
Article:
da Rosa, Gabriela, Grille, Leandro, Calzada, Victoria et al. (23 more authors) (2021) Sequence-dependent structural properties of B-DNA:what have we learned in 40 years? Biophysical Reviews. pp. 995-1005. ISSN 1867-2450

https://doi.org/10.1007/s12551-021-00893-8

Sequence-Dependent Structural Properties of B-DNA: What Have We Learned in 40 Years?

Gabriela da Rosa[1,2, ⋔,§], Leandro Grille[1,2, §], Victoria Calzada[3, §], Katya Ahmad[4,⋔,ⵝ], Juan Pablo Arcon[5,⋔,ⵝ], Federica Battistini[5,⋔,ⵝ], Genís Bayarri[5,⋔,ⵝ], Thomas Bishop[6,⋔,ⵝ], Paolo Carloni[4,⋔,ⵝ], Thomas Cheatham III[7,⋔,ⵝ], Rosana Collepardo-Guevara[8,⋔,ⵝ], Jacek Czub[9,⋔,ⵝ], Jorge R. Espinosa[8,⋔,ⵝ], Rodrigo Galindo-Murillo[7,⋔,ⵝ], Sarah A. Harris[10,⋔,ⵝ], Adam Hospital[5,⋔,ⵝ], Charles Laughton[11,⋔,ⵝ], John Maddocks[12,⋔,ⵝ], Agnes Noy[13,⋔,ⵝ], Modesto Orozco[5,⋔,ⵝ], Marco Pasi[14,⋔,ⵝ], Alberto Perez[15,⋔,ⵝ], Daiva Petkevičiūté[16,⋔,ⵝ], Rahul Sharma[12,⋔,ⵝ], Ran Sun[6,⋔,ⵝ] and Pablo D. Dans[1,2,5, ⋔,*]

[1]Computational Biophysics Group, Department of Biological Sciences, CENUR Litoral Norte, Universidad de la República, Salto, Uruguay. [2]Functional Genomics Laboratory, Institut Pasteur de Montevideo, Montevideo, Uruguay. [3]Radiopharmacy Department, Centro de Investigaciones Nucleares, Universidad de la República, Montevideo, Uruguay. [4]Computational Biomedicine, Forschungszentrum Jülich, Jülich, Germany. [5]Molecular Modelling and Bioinformatics group, Institute for Research in Biomedicine, Barcelona, Spain. [6]Theoretical Molecular Biology Lab, Louisiana Tech University, Ruston LA, US. [7]College of Pharmacy, University of Utah, Salt Lake City UT, US. [8]Computational Physics of Chromatin, Department of Physics, University of Cambridge, Cambridge, UK. [9]Deparment of Physical Chemistry, Gdansk University of Technology, Gdansk, Poland. [10]Engineering and Physical Sciences, University of Leeds, leeds, UK. [11]School of Pharmacy, University of Nottingham, Nottingham, UK. [12]Laboratory for Computation and Visualization in Mathematics and Mechanics, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [13]Deparment of Physics, University of York, York, UK. [14]Laboratory of Biology and Applied Pharmacology, University Paris-Saclay, Cachan, France. [15]Department of Chemistry, University of Florida, Gainesville FL, US. [16]Nonlinear Systems Mathematical Research Centre, Kaunas University of Technology, Kaunas, Lithuania.

[⋔]Ascona B-DNA Consortium members. [§]Equally contributed and should be considered co-first authors. [ⵝ]Alphabetical order.

[*]To whom correspondence should be addressed: Pablo D. Dans, e-mail: pdans@pasteur.edu.uy

Keywords

Structural polymorphisms, helical parameters, Drew-Dickerson Dodecamer, Ascona B-DNA Consortium, nearest neighbours.

Abstract

The structure of B-DNA, the physiological form of the DNA molecule, has been a central topic in biology, chemistry, and physics. Far from uniform and rigid, the double helix was revealed as a flexible and structurally polymorphic molecule. Conformational changes that lead to local and global changes in the helix geometry are mediated by a complex choreography of base and backbone rearrangements affecting the ability of the B-DNA to recognize ligands, and consequently on its functionality. In this sense, the knowledge obtained from the sequence-dependent structural properties of B-DNA has always been thought crucial to rationalize how

ligands and, most notably proteins, recognize B-DNA and modulate its activity, *i.e*. the structural basis of gene regulation. Honouring the anniversary of the first high-resolution X-ray structure of a B-DNA molecule, in this contribution, we present the most important discoveries of the last 40 years on the sequence-dependent structural and dynamical properties of B-DNA; from the early beginnings to the current frontiers in the field.

The early beginnings

The Drew-Dickerson first B-DNA crystal structure. After the Watson and Crick double helix theoretical model for B-DNA was described (Watson and Crick 1953), the direct evidence for its structure was based on "ideal" representations using standard bond parameters obtained from low-resolution X-ray diffraction data of crystalline fibres (Langridge et al. 1960; Fuller et al. 1965; Arnott and Hukins 1972; Arnott et al. 1980). It took 28 years from Watson-Crick discovery and significant advances in the DNA synthesis field and X-ray diffraction techniques, to make possible the first B-DNA crystal structure at atomic resolution (Drew et al. 1981). In 1981, Dickerson's group described the conformation and dynamics of a complete turn of right-handed B-DNA with the sequence d(CpGpCpGpApApTpTpCpGpCpG) (Fig. 1a). To date, this famous sequence known as the Drew-Dickerson Dodecamer (DDD) has been determined experimentally hundreds of times. The non-uniform motion of bases and phosphate groups in the helix, the description of some local helical parameters, torsion angles of the backbone, and sugar conformations, led to the foundations 40 years ago (Drew et al. 1981) for the study of the sequence-dependent structural properties of B-DNA.

Calladine rules and Zhurkin's contributions. Based on the findings from Dickerson's group, the Journal of Molecular Biology published in 1982 a letter sent by Calladine (Calladine 1982), suggesting that the specific B conformation observed in the DDD structure was a consequence of steric repulsive forces between purine bases in consecutive basepairs on opposite strands (Fig. 1b). With simple steric considerations and geometric constraints based on a few structural parameters at the base (propeller twist) and basepair (roll and slide) level, a set of rules to describe the 3D structure of B-DNA from sequence was proposed (Calladine 1982; Fratini et al. 1982). One year later, Calladine rules were extended by including additional aspects of the overall helix geometry, sequence-structure relationships, and the effect of hydration on DNA stability (Fig. 1c) (Dickerson 1983; Calladine et al. 2004). In parallel to these initial findings, Zhurkin and colleagues were also pioneer in pointing to differences on DNA's structure/deformability associated with sequence effects (Zhurkin et al. 1979). In those early times, the studies on the sequence-dependent anisotropic flexibility of B-DNA were even able to explain the structural behaviour of a few selected tetranucleotides (Ulyanov and Zhurkin 1984a; Ulyanov and Zhurkin 1984b).

Lavery-Sklenar generalized helical parameters. The increasing need in the community for tools able to standardize the structural analysis of nucleic acids led to the definition of a set of generalized helical parameters and an axis of curvature (Lavery and Sklenar 1988). From the description of the overall DNA conformation to the way of distributing its sequence "irregularities", Lavery and Sklenar presented

a unique algorithm solving the problem of obtaining a rigorous helical description for any given segment. The associated computer program termed *Curves* (Lavery and Sklenar 1988), nowadays *Curves+* (Lavery et al. 2009; Blanchet et al. 2011), was the first algorithm of its kind and is still one of the most used tools in the field. A few years later, Lu and Olson published another comprehensive software package termed *3DNA* (Lu and Olson 2003).

Cambridge and Tsukuba conventions. One year after the definitions of Lavery and Sklenar were implemented in *Curves*, an EMBO workshop was held in Cambridge to reach international consensus in key concepts and define a common structural nomenclature for base, basepair and basepair step parameters in the helical space (Fig. 1d) (Dickerson 1989). The standardization of these parameters was revisited in 1999 at the Tsukuba workshop held in Japan (Olson et al. 2001).

The nearest neighbour approximation

DNA sequence-dependent deformability from crystal structures. In a seminal work, Olson and coworkers studied the DNA deformability associated with the identity of adjacent basepairs from a set of 92 DNA-Protein complexes resolved by X-ray diffraction (Olson et al. 1998). The behaviour of DNA in basepair steps was interrogated by analyzing six helical parameters that affect the planes of successive basepairs: the twist, roll and tilt rotations and the shift, slide and rise translations (Fig. 1d). Their work marked the beginning of several systematic studies focused on uncovering sequence effects in the helical space. At that time, due to the scarcity of structures deposited in the PDB, Olson's main assumption was that different proteins and different environments impose different forces on the DNA, so if averaged over a large number of DNA-protein complexes and environments, the external-induced distortions would cancel each other and the conformational response of DNA would emerge. In this reduced helical space, the deformation energy of the 10 unique basepair steps was expressed for the first time using a harmonic approximation. Among other findings (Fig. 2a), the pyrimidine-purine (YR) steps emerged as the more flexible ones, acting as "hinges" at the Protein-DNA interface. Some key correlations and conformational sub-states interdependences at the dinucleotide level were also described in Olson's PNAS article, leading a few years later to an initial "theory of sequence-dependent DNA elasticity" (Coleman et al. 2003). Also in 1998 and using a similar approach based on the analysis of X-ray crystal structures, Packer and Hunter provided a comprehensive view of the coupling between the conformational properties of the sugar-phosphate backbone and the basepair step parameters (Packer and Hunter 1998). The knowledge gained was then used to build a simple model to compute potential energy surface maps for dinucleotide and tetranucleotide steps as a function of slide and shift (Packer et al. 2000a; Packer et al. 2000b).

DNA sequence-dependent flexibility from MD simulations. Since the experimental data on sequence-dependent DNA deformability was rather limited 20 years ago, Lankas, Pérez, Fujii and their collaborators were among the first to investigate the sequence-dependent flexibility of DNA using Molecular Dynamics (MD) simulations (Lankas et al. 2000; Lankas et al. 2003; Fujii et al. 2007; Pérez et al. 2008). Inspired by Olson and Hunter approach to analysis the helical space of

DNA, a library of sequences containing all unique basepair steps was designed and simulated. Significant sequence-dependence on the torsional stiffness and stretching of DNA segments was observed (Lankas et al. 2000; Lankas et al. 2003), leading nevertheless to a simpler description of the conformational landscape than the one previously reported using crystallographic data (Olson et al. 1998), all assuming a rigid basepair approximation with two, uncoupled, nearest neighbour, interactions. In parallel, and rooted in the same ideas, Gonzalez and Maddocks introduced a rigorous method to compute expectations from MD data for models with rigid units, while properly accounting for the statistical mechanics of the rotational degrees of freedom (Gonzalez and Maddocks 2001). A few years later these ideas were implemented by Lankas and coworkers (Lankas et al. 2009), suggesting that a rigid base approximation, with its associated five nearest neighbour interactions, was a remarkably good fit to the MD data, thereby providing detailed, quantitative support for the earlier ideas of Calladine on base interactions. The quality of the fit of the rigid base, nearest neighbour approximation to MD data, in turn lead to the predictive, sequence-dependent, rigid base, coarse-grained *cgDNA* model which can efficiently predict Gaussian statistics (*i.e.* first and second moments) for double-stranded B-DNA fragments of arbitrary length and sequence (Gonzalez et al. 2013, Petkevičiūtė et al. 2014). In 2007, Sarai's group was among the first to provide a global view of the conformational similarities of all steps in their tetranucleotide environment (Fujii et al. 2007). One year later, Orozco's group published a comparative study showing that MD simulations performed with the *Parmbsc0* force field (Pérez et al. 2007) were able to predict the dynamic and sequence-dependent physical properties of DNA with an accuracy, in terms of local and global descriptors, reasonably close to that obtained with gold-standard NMR and X-ray experiments (Pérez et al. 2008), providing a first consensus view between theory and experiments (Fig. 2b).

The Ascona B-DNA Consortium (ABC). In June 2001, during an informal discussion in a conference on *Atomistic to Continuum Models for Long Molecules* organized by Maddocks and held in Ascona, Switzerland, began an initiative that involved 17 researchers from nine independent laboratories (Beveridge et al. 2012). The original objective of the ABC project was to produce a database of state-of-the-art MD of DNA in which all the tetranucleotides were represented, developing standards and simulation protocols, and obtaining information for the comprehensive study and improved understanding of basepair sequence effects on structure and dynamics. This collaborative and massive simulation effort allowed for a systematic view, highlighting the sequence-induced effects produced by the nearest neighbours (NN) of the d(CpG) basepair step (Beveridge et al. 2004), and later of the 10 unique dinucleotides (Lavery et al. 2010). Using the trajectories made available by the ABC in 2010, a rigid basepair coarse-grained model derived elastic force constants to build an interactive chromatin modeling web server (Stolz and Bishop 2010), while other colleagues focused on the detailed analysis of the sequence-dependent solvation and ion atmosphere of B-DNA (Dixit et al. 2012).

Challenges to the NN approximation. The ABC simulations, based on a library of 39 duplexes containing multiple copies of the 136 unique tetranucleotides, revealed that some helical parameters at the basepair step level were greatly affected by their nearest neighbours and that certain of them did not exhibit a normal distribution

(which is an evidence for the existence of a unique stable conformational sub-state for that specific degree of freedom), but rather showed a bimodal behaviour (Fig. 2c) (Lavery et al. 2010). Furthermore, this bimodality seemed to be intrinsic to some dinucleotide steps in certain tetranucleotide environments, and was particularly notable in the case of d(CpG) steps, where this structural polymorphism emerged from both theoretical and experimental evidence (Fig. 2d) (Dans et al. 2012). These results established the limitations of Olson's original approach proposed more than 10 years earlier based on dinucleotides, and even challenged the current harmonic-based NN models by introducing the existence of bimodal degrees of freedom.

Structural polymorphisms at the base and backbone level

Experimental evidence from high-resolution X-ray structures. It was soon warned that the experimental evidence supporting the existence of structural polymorphisms had been obtained indirectly, by analyzing DNA structures deposited in the PDB (Dans et al. 2012; Pérez et al. 2012). This bioinformatics approach, despite being valid, did not prove the simultaneous existence of the two detected sub-states nor the reversible transitions observed between them in the simulations. In this context, the high-resolution crystal structure published by Hud and Williams groups from Georgia Tech in 2012 (that same year, PDB id 3GGB) was particularly relevant (Maehigashi et al. 2012). It was the first observation by X-ray diffraction of full positional heterogeneity in DNA at the base level and hence, the first direct evidence supporting previous theoretical findings.

The paradigmatic d(CpG) step. A detailed study of the twist degree of freedom for the most polymorphic step, d(CpG), allowed for an in-depth understanding of the structural transitions coupled with the observed 'low twist' (LT) and a 'high twist' (HT) sub-states, and the dramatic influence exerted by the nearest neighbours (Dans et al. 2014). It was shown that the twist polymorphism was coupled with transitions at the backbone level leading to BI/BII conformations at the 3'-side junction of the d(CpG) step, where the canonical HT sub-state is associated with BI, while the LT state correlates with BII. The BI/BII sub-states were defined following the established conventions based on the value of $\varepsilon - \zeta$ torsions (Fig. 3a) (Hartmann et al. 1993), but measured at the 3'-side of the d(CpG) step *i.e*. in the next backbone junction on each strand (Fig. 3b). In this context, RCGY flanking sequences tended to favour almost exclusively the canonical conformation (HT/BI), YCGR mostly favoured the non-canonical sub-state (LT/BII), while RCGR sequences could sample both. The spontaneous twist polymorphism of the d(CpG) step also provided some clues about its extreme prevalence in intercalation complexes (Boer et al. 2009; Dans et al. 2014).

The significant stabilization of the LT/BII sub-state observed for YCGR tetranucleotides was connected to the capacity of these sequences to form non-conventional hydrogen bonds (Hbond) at the 3'-sides (on both strands) of the d(CpG) step (Wahl and Sundaralingam 1997). These Hbonds, which are established by the interaction of the C8H8 group of the R base and the O3* atom of the backbone at the 3'-side (*i.e.* at the level of the following d(GpR) step), were confirmed for all d(RpR) tetranucleotides (Pasi et al. 2014) and later extended to all d(RpY) steps between the same oxygen atom in the backbone and the corresponding C6H6 group

of pyrimidines (Balaceanu et al. 2017). Another key finding was the relationship determined between these coupled sub-states and the sequence-dependent dynamic presence of cations in the minor groove of the d(CpG) step. The presence of Na^+/K^+ was exclusively observed in the minor groove during the LT/BII conformation, following the same sequence-dependent trends described, and thus, reaching the highest concentration for the most polymorphic tetranucleotide: TCGA. This complex choreography of conformational changes was shown to be connected by a cause-and-effect relationship, where the entrance of cations in the minor groove of the d(CpG) step triggers the transition from HT/BI/no-CHO Hbond to LT/BII/CHO Hbond in a few picoseconds (Fig. 4a) (Dans et al. 2014). A largely partial, but direct experimental validation of the unequivocal relationship between the backbone polymorphism and the interaction with cations was provided once again by high-resolution crystals of B-DNA (PDB id 1D8G and 1EN9) (Kielkopf et al. 2000; Chui and Dickerson 2000; Hud and Engelhart 2009). Two years later, this molecular choreography was described again for the GCGA tetranucleotide, by analyzing a multi-microsecond long simulation of the DDD (Dans et al. 2016) performed with the latest *Parmbsc1* force field for DNA (Ivani et al. 2015). The corrections introduced in the *Parmbsc1* force field led to the first sequence-dependent consensus view between MD simulations (Balaceanu et al. 2017) and a vast amount of information on BI/BII structural heterogeneity that was obtained using NMR (Nuclear Magnetic Resonance) techniques. Hartmann's group has been particularly active in this field, providing a sequence-dependent view of BI/BII at the dinucleotide and even tetranucleotide level (Heddi et al. 2006; Abi-Ghanem et al. 2010; Heddi et al. 2008; Heddi et al. 2010) but also putting into perspective NMR results with values obtained from crystal structures or MD simulations (Imeddourene et al. 2015).

A systematic study of tetranucleotide sequence effects. In 2014, after a new ABC collaborative round termed μABC, the detailed study of the complete conformational space at the nearest neighbours level including all structural polymorphisms at the backbone, sugar and base level was published (Pasi et al. 2014). The co-existence of more than one stable conformational sub-state was confirmed in RR steps for shift and twist degrees of freedom, while YR steps displayed bimodal distributions for slide and twist. RY steps appeared to be unimodal for all helical parameters, adopting, in the case of the three polymorphic parameters, typical canonical values (Fig. 4b). The relationship between HT/LT sub-states and BI/BII transitions at the backbone level was also confirmed for all tetranucleotide contexts, highlighting complex sequence-dependent patterns of correlated and anti-correlated movements between base stacking and the backbone conformations, in perfect agreement with early surveys of high-resolution X-ray structures and solid-state ^{31}P-NMR experiments (Gorin et al. 1995; van Dam and Levitt 2000; Tolstorukov et al. 2007; Olson and Zhurkin 2011). The coupling between backbone conformations and LT↔HT transitions was revisited one year later by Zgarbova and coworkers (Zgarbova et al. 2017) using a library of longer B-DNA oligomers and the refined *OL15* force field (Zgarbova et al. 2015). All six nearest phosphates were found to influence the twist of the central step and when an LT/BII sub-state emerges at a certain place, it generates twist changes in its neighbourhoods that tend to compensate locally the produced under-twisting. All this conformational information extracted from MD was integrated along with X-ray data and then used

to predict, based on DNA shape features, binding specificities of transcription factors at the genome-scale (Li et al. 2017). Other colleagues used the μABC trajectories to analyze ion distributions around DNA (Pasi et al. 2015) using an innovative Curvilinear Helicoidal Coordinate (Lavery et al. 2014; Dans et al. 2014).

One key aspect of these transitions is the timescale on which they take place. B-DNA breathing (fast, spontaneous and reversible opening/closing of WC Hbonds), transitions between backbone sub-states, sugar re-puckering, moderate bending and sampling of a wide variety of helical values are fast motions occurring clearly below the μs timescale (Dans et al. 2014; Galindo-Murillo et al. 2014). Beyond the μs timescale, with the exception of terminal basepair fraying (opening of the terminal WC basepairs) and rare sequence-specific kink events (Dans et al. 2019), these faster movements are averaged out and no other motions are actually expected for "naked" B-DNA up to the ms timescale (Galindo-Murillo et al. 2014).

Extending Calladine-Dickerson rules. Integrating, unifying, and putting in perspective all findings of the last decade, made possible for a new and comprehensive extension of Calladine–Dickerson rules (Dans et al. 2019). By defining a small library of B-DNA sequences, termed miniABC, all tetranucleotide contexts were analyzed in different conditions employing microsecond-long MD simulations with *Parmbsc1*. A systematic analysis of 16 helical parameters and the associated backbone/sugar torsional changes was conducted, which included the first prediction of anharmonicity based on sequence context (Fig. 4c). These extended rules were implemented in a web server that predicts the average conformation of any B-DNA sequence, highlighting those steps where structural polymorphisms could occur (Fig. 4c). The knowledge gained and the extensive volume of raw trajectories made available to the community, were then used to develop state-of-the-art coarse-grained models of B-DNA based on rigid bases (De Bruin and Maddocks 2018), rigid basepairs coupled to a multi-modal Hamiltonian and a Metropolis Monte Carlo algorithm (Walther et al. 2020), and a multiscale model of chromatin describing B-DNA at the basepair level and proteins at single amino acid resolution (Farr et al. 2021).

Beyond nearest neighbours?

As discussed above, the structural and dynamical study of all the unique tetranucleotides requires, to reach a complete view, the analysis of 136 sequence combinations. Pushing the conformational analysis of B-DNA statistical mechanical properties to the next level would imply the systematic study of 2,080 unique hexanucleotide sequences, which could be used to detect specific next-to-nearest neighbour (NNN) sequence effects. In the same way, the analysis of possible "long-range" structural correlations at the octanucleotide level would require the simulation of 32,826 unique sequence combinations. The huge number of sequences that must be considered to move the paradigm to the next stage can hardly be covered by a single research group and is even challenging for a whole consortium. Nevertheless, a partial view can be obtained by carefully selecting a given sub-set of representative sequences from the complete sequence space. Using this approach, Balaceanu and coworkers studied using MD simulations the NNN sequence effects affecting the d(TpA) step when embedded in an NCTAGN environment detecting the

existence of multiple simultaneous sub-states (Fig. 4d) (Balaceanu et al. 2019). The results seemed to confirm the existence of complex and correlated patterns through consecutive steps at the NNN level and even beyond (Fig. 4d), capable to modulate on both strands the relative populations of BI/BII sub-states observed at the central d(TpA). This "proof-of-principle" study showed that there is indeed interesting "long-range" structural and dynamical correlations to be discovered, therefore making the grand challenge project to explore it all a worthwhile endeavour.

Closing remarks

40 years after the first B-DNA crystal structure and 20 years since the beginning of the Ascona B-DNA Consortium, seems to be the perfect time to celebrate and embark on the journey through the major discoveries on the sequence-dependent structural effects in B-DNA that we covered in this review. From an "ideal" B-DNA model to the description of subtle conformational sub-states at the base, sugar and backbone level, the gained knowledge has influenced several fields. The new initiative from the ABC, aimed to unravel the sequence-dependent physical properties of B-DNA at the complete NNN level (https://bit.ly/3ywgGV6) and the work of many other colleagues from the field, herald new and exciting chapters in the days to come.

Acknowledgements and funding

Authors want to express their gratitude to the International Union for Pure and Applied Biophysics (IUPAB) for the invitation to write this review and to Prof. Daniel Peluffo (UdelaR, UY) for suggesting our names to the IUPAB council. This work was funded by the Agencia Nacional de Investigación e Innovación (ANII, Uruguay) [FCE_3_2018_1_748945, to PDD], the Comisión Sectorial de Investigación Científica (CSIC, UdelaR, Uruguay), and the Programa de Desarrollo de las Ciencias Básicas (PEDECIBA, UdelaR, Uruguay). This work was supported by FOCEM (Fondo para la Convergencia Estructural del Mercosur) [COF 03/11 to IPMONT]. VC and PDD are SNI (Sistema Nacional de Investigadores, ANII, Uruguay) and PEDECIBA Química researchers.

Conflict of interest statement

None declared.

References

Abi-Ghanem J, Heddi B, Foloppe N, Hartmann B (2010). DNA structures from phosphate chemical shifts, *Nucleic Acids Res*, 38(3):e18. https://doi.org/10.1093/nar/gkp1061

Arnott S, Hukins DWL (1972). Optimized parameters for A-DNA and B-DNA. *Biochem Biophys Res Commun 47*(6):1504–1509. https://doi.org/10.1016/0006-291x(72)90243-4

Arnott S, Chandrasekaran R, Birdsall DL, Leslie AG, Ratliff RL (1980). Left-handed DNA helices. *Nature* 283:743–745. https://doi.org/10.1038/283743a0

Balaceanu A, Pasi M, Dans PD, Hospital A, Lavery R, Orozco M (2017). The Role of

Unconventional Hydrogen Bonds in Determining BII Propensities in B-DNA. *Phys Chem Lett 8*(1):21–28. https://doi.org/10.1021/acs.jpclett.6b02451
Balaceanu A, Buitrago D, Walther J, Hospital A, Dans PD, Orozco M (2019). Modulation of the helical properties of DNA: Next-to-nearest neighbour effects and beyond. *Nucleic Acids Res 47*(9):4418–4430. https://doi.org/10.1093/nar/gkz255
Blanchet C, Pasi M, Zakrzewska K, Lavery R (2011). CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res 39*(2):W68-W73. https://doi.org/10.1093/nar/gkr316
Beveridge DL, Barreiro G, Byun KS, Case DA, Cheatham TE, Dixit SB, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Seibert E, Sklenar H, Stoll G, Thayer KM, Varnai P, Young, MA (2004). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J 87*(6):3799–3813. https://doi.org/10.1529/biophysj.104.045252
Beveridge DL, Cheatham TE, Mezei M (2012). The ABCs of molecular dynamics simulations on B-DNA, circa 2012. *J Biosci 37*(3):379–397. https://doi.org/10.1007/s12038-012-9222-6
Boer DR, Canals A, Coll M (2009). DNA-binding drugs caught in action: The latest 3D pictures of drug-DNA complexes. *Dalton Trans* 21(3):399–414. https://doi.org/10.1039/b809873p
Calladine CR (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. *J Mol Biol 161*(2):343–352. https://doi.org/10.1016/0022-2836(82)90157-7
Calladine CR, Drew H, Luisi B, Travers A (2004). Understanding DNA: The Molecule & How It Works. 3rd edition. Elsevier Academic Press. San Diego, CA, US.
Chiu TK, Dickerson RE (2000). 1 A crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J Mol Biol*, *301*(4), 915-45. https://doi.org/ 10.1006/jmbi.2000.4012.
Coleman BD, Olson WK, Swigon D (2003). Theory of sequence-dependent DNA elasticity. *J Chem Phys* 118(15):7127–7140. https://doi.org/10.1063/1.1559690
Dans PD, Pérez A, Faustino I, Lavery R, Orozco M (2012). Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res 40*(21):10668–10678. https://doi.org/10.1093/nar/gks884
Dans PD, Faustino I, Battistini F, Zakrzewska K, Lavery R, Orozco M (2014). Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res 42*(18):11304–11320. https://doi.org/10.1093/nar/gku809
Dans PD, Danilane L, Ivani I, Drsata T, Lankas F, Hospital A, Walther J, Pujagut RI, Battistini F, Gelpí JL, Lavery R, Orozco M (2016). Long-timescale dynamics of the Drew-Dickerson dodecamer. *Nucleic Acids Res 44*(9):4052–4066. https://doi.org/10.1093/nar/gkw264
Dans PD, Balaceanu A, Pasi M, Patelli AS, Petkevičiūtė D, Walther J, Hospital A, Bayarri G, Lavery R, Maddocks JH, Orozco M (2019). The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. *Nucleic Acids Res 47*(21):11090–11102. https://doi.org/10.1093/nar/gkz905
De Bruin L, Maddocks JH (2018). cgDNAweb: a web interface to the cgDNA

sequence-dependent coarse-grain model of double-stranded DNA. *Nucleic Acids Res 46*:W5–W10. https://doi.org/10.1093/nar/gky351
Dickerson RE (1983). Base sequence and helix structure variation in B and A DNA. *J Mol Biol 166*:419–441. https://doi.org/10.1016/S0022-2836(83)80093-X
Dickerson RE (1989). Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res*, *17*, 1797-1803. https://doi.org/ 10.1093/nar/17.5.1797
Dixit SB, Mezei M, Beveridge DL (2012). Studies of base pair sequence effects on DNA solvation based on all-atom molecular dynamics simulations. *J Biosci 37*(3):399-421. https://doi.org/10.1007/s12038-012-9223-5
Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci USA 78*:2179–2183. https://doi.org/10.1073/pnas.78.4.2179
Fratini AV, Kopka ML, Drew HR, Dickerson RE (1982). Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATT(Br)CGCG. *J Biol Chem 257*(24):14686–14707. https://doi.org/10.1016/s0021-9258(18)33335-0
Farr SE, Woods EJ, Joseph JA, Garaizar A, Collepardo-Guevara R (2021). Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions. *Nat Commun 12*:2883. https://doi.org/10.1038/s41467-021-23090-3
Fujii S, Kono H, Takenaka S, Go N, Sarai A (2007). Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res* 35(18):6063-6074. https://doi.org/10.1093/nar/gkm627
Fuller W, Wilkins MHF, Wilson HR, Hamilton LD, Arnott S (1965). The Molecular Configuration of Deoxyribonucleic Acid IV. X-ray Diffraction Study of the A Form. *J Mol Biol 12*(1):60-80. https://doi.org/10.1016/s0022-2836(65)80282-0
Galindo-Murillo R, Roe D, Cheatham TE (2014). On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nat Commun 5*:5152. https://doi.org/10.1038/ncomms6152
Gonzalez O, Maddocks JH (2001). Extracting parameters for base-pair level models of DNA from molecular dynamics simulations. *Theor Chem Acc 106*:76–82. https://doi.org/10.1007/s002140100256
Gonzalez O, Petkevičiūtė D, Maddocks JH (2013). A sequence-dependent rigid-base model of DNA. *J Chem Phys 138*(5):055122:1-28. https://doi.org/10.1063/1.4789411
Gorin AA, Zhurkin VB, Olson WK (1995). B-DNA twisting correlates with base-pair morphology. *J Mol Biol* 247(1):34-48. https://doi.org/10.1006/jmbi.1994.0120.
Hartmann B, Piazzola D, Lavery R (1993). BI-BII transitions in BDNA. *Nucleic Acids Res 21*(3):561–568. https//doi.org/ 10.1093/nar/21.3.561
Heddi B, Foloppe N, Bouchemal N, Hantz E, Hartmann B (2006). Quantification of DNA BI/BII Backbone States in Solution. Implications for DNA Overall Structure and Recognition. *J Am Chem Soc 128* (28):9170-9177. https://doi.org/10.1021/ja061686j
Heddi B, Foloppe N, Oguey C, Hartmann B (2008). Importance of Accurate DNA Structures in Solution: The Jun–Fos Model. *J Mol Biol* 382(4):956-970. https://doi.org/10.1016/j.jmb.2008.07.047
Heddi B, Oguey C, Lavelle C, Foloppe N, Hartmann B (2010). Intrinsic flexibility of

B-DNA: the experimental TRX scale. *Nucleic Acids Res* 38(3):1034-1047. https://doi.org/ 10.1093/nar/gkp962

Hud NV, Engelhart AE (2009). Sequence-specific DNA–Metal Ion Interactions. In: Hud NV (ed) Nucleic Acid–Metal Ion Interactions, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA pp 75-112.

Imeddourene AB, Elbahnsi A, Guéroult M, Oguey C, Foloppe N, Hartmann B (2015). Simulations Meet Experiment to Reveal New Insights into DNA Intrinsic Mechanics. *PLoS Comput Biol* 11(12):e1004631. https://doi.org/10.1371/journal.pcbi.1004631

Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, Walther J, Andrio P, Goñi R, Balaceanu A, Portella G, Battistini F, Gelpí JL, González C, Vendruscolo M, Laughton CA, Harris SA, Case DA, Orozco M (2015). Parmbsc1: A refined force field for DNA simulations. *Nat Methods 13*(1):55–58. https://doi.org/10.1038/nmeth.3658

Kielkopf CL, Ding S, Kuhn P, Rees DC (2000). Conformational flexibility of B-DNA at 0.74 Å resolution: d(CCAGTACTGG)2. *J Mol Biol 296*(3):787–801. https://doi.org/10.1006/jmbi.1999.3478

Langridge R, Wilson HR, Hooper CW, Wilkins MHF, Hamilton LD (1960). The molecular configuration of deoxyribonucleic acid: I. X-ray diffraction study of a crystalline form of the lithium salt. *J Mol Biol 2*(1):19-37. https://doi.org/10.1016/S0022-2836(60)80004-6

Lankas F, Šponer J, Hobza P, Langowski J (2000). Sequence-dependent elastic properties of DNA. *J Mol Biol 299*(3):695–709. https://doi.org/10.1006/jmbi.2000.3781

Lankas F, Šponer J, Langowski J, Cheatham TE (2003). DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations. *Biophys J 85*(5):2872–2883. https://doi.org/10.1016/S0006-3495(03)74710-9

Lankas F, Gonzalez O, Heffler LM, Stoll G, Moakher M, Maddocks JH (2009). On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations. *Phys Chem Chem Phys*, *11*:10565-10588. https://doi.org/10.1039/b919565n

Lavery R, Sklenar H (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn 6*(1):63–091. https://doi.org/10.1080/07391102.1988.10506483

Lavery R, Moakher M, Maddocks JH, Petkevičiūtė D, Zakrzewska K (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res* 37:5917-5929. https//doi.org/10.1093/nar/gkp608

Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham TE, Dixit S, Jayaram B, Lankas F, Laughton C, Maddocks JH, Michon A, Osman R, Orozco M, Perez A, Singh T, Spackova N, Sponer J (2010). A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in in B-DNA. *Nucleic Acids Res* 38:299–313. https//doi.org/10.1093/nar/gkp834.

Lavery R, Maddocks JH, Pasi M, Zakrzewska K (2014). Analyzing ion distributions around DNA. *Nucleic Acids Res 42*(12):8138-8149. https://doi.org/10.1093/nar/gku504

Li J, Sagendorf JM, Chiu T-P, Pasi M, Pérez A, Rohs R (2017). Expanding the repertoire of DNA shape features for genome-scale studies of transcription factor binding. *Nucleic Acids Res 45*(22):12877-12887.

https://doi.org/10.1093/nar/gkx1145
Lu X-J, Olson WK (2003). 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res 31*(17): 5108–5121. https://doi.org/10.1093/nar/gkg680
Maehigashi T, Hsiao C, Kruger Woods K, Moulaei T, Hud NV, Williams LD (2012). B-DNA structure is intrinsically polymorphic: Even at the level of base pair positions. *Nucleic Acids Res 40*(8):3714–3722. https://doi.org/10.1093/nar/gkr1168
Olson WK, Gorin AA, Lu X-J, Hock LM, Zhurkin VB (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci USA*, *95*(19):11163-11168. https://doi.org/10.1073/pnas.95.19.11163
Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu X-J, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolberger C, Berman HM (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol 313*:229–237. https//doi.org/ 10.1006/jmbi.2001.4987
Olson WK, Zhurkin VB (2011). Working the kinks out of nucleosomal DNA. *Curr Opin Struct Biol 21:*348-57. https://doi.org/10.1016/j.sbi.2011.03.006.
Packer MJ, Hunter CA (1998). Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J Mol Biol 280*(3):407-420. https://doi.org/10.1006/jmbi.1998.1865.
Packer MJ, Dauncey MP, Hunter CA (2000a). Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol 295*(1):71-83. https://doi.org/10.1006/jmbi.1999.3236.
Packer MJ, Dauncey MP, Hunter CA (2000b). Sequence-dependent DNA structure: tetranucleotide conformational maps*. J Mol Biol 295*(1):85-103. https://doi.org/10.1006/jmbi.1999.3237
Pasi M, Maddocks JH, Beveridge D, Bishop TC, Case DA, Cheatham TE, Dans PD, Jayaram B, Lankas F, Laughton C, Mitchell J, Osman R, Orozco M, Pérez A, Petkevičiūtė D, Spackova N, Sponer J, Zakrzewska K, Lavery R (2014). μABC: A systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res 42*(19):12272–12283. https://doi.org/10.1093/nar/gku855
Pasi M, Maddocks JH, Lavery R (2015). Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res 43*(4):2412-2423. https://doi.org/10.1093/nar/gkv080
Pérez A, Luque FJ, Orozco M (2007). Dynamics of B-DNA on the microsecond time scale. *J Am Chem Soc 129*(47):14739–14745. https://doi.org/10.1021/ja0753546
Pérez A, Lankas F, Luque FJ, Orozco M (2008). Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res 36*(7):2379–2394. https://doi.org/10.1093/nar/gkn082
Pérez A, Castellazzi CL, Battistini F, Collinet K, Flores O, Deniz O, Ruiz ML, Torrents D, Eritja R, Soler-López M, Orozco M (2012). Impact of methylation on the physical properties of DNA. *Biophys J 102*(9):2140–2148. https://doi.org/10.1016/j.bpj.2012.03.056
Petkevičiūtė D, Pasi M, Gonzalez O, Maddocks JH (2014). cgDNA: a software

package for the prediction of sequence-dependent coarse-grain free energies of B-form DNA. *Nucleic Acids Res 42*(20):e153. https://doi.org/10.1093/nar/gku825

Stolz RC, Bishop TC (2010). ICM Web: the interactive chromatin modeling web server. *Nucleic Acids Res 38*: W254-W261. https://doi.org/10.1093/nar/gkq496

Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB (2007). A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J Mol Biol* 371(3):725-738. https://doi.org/10.1016/j.jmb.2007.05.048

Ulyanov NB, Zhurkin VB (1984a). Sequence-Dependent Anisotropic Flexibility of B-DNA A Conformational Study. *J Biomo. Struct Dyn 2*(2):361-385. https://doi.org/10.1080/07391102.1984.10507573

Ulyanov NB, Zhurkin VB (1984b). Anisotropic flexibility of DNA depends on the base sequence. Conformation calculations of double-stranded tetranucleotides AAAA:TTTT, (AATT)2, (TTAA)2, GGGG:CCCC, (GGCC)2, (CCGG)2. *Mol Biol (Mosk)18*(6):1664–1685.

van Dam L, Levitt MH (2000) BII Nucleotides in the B and C Forms of Natural-sequence Polymeric DNA: A New Model for the C Form of DNA. *J Mol Biol* 304(4):541-561. https://doi.org/10.1006/jmbi.2000.4194

Wahl MC, Sundaralingam M (1997). C-H...O hydrogen bonding in biology. *Trends Biochem Sci* 22(3):97-102. https://doi.org/ 10.1016/s0968-0004(97)01004-9.

Walther J, Dans PD, Balaceanu A, Hospital A, Bayarri G, Orozco M (2020). A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res 48*:e29. https://doi.org/10.1093/nar/gkaa015

Watson JD, Crick FH (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature 171*:737-738. https://doi.org/ 10.1038/171737a0

Zgarbová M, Jurečka P, Lankas F, Cheatham TE, Šponer J, Otyepka M (2017). Influence of BII Backbone Substates on DNA Twist: A Unified View and Comparison of Simulation and Experiment for All 136 Distinct Tetranucleotide Sequences. *J Chem Inf Model 57*(2):275–287. https://doi.org/10.1021/acs.jcim.6b00621

Zhurkin VB, Lysov YP, Ivanov VI (1979). Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res 6*(3):1081-1096. https://doi.org/10.1093/nar/6.3.1081

Figure captions

Figure 1. The early beginnings. a) Representation of the first B-DNA crystal structure, the Drew-Dickerson Dodecamer (DDD). Adapted from Drew et al. 1981. b) Cross-chain stacking of purines (vertical dashed lines) and steric clashes (*) could explain the sequence-dependent values adopted by slide, roll and propeller twist parameters. Adapted from Calladine et al. 2004. c) Prediction of the twist parameter for the DDD by applying Calladine rules. Adapted from Dickerson 1983. d) Helical parameters approved by the Cambridge convention. Adapted from Lu and Olson 2003.

Figure 2. The nearest neighbour approximation. a) Roll-twist correlation for YR steps as revealed by Olson and coworkers in 1998. Adapted from Olson et al. 1998. b) Twist and roll average values for the 10 unique dinucleotides comparing MD simulations and X-ray crystal structures. Adapted from Pérez et al. 2008. c) Average twist values (black circles) for the 10 unique dinucleotides. The analysis of the 136 unique combinations emerging from considering all nearest neighbour nucleotides has revealed the existence of stable conformational sub-states (black triangles). Adapted from Lavery et al. 2010. d) A consensus view on the existence of low-twist (LT) and high-twist (HT) structural polymorphism for d(CpG) steps and the twist degree of freedom. The Bayesian Information Criterion was used for the deconvolution of twist distributions in LT and HT populations. Adapted from Dans et al. 2012.

Figure 3. Definition of backbone sub-states and CH---O interactions. a) BI/BII sub-states are defined according to the established conventions based on the value of ε–ζ torsions. b) HT/LT transitions in YR steps (step i) are coupled to backbone transitions in the following 3'-junction (step i+1), denoted $BI_{3'}$/$BII_{3'}$. c) Formation of the C8H8---O3' hydrogen bond in RR steps in a BII conformation. d) Formation of the C6H6---O3' hydrogen bond in RY steps in a BII conformation.

Figure 4. Structural polymorphisms at the nearest neighbour level and beyond. a) Conformational changes triggered by the entrance of cations in the minor groove of the d(CpG) step that leads to the HT/BI↔LT/BII transition. Adapted from Dans et al. 2014. b) ABC representation of the full polymorphic helical space obtained from the analysis of MD simulations of all the unique 136 tetranucleotides. All unique tetranucleotides are grouped according to the nature (R = purine, Y = pyrimidine) of the central bps. $BI_{3'}$ and $BII_{3'}$ refers to the presence of a BI/BII substate at the 3' junction of a given YR step (see Fig. 3). Adapted from Pasi et al. 2014. c) Patterns and correlations between helical parameters and backbone sub-states extracted from MD, enabled the formulation of extended Calladine-Dickerson rules. The rules allow for the structural prediction of complex conformational sub-states that could occur simultaneously for a given tetranucleotide, as observed experimentally. Adapted from Dans et al. 2019. d) Multi-states observed for d(TpA), and the BI (CH-O broken)/BII (CH-O formed) correlation between consecutive basepair steps. Adapted from Balaceanu et al. 2019.