Short-Range Imbalances in the AMBER Lennard-Jones Potential for (Deoxy)Ribose...Nucleobase Lone-pair...π Contacts in Nucleic Acids.

Klaudia Mrazikova,1,2 Jiri Sponer,1,3 Vojtech Mlynsky,1,3 Pascal Auffinger,4,* Holger Kruse1,4*
1 Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic
2 National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic
3 Regional Centre of Advanced Technologies and Materials, Czech Advanced Technology and Research Institute (CATRIN), Palacky University Olomouc, Šlechtitelů 241/27, 783 71, Olomouc – Holice, Czech Republic
4 Architecture et Réactivité de l’ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg, 67084, France

Table of content:
1) Parameters for the vertical scans of the nucleobase surfaces.
2) Surface scan limitations.
3) DME...G/C versus deoxyribose...G/C calculations.
4) Rotation of DME above a G nucleobase.
5) Discussion related to interaction energy surfaces (IE).
6) MD equilibration protocols at 100 K and 300 K.
7) Table S1 Harmonic penalty restraints for geometry optimizations.
8) Table S2 Interpolated global minima for the DHDF-D3 and AFF minEd and IE surfaces.
9) Table S3 The minEd and IE minima values for DME and deoxyribose...G/C systems.
10) Table S4 SAPT2+(3)5MP2/jun-cc-pVTZ energy values for different orientations of DME above G nucleobase.
11) Table S5 AMBER atom types and Lennard-Jones parameters for nucleobase atoms.
12) Table S6 MD simulation analysis of CpG steps in Z-DNA and r(UUCG) that adopt a Z-step conformation.
13) Figure S1 Schematic representation explaining the construction of the minEd and IE surfaces.
14) Figure S2 Comparison of DHDF-D3 and AFF vertical scans for DME and deoxyribose...G/C models.
15) Figure S3 SAPT2+(3)5MP2/jun-cc-pVTZ of DME...C for O4'...C(C2) and O4'...C(C5) positions.
16) Figure S4 Comparison of DHDF-D3 and AFF IE surfaces for DME...G/A.
17) Figure S5 Comparison of DHDF-D3 and AFF IE surfaces for DME...C/U.
18) Figure S6 Rotation of DME by 90° steps with the O4'(DME) atom located above the guanine pyrimidine ring centroid.
19) References
1. Parameters for the vertical scans of the nucleobase surfaces

List of vertical separation distances (in Å):
2.50, 2.60, 2.70, 2.80, 2.90, 3.00, 3.10, 3.15, 3.20, 3.30, 3.40, 3.50, 3.60, 3.80, 4.00, 4.20 (20 distances); note the smaller 0.05 Å steps around 3.00 Å and the larger 0.2 Å steps for the vertical separation above 3.60 Å.

2. Surface scan limitations

Here, we address some limitations of our energy scans that are also encountered in most modelling approaches. First, we could have used ribose or deoxyribose sugars instead of DME. However, for a rigid 3D surface scan, using a smaller model is advantageous as moving a sugar probe across the nucleobase could easily cause unphysical steric clashes. Hence, it might become less clear if differences in the 3D surfaces are due to a lack of geometry relaxation (steric clashes) or originate from the use of a different molecular probe. Therefore, it is generally advised to limit the occurrence of steric clashes when investigating the performances of computational methods. Nonetheless, we carried out additional vertical separation calculations for deoxyribose...G/C models (Figure S2; equivalent to Figure 3). Details of these calculations that support the use of the smaller DME model are given in the section DME...G/C versus deoxyribose...G/C calculations (see below).

The second issue is related to the molecular probe’s orientation and rigidness. We chose for DME the deoxyribose orientation of Z-steps in Z-DNA (Figure 1). However, other choices are equally viable. The (deoxy)ribose...nucleobase stacking in other nucleic acid motifs and with other nucleobases show variations in tilt angle towards the nucleobase plane as well as in rotations. The section Rotation of DME above a G nucleobase (see below) is dedicated to rotating DME in 90° steps in a DME...G model system (Figure S6). As repeating the surface computations would have been too computationally expensive, we only calculated SAPT values to analyze changes in the interaction energy. The interaction energies range from −3.6 to −5.1 kcal.mol⁻¹ and mostly change in their electrostatic component (Table S4). These values suggest that the CpG Z-step orientation is the least energetically favored one. Thus, this orientation is likely adopted due to steric constraints imposed by the Z-DNA context.

3. DME...G/C versus deoxyribose...G/C calculations

As mentioned above, we computed vertical interaction energy scans for larger deoxyribose...nucleobase models (Figure S2). The deoxyribose...G model was derived from a CpG Z-step extracted from a Z-DNA X-ray structure as described in Figure 1 (see main text). The deoxyribose...C model was derived from the deoxyribose...G model by modification of G to C. The model systems were optimized as described in the QM structure optimization section (see main text) using harmonic penalty restraints listed in Table S1. The position of deoxyribose in the deoxyribose...G model was modified to place the O4’ atom above the centroid of the guanine pyrimidine ring corresponding to the DME...G minEd global minimum (Figure 4). For deoxyribose...C, the O4’ atom was placed above the C2 atom corresponding to the DME...C minEd minimum (Figure 5). Vertical interaction energy scans were performed at O4’(deoxyribose)...nucleobase distances listed in the sub-section Parameters for the vertical
scans of the nucleobase surfaces (see above) using both DHDF-D3 and AFF methods as described in Computational details (see main text).

The optimal DHDF-D3 distances are 2.99 Å for deoxyribose...G (DME...G: 2.97 Å) and 2.87 Å for deoxyribose...C (DME...C: 2.91 Å). The DHDF-D3 interaction energy for deoxyribose...C slightly weakens compared to DME...C leading to a flatter potential curve that allows a shorter deoxyribose...nucleobase distance (Figure S2 and Table S3). The AFF data for deoxyribose models shows qualitatively a similar shift to longer distances as with the DME probe (Table S3). Interpolated AFF minima are 3.12 Å for deoxyribose...G (DME...G = 3.10 Å) and 3.06 Å for deoxyribose...C (DME...C = 3.15 Å). We emphasize that the energy scans were performed at the minEd global minima of the DME...nucleobase minEd surfaces that might differ from hypothetical deoxyribose...nucleobase surfaces, as discussed in section Surface scan limitations (see above) and elsewhere.1

For G, the DHDF-D3 curve has a stronger interaction energy compared to the AFF one, also in agreement with the DME...nucleobase model. However, this is not the case for C. Explanation of this difference is not straightforward since various steric effects might be present due to the rigid monomer approach. Nevertheless, one more penalty restraint was needed for the deoxyribose...C model to keep the desired orientation of deoxyribose during optimization, possibly leading to a smaller relaxation of deoxyribose...C geometry compared to deoxyribose...G (Table S1).

4. Rotation of DME above a G nucleobase

In the main text, we considered only Z-like steps with the deoxyribose...nucleobase orientation of the Z-DNA CpG step. However, we suggest that sufficiently accurate description of deoxyribose...nucleobase stacking by the AFF is necessary also for systems possessing different orientation of the sugar with respect to the nucleobase.

To obtain an idea about how re-orientation of the deoxyribose impacts the energetics of this interaction, we used SAPT calculations to analyze changes in magnitude and decomposition of the interaction energy while varying the orientation of DME above G. Here, we are using a rigid approximation, keeping the O4’...G separation at 2.95 Å. The DME probe located above the centroid of the pyrimidine ring was rotated in 90° steps around the z-axis (Figure S6). The London dispersion, which is the least sensitive to the twist angle, dominates the stabilization in all the four cases. The strongest interaction energy is found for the first 90° step (DME...G (90°); Table S4) followed by the 180° twist. In both cases the electrostatics term dominates the changes in interaction energy upon the twist, analogously to base-base stacking.2–4 These preliminary calculations indicate that (deoxy)ribose...nucleobase stacking interactions in different interaction contexts are of similar origin.

5. Discussion related to interaction energy surfaces (IE)

The interaction energy (IE) surfaces show interaction energies at the distances defining the minEd surfaces (Figure S1). Thus, the distance at which the interaction energy is determined varies across the IE surface allowing to scan energetics of the area above the nucleobase more precisely compared to potential energy surfaces calculated at one fixed distance.
The shapes of IE surfaces differ from minEd surfaces because a smaller interaction distance does not necessarily correspond to a stronger, i.e., more negative, interaction energy and vice versa. The DHDF-D3 interaction energies at the surface minima are around −4 kcal.mol\(^{-1}\). Note that any calculated interaction energy depends to a certain extent on the computational protocol used for obtaining geometries. One generally expects a stronger interaction energy if the system is partially or fully relaxed, e.g., in our case without restricting the nucleobases to be planar and allowing a different spatial orientation of the molecular probe (see Surface scan limitations section). For our purpose, i.e., analyzing the limitations of AFF, we consider the chosen model system and scan strategy as a good compromise.

Overall, we consider the IE surfaces to be less important than the minEd surfaces for the evaluation of the AFF’s performance. Formation of (deoxy)ribose...nucleobase stacking is often a result of the molecular environment and is not solely due to the strength of the interaction per se especially when the interaction energy is weak. This is supported by the fact that IE surfaces are quite homogeneous across the nucleobase surface (Figure S4 and Figure S5). In contrast, imbalance in the short-range repulsion of the AFF description can affect MD simulations to a larger extent.

6. MD equilibration and simulation protocols

Z-DNA. The 300 K equilibration protocol involves first minimizing the energy of the solvent shell with a 500 kcal.mol\(^{-1}\)Å\(^{-2}\) restraint on the solute and was followed by minimization of the entire system with a 25 kcal.mol\(^{-1}\)Å\(^{-2}\) solute restraint. Then, the minimized structure was used for three independent MD simulations, where the system was heated to 100 K during a 20 ps MD simulation at constant volume with a 25 kcal.mol\(^{-1}\)Å\(^{-2}\) restraint on the solute, and to 300 K during a 25 ps MD simulation at 1 bar. The final equilibration step consisted of 100 ps MD simulation at 300 K and 1 bar. In order to control temperature and pressure the Langevin thermostat and Berendsen barostat were used.\(^{5,6}\)

For the 100 K simulations, after following the minimization procedure described above, system restraints of 25 kcal.mol\(^{-1}\)Å\(^{-2}\) were applied over three 50 ps NVT MD simulations at 300 K where the first 25 ps were used for slowly heating the system to the target temperature. Then the restraints were reduced to 10 kcal.mol\(^{-1}\)Å\(^{-2}\) and MD simulations at 1 bar were run for 50 ps. The temperature was slowly reduced to 100 K over the first 25 ps. The final equilibration step consisted in a 400 ps NPT MD simulation at 100 K and 1 bar. The 300 K heating procedure is an attempt to escape local minima. The total size of the system was ~4600 atoms for each MD simulation. 1 μs simulation of system (i.e., 5·10\(^8\) steps) of this size takes around ~2 days using one GeForce GTX 1080 Ti GPU card. Convergences of pressure, temperature and volume were checked for all simulations.

r(UUCG). First step of the equilibration protocol consisted of the energy minimization of the solvent shell with a 500 kcal.mol\(^{-1}\)Å\(^{-2}\) restraints on the solute. Then, the solvent shell was relaxed by the 500 ps MD simulation at 1 bar and 298 K. In next steps, energy of the solute was minimized in several runs with the restraining constant applied on the sugar-phosphate backbone decreasing from 1000 kcal.mol\(^{-1}\)Å\(^{-2}\) to 25 kcal.mol\(^{-1}\)Å\(^{-2}\). Then, the system was heated to 298 K in 100 ps MD simulation at constant volume. Another 100 ps of MD simulation at 1 bar and 298 K was applied to relax the total density of the system. Production runs were performed by the CUDA-driven pmemd module (SPFP-implementation) of Amber with standard SHAKE restraints\(^8\) under NPT (a 2 fs time step) or NVT (with a 4-fs integration time
step using hydrogen mass repartitioning (HMR)\textsuperscript{9}) conditions for simulation 1 (MD1) and simulations 2 and 3 (MD2-3), respectively, at 298 K using Berendsen thermostat (and barostat for NPT) with coupling time of 1 ps.\textsuperscript{5} We used cubic OPC\textsuperscript{10} explicit water periodic box with a 12 Å buffer region and the Joung – Cheatham ion parameters for TIP4P/EW water.\textsuperscript{11} We used two different ion concentration, namely ~1.0 M KCl salt-excess concentrations (112 K\textsuperscript{+} and 99 Cl\textsuperscript{−} added) for MD1 and ~0.15 M KCl (25 K\textsuperscript{+} and 12 Cl\textsuperscript{−} added) for MD2-3. Every 10\textsuperscript{th} snapshot of the MD simulations was analyzed. In total, 5\cdot10\textsuperscript{9} (standard 2 fs time step) and 2.5\cdot10\textsuperscript{9} (the HMR\textsuperscript{9} allowing a 4-fs integration time step) simulation steps were performed with a total size of the system of ~16,000 and ~19,000 atoms for MD1 and MD2-3, respectively. 1 \mu s-long MD simulation of this tetraloop system takes around ~2 days using one GeForce RTX 2080 Ti GPU card. For more details, see Mráziková et al.\textsuperscript{12}
Table S1. Harmonic penalty restraints for geometry optimizations. Penalty function restraints used to keep the nucleobases planar and to keep the monomer orientation as in the CpG Z-DNA fragment. The goal of the restraints is to relax the monomers internally as much as possible while keeping their intermolecular target conformation.\textsuperscript{14}

| Type       | Atom 1 | Atom 2 | Atom 3 | Atom 4 | E (r) / a.u. |
|------------|--------|--------|--------|--------|--------------|
| DME…G\textsuperscript{\alpha} | torsion C8 O4' (DME) C (DME) N1 | 0.1 |
| torsion C6 C5 N7 C8 | 1.0 |
| torsion C6 N1 C2 N3 | 1.0 |
| torsion C2 N3 C4 N9 | 1.0 |
| torsion C2 H21 H22 N2 | 1.0 |
| torsion N2 C2 N1 H(N1) | 1.0 |
| torsion C4 N9 C8 H(C8) | 1.0 |
| torsion C4 N9 C8 | 1.0 |
| torsion C6 C5 C4 N9 | 1.0 |
| torsion H9 N9 C8 H8 | 1.0 |
| torsion O6 C6 N1 H(N1) | 1.0 |

| DME…A |torsion N1 O4' (DME) C (DME) O4' (DME) N7 | 0.1 |
| torsion C6 N1 C2 N3 | 1.0 |
| torsion N3 C4 N9 H9 | 1.0 |
| torsion N9 C8 N7 C5 | 1.0 |
| torsion N7 C5 C4 N9 | 1.0 |
| torsion H9 N9 C8 H8 | 1.0 |
|torsion N6 C6 C5 N7 | 1.0 |
| torsion N6 C6 C5 N7 | 1.0 |
| torsion H62 N6 C6 C5 | 1.0 |
| torsion H61 N6 C6 N1 | 1.0 |

| DME…U |torsion N3 O4' (DME) C (DME) O4' (DME) C3' | 0.1 |
| torsion C4 N3 C2 N1 | 1.0 |
| torsion C2 N1 C6 C5 | 1.0 |
| torsion O2 C2 N1 H1 | 1.0 |
| torsion H1 N1 C6 H6 | 1.0 |
| torsion H6 C6 C5 H5 | 1.0 |
| torsion H6 C6 C5 H5 | 1.0 |
| torsion H41 N4 C4 N3 | 1.0 |

| DME…C |torsion N3 O4' (DME) C (DME) O4' (DME) C3' | 0.1 |
|torsion C4 N3 C2 N1 | 1.0 |
| torsion O2 C2 N1 H1 | 1.0 |
| torsion H1 N1 C6 H6 | 1.0 |
|torsion H6 C6 C5 H5 | 1.0 |
|torsion H6 C6 C5 H5 | 1.0 |
| torsion H41 N4 C4 N3 | 1.0 |

| Deoxyribose…G | torsion (γ) O5' C5' C4' C3' O3' | 0.1 |
|torsion (δ) C5' C4' C3' C6 | 0.1 |
| torsion N3 O4' C4' C5 | 0.1 |
| torsion C2 C1' O4' C5 | 0.1 |
|torsion N3 C1' O4' C6 | 0.1 |

| Deoxyribose…C | torsion H41 N4 C4 C5 | 1.0 |
|torsion H42 N4 C4 C5 | 1.0 |
|torsion H42 N4 C4 C5 | 1.0 |
|torsion H41 N4 C4 N3 | 1.0 |

| Deoxyribose…C | torsion (γ) O5' C5' C4' C3' O3' | 0.1 |
|torsion (δ) C5' C4' C3' C6 | 0.1 |
|torsion N1 O4' C4' N9 | 0.1 |
|torsion C2 C1' O4' N7 | 0.1 |

\textsuperscript{a} data taken from Kruse et al.\textsuperscript{1}
Table S2. Interpolated global minima for the DHDF-D3 and AFF miniEd and IE surfaces. The AFF minima values for the miniEd surfaces are larger compared to those derived from DHDF-D3 surfaces. The DHDF-D3 and AFF minima values for IE surfaces are similar although they are weaker for AFF.

| Method | DME...G | DME...A | DME...C | DME...U |
|--------|---------|---------|---------|---------|
|        | miniEd surface minima (Å) |         |         |         |
| DHDF-D3 | 2.97<sup>a</sup> | 2.95 | 2.91 | 2.92 |
| AFF     | 3.10 | 3.10 | 3.12 | 3.10<sup>b</sup> |
|        | IE surface minima (kcal.mol<sup>-1</sup>) |         |         |         |
| DHDF-D3 | −3.9 | −4.2 | −4.4 | −4.0 |
| AFF     | −3.5 | −3.9 | −3.8 | −3.6 |

<sup>a</sup>Taken from Kruse et al.<sup>1</sup>, <sup>b</sup>the lowest value on the miniEd surface, minimum outside of uracil and undefined (see "AFF miniEd surfaces are too repulsive and shallow compared to the DHDF-D3 reference" in the main text).
Table S3. The $\textit{minEd}$ and $\textit{IE}$ minima values for DME and deoxyribose...G/C systems. DHDF-D3 and AFF $\textit{minEd}$ and $\textit{IE}$ minima values of the vertical interaction energy scans performed above the pyrimidine ring centroid for DME...G and deoxyribose...G, and above the C2 atom for DME...C and deoxyribose...C, respectively.

| Method | DME...G | Deoxyribose...G | DME...C | Deoxyribose...C |
|--------|--------|-----------------|--------|-----------------|
|        | $\textit{minEd}$ values (Å) | $\textit{IE}$ values (kcal.mol$^{-1}$) |        |        |
| DHDF-D3 | 2.97$^a$ | 2.87 |        |        |
| AFF    | 3.10 | 3.12 | 3.15 | 3.06 |
| DHDF-D3 | −3.8 | −4.5 | −3.2 | −2.9 |
| AFF    | −3.5 | −4.3 | −2.3 | −2.5 |

$^a$ taken from Kruse et al.$^1$
Table S4. SAPT2+(3)δMP2/jun-cc-pVTZ energy values for different orientations of DME above the G nucleobase. The value of the clockwise rotation around the z-axis is given in brackets, e.g., DME...G (0°): original structure without rotation. Models are shown in Figure S6. Energy values are in kcal.mol\(^{-1}\). The percentages indicate the contributions of the attractive components. The DME oxygen atom is located above the pyrimidine ring centroid.

| Interaction type | DME...G (0°) \(\Delta E\) | DME...G (90°) \(\Delta E\) | DME...G (180°) \(\Delta E\) | DME...G (270°) \(\Delta E\) |
|-----------------|--------------------------|--------------------------|--------------------------|--------------------------|
| \(\Delta E_{\text{elct}}\) | -2.5 (22 %) | -4.4 (32 %) | -3.2 (27 %) | -2.9 (26 %) |
| \(\Delta E_{\text{disp}}\) | -8.3 (73 %) | -8.4 (62 %) | -8.1 (67 %) | -7.9 (69 %) |
| \(\Delta E_{\text{ind}}\) | -0.6 (5 %) | -0.7 (5 %) | -0.7 (6 %) | -0.6 (5 %) |
| \(\Delta E_{\text{exch-rep}}\) | +7.7 | +8.3 | +7.6 | +7.8 |
| \(\Delta E_{\text{total SAPT}}\) | \(-3.6\) | \(-5.1\) | \(-4.4\) | \(-3.6\) |
Table S5. AMBER atom types and Lennard-Jones parameters for nucleobase atoms. Data are taken from AmberTools21\textsuperscript{14} for AMBER force field OL15\textsuperscript{15} (DNA) and OL3\textsuperscript{16} (RNA). The Lennard-Jones parameters in these force fields are identical to the original Cornell et al. AMBER parameters.\textsuperscript{17} The atom types from the original AMBER parameterization are written in the parenthesis in the cases where they differ from the OL15 or OL3 force fields. The carbon atom parameters are highlighted in grey. Like the nitrogen and oxygen atom parameters, all carbon atom parameters are identical. Thus, only three couples of R and ε values, designed in 1995,\textsuperscript{17} are associated with the many AMBER atom types used to describe the N, C and O nucleobase atoms. Note that the nucleobase carbon LJ parameters are also identical to those associated with the CA atom type used for benzene rings.\textsuperscript{17} Note also that the sp\textsuperscript{2} carbon atoms have larger vdW radii than the nitrogen atoms.

| Nucleobase atom | AMBER atom type | R   | ε   |
|-----------------|-----------------|-----|-----|
| Guanine         |                 |     |     |
| N1              | NA              | 1.8240 | 0.1700 |
| C2              | CA              | 1.9080 | 0.0860 |
| N3              | NC              | 1.8240 | 0.1700 |
| C4              | CB              | 1.9080 | 0.0860 |
| C5              | CB              | 1.9080 | 0.0860 |
| C6              | C               | 1.9080 | 0.0860 |
| N7              | NB              | 1.8240 | 0.1700 |
| C8              | CK/CP\textsuperscript{a} (CK) | 1.9080 | 0.0860 |
| N9              | NB\textsuperscript{a} | 1.8240 | 0.1700 |
| N2              | N2              | 1.8240 | 0.1700 |
| O6              | O               | 1.6612 | 0.2100 |
| Adenine         |                 |     |     |
| N1              | NC              | 1.8240 | 0.1700 |
| C2              | CQ              | 1.9080 | 0.0860 |
| N3              | NC              | 1.8240 | 0.1700 |
| C4              | CB              | 1.9080 | 0.0860 |
| C5              | CB              | 1.9080 | 0.0860 |
| C6              | CA              | 1.9080 | 0.0860 |
| N7              | NB              | 1.8240 | 0.1700 |
| C8              | C2/C4\textsuperscript{a} (CM) | 1.9080 | 0.0860 |
| N9              | N\textsuperscript{a} | 1.8240 | 0.1700 |
| N6              | N2              | 1.8240 | 0.1700 |
| Cytosine        |                 |     |     |
| N1              | N\textsuperscript{a} | 1.8240 | 0.1700 |
| C2              | C               | 1.9080 | 0.0860 |
| N3              | NC              | 1.8240 | 0.1700 |
| C4              | CA\textsuperscript{a} | 1.9080 | 0.0860 |
| C5              | C1/C4\textsuperscript{a} (CM) | 1.9080 | 0.0860 |
| C6              | C1/C4\textsuperscript{a} (CM) | 1.9080 | 0.0860 |
| O2              | O               | 1.6612 | 0.2100 |
| N4              | N2              | 1.8240 | 0.1700 |
| Uracil          |                 |     |     |
| N1              | N\textsuperscript{a} | 1.8240 | 0.1700 |
| C2              | C               | 1.9080 | 0.0860 |
| N3              | NA              | 1.8240 | 0.1700 |
| C4              | C               | 1.9080 | 0.0860 |
| C5              | C5 (CM)         | 1.9080 | 0.0860 |
| C6              | C5 (CM)         | 1.9080 | 0.0860 |
| O2              | O               | 1.6612 | 0.2100 |
| O4              | O               | 1.6612 | 0.2100 |
| Benzene         |                 |     |     |
| C               | CA              | 1.9080 | 0.0860 |

\textsuperscript{a}Atom types differ for DNA and RNA AFFs, data are written as OL15\textsuperscript{15} (DNA)/OL3\textsuperscript{16} (RNA).
Table S6. MD simulation analysis of CpG steps in Z-DNA and r(UUCG) that adopt a Z-step conformation. Populations (Pop.) in % and average O4’…Guanine distances (Avg. d.) in Å calculated for these populations are listed in the table for those CpG dinucleotides (for Z-DNA only C<sub>3</sub>pG<sub>4</sub> and C<sub>9</sub>pG<sub>10</sub> are considered) which satisfy Z-step criteria and dihedrals criteria. The Z-step criteria include i) glycosidic χ angle of the guanosine is in range 30°-90°, ii) cytidine (deoxy)ribose pucker in C2’-endo conformation and iii) O4’…Guanine distance smaller than 3.5 Å. Dihedrals criteria include backbone dihedrals values close to the experimental values. The Z-DNA data are obtained from six 1 μs-long MD trajectories (100 K and 300 K MD1, MD2 and MD3). The r(UUCG) data are obtained from 4.5 μs (MD1), 2.4 μs (MD2) and 2.8 μs (MD3) long parts of MD trajectories.

|                | Z-DNA              | r(UUCG)          |
|----------------|--------------------|-----------------|
|                | 100 K              | 300 K           |
|                | MD1                | MD1             |
| 3-4            | 100                | 86.7            |
| 9-10           | 100                | 85.1            |
| Avg.           | 100                | 85.4            |
| Z-step criteria & dihedrals criteria (χ of G close to exp. & C2’-endo & Dihedral close to exp. & < 3.5 Å) | 100% of the analyzed CpG steps |
| α of guanosine close to exp. value | β pseudorotation angle 144°-180° |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
| Avg.           | 3-4                | 9-10            |

|                | Z-DNA              | r(UUCG)          |
|----------------|--------------------|-----------------|
|                | 100 K              | 300 K           |
|                | MD1                | MD1             |
| 3-4            | 98.2               | 86.7            |
| 9-10           | 97.3               | 85.1            |
| Avg.           | 98.2               | 85.4            |
| Z-step criteria: • χ of G close to exp.; • C2’-endo; • < 3.5 Å; | Z-step criteria: • χ of G close to exp. & C2’-endo & Dihedral close to exp. & < 3.5 Å |
|               | 100% of the analyzed CpG steps |
| O4’...guanine distance < 3.5 Å |  
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
| Avg.           | 3-4                | 9-10            |

|                | Z-DNA              | r(UUCG)          |
|----------------|--------------------|-----------------|
|                | 100 K              | 300 K           |
|                | MD1                | MD1             |
| 3-4            | 98.3               | 86.7            |
| 9-10           | 98.5               | 85.1            |
| Avg.           | 98.4               | 85.4            |
| Cytidine pucker C2’-endo |  
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
| Avg.           | 3-4                | 9-10            |

|                | Z-DNA              | r(UUCG)          |
|----------------|--------------------|-----------------|
|                | 100 K              | 300 K           |
|                | MD1                | MD1             |
| 3-4            | 98.5               | 86.7            |
| 9-10           | 98.4               | 85.1            |
| Avg.           | 98.4               | 85.4            |
| α of guanosine close to exp. value | β pseudorotation angle 144°-180° |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
| Avg.           | 3-4                | 9-10            |

|                | Z-DNA              | r(UUCG)          |
|----------------|--------------------|-----------------|
|                | 100 K              | 300 K           |
|                | MD1                | MD1             |
| 3-4            | 98.2               | 86.7            |
| 9-10           | 97.3               | 85.1            |
| Avg.           | 98.2               | 85.4            |
| Cytidine pucker C2’-endo |  
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
| Avg.           | 3-4                | 9-10            |

|                | Z-DNA              | r(UUCG)          |
|----------------|--------------------|-----------------|
|                | 100 K              | 300 K           |
|                | MD1                | MD1             |
| 3-4            | 98.3               | 86.7            |
| 9-10           | 98.5               | 85.1            |
| Avg.           | 98.4               | 85.4            |
| O4’...guanine distance < 3.5 Å |  
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
|                | 3-4                | 9-10            |
| Avg.           | 3-4                | 9-10            |

—exp. value taken from ref. 14 ± 30°, —pseudorotation angle 144°–180°
**Figure S1.** Schematic representation explaining the construction of the \textit{minEd} and \textit{IE} surfaces. 

\textbf{a)} Vertical interaction energy scan, i.e., interaction energy dependence on the vertical distance between O4’ atom and nucleobase plane, in this case shown for the O4’ localized over the G(centroid). The value of the interaction energy minimum (−3.79 kcal.mol$^{-1}$) is used for the construction of the \textit{IE} surface, while the distance at which the interaction energy minimum occurs (\textit{minEd} value; 2.97 Å) is used for the construction of the \textit{minEd} surface (see also Figure 3). 

\textbf{b)} Horizontal O4’(DME) probe displacements (red dots) showing the in-plane points chosen for the vertical scans. Only ring atoms of purine (up) and pyrimidine (down) bases are shown and labeled. For pyrimidine bases, one additional point near the C2 atom shown by an arrow is used for DHDF-D3 scanning to seek the position of the \textit{minEd} surface minima.
Figure S2. Comparison of DHDF-D3 and AFF vertical scans for DME and deoxyribose...G/C models. The DME...nucleobase vertical scans are taken from Figure 3. The DSD-BLYP-D3/ma-def2-QZVPP method (DHDF-D3; red) and AFF (blue) used for the construction of miniEd and IE surfaces, and the estimated FNO-CCSD(T)/CBS method (black) used as QM reference are shown in the graph. The deoxyribose...G/C scans are indicated by dashed lines. The short-range difference between AFF and DHDF-D3 calculations is visible for both DME and deoxyribose models.
Figure S3. Comparison of the SAPT2+(3)δMP2/jun-cc-pVTZ of DME...C for O4'...C(C2) and O4'...C(C5) positions. These SAPT calculations show difference in the exchange-repulsion term (red) for the two O4' positions. The much smaller exchange-repulsion for the O4'...C(C2) compared to the O4'...C(C5) suggests significant deviation between the C2 and C5 atomic volumes, supporting the conclusions based on AIM α(0) polarizabilities that the C2 atomic volume is smaller than that of the C5 atom. The red dotted lines and red arrows are added to emphasize the large difference in vertical distances at which the exchange-repulsion curve reaches 12 kcal.mol$^{-1}$. Additionally, note that the smaller London dispersion term (violet; related to polarizabilities of the interacting atoms) for the C2 atom corresponds to the smaller polarizability value of the C2 atom compared to the C5 atom.
Figure S4. Comparison of DHDF-D3 and AFF IE surfaces for DME...G/A. Nucleobase coordinates are identical to those used for minIEd surfaces, i.e., the nucleobases lie in the xy plane. The color scale represents the interaction energy between the nucleobase and DME at vertical distances represented by minIEd surfaces (see Figure 4), the contours represent isoline stepping of 0.25 kcal.mol$^{-1}$. Color scale is identical for both plots to ease visual comparison. Black arrows and energy values point to minima of the IE surfaces that are marked by red dots; small black dots mark the points where the vertical scans were calculated. DME...G data are taken from Kruse et al.$^1$
Figure S5. Comparison of DHDF-D3 and AFF IE surfaces for DME...C/U. Nucleobase coordinates are identical to those used for miniEd surfaces, i.e., the nucleobases lie in the xy plane. The color scale represents the interaction energy between the nucleobase and DME at vertical distances represented by miniEd surfaces (see Figure 5), the contours represent isoline stepping of 0.25 kcal.mol$^{-1}$. Color scale is identical for both plots to ease visual comparison. Black arrows and energy values point to minima of the IE surfaces that are marked by red dots; small black dots mark the points where the vertical scans were calculated.
Figure S6. Rotation of DME by 90° steps with the O4'(DME) atom located above the guanine pyrimidine ring centroid. Guanine atoms lie in xy-plane; O4'(DME) and centroid of pyrimidine ring lie on the z-axis. DME is rotated about 90°, 180° and 270°. Vertical distance of O4' from centroid (xy-plane) is 2.95 Å. a) DME...G (0°); b) DME...G (90°); c) DME...G (180°); d) DME...G (270°).
References

(1) Kruse, H.; Mrazikova, K.; D’Ascenzo, L.; Sponer, J.; Auffinger, P. Short but Weak: The Z-DNA Lone-Pair⋯π Conundrum Challenges Standard Carbon Van der Waals Radii. *Angew. Chemie - Int. Ed.* **2020**, *59*, 16553–16560. https://doi.org/10.1002/anie.202004201.

(2) Florian, J.; Sponer, J.; Warshel, A. Thermodynamic Parameters for Stacking and Hydrogen Bonding of Nucleic Acid Bases in Aqueous Solution: Ab Initio/Langevin Dipoles Study. *J. Phys. Chem. B* **1999**, *103*, 884–892.

(3) Sponer, J.; Sponer, J. E.; Mladek, A.; Jurecka, P.; Banas, P.; Otyepka, M. Nature and Magnitude of Aromatic Base Stacking in DNA and RNA: Quantum Chemistry, Molecular Mechanics, and Experiment. *Biopolymers* **2013**, *99*, 978–988. https://doi.org/10.1002/bip.22322.

(4) Kruse, H.; Šponer, J. Revisiting the Potential Energy Surface of the Stacked Cytosine Dimer: FNO-CCSD(T) Interaction Energies, SAPT Decompositions, and Benchmarking. *J. Phys. Chem. A* **2019**, *123*, 9209–9222. https://doi.org/10.1021/acs.jpca.9b05940.

(5) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690. https://doi.org/10.1063/1.448118.

(6) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylglycine-N′-methylamide. *Biopolymers* **1992**, *32*, 523–535. https://doi.org/10.1002/bip.360320508.

(7) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380. https://doi.org/10.1016/j.cpc.2012.09.022.

(8) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(9) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874. https://doi.org/10.1021/ct5010406.

(10) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: A different approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871. https://doi.org/10.1021/jz501780a.

(11) Joung, J. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041. https://doi.org/10.1021/jp8001614.

(12) Mráziková, K.; Mlynšký, V.; Kührová, P.; Pokorná, P.; Kruse, H.; Krepl, M.; Otyepka, M.; Banáš, P.; Šponer, J. UUGC RNA Tetraloop as a Formidable Force-Field Challenge for MD Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 7601–7617.

(13) Kruse, H.; Sponer, J. Towards biochemically relevant QM computations on nucleic acids: controlled electronic structure geometry optimization of nucleic acid structural
motifs using penalty restraint functions. *Phys. Chem. Chem. Phys.* **2015**, *17*, 1399–1410. https://doi.org/10.1039/C4CP04680C.

(14) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O’Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. Amber 2021. University of California, San Francisco 2021.

(15) Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham III, T. E.; Galindo-Murillo, R.; Jurecka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736. https://doi.org/10.1021/acs.jctc.5b00716.

(16) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham III, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902. https://doi.org/10.1021/ct200162x.

(17) 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. https://doi.org/10.1021/ja00124a002.

(18) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res.* **2008**, *36*, 3690–3706. https://doi.org/10.1093/nar/gkn260.