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 Kruse[1,*]

[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

Abstract

The lone-pair…π (lp…π) (deoxy)ribose…nucleobase stacking is a recurring structural motif in Z-DNA and RNAs that is characterized by sub-van der Waals lp…π contacts (<3.0 Å). It is part of the structural signature of the CpG Z-steps in Z-DNA and r(UNCG) tetraloops. These nucleic acid structures are poorly behaving in molecular dynamics (MD) simulations. Although the exact origin of these issues remains unclear, a significant part of the problem might be due to an imbalanced description of non-bonded interactions including the characteristic lp…π stacking. To gain insights into the links between lp…π stacking and MD issues, we present an in-depth comparison between accurate large-basis-set double-hybrid Kohn-Sham density functional theory calculations DSD-BLYP-D3/ma-def2-QZVPP (DHDF-D3) and data obtained with the non-bonded potential of the AMBER force field (AFF) for NpN Z-steps (N = G, A, C, U). Among other differences, we found that the AFF overestimates the DHDF-D3 lp…π distances by ~0.1-0.2 Å while the deviation between the DHDF-D3 and AFF descriptions sharply increases in the short-range region of the interaction. Based on atom-in-molecule (AIM) polarizabilities and SAPT analysis, we inferred that the DHDF-D3 vs. AFF differences partly originate in the Lennard-Jones (LJ) parameters that are identical for nucleobase carbon atoms despite the presence/absence of connected electron withdrawing groups that lead to different effective volumes or vdW radii. Thus, to precisely model the very short CpG lp…π contact distances, we recommend revision of the nucleobase atom LJ parameters. Additionally, we suggest that the large discrepancy between DHDF-D3 and AFF short-range repulsive part of the interaction energy potential may significantly contribute to the poor performances of MD simulations of nucleic acid systems containing Z-steps. Understanding where, and if possible why, the point-charge-type effective potentials reach their limits is vital for developing next-generation FFs and for addressing specific issues in contemporary MD simulations.

Introduction

DNA and RNA are key biological molecules whose shapes are primarily defined by nucleobase…nucleobase,[1] nucleobase…phosphate[2] and (deoxy)ribose…nucleobase[3] interactions. A sub-category of the latter interactions involves a T-shaped (deoxy)ribose…nucleobase stacking.[4,5] This interaction has been first identified in Z-DNA CpG steps[6] and was subsequently named lone-pair…π or lp…π interaction[7] while the associated CpG steps were called Z-steps.[4] A typical CpG Z-step involves a T-shaped stacking arrangement of the cytidine (deoxy)ribose with the guanine base[4] (**Figure 1**). Recent DNA and RNA PDB surveys[4,5] revealed the occurrence of 'Z-like' steps that correspond to dinucleotide sequences involving any combination of the four nucleobases and possessing structural characteristics similar to those of a CpG Z-step, i.e. a 3'-nucleotide in a *syn* conformation, a 5'-nucleotide (deoxy)ribose with a *C2'-endo* pucker and a sub-van der Waals (sub-vdW) lp…π contact below 3.2 Å. The recurrence of short T-shaped (deoxy)ribose…nucleobase contacts that occur in Z-like steps or between non-consecutive nucleotides stresses the structural relevance of lp…π motifs in nucleic acids.[4,5,8]

Detailed PDB surveys of Z-like steps have shown that in some instances the O4' atom points towards the nucleobase with contact distances ≤ 3.0 Å and sometimes close to 2.8 Å.[4,8] These salient stacking distances, which are among the shortest distances observed between an oxygen atom and a planar ring, raised questions regarding the strength and nature of the (deoxy)ribose…nucleobase stacking and in general of lp…π contacts.[4,5,7,9,10] Do orbital interactions contribute to their stabilization or is the common non-covalent interaction framework sufficient to describe lp…π contacts?[9–17]

Recently, we performed energy decomposition analysis via high-level symmetry-adapted perturbation theory (SAPT)[17] on a deoxyribose…guanine system that showed that the lp…π contact is mainly stabilized by London dispersion and to a lesser extent by electrostatic interactions. Thus, it became clear that the deoxyribose…nucleobase stacking is essentially a common non-covalent interaction. Moreover, we found that the origins of the observed sub-vdW contacts cannot be explained by orbital effects as the "lp…π interaction" terminology would imply.[17] Instead, we established that the sub-vdW lp…π contacts could be rationalized through atom-in-molecule (AIM) α(0) polarizabilities, a measure that provides an estimate of effective atomic volumes.[18,19] The calculated α(0) polarizabilities revealed that atomic volumes of the sp^2 carbon atoms of a guanine nucleobase do vary as a function of their chemical environment and that the carbon atoms connected to electron withdrawing groups are among the smallest of the nucleobase. Indeed, a small effective vdW radius for a sp^2 carbon may explain a lp…π contact distance in the 2.8–3.0 Å range that is much shorter than the 3.3–3.4 Å vertical distance between stacked base pairs in canonical double helical systems,[20] the latter distance being twice the 1.70 Å vdW radius of an sp^3 carbon mentioned in the Bondi tabulation dated from 1964 and still in use.[17,21–23] The smaller vdW radius of some sp^2 carbons may also explain the absence of orbital effects.

Understanding vdW radii variations has far reaching consequences, one of them being the calibration of biomolecular force fields (FF) used for molecular dynamics (MD) simulations. The AMBER FF (AFF in the following), based on the 1995 Cornell et al. parametrization,[24] is currently one of the most popular nucleic acid force fields.[25–30] Most of the recent AFF modifications focused on adapting the backbone dihedral terms, which is one of the most straightforward tuning methods.[25] Additional modifications of non-bonded terms have been suggested, including direct modifications of vdW parameters of phosphate group

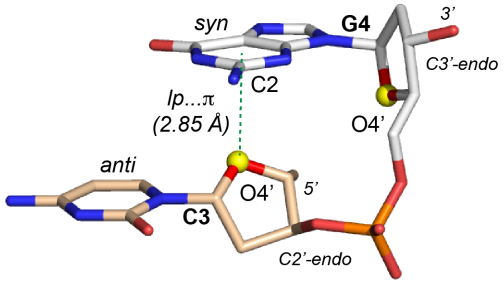

Figure 1. Typical CpG dinucleotide Z-step with a T-shaped deoxyribose…nucleobase stacking arrangement as occurring in Z-DNA and r(UNCG) tetraloops.[4,8] The represented C3pG4 step is extracted from the ultra-high resolution (0.54 Å) Z-DNA structure (PDB 3P4J).[17,72] In this arrangement, the lp…π stacking contact distance (or O4' atom to guanine plane distance) is 2.85 Å and is represented by a dashed line.

oxygen atoms, adjustments of Lorentz-Berthelot combination rules for Lennard-Jones (LJ) solute-solute and solute-solvent interactions as well as the addition of simple H-bond interaction tuning terms.[31–38] However, none of these modifications have been widely tested and no attempts have been made to address lp…π parametrizations in nucleic acids.

Among nucleic acid systems, the Z-steps containing Z-DNA helices and r(UNCG) tetraloops are especially challenging for MD simulations.[32,34,37,39–45] Previous studies exploring the behavior of Z-DNA in MD simulations focused mainly on the dynamics of backbone dihedrals and helical parameters.[39,41,46–48] A recent study that analyzed the disruption of the r(UUCG) tetraloop canonical structure occurring in μs long MD simulations focused more precisely on intramolecular interactions within the loop region and traced several potential FF issues, the ribose…guanine lp…π stacking being one of those.[40] To this point, it is unknown to what extent a potentially imbalanced description of lp…π contacts causes MD simulations of Z-step-containing systems to ill-behave.

The purely non-covalent nature of (deoxy)ribose…nucleobase lp…π contacts indicates that they could be reasonably modelled by the AFF, as previously suggested for base…base stacking.[49,50] However, the (deoxy)ribose…nucleobase stacking includes short lp…π contacts where the simple LJ potential starts to deviate from more rigorous descriptions.[51,52] Thus, the approximate nature of the empirical potential might lead to errors affecting the outcome of the simulations as observed for some specific base…base stacking motifs.[46,50,53] A few well-known issues arising from the oversimplified non-bonded potential of the pairwise-additive FFs are: *i)* the use of a r^{-12} term for the repulsive part of LJ potential that lead to an overrepulsive short-range region[51,52] instead of a more accurate exponential distance-dependence as found in a Buckingham potential;[54] *ii*) missing atomic anisotropicity; *iii)* missing multi-body effects for the London dispersion component of the LJ term; and *iv)* missing (explicit) polarization and charge-penetration terms as a consequence of using fixed atomic point charges.[30,55–60] Regarding deficiencies of the fixed point-charge based FFs for short-range electrostatics, various correction schemes were proposed.[61–64] Notable advances for polarizable FFs have been made.[61,65–67] Recent QM derived FF approximations provided excellent results for the treatment of short-range electrostatics.[68,69] However, these rather computationally demanding FFs are not suited for explicit-solvent MD simulations of large biomolecules.

Herein, we extend our previous QM characterization of the (deoxy)ribose…nucleobase stacking in CpG Z-steps[17] to all NpN (N = G, A, C, U) Z-like steps by varying the 5' nucleobase

(the 3' nucleotide can be of any type since it establishes a non-specific contact to the 5' nucleobase through its O4' (deoxy)ribose atom). We use highly-accurate dispersion-corrected double-hybrid density functional (DHDF-D3) calculations to evaluate the performance of the non-bonded potential of the AFF by constructing *minimum interaction energy distance* (*minIEd*) surfaces. For that purpose, and as described earlier,[17] we approximated the full (deoxy)ribose by a dimethylether (DME) molecular probe. Rationalization of the DHDF-D3 surface topologies is made by using atom-in-molecule (AIM) polarizabilities.[70,71] To further assess the performance of the AFF, these data are compared against symmetry-adapted perturbation theory (SAPT) computations. Lastly, we analyze the behavior of Z-steps in MD simulations of Z-DNA and r(UUCG) tetraloop systems by employing the latest AFF versions. We found that the massively sampled short O4'…guanine plane distances are likely to cause imbalances in the simulations because of the overestimated short-range repulsion originating from the LJ potential of the AMBER FF. This led us to suggest that the usage of unified LJ parameters for all nucleic acids carbon atoms contribute to a yet unrecognized degree to the AMBER FF errors.

Computational Details

Structural models: As described earlier,[17] a CpG Z-step (see **Figure 1**) was extracted from a 0.54 Å ultra-high resolution Z-DNA crystallographic structure (PDBid: 3P4J;[72] only the deoxyribose of residue 3 and the guanine of residue 4 were kept). Then, the deoxyribose was changed into dimethylether (DME) while keeping the C-O4'-C atoms fixed in space to preserve the X-ray deoxyribose…guanine orientation (in the following, the DME oxygen atom will be named O4'). The DME orientation was maintained for all nucleobases (note that the DME orientation may differ for lp…π contacts involving non-consecutive nucleotides).[5] To study all (deoxy)ribose…nucleobase stacking types, G was substituted by each of the three A, C or U nucleobases (**Figure 2**). Most of the computational protocol followed procedures described in earlier work.[17]

QM structure optimization: Prior to making rigid-monomer interaction energy scans, the DME-nucleobase systems were gradient-optimized by the hybrid density functional approximation B3LYP[73] with DFT-D3(BJ)[74,75] dispersion correction and the def2-QZVP[76] atomic orbital basis set (B3LYP-D3/def2-QZVP). Harmonic penalty restraints as implemented in our in-house optimizer[77,78] were used to keep the initial DME…nucleobase fragment orientation (restraints are listed in **Table S1** of the **Supporting Information**). Turbomole[79,80] energy and gradient calculations employed an energy change threshold of 10^{-7} E_h for the SCF and the total energy convergence. The RI-JK integral approximation (density-fitting) was used with the m4 DFT quadrature grid.[81,82] The XYZ coordinates for the optimized structures are provided in the **Supporting Information**.

Reference *ab initio* CCSD(T) and SAPT calculations: PSI4[83] was employed for density fitted FNO-CCSD(T)[84] and density fitted SAPT2+(3)δMP2[85–87] computations using the frozen core approximation. Extrapolation to complete basis set (CBS) FNO-CCSD(T) computations was automated with PSI4 using the formula:

$$E_{CCSD(T)/CBS} = E_{HF/aQ} + E_{MP2/CBS(aT,aQ)} + E_{FNO-CCSD(T)/junTZ} - E_{MP2/junTZ} \qquad (1)$$

where aT, aQ and $junTZ$ represent the aug-cc-pVTZ, aug-cc-pVQZ and jun-cc-pVTZ basis sets, respectively.[88–90] *MP2/CBS(at,aQ)* indicates the standard 2-point extrapolation following Halkier et al.[91] SAPT2+(3)δMP2 calculations also employed the jun-cc-pVTZ basis set. Both,

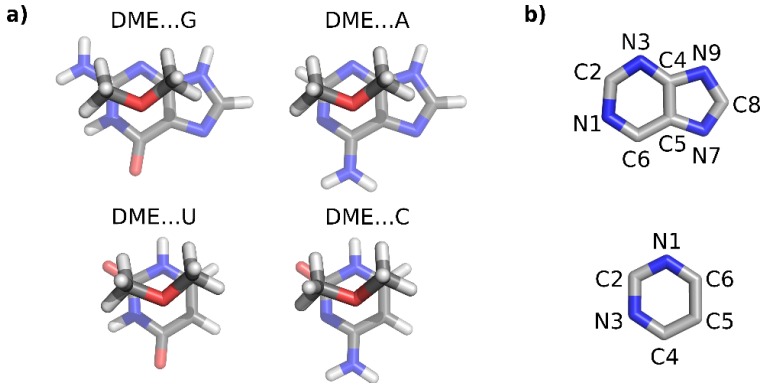

Figure 2. a) Top-down view on model systems showing DME…G,[17] DME…A, DME…U and DME…C with DME located above the pyrimidine ring centroid. **b)** Purine (top) and pyrimidine (bottom) ring atom numberings.

CCSD(T) and the mentioned SAPT theory level are highly accurate QM methods if appropriate basis sets are used and are labelled as 'gold standard' for the evaluation of non-covalent interactions.[87,92] The CCSD(T) data were used to validate the DHDF-D3 approach (see below). SAPT calculations were used to decompose the interaction energy into various components.

QM DHDF-D3 single point energy computations: Energies were computed using the double-hybrid density functional approximation DSD-BLYP-D3[93] with D3(BJ)[94] correction, the minimally-augmented ma-def2-QZVPP[76,95] basis set (we use the abbreviation DHDF-D3 for DSD-BLYP-D3/ma-def2-QZVPP in following sections) and the ORCA program.[96] The "GRID5" quadrature grid was used for DFT calculations as well as the RI approximation for Coulomb integrals (RI-J) and the COSX (chain-of-spheres exchange) approximation for exchange integrals with "GRIDX6" grid.[97] Auxiliary basis sets were automatically constructed (AutoAux keyword).[98] A tight energy change threshold for SCF convergence was used (*TightSCF* keyword; 10^{-8} E_h). The double-hybrid calculations used the frozen core approximation for the RI-MP2 contribution.

AMBER FF single point energy computations: The in-house program *bff*[99] was used for calculating AFF interaction energies on QM-optimized geometries of DME…nucleobase and deoxyribose…nucleobase models. Note that for interaction energies, only the non-bonded potential of the AFF is needed. The AFF interaction energy was calculated as

$$E_{int(AFF)} = 332 \sum \frac{q_i q_j}{r_{ij}} + \sum \varepsilon_{ij} \left[\left(\frac{R_{\min(ij)}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min(ij)}}{r_{ij}} \right)^{6} \right] \quad (2)$$

where the first term is the electrostatic potential describing the pairwise Coulomb interactions between all intermolecular atomic pairs i and j by using q_i and q_j partial atomic charges and r_{ij} interatomic distance. The 332 factor takes care of the conversion to kcal.mol^{-1} from charges in atomic units and distances in Ångström. The second term is the Lennard-Jones (LJ) potential composed of a repulsive ($r_{ij}{}^{-12}$) and an attractive ($-r_{ij}{}^{-6}$) part. The LJ parameters are represented by the $R_{\min(ij)}$ optimal interatomic distance and by the ε_{ij} potential well depth.

Partial charges were calculated by using the Antechamber[100] automated RESP[101] procedure at the recommended HF/6-31G* level of theory. LJ parameters for the

"parm99bsc0[102] AMBER variant" were taken from the AmberTool library files of AMBER16.[103] These LJ parameters are identical to those of the original Cornell et al. AMBER parametrization[24] and to those used in MD simulations of Z-DNA and r(UUCG) (see *AMBER MD simulations* and **Table S5**). The atom types and partial charges used for the AFF calculations are available as text files with '.resp' ending in the **Supporting Information**.

***Minimum Interaction Energy distance* (*minIEd*) and *Interaction Energy* (*IE*) surfaces:** To compute these surfaces, we have performed vertical interaction-energy scans for a series of positions of the DME probe covering the entire purine and pyrimidine nucleobase rings. We define the "*minimum interaction energy distance"* (*minIEd*) as the O4'…nucleobase distance at which the lp…π DME…nucleobase interaction energy is at its minimum. The complementary *interaction energy* (*IE*) surface shows the corresponding interaction energy minima to the *minIEd* surface. Thus, both surfaces use the same geometry for a given xyz coordinate; the *IE* surface displays the value of the interaction energy minimum and the *minIEd* surface displays the corresponding lp…π distance. Note that the *minIEd* distance is not systematically associated with the interaction energy strength as the balance of interactions differs depending on the local environment of the probed point. We found that it is useful to construct both *minIEd* and *IE* surfaces for the DHDF-D3 vs. AMBER comparison. The *minIEd* surface allows to compare the vertical distances while the *IE* surface shows the energetics of the interaction. A schematic that explains how the *minIEd* and *IE* values are derived from vertical scans is provided in **Figure S1** of the **Supporting Information**.

Note that we changed the previous term *closest-contact* surface[17] to the *minIEd* surface, since the former term could be perceived as ambiguous outside the structural biology community.

The four-dimensional data sets (xyz coordinates and interaction energy) used for the construction of the surfaces were collected as follows: first, the O4' atom of the optimized DME…nucleobase system was placed above the centroid of the pyrimidine ring at a 2.5 Å vertical distance. Then, DME was moved as a molecular probe in horizontal and vertical directions with respect to the nucleobase to sample 29 (horizontal scan) x 20 (vertical scan) = 580 points for purines and 20 (horizontal scan) x 20 (vertical scan) = 400 points for pyrimidines to cover the whole rings (see **Figure S1** and *Parameters for the vertical scans of the nucleobase surfaces* in the **Supporting Information**).

One additional horizontal point was added for the DHDF-D3 surfaces of C and U pyrimidine nucleobases outside the ring next to the C2 atom (along the centroid-C2 vector) to localize more precisely the *minIEd* surface minima (**Figure S1**). Gnuplot was used to create the surfaces using its internal algorithm for data interpolation.[104] Limitations to this modelling approach are described in the *Surface scan limitations* section of the **Supporting Information**.

Atom-in-molecule (AIM) α(0) polarizabilities calculated with DFT-D4 dispersion theory using Grimme's *dftd4* code were used to rationalize DHDF-D3 surface shapes.[17,71,105]

AMBER MD simulations: Molecular dynamics (MD) simulations of a Z-DNA $[(CpG)_3]_2$ hexamer duplex (PDBid: 3P4J X-ray structure)[72] were performed in AMBER18[106] by employing the AMBER OL15 force field[39] in an NPT ensemble using the Langevin thermostat with a collision frequency of 2 ps^{-1} and the Berendsen (weak-coupling) barostat.[107,108] A total simulation time of 3 μs (3x1 μs trajectories with randomized initial velocities) was achieved at both 100 K and 300 K temperatures. The CUDA-driven *pmemd* module (SPFP-implementation) of Amber was used,[109] along with a 2 fs time step and standard SHAKE restraints.[110] A SPC/E[111] water box with a 10 Å buffer region and with an octahedral periodicity was used. Simulations were done

in Na^+ excess salt conditions of 0.2 M NaCl resulting from the addition of 15 Na^+ and 5 Cl^-.[112] Ions were placed using AmberTool's *leap* module that uses basic electrostatic mapping.

MD simulations of the r(ggcacUUCGgugcc) tetraloop hairpin were initiated from the NMR structure (PDBid: 2KOC[113]) and data were taken from our previous work.[40] Simulations were performed with the AMBER OL3 (i.e., ff99bsc0χ_{OL3})[114] RNA force field with modified LJ parameters for phosphate oxygens,[31] associated dihedral adjustments,[115] and the external gHBfix potential.[37] Trajectory snapshots were saved every 10 ps for both Z-DNA and r(UUCG) tetraloop. Equilibration protocols and further details to r(UUCG) production runs are provided in the **Supporting Information**.

For Z-DNA, the whole 6 μs trajectories were analyzed. In contrast, we used only the initial 4.5 μs (MD1), 2.4 μs (MD2) and 2.8 μs (MD3) segments extracted from 3x10 μs trajectories of r(UUCG) tetraloop.[40] In these segments the majority of the signature interactions of the r(UNCG) tetraloop is preserved.[40] The remaining MD segments were not analyzed since the r(UUCG) tetraloop denaturates and deviates significantly from the canonical structure. Further, the ribose…guanine lp…π contact is totally lost because the guanine nucleobase flips out of the loop.

Additional software: Structures were visualized by Molden[116] and VMD[117]; figures were prepared with PyMOL[118], Jmol[119] and VMD; graphs were prepared by Gnuplot[104] and Xmgrace, MD analyses were performed using CPPTRAJ[120] obtained from AmberTools18[106].

Results and Discussion

Vertical energy scans reveal substantial differences between DHDF-D3 and AFF lp…π descriptions.

First, we discuss two representative DME…nucleobase vertical energy scans (**Figure 3**; for deoxyribose…nucleobase scans see *DME…G/C versus deoxyribose…G/C calculations* and **Figure S2** in the **Supporting Information**). For DME…G, the O4' atom is located above the pyrimidine ring centroid; for DME…C, the O4' atom is positioned above the C2 atom. Both positions are near the global minimum on the respective *minIEd* surfaces, i.e., the DME positions leading to the closest optimal contact distance (see below). Additionally, **Figure 3** displays FNO-CCSD(T)/CBS (abbreviated as "CCSD(T)") reference data to validate the DHDF-D3 results. The excellent agreement between the less computationally demanding DHDF-D3 (in red) and the more accurate CCSD(T) 'gold standard' calculations (in black; **Figure 3**) supports the use of the DHDF-D3 approximation.

For DME…G, the DHDF-D3 interaction energy minimum at 2.97 Å versus 3.10 Å for the AFF interaction energy curve displays a clear +0.13 Å shift of the AFF minimum towards longer distances. For DME…C, the DHDF-D3 and AFF curves show a greater distance shift (+0.24 Å) with interaction energy minima at 2.91 Å and 3.15 Å, respectively.

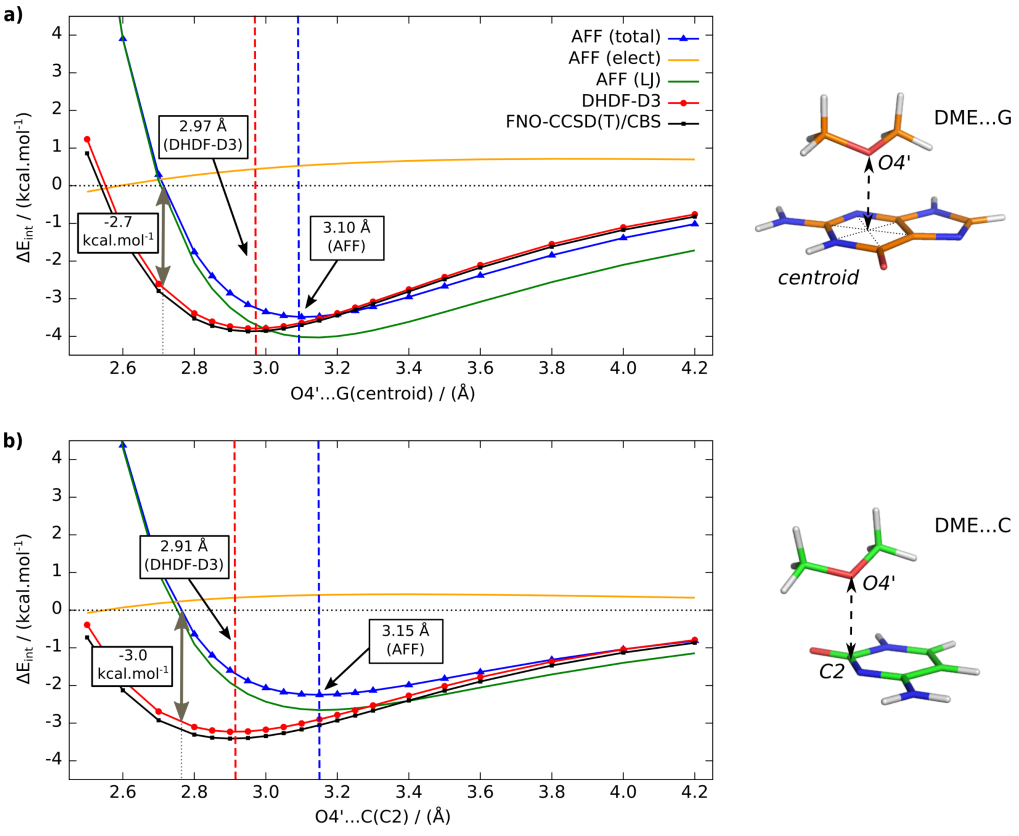

Figure 3. Comparison of AFF (AMBER FF; blue) and DHDF-D3 (DSD-BLYP-D3/ma-def2-QZVPP; red) vertical scans of the interaction energy along the O4'…nucleobase distance for **a)** DME…G and **b)** DME…C. The 'gold-standard' FNO-CCSD(T)/CBS calculations (black) are shown for reference. The decomposition of the non-bonded AFF potential in electrostatic (orange) and Lennard-Jones (green) components is also shown. AFF and DHDF-D3 minima distances are indicated. The grey double arrows mark the distances where the AFF becomes positive; the associated box displays the interaction energy difference between AFF and DHDF-D3.

AFF overestimates short-range repulsion for lp…π contacts compared to the DHDF-D3 reference.

To assess the importance of the AFF limitations and their implications for MD simulations, it is crucial to consider not only the position of the minima, but also the gradients of the energy curves in the regions commonly sampled in MD simulations. The too steep short-range AFF repulsive regions are a source of spurious forces that could affect the outcome of MD simulations. **Figure 3** highlights large AFF energy gradients at distances below the energy minima and a sharp increase in the DHDF-D3 vs. AFF difference for intermonomer distances below 2.9 Å, where the LJ approximation for the short-range repulsion starts to break down (see *AFF electrostatic and LJ energies show substantial deviations from SAPT especially for compressed structures*). For instance, the AFF energies become positive below 2.7 Å, while the QM energies are still around −3.0 kcal.mol^{-1} at this distance.

AFF *minIEd* surfaces are too repulsive and shallow compared to the DHDF-D3 reference.

***DME…purines*.** The global minimum of the DHDF-D3 *minIEd* surface for both purine nucleobases is located in the middle of the pyrimidine rings at 2.97 Å (G) and 2.95 Å (A) (**Figure 4** and **Table S2**). The surface remains shallow towards the C2 and C6 atoms. A secondary minimum is located above the imidazole rings at 3.04 Å (G) and 3.01 Å (A). For G, the location of the global *minIEd* minimum is in agreement with PDB database analysis of NpG Z-like steps that show that the O4' atom is mainly distributed above a region located between the C2 and C6 atoms and the pyrimidine ring centroid.[4,17] The data is also consistent with a PDB database analysis taking into account (deoxy)ribose…nucleobase stacking contacts between non-consecutive nucleotides.[5]

The AFF *minIEd* surfaces recover some features of the DHDF-D3 surfaces, i.e., the surface minima are located close to the centroid of the pyrimidine ring (**Figure 4**). However, the minima lie at ≈3.10 Å, a notable deviation of 0.13 Å (A) and 0.15 Å (G) from the DHDF-D3 surface minima (**Figure 4** and **Table S2**). Overall, the AFF *minIEd* surfaces are too repulsive over the whole surface area and are shallower than the DHDF-D3 *minIEd* surfaces. This is likely a consequence of the simple form of the AFF non-bonded potential that is unable to capture short-range electronic-structure effects like polarization, charge-penetration, and exchange-repulsion. AFF also fails to correctly describe the area between the two rings and the second minimum within the imidazole ring. The latter point should be of limited practical importance for MD simulations of systems with Z-step-like lp…π stacking since the O4' atom does not occur in positions above the imidazole ring in these structures, in contrast to what is observed for some ribose…nucleobase stacks occurring between non-neighboring residues in RNAs.[4,5]

***DME…pyrimidines*.** Z-like steps with pyrimidine instead of purine nucleobases are less frequent. Few cases were reported for RNA and only one for DNA.[4] However, in RNA different, non-Z-step-like ribose…nucleobase stacking types were observed in crystallographic structures with uracil or cytosine as a nucleobase.[5]

The DHDF-D3 *minIEd* surfaces of both pyrimidine nucleobases are similar to each other (**Figure 5 left**). The minima are located close to the C2 atom and not close to the ring centroid as one would expect from the purine nucleobases data (**Figure 4**). This dissimilarity may result from the absence of the pyrimidine/imidazole junction that allows for smaller

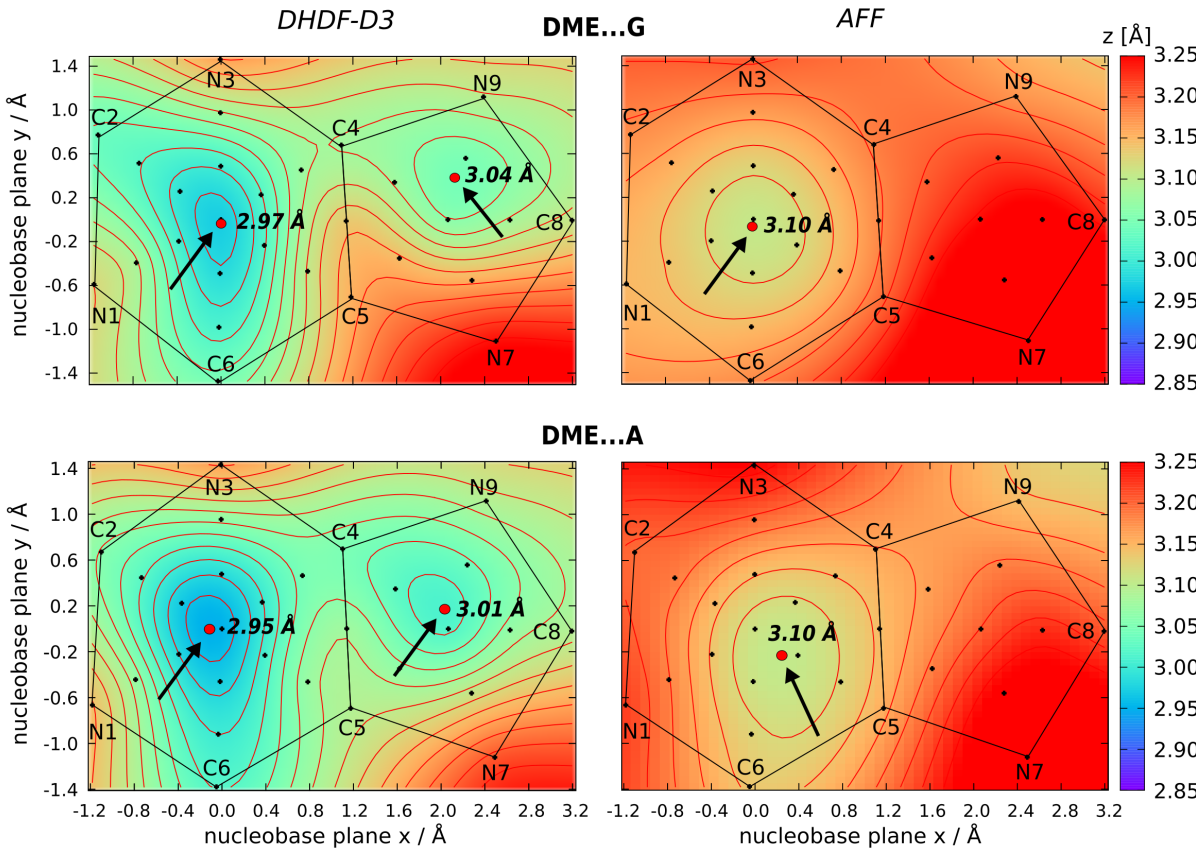

Figure 4. DME…G and DME…A *minIEd* surfaces for DHDF-D3 and AFF. Guanine and adenine nucleobases lie in the xy plane. Color represents the 'z' coordinate of the O4' DME atom. The color scale is the same for all plots to ease visual comparison. This causes a region of the AMBER surface around the N7 atom to be imprecisely represented as it slightly exceeds the scale. Black arrows and distance values point to minima of the *minIEd* surfaces marked by red dots; small black dots mark nucleobase points where the vertical scans were calculated (see also **Figure S1**). The AFF vs. DHDF-D3 difference range is 0.03–0.18 Å and 0.05–0.19 Å for G and A, respectively. DHDF-D3 data for G are taken from Kruse et al.[17]

atomic volumes of the purine C4/C5 carbon atoms when compared to the "equivalent" pyrimidine C6/C5 atoms due to the presence of the electron-withdrawing nitrogen atoms of the imidazole ring. For instance, the minimum distance of approach to the C5 atom in pyrimidines is ≈3.22 Å while it is close to 3.15 Å (G) and 3.09 Å (A) in purines (see *AIM polarizabilities* α(0) *explain shapes of minIEd surfaces* and **Figure 6**). Interestingly, our DHDF-D3 results are in agreement with a PDB database study of (deoxy)ribose…nucleobase stacking that shows that the O4' atom is mostly positioned above the C2 atom of a pyrimidine nucleobase.[5]

Overall, the agreement between DHDF-D3 and AFF for the pyrimidine systems is poor. The AFF *minIEd* surfaces of the pyrimidine systems (**Figure 5**) are barely capturing any of the essential features of their DHDF-D3 counterparts. For DME…C, the AFF surface shows a spread-out minimum covering a large part of the left side of the surface plot and the lowest value is by ~0.21 Å larger compared to the DHDF-D3 minimum (**Figure 5** and **Table S2**). As a reminder, **Figure 3b** shows the vertical scan for the position above the C2 atom and a 0.23 Å difference between DHDF-D3 and AFF for the minimum. The lowest point of the DME…U surface lies outside the ring near the N3 atom. Adding another nucleobase point outside of

N3 was not sufficient to determine an unambiguous minimum. In lieu of a proper minimum the shortest distance of the AMBER *minIEd* surface is used as a pseudo-minimum for DME...U in **Figure 5** and **Table S2.**

AIM polarizabilities α(0) explain the *minIEd* surface shapes.

Figure 6 pictures atom-in-molecules (AIM) polarizabilities α(0) of the four nucleobases that

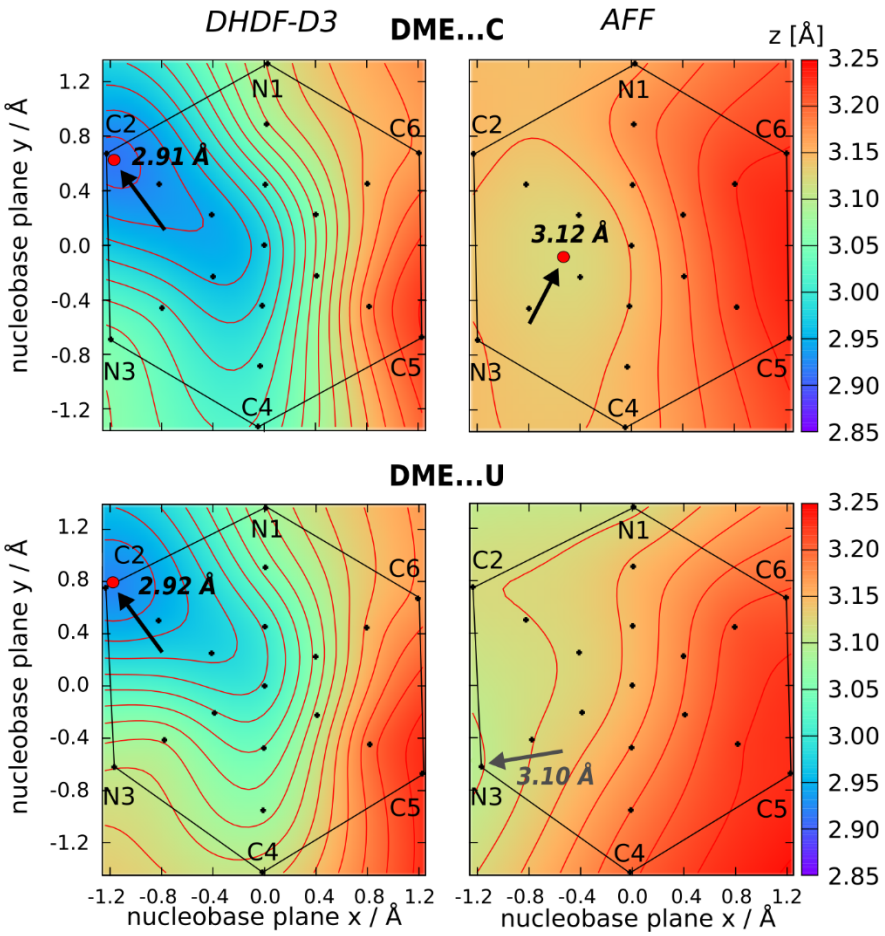

Figure 5. DME…C and DME…U *minIEd* surfaces for DHDF-D3 and AFF. Cytosine and uracil nucleobases lie in the xy plane. Color represents the 'z' coordinate of the O4' DME atom. The color scale is the same for all plots to ease visual comparison. Black arrows and distance values point to minima of the *minIEd* surfaces marked by red dots; small black dots mark points where the vertical scans were calculated (see also **Figure S1**). Note that for DME…U no AFF *minIEd* minimum is defined as it is outside the ring beyond the N3 atom. In this case, a grey arrow points to the N3 atom and the *minIEd* distance from vertical scan above the N3 atom is written in the graph. The AFF vs. DHDF-D3 difference range is 0.01–0.23 Å and 0.00–0.20 Å for C and U, respectively.

we calculated to rationalize the shapes of DHDF-D3 *minIEd* surfaces (see below). The AIM polarizabilities obtained from DFT-D4 dispersion theory[70,71] approximate atomic volumes[18,19] and thus can serve as a qualitative surrogate measure for the elusive 'real' vdW surface, although atomic anisotropy information is missing.[17]

The shape of the guanine *minIEd* surface fits nicely with the relative magnitudes of the AIM α(0) polarizabilities (**Figure 6** and Kruse et al.[17]), especially for the pyrimidine ring. The C2 and C6 atoms, being the two smallest atoms of the guanine rings, allow for the closest

O4'...ring atom contacts. For DME…A, the *minIEd* surface minimum spreads to the C4 and C6 atoms which are the smallest adenine rings atoms.

AIM polarizabilities explain also the *minIEd* surfaces of pyrimidine nucleobases where the smallest C2 and the biggest C5 atom show the shortest and the largest O4'...ring atom contact distances, respectively. Moreover, the significant shift of the *minIEd* minimum from the center of the six-atom rings for purines to the C2 atom for pyrimidines can be explained by the AIM α(0) polarizability values of C5 and C6 atoms (right side of the ring) that are significantly larger compared to the corresponding C4 and C5 atoms of purine nucleobases (**Figure 6**).

LJ parameters for nucleobase carbon atoms should not be identical in force fields.

The large range of DHDF-D3 vs. AFF differences shows that some minima locations on nucleobase surfaces are better described by AFF than others. The largest range is observed for DME…C where the difference between the DHDF-D3 and AFF *minIEd* distance is negligible

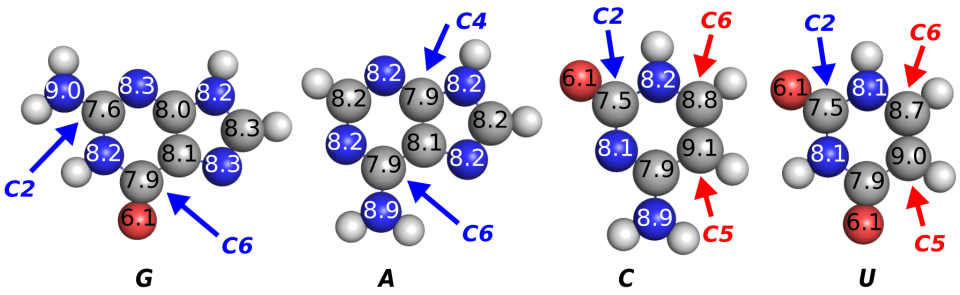

Figure 6. Atom-In-Molecule (AIM) polarizabilities α(0) representing the volume of atoms in atomic units of cubic bohr (a_0^3) for the four nucleobases guanine (G), adenine (A), cytosine (C) and uracil (U). The carbons discussed in the text are marked by blue (smaller carbon atoms) or red (larger carbon atoms) arrows.

for O4'…C(C5), while the deviation is large for O4'…C(C2) (0.23 Å). This appears to be correlated with the differences in AIM polarizabilities between the carbons (small C2, large C5; **Figure 6**). It suggests that the vdW radii for the nucleobase carbon atoms — currently represented by an identical set of LJ parameters in the AFF (**Table S5**) — might be too large for the atoms connected to electron withdrawing groups like the guanine C2 atom.[17] Note that the AFF LJ parameters for the nucleobase sp^2 carbons are derived from Monte Carlo simulations on liquid benzene and are thus identical to those of benzene rings devoid of electro-attractive-groups (**Table S5**).[24] Thus, it can be assumed that the pyrimidine C5 and C6 atoms that display the largest AIM volumes are similar to the benzene sp^2 carbon atoms and that the same AFF LJ parameters can be used. But it is clearly not the case for the carbon atoms connected to electron withdrawing groups and it is timely to add some diversity to the set of nucleobase LJ parameters.

To support our rationalizations based on AIM polarizabilities, we also calculated SAPT vertical scans with (DME)O4' located above the smallest (C2) and largest (C5) cytosine atom. Results show that exchange-repulsion (essentially Pauli repulsion) as well as the London dispersion (related to atomic polarizabilities) are notably smaller when O4' is located above the C2 atom (**Figure S3** in **Supporting Information**). This agrees with a carbon atom with a smaller effective vdW radius and a smaller α(0) polarizability value.

The DHDF-D3 and AFF *minIEd* surface comparison and the different AIM polarizability volumes of the ring carbon atoms indicate that (deoxy)ribose…nucleobase stacking cannot be precisely described at a quantitative level by the current uniform carbon atom AMBER LJ

parameters. On one side, FFs should be parametrized with limited number of atom types to avoid over-fitting, as discussed in Schauperl et al.[121] On the other side, describing all nucleobase ring carbons by the same LJ parameters is a very rough approximation as becomes apparent from our data. This may be especially important for systems where the distance of an oxygen atom to a carbon atom is among the shortest known. Thus, we tentatively propose that LJ radii for nucleobase carbon atoms should adhere to the following size-relation: C2<C6<C4~C5 for G, C4~C6<C2~C5 for A, and C2<C4<C6<C5 for C and U and be smaller or equal to those of benzene rings. Further evaluation using either a larger dataset or following a different approach, like obtaining AIM electron density-based LJ parameters unique for each atom prior to MD simulation as proposed by Kantonen et al.[122], is needed to precisely parametrize LJ radii values for sp^2 carbon atoms. It can be noted that earlier we calculated α(0) polarizabilities for the indole rings[17] and that the values for the two junction carbons are large and close to those calculated for pyrimidine C5 and C6 nucleobase atoms.

Interaction energies for the lp…π DME…nucleobase stacking are equally weak for all the nucleobases.

Interaction energy (*IE*) surfaces, defined as the strongest (minimum) interaction energies of the vertical scans, are discussed in detail in the **Supporting Information**. **Table S2** shows that DME…G is the system with the weakest DHDF-D3 interaction energy (−3.9 kcal.mol^{-1}) and DME…C the strongest (−4.4 kcal.mol^{-1}). A cautious interpretation of the data is that all our systems are very close in energy, i.e., no nucleobase stands out as particularly suitable or unsuitable for (deoxy)ribose…nucleobase interactions and all the interaction energies are weak compared to typical nucleobase...nucleobase interactions (−6.0 kcal.mol^{-1}).[53] The AFF reproduces the DHDF-D3 energy range around the surface minima relatively well although it fails to reproduce important structural features and short-range energies (**Figure 3**). We also note that interaction energies cannot be straightforwardly related to free energies in nucleic acids.[123]

AFF electrostatic and LJ energies show substantial deviations from SAPT especially for compressed structures.

It has been suggested that short (deoxy)ribose…nucleobase contacts may be subtly vertically compressed by the surrounding structural context, because lp…π contacts in nucleic acids often occur in locally strained molecular topologies as in Z-DNA and in r(UNCG) tetraloops.[17,124,125] To evaluate the range of potential compression effects, we performed SAPT computations that decompose the interaction energy into electrostatics, London dispersion, induction, and exchange-repulsion at three vertical separations: *(i)* the minimum of the corresponding *minIEd* surface *(ii)* extended by 0.3 Å and *(iii)* compressed by 0.2 Å. For all three separations SAPT shows dominating London dispersion followed by strong electrostatics (**Table 1**), in agreement with previous works.[16,17] Although extending or compressing the stacking distance does not change the interaction energy substantially, as typical for non-covalent interactions with naturally flat potentials near their minimum, it does change the relative composition of the interaction energy (**Table 1** and Figure 4 of Kruse et al.[17]). At longer distances the electrostatic interactions contribute less and dispersion

becomes more important, while at shorter distances, where the electron clouds overlap more, the electrostatic contribution increases as charge penetration effects come to play.[60]

The AFF decomposition shows that the LJ potential dominates the interaction energy curve as the Coulomb FF electrostatics remains flat in the equilibrium region and almost without a contribution to the AFF curve (**Table 1** and **Figure 3**). It is known that the AMBER electrostatics is fairly insensitive for nucleobase…nucleobase stacking vertical separation.[50,52] We show that this is also the case for the lp…π (deoxy)ribose…nucleobase stacking. Point-charge schemes are known to be deficient at short-range due to the lack of explicit charge-cloud penetration effects that become important once the electron densities of both molecules start to overlap ('interpenetrate') and are usually attractive for charge-neutral molecular densities.[56–60] However, most classical non-bonded FF potentials should be understood as *effective potentials*, where the accuracy of the individual parts is less important

Table 1. SAPT2+(3)δMP2/jun-cc-pVTZ (*SAPT*) interaction energy decomposition into electrostatics (ΔE_{elect}), London dispersion (ΔE_{disp}), induction (ΔE_{ind}), exchange-repulsion ($\Delta E_{exch\text{-}rep}$) as well as AFF electrostatic (ΔE_{elect}) and Lennard-Jones (ΔE_{LJ}) terms. The percentages indicate the contributions of the attractive components.

Interaction component type	DME…G	DME…A	DME…C	DME…U
SAPT at *minIEd* minimum (kcal.mol^{-1})				
ΔE_{elect}	−2.3 (21 %)	−2.7 (24 %)	−2.8 (30 %)	−3.4 (35 %)
ΔE_{disp}	−8.0 (74 %)	−7.7 (71 %)	−5.8 (62 %)	−5.8 (59 %)
ΔE_{ind}	−0.5 (5 %)	−0.5 (5 %)	−0.8 (8 %)	−0.6 (6 %)
$\Delta E_{exch\text{-}rep}$	7.3	7.2	6.3	6.3
ΔE_{total_SAPT}	**−3.6**	**−3.7**	**−3.2**	**−3.6**
AFF (kcal.mol^{-1})				
ΔE_{elect}	0.4	0.4	0.2	−0.7
ΔE_{LJ}	−3.7	−3.3	−2.1	−2.0
ΔE_{total_AFF}	**−3.3**	**−2.9**	**−1.8**	**−2.7**
SAPT at *minIEd* minimum +0.3 Å (extension; kcal.mol^{-1})				
ΔE_{elect}	−0.5 (8 %)	−0.8 (14 %)	−1.0 (20 %)	−1.5 (28 %)
ΔE_{disp}	−5.1 (87 %)	−4.9 (82 %)	−3.7 (72 %)	−3.7 (67 %)
ΔE_{ind}	−0.3 (5 %)	−0.3 (4 %)	−0.4 (9 %)	−0.3 (6 %)
$\Delta E_{exch\text{-}rep}$	2.7	2.7	2.3	2.3
ΔE_{total_SAPT}	**−3.2**	**−3.3**	**−2.9**	**−3.3**
AFF (kcal.mol^{-1})				
ΔE_{elect}	0.6	0.5	0.3	−0.4
ΔE_{LJ}	−3.9	−3.6	−2.7	−2.7
ΔE_{total_AFF}	**−3.3**	**−3.2**	**−2.3**	**−3.1**
SAPT at *minIEd* minimum −0.2 Å (compression; kcal.mol^{-1})				
ΔE_{elect}	−4.9 (29 %)	−5.3 (32 %)	−5.4 (37 %)	−6.1 (40 %)
ΔE_{disp}	−10.9 (66 %)	−10.4 (63 %)	−8.0 (54 %)	−8.0 (53 %)
ΔE_{ind}	−0.9 (5 %)	−0.9 (5 %)	−1.3 (9 %)	−1.1 (7 %)
$\Delta E_{exch\text{-}rep}$	13.9	13.7	12.3	12.4
ΔE_{total_SAPT}	**−2.8**	**−2.8**	**−2.4**	**−2.8**
AFF (kcal.mol^{-1})				
ΔE_{elect}	0.2	0.3	0.1	−0.9
ΔE_{LJ}	−1.5	−1.1	0.7	0.8
ΔE_{total_AFF}	**−1.3**	**−0.8**	**0.8**	**−0.1**

than their sum.[57] For example, the exponential distance-dependence of the charge penetration and the short-range exchange-repulsion allow for the systematic cancelling of errors between the missing charge penetration effects and the approximate r^{-12} distance-

dependence of the repulsive part of the LJ potential in AFF.[57] Thus, the short-range charge-penetration part of QM electrostatics is effectively included in the vdW term of AFF.

While these schemes work relatively well at typical vdW distances, the erroneous behavior of the too steep r^{-12} term of the LJ potential dominates the short-range region entirely, as also shown by a significant weakening of AFF interaction energy in the compressed lp…π structures (**Table 1**). For instance, while the difference for the DME…G model between the SAPT and AFF energy at the *minIEd* minimum is -0.3 kcal.mol^{-1}, it is -1.5 kcal.mol^{-1} when compressed by only 0.2 Å. For DME…C, the AFF interaction energy is even repulsive (+0.8 kcal.mol^{-1}) above the minimum with a compression of 0.2 Å. Similar observations of a systematically too repulsive AFF potential in the short-range region of interatomic contacts have been reported before.[40,50–52,57,126] The awareness about the too steep repulsion of the LJ potential goes all the way back to Buckingham.[54] However, the r^{-12} form (equation (2) in *Computational details*) has been favored in all major biomolecular FFs over the Buckingham exponential form for computational efficiency reasons.

Sampling lp…π distances in the repulsive short-range region by MD simulations is problematic due to inaccurate LJ potentials.

Above, we have shown that AFF overestimates the *minIEd*s in lp…π (deoxy)ribose…nucleobase stacking compared to the reference DHDF-D3 method (**Figure 4** and **5**). In addition, for the short-range repulsion part, the agreement between AFF and DHDF-D3 worsens due to the steepness of the r^{-12} LJ repulsive term at small contact distance (**Figure 3**). This means that the shorter the lp…π distances below the QM minimum, the larger the bias introduced by the AFF approximation. To evaluate the extent to which the short-range regions are sampled by MD simulations, we analyzed the (deoxy)ribose…G contacts in MD simulations of Z-DNA and r(UNCG) tetraloops systems.

Z-DNA. It has been shown that MD simulations of Z-DNA using currently available AFFs lead to an overall unsatisfactory description of the double-helical geometry.[39,46–48] Although not fully satisfactory regarding description of the Z-DNA backbone substates, the best performance so far has been obtained with the OL15 version of the AMBER FF.[39,48]

We analyzed Z-DNA MD simulations conducted with OL15 at 100 K and 300 K. The 300 K simulations sampled various backbone substates similar to those described by Zgarbová et al.[39] including flipping of the native α_{g+}/γ_t to the non-native α_t/γ_{g+} conformation in CpG steps. The 100 K simulations are mirroring the situation of the crystal structure more closely and retain well the crystal structure conformation. We noticed that the cytidine deoxyribose changes its pucker from C2'-*endo* to C1'-*exo* at both temperatures and even moves to O4'-*endo* in the 300 K simulation.

The average O4'…G-plane distance (**Figure 7**) analyzed for the middle CpG dinucleotides of Z-DNA strands is 3.03 Å (300 K). If selecting only the snapshots with a preserved canonical CpG Z-step (around 18 % of all the snapshots), the average distance is reduced by 0.07 Å to 2.96 Å (see **Table S6**). Interestingly, sampling of short deoxyribose…G distances is correlated with a moderate deformation of the deoxyribose ring (the above-mentioned shift of deoxyribose pucker to C1'-*exo* or O4'-*endo* conformations). The deoxyribose deformation appears to ease the atom-atom steric conflict and leads to extension of the O4'…G distances. While the average O4'…G distance over C_3pG_4 and C_9pG_{10} in the 3P4J X-ray structure (2.87 Å)[17] is exceeded by the 300 K MD, it is well reproduced by the 100 K simulation (2.91 Å; this distance is reduced by 0.03 Å to 2.88 Å if only the C_3pG_4 and

C_9pG_{10} satisfying the Z-step criteria – 36 % of all the snapshots – are considered; see **Table S6**). At the same time, the average MD distances are smaller than the AFF *minIEd* values for the model system indicating that the repulsive region of the non-bonded potential for the deoxyribose…G stacking is sampled in both the 100 K and 300 K simulations. However, in the 100 K simulations the Z-DNA conformation remains stable, likely due to the reduced temperature preventing crossing of conformational energy barriers, which is also the reason for the well reproduced averaged O4'…G distance. Also, it agrees with the distribution of the O4' projection points on the G nucleobase (**Figure 7**). While in the 300 K simulations the O4' projection points form a single cloud spreading from the center of the pyrimidine ring to the ring-atoms, there is a group of six clouds in the 100 K simulations (one for each CpG step) as the O4' occupies a smaller region of space above the nucleobase in the locked Z-DNA geometry. **Figure 7** also shows no preference for any of the ring atoms by the O4' atom in 300 K simulation, which is in agreement with the spuriously flat AFF *minIEd* surface (**Figure 4**). However, in X-ray data[72] and to some extent in the 100 K simulations, the C2 atom contact is preferred.[17]

The r(UUCG) tetraloop. Tracing the origins of systematic AFF shortcomings from real-system MD simulations is challenging. Inaccuracies in the approximate potentials may accumulate and are often mutually interrelated as highlighted by a recent study of the r(UUCG) tetraloop.[40] Despite efforts involving a true multi-scale approach employing small QM models, large QM/MM systems and long time-scale MD simulations no singular or conceptually simple force-field aspect leading to the observed instabilities could be identified. It was concluded that the difficulties in describing r(UNCG) tetraloops are likely due to multiple force-field imbalances that are magnified by the unusually confined conformational space of the native r(UNCG) free-energy basin.

Here, we used three 10 μs-long MD simulations of the r(ggcacUUCGgugcc) tetraloop NMR structure[113] taken from Mráziková et al.[40] to analyze O4'…G distances in three simulations segments where the tetraloop structure is preserved (see "*Computational details*").[40] We observe again a high sampling of short O4'…G distances (**Figure 7 right**). The average from all analyzed segments is 2.93 Å, which is shorter by 0.10 Å compared to Z-DNA

MD data but close to the ≈2.90 Å experimental O4'…G distances derived from both the NMR (PDBid: 2KOC[113]) and an ensemble of X-ray structures.[4] In contrast to Z-DNA, the O4' is frequently located directly over the ring and close to the C2 atom, as displayed by the O4'

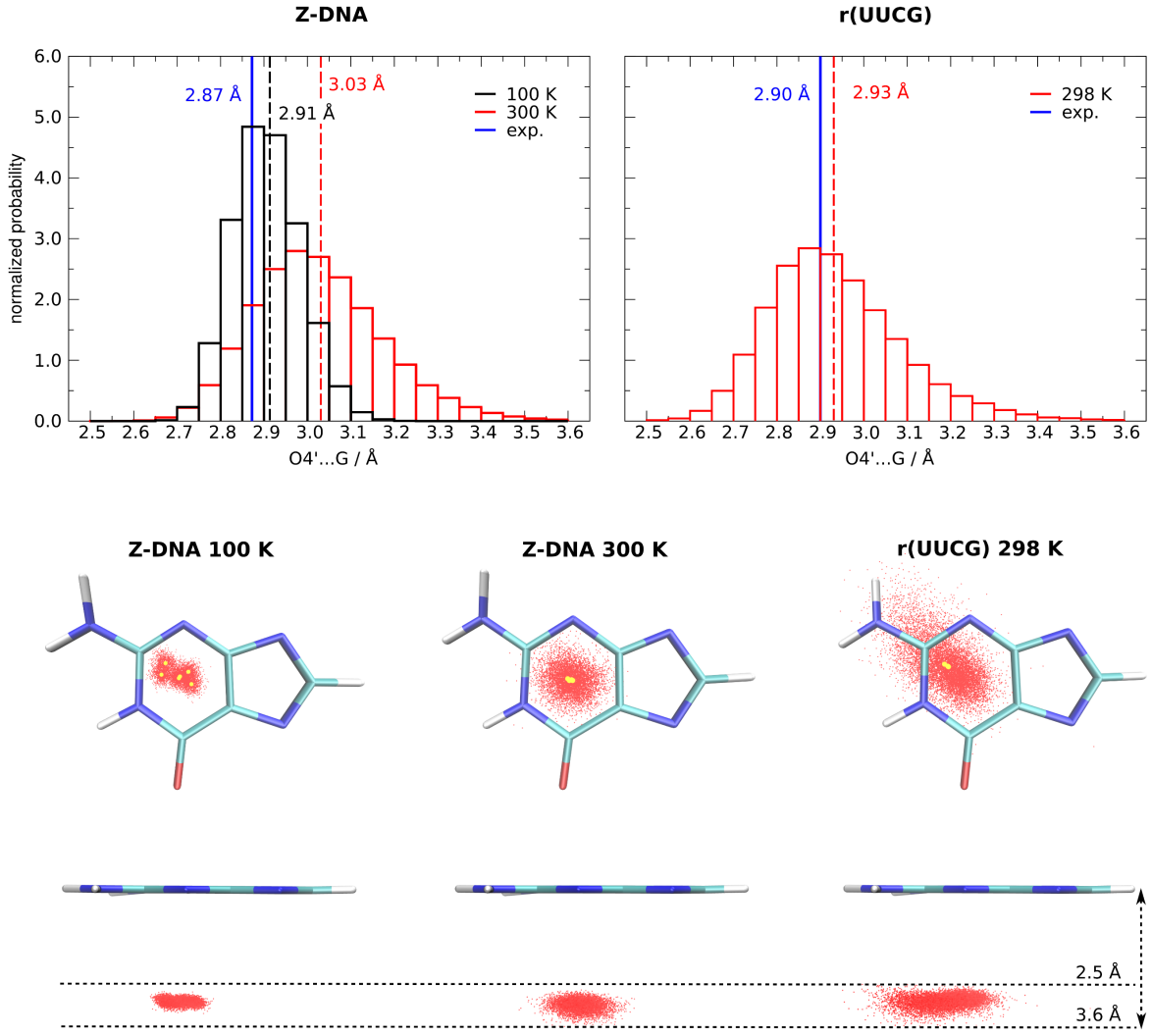

Figure 7. Histograms of O4'…G (lp-π) distances and top-down/side views on the O4'…G nucleobase from AFF MD simulations of Z-DNA and r(UUCG) tetraloop models. In the histograms the time-averaged O4'…G distances are marked by vertical dashed lines (black/100 K, red/300 K) while experimental values for Z-DNA (for 3P4J C_3pG_4 and C_9pG_{10} steps) and r(UUCG) (2KOC, average from 20 models) are marked by blue solid lines. Projection points of the O4' atom on the G nucleobase are displayed as red points and averages for each combination of one CpG dinucleotide (for Z-DNA either 3-4 or 9-10) with one MD simulation are marked as yellow dots. In the top-down/side views on the O4'…G nucleobase every 100th snapshot is visualized. Note that for Z-DNA MD simulations the picture consists of the six independent O4'…G distance calculations for both the 100 K and 300 K simulations (3 MD times two dinucleotides). While in the 300 K simulations the O4' projection points are equally distributed for the six cases, there is a group of six clouds in the 100 K simulations because of the locked-in starting conformations of the CpG steps. For r(UUCG) tetraloop MD simulations, the picture consists of the three independent O4'…G distances calculations (3 MD).

projections on the G nucleobase (**Figure 7**). It indicates that the r(UUCG) tetraloop is too stiff to ease the steric repulsion between the O4' and the ring-atoms, which eventually leads to disruption of its native conformation.

In conclusion, distances below the *minIEd* surface minima are sampled in both Z-DNA and r(UUCG) tetraloop trajectories. It indicates that the overall Z-DNA and UNCG tetraloop structures tend to compact the (deoxy)ribose…G stacking. Thus, due to thermal fluctuations, the simulations face an exaggerated short-range LJ potential repulsion that inevitably biases the statistical MD sampling and has the potential to destabilize the native state. This may be one of the reasons leading to a problematic description of Z-step containing nucleic acid structures in current MD simulations.

Conclusion

Herein, we evaluated the ability of the AMBER FF (AFF) non-bonded potential to describe the (deoxy)ribose…nucleobase lp…π stacking through comparisons with the highly accurate double-hybrid DFT (DHDF-D3) method. Vertical interaction energy scans calculated using DME…nucleobase (G, A, C, U) models reveal a clear over-repulsiveness of the AFF especially in the short-range region of the interaction. Moreover, the over-repulsiveness of the AFF is non-systematic and varies for different positions of the O4' atom above the nucleobase. For instance, while the difference in DHDF-D3 and AFF *minimum interaction energy distances* (*minIEd*s) is significant for the C2 atom for all the nucleobases, the AFF behaves correctly for the C5 atoms of pyrimidine nucleobases. This is a consequence of notable differences in atomic volumes of the nucleobase carbon ring-atoms, as indicated by AIM α(0) polarizability calculations. This implies that using the same vdW radius for all ring-carbon atoms is not the best choice for nucleobase systems. Based on our results we suggest that nucleobase carbon atom LJ parameters, which are all identical in the AFF, should follow the size-relationships: C2<C6<C4~C5 (for G), C4~C6<C2~C5 (for A), and C2<C4<C6<C5 (for C/U) and recommend that the LJ parameters of N and O atoms in aromatic rings should also be reevaluated.

Since the (deoxy)ribose…nucleobase lp…π stacking occurs commonly in conformationally highly strained structures, the AMBER MD simulations may either sample short (deoxy)ribose…nucleobase distances, contributing to a structural strain caused by the over-estimated short-range repulsion, or avoid them, leading to disruption of local conformations. MD distance sampling of a Z-DNA helix and a r(UUCG) tetraloop strongly suggests that simulations of nucleic acid structures with Z-steps are affected by an exaggerated short-range repulsion of the O4'…G contacts. The excessive short-range LJ repulsion prevents further compaction of the O4'…G interaction that the real system would use to achieve an overall balance of the intricate Z-step conformation. Therefore, the too 'large' AFF nucleobase carbon atoms distort any structural sampling of Z-steps due to an increased strain in these highly specific and apparently stiff (and thus also brittle) conformations. The mere modification of the AFF atomic radii would not necessarily eliminate the problem, as the spurious behavior of the AFF short-range repulsion region stems from the oversimplified form of the LJ term. Thus, it is necessary to work on both issues: improving the vdW description of nucleobase atoms and finding a way to flatten the AFF lp…π interaction potential. Current computational means would allow the use of Buckingham like potentials, although this would imply significant reparameterization efforts.

Associated content

Supporting Information

Additional computational details, surface scan limitations, deoxyribose…G and deoxyribose…C calculations, SAPT for DME…guanine with rotated DME, discussion to *IE* surfaces, MD equilibration and simulations protocols, AMBER atom types and LJ parameters for nucleobase atoms, additional MD analyses, SAPT comparison between O4'…C(C2) and O4'…C(C5) (PDF)

Optimized geometries of DME…nucleobase and deoxyribose…nucleobase models; input files for *bff* calculations containing XYZ coordinates of the optimized geometries of DME…nucleobase and deoxyribose…nucleobase models, RESP charges and atom types; surface data containing XYZ files used for single-point energy calculations and files containing O4'…nucleobase distance and the associated interaction energy for both AFF and DHDF-D3 methods and for each nucleobase (ZIP)

Data and Software Availability

The datasets generated during the current study are contained within the manuscript or Supporting Information. Raw trajectory data are available from the authors on request.

The in-house program used for restrained geometry optimizations can be downloaded from https://github.com/hokru/xopt (Xopt) and the QM software coupled to Xopt used for energy and gradient calculations can be purchased from https://www.turbomole.org/ (Turbomole version 7.3). The software used for DHDF-D3 single-point calculations for surface scans is freely available after registration at https://orcaforum.kofo.mpg.de/app.php/portal (ORCA version 4.0.1), the software used for CCSD(T) calculations and SAPT analysis is freely available at https://psicode.org/installs/v14/ (PSI4 version 1.4). The software used for the electrostatic potential maps calculations further used for generating RESP charges by Antechamber can be purchased from http://gaussian.com/ (Gaussian 09). Antechamber itself, as well as CPPTRAJ used for analysis of MD trajectories, is a part of freely available AmberTools found at https://ambermd.org/GetAmber.php. The MD software AMBER can be purchased from the same website. The in-house program used for single-point AFF calculations for surface scans can be downloaded from https://github.com/hokru/BrnoFF (*bff*). A program used for calculations of AIM polarizabilities is freely available at https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dftd4 (dftd4).

The used visualization programs are freely available at https://www3.cmbi.umcn.nl/molden/ (Molden), https://www.ks.uiuc.edu/Research/vmd/ (VMD version 1.9.3), http://pymol.org/ (PyMOL version 2.5.0) and http://jmol.sourceforge.net/download/ (JMOL version 14.6.4). The software used for preparing graphs is freely available at http://www.gnuplot.info/ (Gnuplot version 4.6) and http://plasma-gate.weizmann.ac.il/Grace/ (Xmgrace version 5.1.22).

Acknowledgement

This work was supported by the project SYMBIT reg. number CZ.02.1.01/0.0/0.0/15_003/0000477 financed by the ERDF (HK, JS) and by project 21-23718S by the Czech Science Foundation (KM, VM, JS).

References

(1) Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B. Quantum Chemical Studies of Nucleic Acids: Can We Construct a Bridge to the RNA Structural Biology and Bioinformatics Communities? *J. Phys. Chem. B* **2010**, *114* (48), 15723–15741. https://doi.org/10.1021/jp104361m.Quantum.

(2) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. Classification and Energetics of the Base-Phosphate Interactions in RNA. *Nucleic Acids Res.* **2009**, *37* (15), 4898–4918. https://doi.org/10.1093/nar/gkp468.

(3) Sweeney, B. A.; Roy, P.; Leontis, N. B. An Introduction to Recurrent Nucleotide Interactions in RNA. *Wiley Interdiscip. Rev. RNA* **2015**, *6* (1), 17–45. https://doi.org/10.1002/wrna.1258.

(4) D'Ascenzo, L.; Leonarski, F.; Vicens, Q.; Auffinger, P. "Z-DNA like" Fragments in RNA: A Recurring Structural Motif with Implications for Folding, RNA/Protein Recognition and Immune Response. *Nucleic Acids Res.* **2016**, *44* (12), 5944–5956. https://doi.org/10.1093/nar/gkw388.

(5) Chawla, M.; Chermak, E.; Zhang, Q.; Bujnicki, J. M.; Oliva, R.; Cavallo, L. Occurrence and Stability of Lone Pair-π Stacking Interactions between Ribose and Nucleobases in Functional RNAs. *Nucleic Acids Res.* **2017**, *45* (19), 11019–11032. https://doi.org/10.1093/nar/gkx757.

(6) Wang, A. J.; Quigley, G. J.; Kolpak, F. J.; van der Marel, G.; van Boom, J. H.; Rich, A. Left-Handed Double Helical DNA: Variations in the Backbone Conformation. *Science (80-.).* **1981**, *211*, 171–176.

(7) Egli, M.; Gessner, R. V. Stereoelectronic Effects of Deoxyribose O4' on DNA Conformation. *Proc. Natl. Acad. Sci.* **1995**, *92* (1), 180–184. https://doi.org/10.1073/pnas.92.1.180.

(8) D'Ascenzo, L.; Leonarski, F.; Vicens, Q.; Auffinger, P. Revisiting GNRA and UNCG Folds: U-Turns versus Z-Turns in RNA Hairpin Loops. *Rna* **2017**, *23* (3), 259–269. https://doi.org/10.1261/rna.059097.116.

(9) Egli, M.; Sarkhel, S. Lone Pair-Aromatic Interactions: To Stabilize or Not to Stabilize. *Acc. Chem. Res.* **2007**, *40* (3), 197–205. https://doi.org/10.1021/ar068174u.

(10) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. Base-Base and Deoxyribose-Base Stacking Interactions in B-DNA and Z-DNA: A Quantum-Chemical Study. *Biophys. J.* **1997**, *73* (1), 76–87. https://doi.org/10.1016/S0006-3495(97)78049-4.

(11) Kozelka, J. Lone Pair–π Interactions in Biological Systems: Occurrence, Function, and Physical Origin. *Eur. Biophys. J.* **2017**, *46* (8), 729–737. https://doi.org/10.1007/s00249-017-1210-1.

(12) Gallivan, J. P.; Dougherty, D. A. Can Lone Pairs Bind to a Pi System? The Water...Hexafluorobenzene Interaction. *Org. Lett.* **1999**, *1* (1), 103–105. https://doi.org/10.1021/ol990577p.

(13) Gung, B. W.; Zou, Y.; Xu, Z.; Amicangelo, J. C.; Irwin, D. G.; Ma, S.; Zhou, H. C.

Quantitative Study of Interactions between Oxygen Lone Pair and Aromatic Rings: Substituent Effect and the Importance of Closeness of Contact. *J. Org. Chem.* **2008**, *73* (2), 689–693. https://doi.org/10.1021/jo702170j.

(14) Zhao, R.; Zhang, R.-Q. Beyond the Electrostatic Model: The Significant Roles of Orbital Interaction and the Dispersion Effect in Aqueous–π Systems. *Phys. Chem. Chem. Phys.* **2017**, *19* (2), 1298–1302. https://doi.org/10.1039/C6CP07698J.

(15) Ran, J.; Hobza, P. On the Nature of Bonding in Lone Pair···π-Electron Complexes: CCSD(T)/Complete Basis Set Limit Calculations. *J. Chem. Theory Comput.* **2009**, *5*, 1180–1185.

(16) Novotný, J.; Bazzi, S.; Marek, R.; Kozelka, J. Lone-Pair–π Interactions: Analysis of the Physical Origin and Biological Implications. *Phys. Chem. Chem. Phys.* **2016**, *18* (28), 19472–19481. https://doi.org/10.1039/C6CP01524G.

(17) 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* (38), 16553–16560. https://doi.org/10.1002/anie.202004201.

(18) Politzer, P.; Jin, P.; Murray, J. S. Atomic Polarizability, Volume and Ionization Energy. *J. Chem. Phys.* **2002**, *117* (18), 8197–8202. https://doi.org/10.1063/1.1511180.

(19) Fedorov, D. V; Sadhukhan, M.; Stöhr, M.; Tkatchenko, A. Quantum-Mechanical Relation between Atomic Dipole Polarizability and the van Der Waals Radius. *Phys. Rev. Lett.* **2018**, *121*, 183401–183406. https://doi.org/10.1103/PhysRevLett.121.183401.

(20) Chandrasekaran, R.; Arnott, S. The Structure of B-DNA in Oriented Fibers. *J. Biomol. Struct. Dyn.* **1996**, *13* (6), 1015–1027.

(21) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441–451. https://doi.org/10.1021/j100785a001.

(22) Schiemenz, G. The Sum of van Der Waals Radii--A Pitfall in the Search for Bonding. *Z. Naturforsch. B* **2007**, *62* (2), 235–243. https://doi.org/10.1515/znb-2007-0215.

(23) Hu, S. Z.; Zhou, Z. H.; Xie, Z. X.; Robertson, B. E. A Comparative Study of Crystallographic van Der Waals Radii. *Z. Krist. Cryst. Mater.* **2014**, *229* (7), 517–523. https://doi.org/10.1515/zkri-2014-1726.

(24) 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* (19), 5179–5197. https://doi.org/10.1021/ja00124a002.

(25) Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118* (8), 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.

(26) Nerenberg, P. S.; Head-Gordon, T. New Developments in Force Fields for Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129–138.

https://doi.org/10.1016/j.sbi.2018.02.002.

(27) Vangaveti, S.; Ranganathan, S. V.; Chen, A. A. Advances in RNA Molecular Dynamics: A Simulator's Guide to RNA Force Fields. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (2). https://doi.org/10.1002/wrna.1396.

(28) Smith, L. G.; Zhao, J.; Mathews, D. H.; Turner, D. H. Physics-Based All-Atom Modeling of RNA Energetics and Structure. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (5), 1–45. https://doi.org/10.1002/wrna.1422.Physics-Based.

(29) Salsbury, A. M.; Lemkul, J. A. Recent Developments in Empirical Atomistic Force Fields for Nucleic Acids and Applications to Studies of Folding and Dynamics. *Curr. Opin. Struct. Biol.* **2020**, *67*, 9–17. https://doi.org/10.1016/j.sbi.2020.08.003.

(30) Mackerell, A. D. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25* (13), 1584–1604. https://doi.org/10.1002/jcc.20082.

(31) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405–4412. https://doi.org/10.1021/ct300613v.

(32) Chen, A. A.; Garcia, A. E. High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci.* **2013**, *110* (42), 16820–16825. https://doi.org/10.1073/pnas.1309392110.

(33) Bergonzo, C.; Cheatham III, T. E. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11* (9), 3969–3972. https://doi.org/10.1021/acs.jctc.5b00444.

(34) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham III, T. E. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21* (9), 1578–1590. https://doi.org/10.1261/rna.051102.115.

(35) Darré, L.; Ivani, I.; Dans, P. D.; Gómez, H.; Hospital, A.; Orozco, M. Small Details Matter: The 2'-Hydroxyl as a Conformational Switch in RNA. *J. Am. Chem. Soc.* **2016**, *138* (50), 16355–16363. https://doi.org/10.1021/jacs.6b09471.

(36) Kuhrova, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banas, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12* (9), 4534–4548. https://doi.org/10.1021/acs.jctc.6b00300.

(37) Kuhrova, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Sponer, J.; Banas, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15* (5), 3288–3305. https://doi.org/10.1021/acs.jctc.8b00955.

(38) Mlynsky, V.; Kuhrova, P.; Kuhr, T.; Otyepka, M.; Bussi, G.; Banas, P.; Sponer, J. Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides. *J. Chem. Theory Comput.* **2020**, *16*, 3936–3946. https://doi.org/10.1021/acs.jctc.0c00228.

(39) 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* (12), 5723–5736. https://doi.org/10.1021/acs.jctc.5b00716.

(40) Mráziková, K.; Mlýnský, V.; Kührová, P.; Pokorná, P.; Kruse, H.; Krepl, M.; Otyepka, M.; Banáš, P.; Šponer, J. UUCG RNA Tetraloop as a Formidable Force-Field Challenge for MD Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 7601–7617. https://doi.org/, is it.

(41) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Refinement Using Quantum Mechanical Base Stacking and Conformational Energetics. *J. Chem. Theory Comput.* **2017**, *13* (5), 2053–2071. https://doi.org/10.1021/acs.jctc.7b00067.

(42) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. *J. Chem. Theory Comput.* **2017**, *13* (5), 2072–2085. https://doi.org/10.1021/acs.jctc.7b00068.

(43) Bottaro, S.; Banas, P.; Sponer, J.; Bussi, G. Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J. Phys. Chem. Lett.* **2016**, *7* (20), 4032–4038. https://doi.org/10.1021/acs.jpclett.6b01905.

(44) Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banas, P.; Sponer, J.; Bussi, G. Fitting Corrections to an RNA Force Field Using Experimental Data. *J. Chem. Theory Comput.* **2019**, *15* (6), 3425–3431. https://doi.org/10.1021/acs.jctc.9b00206.

(45) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8* (1), 348–362. https://doi.org/10.1021/ct200723y.

(46) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8* (7), 2506–2520. https://doi.org/10.1021/ct300275s.

(47) Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E.; Otyepka, M.; Jurečka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9* (5), 2339–2354. https://doi.org/10.1021/ct400154j.

(48) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127. https://doi.org/10.1021/acs.jctc.6b00186.

(49) Sponer, J.; Leszczynski, J.; Hobza, P. Nature of Nucleic Acid-Base Stacking: Nonempirical Ab Initio and Empirical Potential Characterization of 10 Stacked Base Dimers. Comparison of Stacked and H-Bonded Base Pairs. *J. Phys. Chem.* **1996**, *100* (13), 5590–5596. https://doi.org/10.1021/jp953306e.

(50) 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* (42), 9209–9222. https://doi.org/10.1021/acs.jpca.9b05940.

(51) Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. Balance of Attraction and Repulsion in Nucleic-Acid Base Stacking: CCSD(T)/Complete-Basis-Set-Limit Calculations on Uracil Dimer and a Comparison with the Force-Field Description. *J. Chem. Theory Comput.* **2009**, *5* (6), 1524–1544. https://doi.org/10.1021/ct9000125.

(52) Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. Reference MP2/CBS and CCSD(T) Quantum-Chemical Calculations on Stacked Adenine Dimers. Comparison with DFT-D, MP2.5, SCS(MI)-MP2, M06-2X, CBS(SCS-D) and Force Field Descriptions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3522–3534. https://doi.org/10.1039/b924461a.

(53) Kruse, H.; Banas, P.; Sponer, J. Investigations of Stacked DNA Base-Pair Steps: Highly Accurate Stacking Interaction Energies, Energy Decomposition, and Many-Body Stacking Effects. *J. Chem. Theory Comput.* **2019**, *15* (1), 95–115. https://doi.org/10.1021/acs.jctc.8b00643.

(54) Buckingham, R. A. The Classical Equation of State of Gaseous Helium, Neon and Argon. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1938**, *168* (933), 264–283. https://doi.org/10.1098/rspa.1938.0173.

(55) Kaplan, I. G. *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*; John Wiley & Sons, 2006.

(56) Cisneros, G. A.; Darden, T. A.; Gresh, N.; Pilme, J.; Reinhardt, P.; Parisel, O.; Piquemal, J.-P. Design Of Next Generation Force Fields From AB Initio Computations: Beyond Point Charges Electrostatics. In *Multi-scale Quantum Models for Biocatalysis. Challenges and Advances in Computational Chemistry and Physics*; York, D. M., Lee, T. S., Eds.; Springer, Dordrecht, 2009; pp 137–172. https://doi.org/10.1007/978-1-4020-9956-4_6.

(57) Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. Large-Scale Compensation of Errors in Pairwise-Additive Empirical Force Fields: Comparison of AMBER Intermolecular Terms with Rigorous DFT-SAPT Calculations. *Phys. Chem. Chem. Phys.* **2010**, *12* (35), 10476–10493. https://doi.org/10.1039/c002656e.

(58) Sherrill, C. D. Energy Component Analysis of π Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 1020–1028. https://doi.org/10.1021/ar3001124.

(59) Parker, T. M.; Sherrill, C. D. Assessment of Empirical Models versus High-Accuracy Ab Initio Methods for Nucleobase Stacking: Evaluating the Importance of Charge Penetration. *J. Chem. Theory Comput.* **2015**, *11* (9), 4197–4204. https://doi.org/10.1021/acs.jctc.5b00588.

(60) Gryn'ova, G.; Corminboeuf, C. Steric "Attraction": Not by Dispersion Alone. *Beilstein J. Org. Chem.* **2018**, *14*, 1482–1490. https://doi.org/10.3762/bjoc.14.125.

(61) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-up Strategy. *J. Chem. Theory Comput.* **2007**, *3*

(6), 1960–1986. https://doi.org/10.1021/ct700134r.

(62) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chem. Rev.* **2014**, *114* (1), 779–814. https://doi.org/10.1021/cr300461d.

(63) Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; Van Der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **2017**, *13* (3), 1034–1043. https://doi.org/10.1021/acs.jctc.7b00169.

(64) Wang, B.; Truhlar, D. G. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.* **2014**, *10*, 4480–4487. https://doi.org/10.1021/ct5005142.

(65) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An Optimized Charge Penetration Model for Use with the AMOEBA Force Field. *Phys. Chem. Chem. Phys.* **2017**, *19* (1), 276–291. https://doi.org/10.1039/c6cp06017j.

(66) Lemkul, J. A.; MacKerell Jr, A. D. Polarizable Force Field for RNA Based on the Classical Drude Oscillator. *J. Comput. Chem.* **2018**, *39* (32), 2624–2646. https://doi.org/10.1002/jcc.25709.

(67) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. **2019**, *48*, 371–394. https://doi.org/10.1146/annurev-biophys-070317-033349.Polarizable.

(68) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A General Intermolecular Force Field Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Phys.* **2017**, *147* (16), 161708. https://doi.org/10.1063/1.4991798.

(69) Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13* (1), 161–179. https://doi.org/10.1021/acs.jctc.6b00969.

(70) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 Dispersion Coefficient Model. *J. Chem. Phys.* **2017**, *147* (3). https://doi.org/10.1063/1.4993215.

(71) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. *J. Chem. Phys.* **2019**, *150* (15). https://doi.org/10.1063/1.5090222.

(72) Brzezinski, K.; Brzuszkiewicz, A.; Dauter, M.; Kubicki, M.; Jaskolski, M.; Dauter, Z. High Regularity of Z-DNA Revealed by Ultra High-Resolution Crystal Structure at 0.55. *Nucleic Acids Res.* **2011**, *39* (14), 6238–6248. https://doi.org/10.1093/nar/gkr202.

(73) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.

(74) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.

(75) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. https://doi.org/10.1002/jcc.21759.

(76) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297. https://doi.org/10.1039/b508541a.

(77) 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* (2), 1399–1410. https://doi.org/10.1039/C4CP04680C.

(78) Kruse, H. Xopt https://github.com/hokru/xopt (development version).

(79) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C.; Hellweg, A.; Helmich-Paris, B.; Holzer, C.; Huniar, U.; Kaupp, M.; Marefat Khah, A.; Karbalaei Khani, S.; Müller, T.; Mack, F.; Nguyen, B. D.; Parker, S. M.; Perlt, E.; Rappoport, D.; Reiter, K.; Roy, S.; Rückert, M.; Schmitz, G.; Sierka, M.; Tapavicza, E.; Tew, D. P.; van Wüllen, C.; Voora, V. K.; Weigend, F.; Wodyński, A.; Yu, J. M. TURBOMOLE: Modular Program Suite for Ab Initio Quantum-Chemical and Condensed-Matter Simulations. *J. Chem. Phys.* **2020**, *152* (18), 184107. https://doi.org/10.1063/5.0004635.

(80) TURBOMOLE V7.3 2018, a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; Available from https://www.turbomole.org.

(81) Weigend, F. A Fully Direct RI-HF Algorithm: Implementation, Optimised Auxiliary Basis Sets, Demonstration of Accuracy and Efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4* (18), 4285–4291. https://doi.org/10.1039/b204199p.

(82) Treutler, O.; Ahlrichs, R. Efficient Molecular Numerical Integration Schemes. *J. Chem. Phys.* **1995**, *102* (1), 346–354. https://doi.org/10.1063/1.469408.

(83) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F.; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E.; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. Psi4 1.4: Open-Source Software for High-Throughput Quantum Chemistry. *J. Chem. Phys.* **2020**, *152* (18), 184108. https://doi.org/10.1063/5.0006002.

(84) DePrince, A. E.; Sherrill, C. D. Accurate Noncovalent Interaction Energies Using Truncated Basis Sets Based on Frozen Natural Orbitals. *J. Chem. Theory Comput.* **2013**, *9* (1), 293–299. https://doi.org/10.1021/ct300780u.

(85) Hohenstein, E. G.; Sherrill, C. D. Density Fitting of Intramonomer Correlation Effects in Symmetry-Adapted Perturbation Theory. *J. Chem. Phys.* **2010**, *133* (1), 0–12. https://doi.org/10.1063/1.3451077.

(86) Hohenstein, E. G.; Sherrill, C. D. Efficient Evaluation of Triple Excitations in Symmetry-Adapted Perturbation Theory via Second-Order Møller-Plesset Perturbation Theory Natural Orbitals. *J. Chem. Phys.* **2010**, *133* (10). https://doi.org/10.1063/1.3479400.

(87) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of Symmetry Adapted Perturbation Theory (SAPT). I. Efficiency and Performance for Interaction Energies. *J. Chem. Phys.* **2014**, *140* (9). https://doi.org/10.1063/1.4867135.

(88) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96* (9), 6796–6806. https://doi.org/10.1063/1.462569.

(89) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023. https://doi.org/10.1063/1.456153.

(90) Papajak, E.; Truhlar, D. G. Convergent Partially Augmented Basis Sets for Post-Hartree-Fock Calculations of Molecular Properties and Reaction Barrier Heights. *J. Chem. Theory Comput.* **2011**, *7* (1), 10–18. https://doi.org/10.1021/ct1005533.

(91) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286* (3–4), 243–252. https://doi.org/10.1016/S0009-2614(98)00111-0.

(92) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Appointing Silver and Bronze Standards for Noncovalent Interactions: A Comparison of Spin-Component-Scaled (SCS), Explicitly Correlated (F12), and Specialized Wavefunction Approaches. *J. Chem. Phys.* **2014**, *141* (23). https://doi.org/doi:http://dx.doi.org/10.1063/1.4903765.

(93) Kozuch, S.; Gruzman, D.; Martin, J. M. L. DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction. *J. Phys. Chem. C* **2010**, *114* (48), 20801–20808. https://doi.org/10.1021/jp1070852.

(94) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. https://doi.org/10.1002/jcc.21759.

(95) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally Augmented Karlsruhe Basis Sets. *Theor. Chem. Acc.* **2011**, *128* (3), 295–305. https://doi.org/10.1007/s00214-010-0846-z.

(96) Neese, F. Software Update: The ORCA Program System, Version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8* (1), 4–9. https://doi.org/10.1002/wcms.1327.

(97) Izsak, R.; Neese, F.; Izsák, R.; Neese, F. An Overlap Fitted Chain of Spheres Exchange Method. *J. Chem. Phys.* **2011**, *135* (14), 144105. https://doi.org/10.1063/1.3646921.

(98) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. *J. Chem. Theory Comput.* **2017**, *13* (2), 554–562. https://doi.org/10.1021/acs.jctc.6b01041.

(99) Kruse, H. BrnoFF https://github.com/hokru/BrnoFF.

(100) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type

Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005.

(101) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. https://doi.org/10.1021/j100142a004.

(102) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829. https://doi.org/10.1529/biophysj.106.097782.

(103) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. AMBER 2016. University of California, San Francisco 2016.

(104) Williams, T. Kelley, C. Gnuplot 4.6 http://www.gnuplot.info/.

(105) Caldeweyher, E.; Ehlert, S.; Grimme, S. DFT–D4 Standalone https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dftd4.

(106) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 2018. University of California, San Francisco 2018.

(107) 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* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(108) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32* (5), 523–535. https://doi.org/10.1002/bip.360320508.

(109) 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* (2), 374–380. https://doi.org/10.1016/j.cpc.2012.09.022.

(110) 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* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(111) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271. https://doi.org/10.1021/j100308a038.

(112) Joung, I. 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* (30), 9020–9041. https://doi.org/10.1021/jp8001614.

(113) Nozinovic, S.; Fürtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-Resolution NMR Structure of an RNA Model System: The 14-Mer CUUCGg Tetraloop Hairpin RNA. *Nucleic Acids Res.* **2009**, *38* (2), 683–694. https://doi.org/10.1093/nar/gkp956.

(114) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham III, T. E.; Jurecka, 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* (9), 2886–2902. https://doi.org/10.1021/ct200162x.

(115) Mlynsky, V.; Kuhrova, P.; Zgarbova, M.; Jurecka, P.; Walter, N. G.; Otyepka, M.; Sponer, J.; Banas, P. Reactive Conformation of the Active Site in the Hairpin Ribozyme Achieved by Molecular Dynamics Simulations with ε/ζ Force Field Reparametrizations. *J. Phys. Chem. B* **2015**, *119* (11), 4220–4229. https://doi.org/10.1021/jp512069n.

(116) Schaftenaar, G.; Noordik, J. H. Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures. *J. Comput. Aided. Mol. Des.* **2000**, *14* (2), 123–134. https://doi.org/10.1023/A:1008193805436.

(117) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(118) DeLano, W. L. The PyMol Molecular Graphics System, Version 2.0.

(119) Jmol: An Open-Source Java Viewer for Chemical Structures in 3D.

(120) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(121) Schauperl, M.; Kantonen, S. M.; Wang, L. P.; Gilson, M. K. Data-Driven Analysis of the Number of Lennard–Jones Types Needed in a Force Field. *Commun. Chem.* **2020**, *3* (1). https://doi.org/10.1038/s42004-020-00395-w.

(122) Kantonen, S. M.; Muddana, H. S.; Schauperl, M.; Henriksen, N. M.; Wang, L. P.; Gilson, M. K. Data-Driven Mapping of Gas-Phase Quantum Calculations to General Force Field Lennard-Jones Parameters. *J. Chem. Theory Comput.* **2020**, *16* (2), 1115–1127. https://doi.org/10.1021/acs.jctc.9b00713.

(123) 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* (12), 978–988. https://doi.org/10.1002/bip.22322.

(124) Gavezzotti, A. The Lines-of-Force Landscape of Interactions between Molecules in Crystals; Cohesive versus Tolerant and Collateral Damage Contact. *Acta Crystallogr.*

Sect. B Struct. Sci. **2010**, *66* (3), 396–406. https://doi.org/10.1107/S0108768110008074.

(125) Dance, I. Distance Criteria for Crystal Packing Analysis of Supramolecular Motifs. *New J. Chem.* **2003**, *27* (1), 22–27. https://doi.org/10.1039/b206867b.

(126) Zhang, Z.; Vögele, J.; Mráziková, K.; Kruse, H.; Cang, X.; Wöhnert, J.; Krepl, M.; Šponer, J. Phosphorothioate Substitutions in RNA Structure Studied by Molecular Dynamics Simulations, QM/MM Calculations, and NMR Experiments. *J. Phys. Chem. B* **2021**, *125* (3), 825–840. https://doi.org/10.1021/acs.jpcb.0c10192.