W55 (2013).
- [36] Orozco, M., Pérez, A., Noy, A. & Luque, F. J. *Chem. Soc. Rev.***32**, 350–364 (2003); and references therein
- [37] Lankaš, F., Šponer, J., Hobza, P. & Langowski, J. *J. Mol. Biol.***299**, 695–709 (2000).
- [38] Noy, A. & Golestanian, R. *Phys. Rev. Lett.* **109**, 228101 (2012).
- [39] Zheng, G., Czaplá, L., Srinivasan, A. R. & Olson, W. K. *Phys. Chem. Chem. Phys.***12**, 1399–1406 (2010).
- [40] Cuervo, A. *et al.* *Proc. Natl. Acad. Sci.***111**, E3624–E3630 (2014).

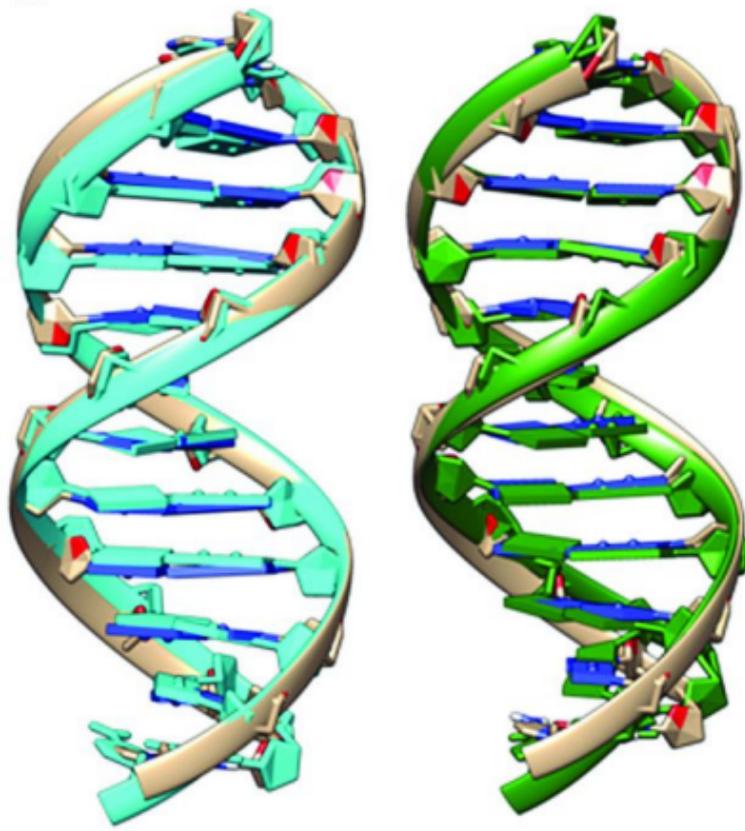
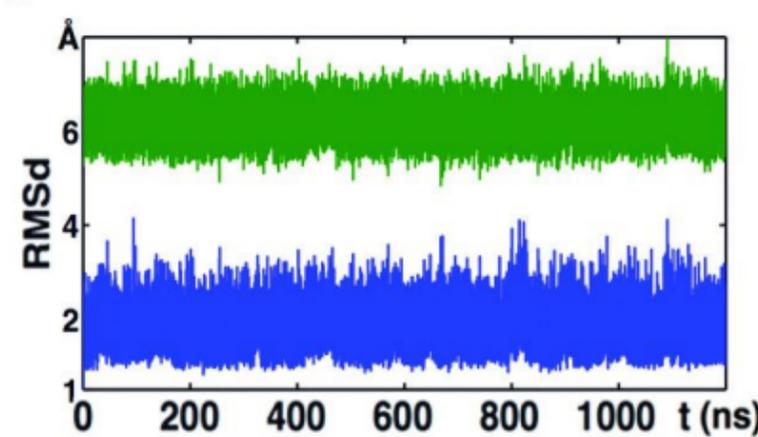
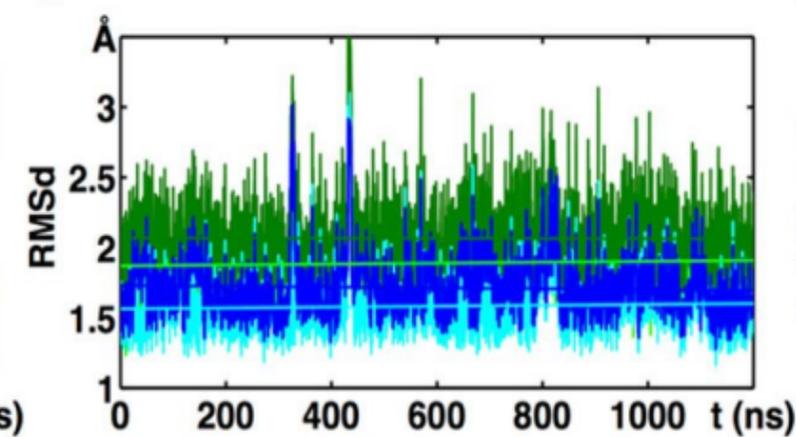
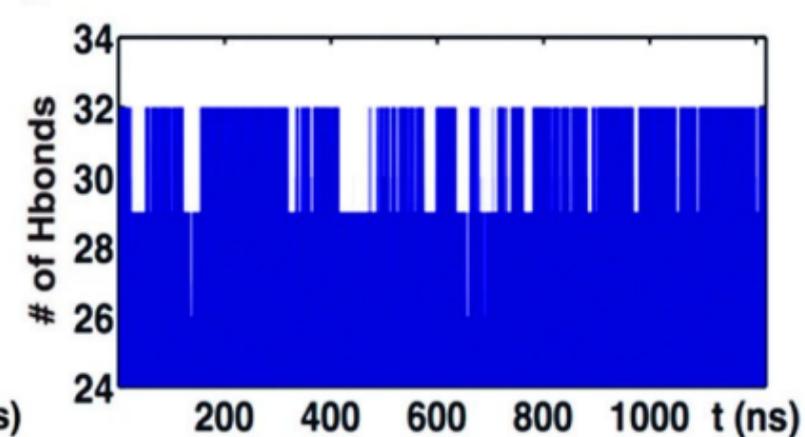
- [41] Yang, L., Weerasinghe, S., Smith, P.E. & Pettitt, P.M. *Bioph. J.***69**, 1519-1527 (1995).
- [42] Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. *Nucleic Acids Res.***37**, 5917–5929 (2009).
- [43] Arnott, S. Hukins, D.W.L. *Biochem. Biophys. Res. Comm.* **47**, 1504-1509 (1972).
- [44] Mobley, D.L., Chodera, J.D, Dill, K. A., *J. Chem. Phys.* **125**, 084902 (2006).
- [45] Sousa da Silva, A.W., Vranken, W.F., *BMC Res Notes***5**, 367 (2012).
- [46] Ulrich, E. L. *et al. Nucleic Acids Res.***36**, D402–D408 (2008).
- [47] Cornell, W. D. *et al. J. Am. Chem. Soc.***117**, 5179–5197 (1995).
- [48] Cheatham III, T. E., Cieplak, P. & Kollman, P. A. *A J. Biomol. Struct. Dyn.***16**, 845–862 (1999).
- [49] Hart, K. *et al. J. Chem. Theory Comput.***8**, 348–362 (2011).
- [50] Cheng, A.A., Garcia, A.E. *Proc. Natl. Acad. Sci. USA***110**, 16820-25 (2013).

FIGURE CAPTIONS

Figure1 |Analysis of the Drew-Dickerson dodecamer. (a) Visual comparison of MD average structure (brown) and NMR structure (PDB id: 1NAJ) (light blue) and X-ray structure (PDB id: 1BNA) (green). (b) RMSd of 1.2 μ s trajectory of DDD compared with B-DNA (blue) and A-DNA (green) form (coming from Fiber). (c) RMSd compared to experimental structures (with (dark) and without (light) ending base-pairs): X-ray (green) and NMR (blue). Linear fits of all RMSd curves are plotted on top. (d) Evolution of total number of hydrogen bonds formed between base pairs in the whole duplex. (e) Helical rotational parameters (twist, roll, and tilt) comparison of average values per base-pair step (standard deviations are shown by error bars) coming from NMR (light blue), X-ray (green), 1 μ s parmbsc0 trajectory⁷ (black) and 1.2 μ s parmbsc1 trajectory (red).

Figure2 |Analysis of non-canonical DNA structures. (a) Comparison of Z-DNA (PDB id: 1I0T) simulations in neutralized conditions (green) and in 4 M solution of Na⁺Cl⁻ (blue). Structural comparisons at given time points are shown above the RMSd curves. (b) Simulation of anti-parallel H-DNA (PDB id: 2AF1) showing deviation of the structure over time (highlighted in red). RMSd of (c) parallel d(T-A•T)₁₀, (d) parallel d(G-G•C)₁₀, and (e) antiparallel d(G-G•C)₁₀ triplexes. (f) Parallel (PDB id: 352D) and (g) anti-parallel (PDB id: 156D) quadruplex showing stable structures over time. (h) Structural stability of d(GCGAAGC) hairpin (PDB id: 1PQT) and (i) OxyQ quadruplex (PDB id: 1JRN) with ions, over time. (j) Human Telomeric Quadruplex (PDB id: 1KF1) with highlighted loops. RMSd of HTQ backbone, loop 1, loop 2 and loop 3 regions are shown below. In all panels, parmbsc1 (final, averaged or at a given trajectory point) structures (light blue; also green for Z-DNA) are overlapped over experimental structure (grey) for comparison. See Table S1 for information on the PDB structures.

Figure3 |Analysis of DNA-protein complexes. Structural details of microsecond trajectories of four complexes with PDB id: 1TRO (**a**), 2DGC (**b**), 3JXC (**c**) and 1KX5 (**d**) (500 ns trajectory). Each plot shows overlap of the MD starting (red) and final (blue) structures, time dependent mass-weighted root mean square deviation (RMSD in Å) of all DNA (red) and protein (cyan) heavy atoms, and comparison of the values of rotational helical parameter roll (in degrees) at each base pair step calculated from the X-ray crystal structure (cyan) and averaged along the MD simulation (red line with the standard deviation envelope in light red). For clarity, in the 1KX5 plot of the roll value, the base pair steps are defined by the number of the position along the DNA strand and not by the base pair step name.

a**NMR****X-ray****b****c****d****e**