The Role of Unconventional Hydrogen Bonds in Determining BII Propensities in B-DNA

Alexandra Balaceanu,[†,‡,∇] Marco Pasi,[§,‖,∇] Pablo D. Dans,[†,‡,∇] Adam Hospital,[†,‡] Richard Lavery,*[,§] and Modesto Orozco*[,†,‡,⊥]

[†]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, Barcelona 08028, Spain

[‡]Joint BSC-IRB Program in Computational Biology, Institute for Research in Biomedicine, Baldiri Reixac 10-12, Barcelona 08028, Spain

[§]MMSB, Univ. Lyon I/CNRS UMR 5086, Institut de Biologie et Chimie des Protéines, 7 Passage du Vercors, Lyon 69367, France

[‖]School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, University Park NG7 2RD, U.K.

[⊥]Department of Biochemistry and Biomedicine, Faculty of Biology, University of Barcelona. Diagonal 643, Barcelona 08028, Spain

Ⓢ *Supporting Information*
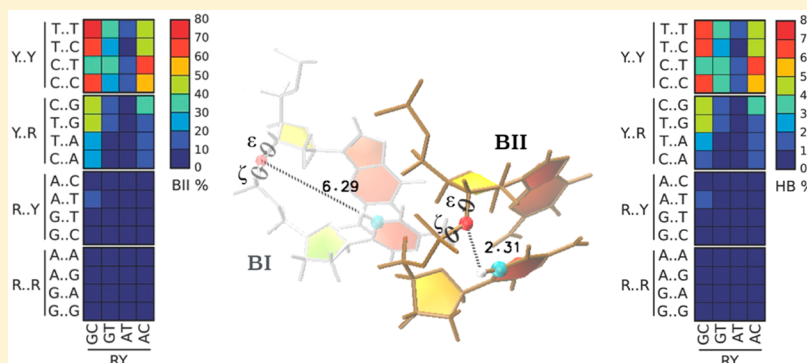
ABSTRACT: An accurate understanding of DNA backbone transitions is likely to be the key for elucidating the puzzle of the intricate sequence-dependent mechanical properties that govern most of the biologically relevant functions of the double helix. One factor believed to be important in indirect recognition within protein−DNA complexes is the combined effect of two DNA backbone torsions (ε and ζ) which give rise to the well-known BI/BII conformational equilibrium. In this work we explain the sequence-dependent BII propensity observed in RpY steps (R = purine; Y = pyrimidine) at the tetranucleotide level with the help of a previously undetected C−H···O contact between atoms belonging to adjacent bases. Our results are supported by extensive multimicrosecond molecular dynamics simulations from the Ascona B-DNA Consortium, high-level quantum mechanical calculations, and data mining of the experimental structures deposited in the Protein Data Bank.

The fact that DNA's overall conformation is associated with changes in backbone geometry became apparent from the analysis of the first generation of successfully resolved diffraction patterns.[1,2] One of the major backbone structural polymorphisms in B-DNA arises from its ability to populate two distinct conformations, known as BI and BII[3] (see scheme in Figure 1A). Within each DNA strand, the phosphodiester junction between two consecutive bases undergoes fast interconversions between two states defined by specific combinations of rotations around the ε and ζ dihedrals. The canonical state, termed BI, features ε/ζ in a trans/gauche- (t/g-) conformation, while the other state, BII, has ε/ζ in g-/t conformation. To determine BI/BII equilibrium in B-DNA, proton and phosphate NMR experiments,[4−7] molecular dynamics (MD) simulations,[8−10] and data mining of crystal structures from databases[11−13] have been historically used as

the preferred methods. Following initial observations based on crystal structures showing that BI/BII transitions were associated with base destacking and minor groove widening,[14,15] computer MD simulations have shed light on the influence of water and ion dynamics on the propensity of BI/BII states.[8,9,16,17] Destacking of two successive bases along one strand and water migration were found to be necessary, but not sufficient, conditions for the adoption of the BII conformation.[16,18] Also based on crystal structures, successive nucleotides in one strand have been shown to have anticorrelated backbone conformational states.[11,13] From the very beginning, ^{31}P NMR experiments were crucial to identifying sequence
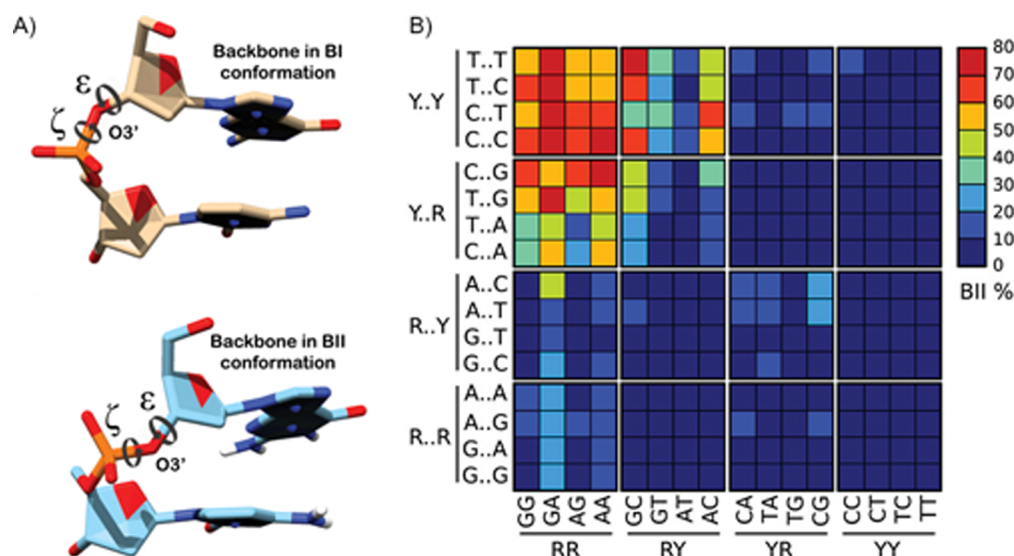
Figure 1. (A) Depiction of B-DNA BI and BII conformers resulting from rotations around the ζ and ε torsions. (B) Sequence dependence of BII backbone conformations. The percentage occurrence of BII backbone states for the phosphodiester junction at the central base step of each of the 256 possible tetranucleotide sequences is shown (BII%), using the color code defined on the right (0% is dark blue; 80% is dark red). The sequences are arranged so that each column represents one of 16 dinucleotide steps, and each row corresponds to one of the 16 possible flanking sequences; columns and rows are further grouped on the basis of base type (R = purine and Y = pyrimidine).

dependence as a key modulator of BI/BII transitions, capable of fine-tuning the population of states. Hartmann's group has been particularly active in this field, providing a sequence-dependent view of BI/BII at the dinucleotide level (and even tetranucleotide level),[4,7,19,20] but also putting into perspective NMR results with values obtained from crystal structures or MD simulations.[21] Moreover, recent MD studies have also shown[10,22,23] that the impact of sequence in BI/BII equilibrium is more complex than anticipated, supporting the importance of the tetranucleotide environment.

The transition between backbone substates can be accurately captured in its entirety only from a cohesive set of data where all 136 unique tetranucleotides are represented with statistical significance, which precludes the use of experimental information deposited in structural databases,[22] instead having to rely on the use of information gained from atomistic molecular dynamics (MD) simulations. To this end, we performed part of the present analysis on trajectories obtained by the Ascona B-DNA Consortium (ABC, bisi.ibcp.fr/ABC/Welcome.html) who created a database of simulations, called μABC, containing multimicrosecond MD simulations of DNA oligomers containing multiple copies of all the 136 distinct tetranucleotides.[9,10] An impressive amount of information was already extracted from this collection of simulations, extending our vision of the conformational landscape of B-DNA at the tetranucleotide level[10] and leading to some interesting observations, including the fact that YpR and YpY dinucleotide steps rarely populate the BII state, and also that the base pairs flanking a given step can significantly modulate the BI/BII equilibrium in RpR and RpY steps. Interestingly, a previously uncharacterized C−H···O hydrogen bond (H-bond) between the C8−H8 atoms of the R base and the O3′ atom of the corresponding 5′ phosphate turned out to be a key player in the stabilization of BII states at steps featuring a purine in the 3′-position (i.e., RpR and YpR).[10,23] It follows that the C8−H8···O3′ H-bond is not expected to have any role in the stabilization of the BII state in RpY and YpY steps, where it cannot be formed. Our analysis of RpY steps from the μABC set has

revealed that a similar contact can be formed between the same O3′ oxygen in the backbone phosphate and C6−H6 atoms of pyrimidines. In this work, we characterize in detail the nature of this previously undetected contact, first as captured by the classical potentials of MD simulations, and second, by carrying out quantum mechanics (QM) calculations at the MP2 level on several representative structures of RpY steps taken from the MD trajectories. The electron density obtained was a posteriori analyzed using the Atoms in Molecules[24] (AIM) approach in order to determine the stabilizing nature of this interaction. Finally, we extracted all the RpY steps from the high-resolution X-ray structures of isolated DNA deposited in the Protein Data Bank (PDB). As suggested by our atomistic simulations, RpY steps in the PDB found in the BII state showed a distance between the O3′ and C6 atoms compatible with the presence of the stabilizing C6−H6···O3′ H-bond. Our results also demonstrate that the occurrence of this interaction is highly time-correlated with the backbone BI → BII transitions and establish it as a stabilizing factor of the BII state, providing a complete view of the BI/BII equilibrium.

We started by analyzing two of the μABC simulations,[10] namely, those referring to the CAAG (5′-GCAGCAAGCAA-GCAAGGC-3′) and TAAG (5′-GCAGTAAGTAAGTAAG-GC-3′) sequences, consisting of 3.5 repeats of each of these two tetranucleotides with GpC base pairs capping each end. We focused our analysis on the centermost G8pC9 (Watson strand) and G24pC25 (Crick strand) steps from the CAAG oligomer, along with G8pT9 (Watson) and A24pC25 (Crick) steps of the TAAG sequence. These were chosen because GpY and ApC steps show the highest BII propensities among RpY steps from MD simulations (75% and 60% BII, respectively; see Figure 1B). The reader should be aware that when NMR, MD simulations, and X-ray structures are compared, differences exist between the specific BII percentages assigned to some base pairs steps.[12,21] The sources of uncertainties from crystal structures have being discussed several times and are clearly related to low resolution (frequently insufficient to define backbone's states), lack of dynamics, and lattice re-
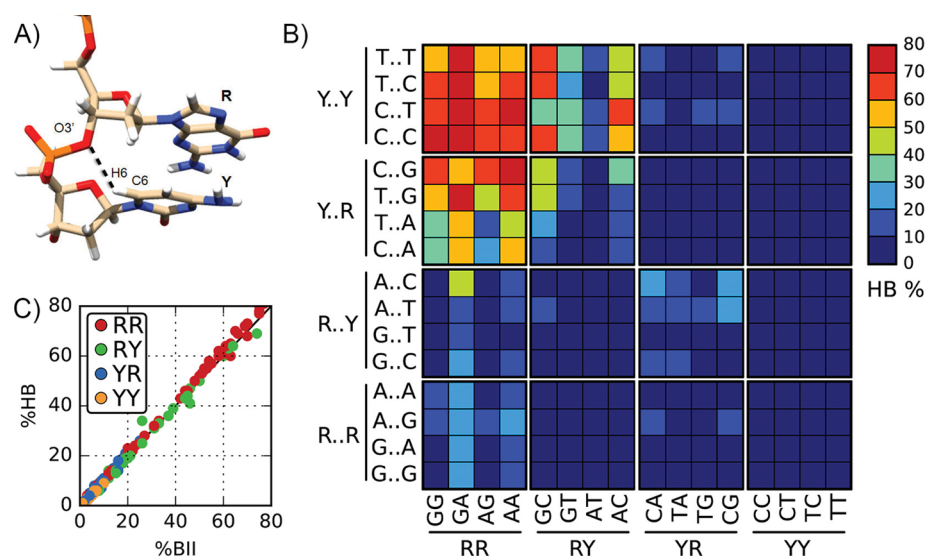
Figure 2. (A) Representation of the C6−H6···O3′ interaction in an RpY step showing the atoms involved. (B) Sequence dependence of C−H···O H-bond formation. The percentage occurrence of either the C6−H6···O3′ or the C8−H8···O3′ H-bond at the central base step of each of the 256 possible tetranucleotide sequences is shown, using the color code defined on the right (0% is dark blue; 80% is dark red). The sequences are arranged so that each column represents one of 16 dinucleotide steps, and each row corresponds to one of the 16 possible flanking sequences; columns and rows are further grouped on the basis of base type (R = purine and Y = pyrimidine). (C) Correlation between the percentage of BII (% BII, horizontal axis) and of occurrence of formation of the C−H···O H-bonds at the central base step of each of the 256 possible tetranucleotide sequences, color-coded according to base type of the central base step. The correlation coefficient is 0.998.

Table 1. C−H···O Hydrogen Bond Parameters (Average Distance, Average Angle, Charges, and Energy) and Electron Density at the Bond Critical Point Computed at the MP2(FC)/6-31G(d,p) Level

seq/dinuc.	distance (Å)	angle (deg)	ρ (au)	∇^2 (au)	E_{Hbond}[a] (kcal/mol)	q_C[b] (e)	q_H (e)	q_O (e)	$U_{Coulomb}$ (kcal/mol)
AGCA (GpC)	3.30 ± 0.02	139	0.011 ± 0.002	0.037 ± 0.004	−3.51	−0.018	0.229	−0.523	−11.6
AGTA (GpT)	3.19 ± 0.06	129	0.011 ± 0.001	0.037 ± 0.005	−3.75	−0.221	0.261	−0.523	−13.8
TACT (ApC)	3.34 ± 0.06	142	0.011 ± 0.001	0.037 ± 0.004	−3.44	−0.018	0.229	−0.523	−11.6
CGGC (GpG)[c]	3.2	138	0.013	0.043	−4.52	0.074	0.200	−0.523	−10.8
CGAC (GpA)[c]	3.0 ± 0.1	136	0.018 ± 0.001	0.059 ± 0.004	−7.19	0.161	0.188	−0.523	−10.0
A-U bp[d]	3.6	−	0.006	0.021	−0.85	0.572	0.060	−0.548	−3.0
U-U bp[e]	3.3	−	0.016	0.047	−5.19	−0.364	0.181	−0.548	−9.7

[a]Computed from the Laplacian values using the linear regression from Cubero et al.[35] [b]Taken directly from the AMBER 14 libraries (as reported in the original parm94 article[32]). [c]Adapted with permission from Dans et al.[23] [d]We reproduced the values for the C2−H2···O2 H-bond as reported in the work of Martin-Pintado et al.[43] [e]Values for the C5−H5···O2 hydrogen bond as reported in the work of Martin-Pintado et al.[43]

straints.[13,25,26] In the same way, changes in NMR's refinement protocols or annealing procedures from one experiment to another,[27,28] the usually low number of restraints due to the low density of protons in DNA, and the frequent overlap of several NOE peaks[20] are the main sources of uncertainties, with the average standard deviation in the prediction of BII percentages being ±8.[21] In addition, both experimental sources of BII values suffer from a sequence bias due to the limited number of tetranucleotide represented in the limited set of sequences available.[20,22,29] For its part, the parmBSC0[30] force field for MD simulations of DNA, the gold standard for the past decade used to produce the μABC data set, is known to produce an overall underestimation of twist and a clear underestimation of BII populations,[28,31] in particular for YpR steps. Despite these considerations, the stabilizing C−H···O contacts are postulated to exist beyond the apparent discrepancies reported between the different methods or the specific value assigned to a specific base pair step. Nevertheless, we decided to support our conclusions on the role of the C−H···O H-bonds, simulating again the CAAG and TAAG

sequences, following the same protocol,[10] but with the latest force field for DNA parmBSC1,[31] which is known to fix the twist and BII issues. These new simulations validate our results and ensure that our conclusions go beyond the use of a specific force field (results obtained with parmBSC1 are presented in the Supporting Information but discussed throughout the text).

To correctly capture the fast interconversion between backbone states[10,23] or related twist/slide states,[23] we extracted information from the simulations every 1 ps leading to conformational ensembles comprising more than 10^6 structures for each of the two sequences. From the ensemble of collected conformations, we identified the existence of an interaction between the O3′ atom of the backbone phosphate of the GpC, GpT, and ApC junctions and the C6−H6 atoms of the pyrimidine (see scheme in Figure 2A). This contact is structurally equivalent to the C8−H8···O3′ H-bond identified in RpR or YpR steps,[10,23] where it has been shown to play a fundamental role in stabilizing the BII state. This equivalence suggests that this new interaction can explain the BII propensity observed for the two remaining classes of steps, namely, RpY
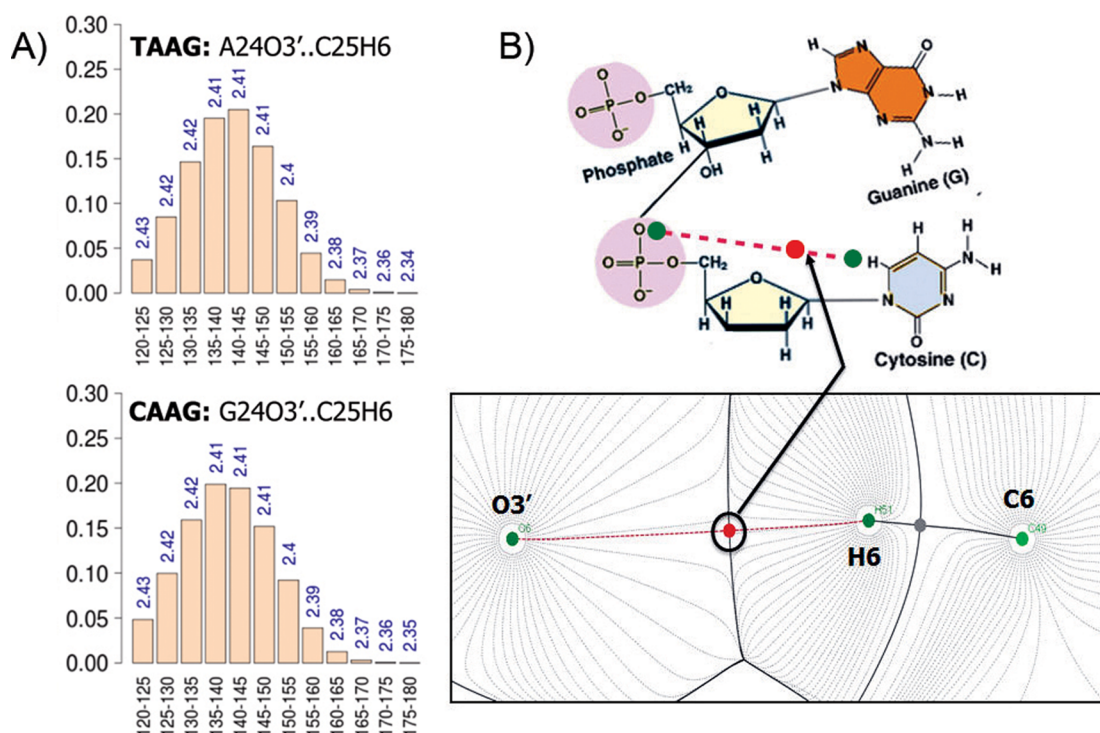
Figure 3. (A) Distribution of C6−H6···O3′ angles in RpY steps. Structures with H6···O3′ bond distances <2.5 Å and C6−H6···O3′ angles >120° were selected, and a histogram was built. The mean distance of the corresponding set is given above each bin bar. The equivalent distributions obtained with the parmBSC1 force field are reported in Figure S1. (B) Hydrogen bond AIM analysis for the GpC dinucleotide in the BII conformation. The bond critical point is indicated by a red dot. The nuclear critical points (located at the position of the nuclei) are indicated by green dots, while the basin paths and the gradient field are shown with gray lines. The bond paths, defined by the chosen two-dimensional projection (plane), are shown with red dotted lines.

and YpY. To evaluate whether this is the case, we turned to the complete μABC data set and found that the occurrence of the C6−H6···O3′ hydrogen bond is indeed in perfect sequence-dependent correlation with the BII state population of RpY and YpY steps (compare Figures 1B with 2B, and see the correlation in Figure 2C). This interaction can reasonably be termed a H-bond because the partial charges assigned in the force field (obtained by QM fitting[32]) to the C6−H6 atoms in thymine or cytosine generate a significant bond dipole (Table 1). Furthermore, the geometrical features of the three atoms, whenever this contact occurs, are consistent with the partial covalent nature of hydrogen bonding[33] as assessed from angle and distance distributions within the ensemble of MD structures (Figure 3A shows the angle distribution of structures with H6···O3′ distances below 2.5 Å). We validate this geometrical arrangement by obtaining results totally equivalent with the last generation force field for DNA[31] (Figure S1).

To further confirm our interpretation, we carried out ab initio MP2 calculations on seven different snapshots from each of the two simulations (see Methods). Five representative structures were taken from the most populated state in the C6−H6···O3′ angle−C6···O3′ distance space, while two other structures belonged to the marginal bins of the distributions depicted in Figure 3A with angles above 170° coupled with donor−acceptor distances below 3.35 Å. The electron densities obtained in this way were analyzed using the AIM approach[24] to determine the stationary points and the gradient paths (obtained from the first derivative of the electron density) connecting them. In particular, we focused on the bond critical points (bcp) generated between hydrogen and acceptor group, as previous studies have demonstrated that canonical H-bonds

(X-H···Z; with X and Z being electronegative atoms) are associated with electron densities at the bcp that vary in the range from 0.002 to 0.034 atomic units (au).[34] Bond critical points and bond paths between the H6 and O3′ atoms were found in all 21 electron densities analyzed (see Figure 3B for a representative scheme and Table 1 for numerical description) supporting the existence of H-bonds. On average, the electron densities at the bcp were found to be around 0.011 au, within the expected range of values for canonical interactions (see Table 1 and ref 34), even though slightly lower than those reported for the C8−H8···O3′ H-bond in RpR steps (two cases are shown for comparison). The positive value of the Laplacian (second derivative of the electron density) at the bcp indicates a depletion of electron density toward the interacting nuclei from a density maximum, another feature consistent with the formation of hydrogen bonds.[24] We calculated the gas-phase stabilization provided by the specific C6−H6···O3′ contact in all cases, by estimating the interaction energy from the linear relation described by Cubero et al.[35] We found that each C6−H6···O3′ hydrogen bond stabilizes the BII state by more than 3 kcal mol[−1], a value only slightly lower compared with the equivalent C−H···O bond described for RpR steps (Table 1).[23] It is worth noting that a 3 kcal mol[−1] stabilizing effect should completely drive the equilibrium to BII, but part of this stabilization will be compensated by hydration effects, because water occupancy at the O3′ group decreases dramatically from 36% in the BI state (when the C−H···O bond is not formed) to only 1.4% in BII (data from the analysis of 10[5] structures taken from the last 100 ns of trajectory filtered according to the BI/BII state, confirmed by both force fields).
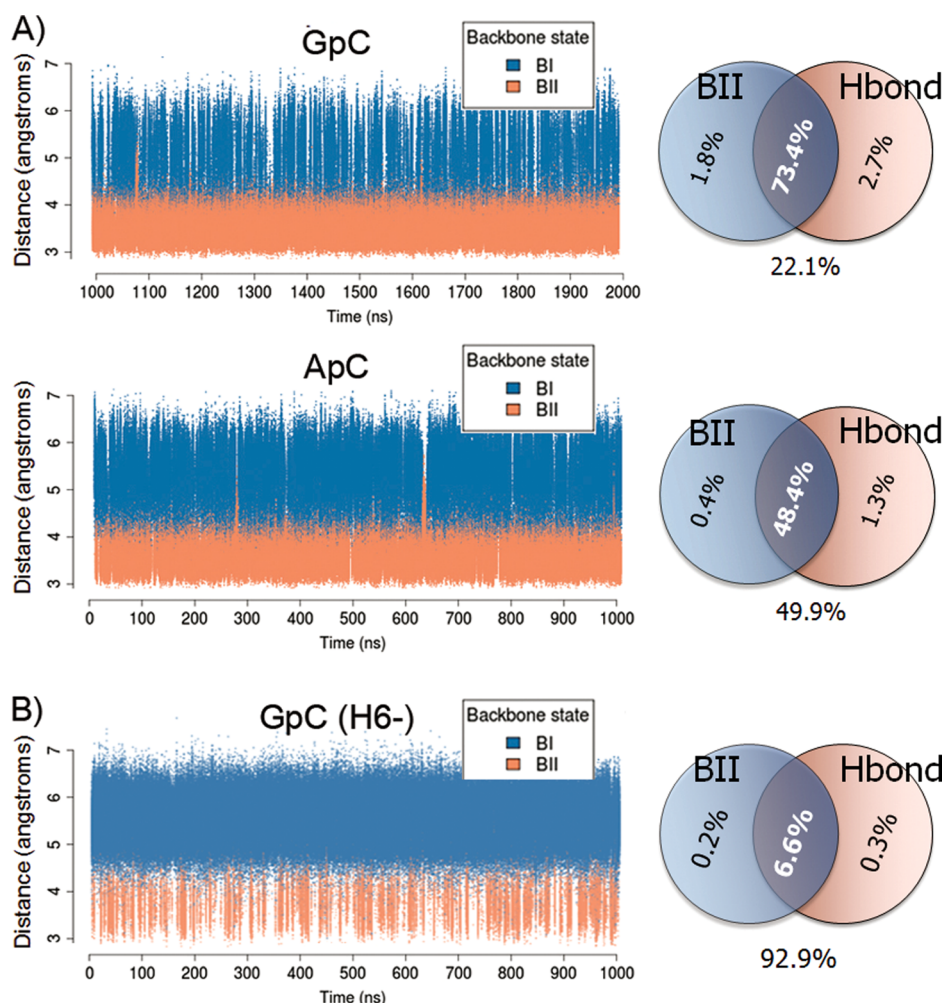
Figure 4. (A) Left: Time evolution of C6···O3′ distance in two RpY steps (GpC and ApC) colored by the backbone conformation of the step. Right: Venn diagrams of occurrences of BII state and C6−H6···O3′ hydrogen bonds at the same RpY steps. An equivalent figure but showing the results obtained with parmBSC1 force field is shown in Figure S2. (B) Same as in panel A for the simulation without the H6 atom in GpC, labeled CAAG(H6-). The results for TAAG(H6-) are presented in Figure S3.

A more detailed analysis of the MD time series shows that the C6−H6···O3′ contact occurs simultaneously with the formation of BII states: our results indicate that almost all BI → BII backbone transitions involve the instantaneous formation of the hydrogen bond (Figure 4A). This sheds new light on the question of whether the hydrogen bond forms prior to the transition, driving the backbone into a BII state slowly, allowing for a period of structural frustration. Our results suggest that this interaction is more of a stabilizing force than a driving element. This is in agreement with the observed average lifetime of hydrogen bond formation (25.6 ps for GpC and 16.5 ps for GpT), compared to the average lifetimes of the corresponding BII states (23.6 and 15.8 ps respectively). These results were confirmed by analyzing the time series obtained with parmBSC1[31] force field. Although the percentage occurrence of the BII state and C6−H6···O3′ hydrogen bond at GpC and ApC steps varies slightly (by less than 10%) when using parmBSC1, the extent to which the two events are time-correlated is substantially unchanged (compare the time series and distributions of Figures 4A and S2).

From a mechanical point of view, the formation of the C8−H8···O3′ H-bond in RpR steps has been shown to be coupled to slide polymorphism at the base level in the same junction.[10,23] In contrast, no helical parameter or torsion angle showed a two-state distribution coupled to the formation of the C6−H6···O3′ hydrogen bond in RpY steps (again confirmed by both force fields). It seems that the mechanical coupling between slide and the backbone with a purine in the 3′ position (which helps to bring closer the C8−H8 atoms to the backbone) is not necessary in the case of pyrimidines in 3′. We also did not find any coupled role of cations in these transitions (in contrast to that reported for CpG steps[23]), while the hydration change around the backbone atoms, synchronized with the BI/BII transition, and the formation of the intramolecular C−H···O hydrogen bond produced a local water migration in agreement with that reported experimentally.[18]

To confirm our results we performed two "proof of concept" simulations, with the same sequences reported above, labeled TAAG(H6-) and CAAG(H6-), in which the H6 atom from the pyrimidine base was removed and its charge was transferred to the C6 atom (see Methods and ref 23). This "alchemical" base is useful in testing the conformational impact of the C6−H6···O3′ H-bond. As expected, in these simulations C6 and O3′ no longer come into close contact and the backbone of RpY steps undergoes significantly fewer transitions to the BII state (up to 69% less BII for GpC), suggesting that without the C−H···O

stabilizing interaction the backbone cannot last in time in the BII state (its average lifetime is decreased to 10.6 ps for GpC and 8.8 ps for GpT, half its normal value), or easily access a g-/t state of the ε/ζ torsions (Figures 4B and S3).

Finally, we performed an analysis of high-resolution experimental structures of isolated DNA to find experimental support for our hypothesis. For this purpose, we extracted all high-resolution X-ray structures of isolated dsDNA oligomers deposited in the PDB ($R < 2.5$ Å, see Table S1); the resulting 554 experimental structures contain information on 3991 RpY dinucleotides (37% GpC, 18% GpT, 19% ApC, and 26% ApT). We found that the distribution of C6···O3′ distances is bimodal, in agreement with simulations, and clearly correlated with the BI/BII state in the backbone (Figure S4). For the GpC case, for which we have better statistics, the C6···O3′ distance decreases in average from 5.03 to 3.48 Å (with a s.d. of 0.4 Å) when moving from BI to BII state in the crystal structures. We repeated the analysis for RpR steps (484 structures analyzed, see Table S1 and Figure S4), finding, analogously, a shortening of the C8···O3′ distance from 5.20 (BI) to 3.80 Å (BII). Equivalent results were obtained for YpR and YpY steps (Table S2 and Figure S4). It should be noted that in spite of the extended set of structures used in this work, BII propensities from crystal structures still seem underestimated when compared with NMR and MD results (Table S2).[4−6,12,20,21] The reasons for the apparent discrepancies between the methods, which are beyond the scope of the present work, are complex, of diverse sources, and have been partially addressed recently.[21] In summary, despite the relative scarcity of experimental structural data, the analysis of crystal structures provides quantitative support for the importance of the C6−H6···O3′ interaction discussed here, as well as for the C8−H8···O3′ H-bond previously reported.[10,23]

The observed sequence-dependent BII propensity in RpY steps, as obtained from MD simulations, can now be explained by taking into account the established hierarchy of bond strength, inferred from the populations of H-bond formation corresponding to each purine−pyrimidine combination (Figure 2B). Considering all possible steps, H-bonds in RpR are the strongest, with GpA being the most favorable, while RpG and RpC interactions are of similar strength. RpT and YpR contacts are rather weaker, but a H-bond in YpY steps is indeed very infrequent. We also observed the hindering effect one H-bond has on the formation of a second H-bond in a neighboring step, in agreement with the known anticorrelation between adjacent BII backbone states. Accordingly, applying these simple considerations in a tetranucleotide context would predict, for example, that a YpRpYpY sequence should result in the highest BII content among all RpY steps in MD simulations, because it is unfavorable to form a H-bond in either of the flanking base steps, while the opposite effect should be observed within an RpRpYpR sequence, where both flanking steps will compete for H-bond formation, leading to conformational frustration. This is an important effect to understand sequence-dependent propensities at the tetranucleotide level, because the crankshaft motion of the backbone ensures almost every time that alternate BI/BII/BI/BII states will be observed in successive junctions in the same strand,[13] leading to conformational frustration when two or more dinucleotides with high BII content are side-by-side. The confirmation of these predictions (Figure 1B), and the extension of our conclusions to RpR and YpR steps,[10] supported by last generation force field, allows us to conclude that the newly detected C−H···O H-bond makes

an important contribution to deciphering the sequence-dependent BII propensity within B-DNA, which in turn has an important role to play in protein−DNA recognition processes.

■ METHODS

Molecular Dynamics Simulations. Sequence dependence analyses are based on the 39 multimicrosecond simulations collected in 2014 by the Ascona B-DNA Consortium that form the μABC data set.[10] Since this analysis, a new version of the Amber force field for DNA named parmBSC1[31] was published. Because this force field has been shown to produce trajectories in even better correlation with experiment,[24,36] we validate parmBSC0-derived conclusions by simulating again with parmBSC1 the two double stranded B-DNA oligomers with sequences 5′-GCAGCAAGCAAGCAAGGC-3′ (labeled CAAG) and 5′-GCAGTAAGTAAGTAAGGC-3′ (labeled TAAG) used in the detailed analysis presented here. Additionally, simulations (using both parmBSC0[30] and parmBSC1[31]) were carried out removing the H6 atom of either the thymine or the two cytosines of interest: labeled TAAG(H6-) and CAAG(H6-), respectively. To maintain the total charge of the system in these model calculations, the H6 charge was transferred to the C6 atom.[23] All simulations were carried out using the protocol described in Pasi et al.[10] The results obtained with parmBSC1 are equivalent to those discussed in the main text and are presented in the Supporting Information.

Analysis of Trajectories. Trajectories were preprocessed with the cpptraj module of the AmberTools15 package.[37] Conformational analysis was performed using the Curves+ and Canal programs,[38] which provide a full set of helical, backbone, and groove geometry parameters, and further dissection of these quantities was done with the use of NaFleX server[39] and in-house tools. We consider a H-bond was formed when the distance between C6/C8 and O3′ was below 4 Å. Trajectories will be deposited in the BigNASim database[40] of the European MuG Virtual Research Environment (www.multiscalegenomics.eu/MuGVRE/).

Quantum Mechanical Calculations. To make a first-principles confirmation of the existence of the C−H···O intramolecular hydrogen-bond, Bader's atoms in molecules (AIM)[24] electron topology analysis was used. Seven representative snapshots from the three selected dinucleotides (GpC, GpT, and ApC) were extracted from the MD simulations to perform single-point MP2 calculations. Waters and ions were removed, and only the dinucleotide step was kept and subjected to single-point calculations at the MP2(FC)/6-31G(d,p) level of theory using Gaussian 09.[41] H atoms were used to complete the valency of the 5′ and 3′ oxygen atoms. The electron density, the gradient, and its Laplacian at the bcp were computed and analyzed using the program AIM-UC.[42]

■ ASSOCIATED CONTENT

ⓈSupporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.6b02451.

> Codes of the structures analyzed from the PDB, analyses from the simulations performed without the H6 atom, and results obtained with the parmBSC1 force field (PDF)

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: richard.lavery@ibcp.fr.
*E-mail: modesto.orozco@irbbarcelona.org.

ORCID ⓘ
Pablo D. Dans: 0000-0002-5927-372X
Modesto Orozco: 0000-0002-8608-3278

Author Contributions
▽A.B., M.P., and P.D.D. contributed equally to this work.

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (19), 11163–11168.

(2) Matthews, B. W. No code for recognition. *Nature* **1988**, *335* (6188), 294–295.

(3) Hartmann, B.; Piazzola, D.; Lavery, R. BI-BII transitions in B-DNA. *Nucleic Acids Res.* **1993**, *21* (3), 561–568.

(4) Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J. Am. Chem. Soc.* **2006**, *128*, 9170–9177.

(5) Schwieters, C. D.; Clore, G. M. A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* **2007**, *46* (5), 1152–1166.

(6) Tian, Y.; Kayatta, M.; Shultis, K.; Gonzalez, A.; Mueller, L. J.; Hatcher, M. E. 31P NMR investigation of backbone dynamics in DNA binding sites. *J. Phys. Chem. B* **2009**, *113*, 2596–2603.

(7) Abi-Ghanem, J.; Heddi, B.; Foloppe, N.; Hartmann, B. DNA structures from phosphate chemical shifts. *Nucleic Acids Res.* **2010**, *38* (3), e18.

(8) René, B.; Masliah, G.; El Antri, S.; Fermandjian, S.; Mauffret, O. Conformations and dynamics of the phosphodiester backbone of a DNA fragment that bears a strong topoisomerase II cleavage site. *J. Phys. Chem. B* **2007**, *111* (16), 4235–4243.

(9) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38* (1), 299–313.

(10) Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* **2014**, *42* (19), 12272–12283.

(11) Djuranovic, D.; Hartmann, B. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* **2003**, *20*, 771–788.

(12) Madhumalar, A.; Bansal, M. Sequence preference for BI/BII conformations in DNA: MD and crystal structure data analysis. *J. Biomol. Struct. Dyn.* **2005**, *23* (1), 13–27.

(13) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res.* **2008**, *36* (11), 3690–3706.

(14) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (4), 2179–2183.

(15) Schneider, B.; Neidle, S.; Berman, H. M. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers* **1997**, *42* (1), 113–124.

(16) Pichler, A.; Rüdisser, S.; Winger, R. H.; Liedl, K. R.; Hallbrucker, A.; Mayer, E. The role of water in B-DNAs BI to BII conformer substates interconversion: a combined study by calorimetry, FT-IR spectroscopy and computer simulation. *Chem. Phys.* **2000**, *258* (2), 391–404.

(17) Winger, R. H.; Liedl, K. R.; Rüdisser, S.; Pichler, A.; Hallbrucker, A.; Mayer, E. B-DNA's BI → BII Conformer Substate Dynamics Is Coupled with Water Migration. *J. Phys. Chem. B* **1998**, *102* (44), 8934–8940.

(18) Grzeskowiak, K.; Yanagi, K.; Privé, G. G.; Dickerson, R. E. The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. The effect of base pair reversals. *J. Biol. Chem.* **1991**, *266* (14), 8861–8883.

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

(20) Heddi, B.; Oguey, C.; Lavelle, C.; Foloppe, N.; Hartmann, B. Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic Acids Res.* **2010**, *38* (3), 1034–1047.

(21) Imeddourene, A. B.; Elbahnsi, A.; Guéroult, M.; Oguey, C.; Foloppe, N.; Hartmann, B. Simulations Meet Experiment to Reveal New Insights into DNA Intrinsic Mechanics. *PLoS Comput. Biol.* **2015**, *11* (12), e1004631.

(22) Dans, P. D.; Pérez, A.; Faustino, I.; Lavery, R.; Orozco, M. Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.* **2012**, *40* (21), 10668–78.

(23) Dans, P. D.; Faustino, I.; Battistini, F.; Zakrzewska, K.; Lavery, R.; Orozco, M. Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.* **2014**, *42* (18), 11304–20.

(24) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: New York, 1994.

(25) Dickerson, R. E.; Grzeskowiak, K.; Grzeskowiak, M.; Kopka, M. L.; Larsen, T.; Lipanov, A.; Prive, G. G.; Quintana, J.; Schultze, P.; Yanagi, K.; et al. Polymorphism, packing, resolution, and reliability in single-crystal DNA oligomer analyses. *Nucleosides Nucleotides* **1991**, *10*, 3–24.

(26) Jain, S.; Richardson, D. C.; Richardson, J. S.. Computational Methods for RNA Structure Validation and Improvement (Chapter 7). In *Methods in Enzymology: Structures of Large RNA Molecules and Their Complexes*; Woodson, S. A., Allain, F. H. T., Eds; Elsevier: Waltham, MA, 2015; Vol. 558, pp 181–212.

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

(28) Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M. How accurate are accurate force-fields for DNA. *Nucleic Acids Res.* **2016**, in press.

(29) Pérez, A.; Luque, F. J.; Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* **2012**, *45*, 196–205.

(30) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for

nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92* (11), 3817−29.

(31) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55−58.

(32) 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. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(33) Grabowski, S. J.; Sokalski, W. A.; Leszczynski, J. The Possible Covalent Nature of N−H⋯O Hydrogen Bonds in Formamide Dimer and Related Systems: An Ab Initio Study. *J. Phys. Chem. A* **2006**, *110* (14), 4772−4779.

(34) Koch, U.; Popelier, P. L. A. Characterization of C-H-O Hydrogen Bonds on the Basis of the Charge Density. *J. Phys. Chem.* **1995**, *99* (24), 9747−9754.

(35) Cubero, E.; Orozco, M.; Hobza, P.; Luque, F. J. Hydrogen bond versus anti-hydrogen bond: a comparative analysis based on the electron density topology. *J. Phys. Chem. A* **1999**, *103*, 6394−6401.

(36) Dans, P. D.; Danilāne, L.; Ivani, I.; Dršata, T.; Lankaš, F.; Walther, J.; Illa Pujagut, R.; Battistini, F.; Gelpí, J. L.; Lavery, R.; Orozco, M. Long-timescale dynamics of the Drew−Dickerson dodecamer. *Nucleic Acids Res.* **2016**, *44* (9), 4052−4066.

(37) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H. et al. *AMBER.* University of California: San Francisco, 2014.

(38) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917−5929.

(39) Hospital, A.; Faustino, I.; Collepardo-Guevara, R.; González, C.; Gelpí, J. L.; Orozco, M. NAFlex: A web server for the study of nucleic acid flexibility. *Nucleic Acids Res.* **2013**, *41* (W1), W47−W55.

(40) Hospital, A.; Andrio, P.; Cugnasco, C.; Codo, L.; Becerra, Y.; Dans, P. D.; Battistini, F.; Torres, J.; Goñi, R.; Orozco, M.; et al. BigNASim: A NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.* **2016**, *44* (D1), D272−D278.

(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian 09*, revision D.01; Gaussian Inc.: Wallingford, CT, 2009.

(42) Vega, D.; Almeida, D. AIM-UC: an application for QTAIManalysis. *J. Comput. Methods Sci. Eng.* **2014**, *14*, 131−136.

(43) Martin-Pintado, N.; Deleavey, G. F.; Portella, G.; Campos-Olivas, R.; Orozco, M.; Damha, M. J.; Gonzalez, C. Backbone FC-H⋯O hydrogen bonds in 2F-substituted nucleic acids. *Angew. Chem., Int. Ed.* **2013**, *52*, 12065−12068.