

# QM Computations on Complete Nucleic Acids Building Blocks: Analysis of the Sarcin–Ricin RNA Motif Using DFT-D3, HF-3c, PM6-D3H, and MM Approaches

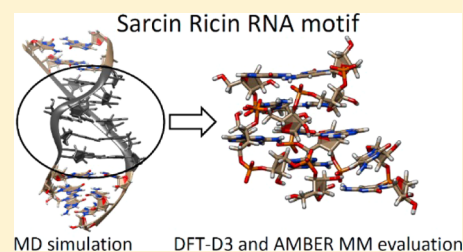
Holger Kruse,<sup>†</sup> Marek Havrila,<sup>†,‡</sup> and Jiří Šponer<sup>\*,†,‡</sup>

<sup>†</sup>CEITEC – Central European Institute of Technology, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

<sup>‡</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

## S Supporting Information

**ABSTRACT:** A set of conformations obtained from explicit solvent molecular dynamics (MD) simulations of the Sarcin–Ricin internal loop (SRL) RNA motif is investigated using quantum mechanical (QM, TPSS-D3/def2-TZVP DFT-D3) and molecular mechanics (MM, AMBER parm99bsc0+ $\chi_{ol3}$  force field) methods. Solvent effects are approximated using implicit solvent methods (COSMO for DFT-D3; GB and PB for MM). Large-scale DFT-D3 optimizations of the full 11-nucleotide motif are compared to MM results and reveal a higher flexibility of DFT-D3 over the MM in the optimization procedure. Conformational energies of the SRL motif expose significant differences in the DFT-D3 and MM energy descriptions that explain difficulties in MD simulations of the SRL motif. The TPSS-D3 data are in excellent agreement with results obtained by the hybrid functionals PW6B95-D3 and M06-2X. Computationally more efficient methods such as PM6-D3H and HF-3c show promising but partly inconsistent results. It is demonstrated that large-scale DFT-D3 computations on complete nucleic acids building blocks are a viable tool to complement the picture obtained from MD simulations and can be used as benchmarks for faster computational methods. Methodological challenges of large-scale QM computations on nucleic acids such as missing solvent–solute interactions and the truncation of the studied systems are discussed.



## INTRODUCTION

RNA plays many different and important roles in biology, and more and more functions of RNA are still being discovered. The RNA is typically single stranded, consisting of A, C, G, and U nucleotide units, and exhibits a high likelihood to fold back onto itself.<sup>1</sup> The 2'-OH group (absent in DNA) is a powerful hydrogen bonding group, allowing formation of many non-canonical RNA structures.<sup>2</sup> Folded RNA structures consist of short Watson–Crick double helices that alternate with noncanonical regions.<sup>3</sup> The noncanonical RNA regions are formally (at the level of secondary 2D structure) unpaired, but they usually fold into precisely shaped tertiary (3D) structures where every single H-bond, stack, or local backbone conformation<sup>4</sup> matters (reviewed in ref 2b). This makes RNA a difficult target for structure predictions<sup>5,1</sup> as well as for molecular modeling and simulations.<sup>6,7</sup>

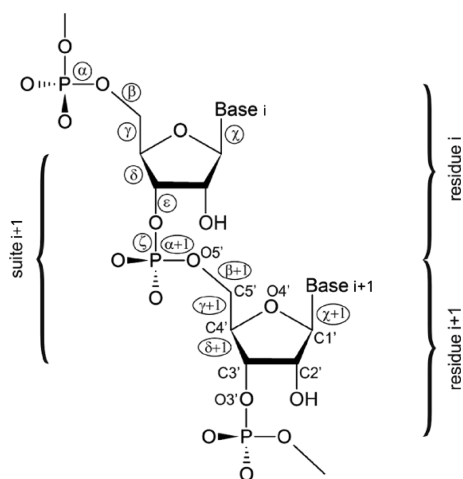
RNA motifs are recurrent, robustly 3D-shaped, RNA building blocks.<sup>2b,3</sup> They are defined as ordered arrays of noncanonical and canonical base pair interactions, complemented by specific consecutive local backbone conformations that support their unique structures. RNA backbone geometries were recently classified using *suites* rather than nucleotides as the basic conformational units<sup>4b</sup> (for a review, see ref 8). A *suite* is the sugar-to-sugar backbone unit including backbone dihedrals between the  $\delta_i$  and  $\delta_{i+1}$  backbone angles of two consecutive nucleotides (Figure 1).

The sarcin–ricin internal loop (SRL) is one of the most salient RNA motifs. It is named after the toxins sarcin and ricin that inactivate ribosome by cleaving key sites present in helix 95 of the large ribosomal subunit.<sup>9</sup> The SR motif has a precisely defined 3D structure formed only by specific sequences and folds autonomously, i.e., in a context-independent manner.<sup>10</sup> SRL includes many unique structural features, such as an absolutely conserved bulged guanosine with a highly uncommon high-*anti*  $\chi$  dihedral value involved in a specific dinucleotide platform forming a nucleobase triad, a unique backbone topology with a nucleotide double flip-over (RNA S-turn), several conserved base-phosphate (BPh) H-bonds,<sup>11</sup> backbone suites with highly uncommon low- $\beta$  dihedral values, and additional intricate H-bonds. SRL is one of the most challenging RNA building blocks to be described by computational methods.

Earlier explicit solvent molecular dynamics (MD) simulations revealed that the SRL motif is rather stiff and suggested that SRL is well described by the molecular mechanics (MM) force field.<sup>12</sup> However, later investigations clarified that although the SRL structure stays rather stable in short simulations, some of its salient local signature features are lost, indicating the likelihood of further degradation in longer simulations.<sup>13</sup> Noncanonical RNA structural motifs are challenging targets

Received: March 3, 2014

Published: April 23, 2014



**Figure 1.** Scheme of a dinucleotide unit showing atom and residue numbering and naming of backbone torsional angles. There are two possibilities of how to define consecutive sugar–phosphate backbone segments. The residue (= nucleotide unit) is defined from phosphate to phosphate and includes six torsional angles ( $\alpha$  to  $\zeta$ ) and the glycosidic torsion angle  $\chi$ . The conformation of the suite excludes the  $\chi$  angles and ranges from  $\delta_i$  to  $\delta_{i+1}$ . The suite is numbered according to the  $i+1$  nucleotide. The suite-based description<sup>4b</sup> of the sugar–phosphate backbone will be used in this paper.

for the current generation of RNA force fields. One of the reasons is the difficulty to provide a balanced description of different backbone conformations using simple pair-additive force fields, commonly tuned by reparametrizations of specific backbone torsional potentials.<sup>14</sup> One of the best performing RNA force fields is the Cornell et al. (AMBER) basic force field<sup>15</sup> supplemented by the  $\chi_{OL3}$ <sup>16</sup> and bsc0<sup>17</sup> corrections. Nevertheless, even this force field is incapable of providing a fully balanced description of the SRL.<sup>18</sup> This force field will be used for all MM computations in the present study. The capability to describe RNA motifs will be a crucial test for future polarizable force fields and fast electronic structure quantum–chemical (QM) methods intended to describe nucleic acids.

Modern QM computations are inherently more accurate compared to MM. However, their applicability to nucleic acids is limited by the size of the systems, difficulties in inclusion of solvent, and lack of conformational sampling. QM studies have, until now, been mostly used to benchmark computations on very small systems.<sup>19</sup> Nevertheless, the fast development in QM methodology<sup>20</sup> has increased chances to tackle large meaningful fragments of nucleic acids.<sup>21</sup> Such calculations obviously cannot replace MM simulations mainly due to conformational sampling, but they can complement the simulation data and refine the picture obtained by MM computations. QM computations including complete nucleic acids building blocks may significantly enrich the computational chemistry of nucleic acids.<sup>18,19a,22,23</sup>

Density functional theory (DFT) is an efficient QM technique for systems over 100 atoms due to its unmatched balance between accuracy and computational costs.<sup>24</sup> Conventional DFT methods are unable to account for long-range correlation effects such as London dispersion interactions.<sup>25</sup> However, modern dispersion-corrected methods such as the approach by Grimme (termed DFT-D3)<sup>26</sup> can very well include dispersion interactions. For an overview of other methodologies, see ref 20a. DFT-D3 provides accurate

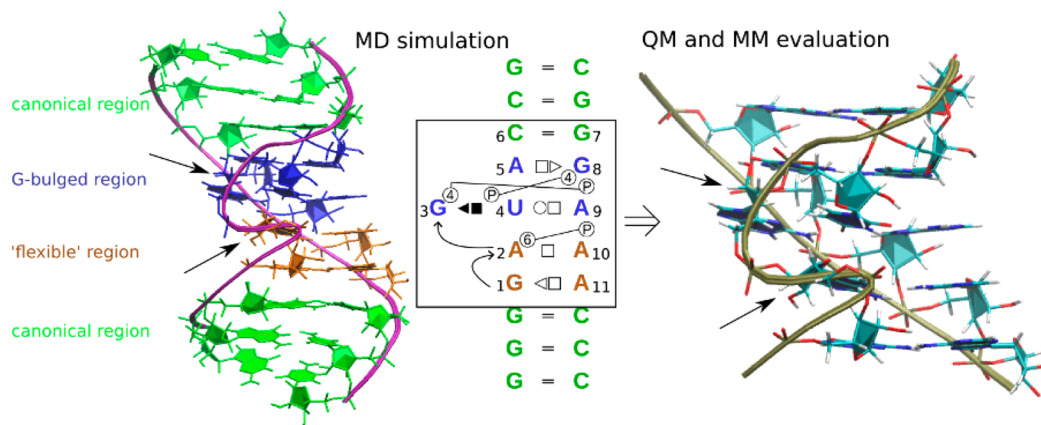
energies<sup>27</sup> and structures<sup>26b,28</sup> and was successfully employed in a study of the guanine quadruplex DNA (G-DNA).<sup>22a</sup> The calculations revealed large differences between the QM and MM data, resulting in a major revision of an earlier prediction<sup>29</sup> of relative free energies of different G-DNA topologies based on the MM-PBSA free energy simulation approach. In contrast to MM, modern DFT-D3 methods accurately describe relative energies of DNA and RNA backbone families.<sup>30</sup>

In this work, we apply for the first time large-scale DFT-D3 computations to a complete RNA building block, namely, the SRL motif. We investigate a set of different conformational substates of the SRL motif based on explicit solvent MM MD simulation trajectories. We perform optimizations at the QM and MM level using implicit solvation models. The comparison of relative stabilities that are predicted from QM and MM gives insights into the performance of the force field and solvation models. Additional information is gained by comparing the optimized structures at the respective level of theories. We also compare the DFT-D3 results with fast QM methods such as HF-3c,<sup>31</sup> TPSS-D3 with a small basis set and the geometrical counterpoise correction<sup>32</sup> (gCP), and the semi-empirical PM6-D3H<sup>33</sup> methods. We discuss the limitations of applying QM methods to complete RNA building blocks and their relation to experimental structural, bioinformatics, and MM simulation data.

## COMPUTATIONAL DETAILS

**MD Simulations.** The AMBER Cornell et al.<sup>15</sup> force field with bsc0<sup>17</sup> and  $\chi_{OL3}$ <sup>16</sup> corrections was applied. We used TIP3P<sup>34</sup> water and Na<sup>+</sup> with a radius of 1.868 Å and 0.0028 kcal/mol well depth. For a detailed discussion of the effects of different water models and the ion parameters, types, and concentrations on RNA simulations, see ref 35. For the purpose of the present study, the choice of water and ion model should have negligible effects on the results, also considering that SRL does not have any salient ion-binding sites.<sup>12</sup> Simulations were performed using the AMBER suite of programs<sup>36</sup> with established protocols.<sup>35b,37–40</sup> For more details, see the Supporting Information.

**MM Optimizations.** MM minimization and single-point calculations were done using the sander module of AMBER.<sup>36</sup> The generalized Born (GB) implicit solvation model of Hawkins, Cramer, and Truhlar<sup>41</sup> using the Tsui et al. parameters<sup>42</sup> and the (linear) Poisson–Boltzmann (PB) implicit solvation model as provided by sander<sup>43</sup> were applied. The adaptive Poisson–Boltzmann solver (APBS) from Baker et al.<sup>44</sup> was also used in conjunction with sander using the iAPBS interface.<sup>45</sup> For the PB method, the grid spacing was reduced to 0.2 Å, while for APBS, the grid spacing for the optimization was set to 0.35 Å and for single points to 0.15 Å. Exact details of the inputs can be found in the Supporting Information. Optimizations with sander use the default optimizer and thresholds with a mixture of conjugate gradient and steepest descent steps. Particularly for the PB and APBS optimization, the minimizer of sander would often stop after about 1000 cycles without reaching full gradient convergence because the line minimization step was unable to find a lower energy structure (a common problem according to the AMBER mailing list<sup>46</sup>). Test calculations with an alternative, group-internal minimizer gave essentially the same structures and energies (energy deviations below 1 kcal/mol). Thus, for our purposes, the structure can be assumed to be sufficiently converged.



**Figure 2.** (Left) Simulated SRL motif, i.e., the noncanonical part (G-bulged and flexible region) sandwiched by three canonical GC base pairs (green) at each end (21 nucleotides). The G-bulged region (blue) is also known as the GpUpA/GpA miniduplex.<sup>4a</sup> The name flexible region (brown) originates from early X-ray studies assuming high flexibility through weak base pairing.<sup>12</sup> The purple-colored backbone features a visible S-like shape caused by the flip-over of the A2 and G3 sugars (black arrows). (Middle) Base pairing and residue numbering of the simulated system. Standard annotation of base pairing<sup>2a</sup> and base-phosphate interactions<sup>11a</sup> are used. (Right) Truncated system used for QM and MM optimizations and energy calculations. It includes the SRL motif alone and one GC base pair on top of the structure (11 nucleotides, default VMD coloring is used). Black arrows point at the A2 and G3 sugars.

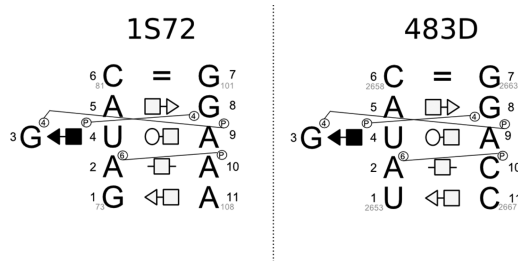
**QM Computations.** DFT calculations were mostly carried out with Turbomole V.6.3 (for exceptions, see below).<sup>47</sup> The meta-GGA TPSS functional<sup>48</sup> is applied together with the DFT-D3 correction<sup>26a</sup> using the Becke–Johnson damping variant<sup>26b</sup> as default damping functions (TPSS-D3). Additional single-point calculations with the hybrid functionals PW6B95-D3<sup>49</sup> and M06-2X,<sup>50</sup> as well as with TPSS-gCP-D3/def2-SVP and HF-3c, were carried out with the ORCA program (V. 3.0.0).<sup>51</sup> The geometrical counterpoise correction<sup>32</sup> (gCP) is calculated within the ORCA program. Unless noted otherwise, the def2-TZVP all-electron basis set is used for all DFT calculations.<sup>52</sup> Computations were speeded up using the resolution of identity (RI-J) approximation<sup>53</sup> for the Coulomb part of the two-electron integrals. ORCA calculations apply the chain of sphere (COSX) approximation<sup>54</sup> for the Fock exchange part of all hybrid functionals. The multipole accelerated resolution of identity approximation (MARLI-J)<sup>55</sup> was additionally used during the geometry optimizations with Turbomole. The Turbomole grid *m4* and the ORCA grid *GRID4* were used for all DFT calculations and the respective programs.<sup>56</sup> Single-point energies were converged until the energy change in subsequent SCF iterations is below  $10^{-7}$  E<sub>h</sub>. Geometries were optimized until the energy change is below  $10^{-7}$  E<sub>h</sub> and the gradient norm below  $10^{-3}$  E<sub>h</sub>. The optimizations were carried out using a group-internal optimizer using approximate normal coordinates and a rational function approach for the step size.<sup>57</sup> The COSMO<sup>58</sup> implicit solvation model was applied using an  $\epsilon$  of 78.4. PM6-D3H calculations were performed using Mopac (version 12)<sup>33a</sup> and the DFT-D3 program<sup>26a</sup> from Grimme's group for the D3H correction. The *PRECISE* keyword was used.

**Backbone Analysis.** Backbone angles were studied using the Dangle and Suitename<sup>59</sup> programs. Molecular structure file format conversion was done using Open Babel.<sup>60</sup> Visualization of structure employed the UCSF Chimera package,<sup>61</sup> Molden,<sup>62</sup> and VMD.<sup>63</sup>

## RESULTS AND DISCUSSION

**Short Description of the SR Motif.** The local internal SR motif consists of five characteristic, consecutive, noncanonical

base pairs flanked by canonical base pairs, resulting into the following base pair interactions: cWW-tSH-tHH-cSH-tWH-tHS-cWW (nucleotides 1–5 and 8–11; Figures 2 and 3 and

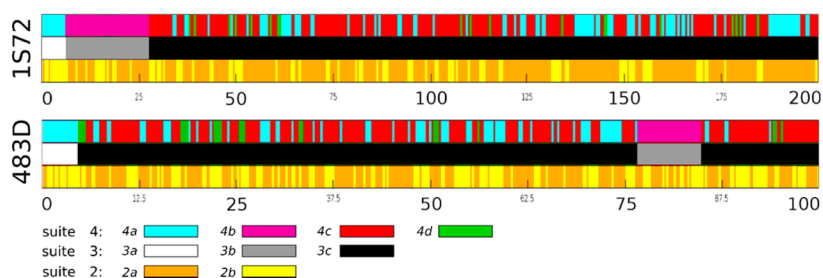


**Figure 3.** Base pairing and residue numbering of the 11nt SRL in structures 1S72<sup>10a</sup> and 483D,<sup>10b</sup> corresponding to the truncated segment in Figure 2. Standard annotation of base pairing<sup>2a</sup> and base-phosphate interactions<sup>11a</sup> is used. The small gray numbers indicate the original residue numbering of the PDB files.

Figure S1, Supporting Information; see ref 2 for the explanation of the base pair classification). The structure further contains base-phosphate (BPh) interactions<sup>11a</sup> (A2(N6)-A10(O2P), G3(N1/N2)-A9(O2P), and G8(N1/N2)-U4(O2P) (Figure 3 and Figure S1, Supporting Information) and cross-strand stacks (A5/A9 and G3/G8). Furthermore, there are U4(O2P)-G3(O2'), G3(O2')-G8(O6), and G9(O2')-A9(O4') interactions resulting in a compact net of hydrogen bonds and base stacks. The truncated and full models further include one and six additional canonical base pairs, respectively (Figures 2 and 3).

The native backbone conformation of SR motifs is generally classified<sup>4b</sup> as **g1aG<sub>1</sub>5zA<sub>2</sub>4sG<sub>3</sub>#aU<sub>4</sub>1aA<sub>5</sub>** (strand I) and **G<sub>7</sub>1aG<sub>8</sub>1eA<sub>9</sub>1aA<sub>10</sub>1aA<sub>11</sub>** (strand II). The first nucleotide in strand I is marked as “g” and belongs to the flanking canonical base pairs that we do not number. The symbols in bold letters define conformational families of the individual backbone suites<sup>4b</sup> around the phosphates linking the two nucleotides (Figure 1 and ref 30a). The SR motif taken from the 1S72 crystal structure has an unclassified conformation in suite 2 (Figure 1) instead of the native **5z** family, which is likely due to





**Figure 4.** Time development of backbone segment with characteristic substates of suites 2 (bottom), 3 (middle), and 4 (top) around the G-bulge in the 1S72 and 483D simulations (Figures 1 and 3). A given overall conformation consists of a combination of these three substates. It can be inferred from the figure using the color strips and is marked in the text via a three-symbol code. Note that the substates *a–d* occurring for different suites are not equivalent, i.e., substate 2*a* is not the same as 3*a*, etc., see the text for exact classification. While substates 3*a* and 4*a* are in agreement with the X-ray structure, the X-ray substate of suite 2 is quickly and irreversibly lost (more details in the text). The X-ray structure is marked as X in the text. See Figures S3 and S4 of the Supporting Information for detailed time development.

the lower resolution (2.4 Å) of the experimental structure. The high-resolution X-ray structure 483D shows the native 5*z* family in suite 2. An exhaustive description of the structural features of the SR motif and backbone conformations can be found elsewhere.<sup>18</sup>

**MD Simulations Sample Non-Native Backbone Conformation of the S-Turn.** Two SR motifs are selected from the PDB database.<sup>64</sup> The first SR motif is taken from the 5S rRNA of *Haloarcula marismortui* (H.m.) structure (PDB code 1S72,<sup>10a</sup> residues 73–83 and 99–108). Its two flanking GU wobble (cWW) base pairs (U82/G100 and G83/U99) are replaced by cWW (cis-Watson–Crick Watson–Crick, canonical) GC base pairs (Figure S1, Supporting Information) to prevent end fraying in simulations and to match the canonical SRL context. Because the structure 1S72 is determined at 2.4 Å resolution, we further simulated the isolated SR motif determined at 1.1 Å resolution (PDB code 483D,<sup>10b</sup> residues 2653–2658 and 2663–2667). Structure 483D has a slightly different sequence, but its shape is within the resolution limit identical to the 1S72 SR motif. The terminal UA base pair was replaced by a cWW CG base pair (Figure S1, Supporting Information) to prevent end fraying.

A 1000 ns long MD simulation of the 1S72 SR motif remained globally stable with a RMSD value of about 1 Å for the noncanonical part (Figure S2, Supporting Information). While all base pairs were entirely stable, we observed reversible disruptions of the signature G8(N1/N2)-U4(O2P) BPh interaction and changes in backbone suites of phosphates 2, 3, and 4 (see Figure 1 for suite definition and Figure 3 for the nucleotide numbering). The first 200 ns provide enough sampling for the purpose of our study.

Two 100 ns simulations were performed for the 483D structure. We observed similar behavior for both simulations, and thus, only one was used for further investigations. The structure was again stable except for small fluctuations of U1/C11 and A2/C10 base pairs (Figure S2, Supporting Information). The observed changes in the backbone conformation were similar to the 1S72 SR motif simulation.

To obtain the initial starting coordinates for the optimizations, the MD trajectory was averaged over a 100–200 ps time windows. Coordinates in the simulation were written each ps, meaning that the averaging includes 100 to 200 snapshots per conformation. The windows of averaging were manually checked for backbone topology consistency to prevent accidental mixing of two conformations (i.e., there were no backbone flips during the averaging periods). Further

information about the MD simulations and selection of the structures can be found in the Supporting Information.

MD simulations always sample a continuum of geometries consisting of the individual snapshots. Rigorous quantitative analysis and classification of all snapshots occurring during MD trajectories is a difficult and essentially unresolved issue. Even when using sophisticated clustering methods, the results may depend on parameters of the clustering. In our study, we needed to select a few geometries that would represent the main structural developments along the trajectories and could be used for the QM computations. We have decided to select these geometries based on semi-quantitative monitoring of all dihedral backbone angles of the key suites 2, 3, and 4 representing the S-turn in strand I.

Figure 4 captures the time development of the key suites 2–4 in a schematic fashion. For 1S72, the suites 2, 3, and 4 change immediately after the simulation start and continue to exhibit a complex dynamics, never adopting the native conformation. The X-ray conformation for 483D is slightly more stable and is sampled during the first 400 ps, before being irreversibly lost. A more detailed view of the angle time development, including the X-ray conformation, can be found in Figure S3 for 1S72 and Figure S4 for 483D of the Supporting Information. Because the sugar–phosphate backbone conformation of the SR motif (see above) is invariably observed in all experimental structures, we have to conclude that the simulation does not reproduce the native sugar–phosphate dihedrals in the SR motif.

We introduce a systematic naming scheme for the structures sampled in the simulations based on the substates of the suites 2, 3, and 4. Each substate is alphabetically indexed. Note that our selected conformations cover most, but not all, possible substate combinations. A given conformation is described by the suite number (2, 3, and 4) and the respective suite substate (*a*, *b*, *c*, or *d*), e.g., the first conformation in the MD simulation of 1S72 is denoted as 2*b*/3*a*/4*a* (see Figure 4 and the color coding). We use italics to distinguish these substates from common backbone suite classification, which is shown in bold in the paper.

Before further analyses, we need to make three comments. First, some suite substates sampled during the simulations do not correspond to any established family in the current RNA backbone classification.<sup>4b,30b</sup> As will be discussed later in the study, it may stem from the accuracy limits of the force field. Second, as we have done our analysis primarily using the 1S72 structure (the 483D data mainly serves for verification purposes), the naming scheme is proposed based on the

**Table 1. QM versus MM Cross-Comparison of Conformational Energies (kcal/mol) between SR Motif Substates of the 1S72 SRL Structure Simulation<sup>a</sup>**

state	MM GB opt.				MM PB opt.				MM APBS opt.				QM COSMO opt.			
	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$
X <sup>b</sup>	19.4	24.9	21.9	23.7	24.2	27.0	24.3	28.0	49.0	67.2	60.1	42.3	35.7 (32.5) <sup>e</sup>	45.3 (41.8) <sup>e</sup>	40.3 (31.3) <sup>e</sup>	21.1 (15.9) <sup>e</sup>
2b/3a/4a <sup>c</sup>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2a/3b/4b	3.9	−0.7	5.9	20.3	3.0	−5.0	2.6	13.8	7.5	5.4	9.3	21.5	4.9 <sup>d</sup>	2.9 <sup>d</sup>	3.2 <sup>d</sup>	1.8 <sup>d</sup>
2b/3b/4b	4.1	−1.9	6.5	19.4	4.0	−0.8	7.4	20.1	3.7	−1.6	5.6	20.7	0.2 <sup>d</sup>	0.1 <sup>d</sup>	−0.4 <sup>d</sup>	−0.2 <sup>d</sup>
2a/3c/4c	1.9	2.3	3.2	14.7	0.2	1.5	2.8	13.2	0.1	1.2	3.2	15.0	−2.4 <sup>d</sup>	0.8 <sup>d</sup>	−2.9 <sup>d</sup>	−7.0 <sup>d</sup>
2a/3c/4a	−0.7	−1.8	−3.5	−4.6	−0.4	−0.1	−1.4	−4.2	−0.7	1.7	−0.3	−3.7	−3.7 (−6.9) <sup>e</sup>	−2.9 (−6.4) <sup>e</sup>	−3.5 (−12.5) <sup>e</sup>	−2.3 (−7.5) <sup>e</sup>
2a/3c/4d	1.7	−2.8	−0.3	4.0	0.9	−1.2	1.8	8.4	1.0	−1.5	0.4	9.8	−0.7 <sup>d</sup>	−0.5 <sup>d</sup>	−0.9 <sup>d</sup>	−2.6 <sup>d</sup>
2b/3c/4c	4.1	1.7	7.5	19.8	4.0	4.8	10.0	19.4	2.8	1.1	6.0	18.7	0.2 <sup>d</sup>	−0.5 <sup>d</sup>	1.1 <sup>d</sup>	−5.3 <sup>d</sup>

<sup>a</sup>MM refers to the bsc0<sub>OL3</sub> force field with GB ( $\Delta E_{GB}$ ), PB ( $\Delta E_{PB}$ ), and APBS ( $\Delta E_{APBS}$ ) implicit solvation models, while QM refers to COSMO-TPSS-D3/def2-TZVP ( $\Delta E_{QM}$ ). Structures were optimized using all four levels of theories (indicated by opt.) and energies were calculated by all methods. See Figures 2 and 3 for the definition of the optimized 11-nt fragment and Figure 4 for the nomenclature of the substates. Starting and final structures are provided in the Supporting Information. <sup>b</sup>X-ray structure. <sup>c</sup>Reference structure. The geometry is taken shortly after the simulation start (4.0–4.1 ns) when the structure is already well equilibrated. With exception of suite 2, the experimental substates are maintained. <sup>d</sup>QM optimization converges suite 4 to the 4a substate (i.e., #a family), thus changing the intended conformation. <sup>e</sup>Alternative energy corrected by the estimate (in kcal/mol: QM = 5.2, GB = 3.2, PB = 3.5, APBS = 9.0) for the terminal 5'-end hydroxyl H-bond absent in this structure. For more details, see the text.

1S72 simulation. When selecting representative geometries using the 483D simulation, we made sure to pick as similar conformations to the structures selected from 1S72 as possible. Third, the naming scheme considers only three (albeit the most interesting) backbone suites, while obviously also other parts of the simulated molecules show dynamic behavior. However, their explicit consideration would make the analysis very cumbersome and complex. Thus, our classification of substates along the trajectories is an imperfect compromise.

In our 1S72 simulation, suite 2 samples two substates. The substate 2a is closest to the native 5z family, except that the  $\alpha$  angle deviates by roughly 60° from the 5z family centroid, reaching a value of ~100°. This backbone topology is close to the 5p family, a rare tentative family that is not yet included in the consensus list of known RNA backbone families.<sup>4b</sup> The family of substate 2b is unclassified (unknown). For suite 3, we differentiate between three substates; 3a can be classified as 4s family, while 3b and 3c do not fit to any known backbone family. For suite 4, we see four different substates. Substate 4a corresponds to #4 family, 4c to the 6g family, and 4d to the 6s family, while 4b is unclassified. The substate 4d differs from 4c by a so-called  $\alpha$ - $\gamma$  flip, i.e., coupled transition of the  $\alpha$  and  $\gamma$  backbone angles (4c and 4d have  $\alpha/\gamma$  values of ~200°/60° and ~75°/175°, respectively).<sup>65</sup> Such an  $\alpha$ - $\gamma$  flip also occurs for the 4b substate in the MD simulation, but it is very short-lived and is not present in any of our selected conformations. Thus, it is not given a systematic name and is not differentiated from 4b in Figure 4. See Figure S3 of the Supporting Information for details of the MD development and Table S4 of the Supporting Information for an overview of the backbone families of the selected conformations.

The first conformation taken from the 1S72 MD simulation is classified as 2b/3a/4a and was obtained by averaged snapshots between 4 and 4.1 ns. It is structurally closest to the X-ray conformation. Suite substates 3a and 4a represent essentially the X-ray conformation with minor (negligible) deviations; however, 2b deviates strongly from the X-ray conformation in several angles. As mentioned above, substate 2a would resemble the X-ray conformation more closely, but it is not sampled within the first few ns of the 1S72 simulation where the 3a/4a combination is stable. Thus, a (hypothetical)

2a/3a/4a conformation would almost correspond to the X-ray structure, but it never occurs in the course of the simulation.

The substates for the different suites are partially coupled. Substate changes in suite 3 occur simultaneously with substate changes in suite 4, e.g., the transition from 3a to 3b is coupled with the transition from 4a to 4b, similar to the 3b/3c and 4b/4c transitions. However, suite 4 also samples substates 4d and 4a without changes in the 3b substate. Changes in suite 2 are not coupled with other suites.

The backbone family classifications were done according to the structural bioinformatics Suitename program based on available RNA X-ray structures.<sup>4b</sup> Obviously, even such a comprehensive bioinformatics approach may miss some real RNA suites. Thus, the MD-derived suites not fitting to the classification might just have been underrepresented in the RNA structure database used to derive the classification. However, it is more likely that the unclassifiable suites reflect inaccuracy of the MM method. We have no tools to differentiate between the two scenarios (for a review discussing the relation between experimental and computed data, see ref 8). Nevertheless, even if the presently unclassified individual MD suite geometries correspond to some real but hitherto unknown backbone families, it is evident that the simulations do not properly describe the S-turn backbone because all simulations quickly deviate from the native combination of SR motif backbone families (see above) and never return back.

**Comparison of MM and QM Relative Energies of Different Substates of SRL.** We first compared conformational energies calculated using QM and MM methods for eight different conformations taken from the X-ray structure and simulation of the 1S72 SR motif (Table 1). We have optimized these structures using the MM force field with three continuum solvent models and with the QM/COSMO approach. Then, for each set of geometries, their relative energies are calculated using all four methods. The computations provide several interesting results but also illustrate challenges in accurate conformational energy computations on nucleic acids building blocks. Note that although we have included implicit solvation models, the energies are far from showing differences in free enthalpies of solvation (as in  $\Delta G_{\text{solv}}$ ).<sup>33b,66</sup>

Table 1 shows a cross-comparison between single-point calculations of all four levels of theory (GB, PB and APBS MM and COSMO-DFT-D3) on structures optimized using these levels of theory.

**The X-ray Structure Is High in Energy.** At first, we need to comment on one peculiar fact. The “experimental” substate *X* has, even after geometry optimization, a very high energy compared to all the other structures. It is much higher in energy than the *2b/3a/4a* structure optimized after short simulation, which remains close to the native S-turn geometry.

The high energy of *X* is observed for all methods across all structures, indicating a systematic structural cause. We first expected that the difference may be caused by the initial assignment of the hydrogen positions. Thus, we inspected their assignment for the X-ray structure as done by the x-leap program and also tried several alternative orientations manually. However, based on GB-MM optimizations, varying initial hydrogen positions has improved the energies of the optimized *X* structure by no more than 5 kcal/mol (data not shown). This indicates that bad H atom positions are not the main cause of the high energy of *X*. Comparing the individual strands of the *X* and *2b/3a/4a* (using the GB-optimized structures) conformations, we see DFT-D3 energy differences of about 20 kcal/mol for strand I, which includes the S-turn, and 10 kcal/mol for the complementary strand. Then, we averaged the first 0 to 50 ps of the production MD simulation. This early MD average shows after GB-MM optimizations and DFT-D3 single-point evaluation an energy just 2.4 kcal/mol higher than the reference structure *2b/3a/4a*. Thus, a short MD was able to essentially eliminate the high energy, which the optimization procedures (all methods) starting directly from the X-ray structure were not capable of doing. Finally, we used the structure directly after the equilibration as the starting geometry for GB optimization. Its DFT-D3 energy was only 2.9 kcal/mol above the *2b/3a/4a* structure. In summary, the high relative energy of the optimized X-ray structure is likely caused by a larger number of high-energy local conformational features present in the starting X-ray structure accumulated over multiple nucleotides from which the optimization procedures cannot escape. A short production MD or even the MD equilibration protocol are, however, capable of relaxing the structure.

The above result may have an important implication for large-scale QM studies of RNA building blocks. As noticed elsewhere,<sup>67</sup> direct utilization of experimental X-ray geometries for energy computations is a risky procedure that often leads to biased energies. The X-ray structures are not intended as input data for accurate energy computations and are affected by the primary experimental data errors and refinement errors. Errors entirely acceptable from the structural point of view may have detrimental effects on subsequently calculated energies, as energies are highly nonlinearly dependent on the coordinates. Our present results further show that for larger nucleic acids fragments even gradient optimization may be in some cases insufficient to bring the necessary relaxation of the experimental structures. In summary, when using X-ray structures for energy computations, the structures should be very carefully inspected and preferably somehow preprocessed. Blindfolded use of the raw experimental coordinates for energy computations is risky. Although we cannot make any definitive conclusion from calculations on just one system, our result might indicate that direct energy calculations using larger fragments of experimental structures of nucleic acids may be quite difficult. Note

that we do not suggest any error in the experimental structure; it looks completely correct from the structural biology point of view.

Because of the high energy of the *X* structure, the *2b/3a/4a* state will be taken as the point of reference. It is still structurally close to the starting X-ray structure. Note that the significance of the reference structure is not the same as in the case of minima in QM studies on small molecular clusters and can be arbitrarily chosen.

**There Is a Visible Difference between the Relative Energies of Substates Predicted by MM and QM Computations.** When comparing the structures, the least dynamical is suite 3. We can consider its *3a* state as characteristic for starting structures, *3b* state characterizes transition structures, and *3c* state represents dominant structures in the later part of the simulation. The GB single-point calculations performed on the GB optimized structures (GB//GB results, i.e., the first column of Table 1) predict the substate *2a/3c/4a* as having the lowest energy, while the other geometries (except for *X*) are within a 5 kcal/mol energy range. These GB//GB results are in basic agreement with the MD simulation, where the similar *2a/3c/4c* and *2a/3c/4a* geometries are highly populated (or, more precisely, appear to be preferred by the simulation method during the trajectory over the reference structure *2b/3a/4a* (Figure 4). One needs to keep in mind that we have just a few geometries representing a very limited sampling. In addition, the solvation model is based on considerable approximations. Thus, we should not over-interpret small energy differences of ~5 kcal/mol. The basic result is that all structures sampled in the MD simulation trajectory have similar energies and are in agreement with their sampling along the simulation. Note that we obviously could calculate more MM points, but as the core of the study is comparison of MM and QM computations, we consistently use the same number of structures for all methods.

The computation of PB energies on the GB structures (PB//GB data, second column of Table 1) gives a slightly different picture. The lowest energy substate is shifted to *2b/3b/4b*, but *2a/3c/4a* and *2b/3a/4a* are essentially equal in energy (below 2 kcal/mol difference). Notable is that the intermediate conformations *2a/3b/4b* and *2b/3b/4b* indicate no energy barrier for the transition to the *2a/3c/4c* substate. At the GB//GB level, these two geometries have energies about 4 kcal/mol higher than *2a/3c/4a*. However, the PB//GB results are consistent with the GB//GB data in the sense that all geometries (except for *X*) have more or less similar energies, still within 5 kcal/mol range. The trend of the APBS results (APBS//GB data, third column of Table 1) agrees with the trend of the GB results with a slightly larger range for the relative energies. Nevertheless, we suggest that also the PB//GB and APBS//GB results can be considered as being consistent with the simulation behavior.

It is obvious that the accuracy of continuum solvent models is limited, and moderate differences (up to 10 kcal/mol) between the data obtained by PB, GB, and APBS methods on the same GB geometries are not surprising. These differences are reflecting common uncertainties stemming from the many approximations inherent to continuum solvent computations of conformational energies of nucleic acids building blocks.<sup>68</sup> Further analysis of the differences in solvation models is beyond the scope of this work, and we plan to provide more analysis in subsequent studies. However, the MM data is clearly quite sensitive to the implicit solvation model as seen from the



data herein and other studies,<sup>69</sup> which suggests that results obtained with a single implicit solvation model should be interpreted with caution.

The QM single points on the GB structures (QM//GB data, fourth column of Table 1) show a sharply different picture. The 2a/3c/4a geometry has still the lowest energy and remains relatively close (difference of less than 5 kcal/mol) to the reference 2b/3a/4a conformation. However, most structures that occur later in the simulation are substantially penalized by the QM method compared to MM. The intermediate conformations 2a/3b/4b and 2b/3b/4b have rather high relative energies of 20 and 19 kcal/mol, and also the 2a/3c/4c conformation has a fairly high relative energy of about 15 kcal/mol. This high relative energy of the highly populated 2a/3c/4c structure is a major difference from the MM results.

The results indicate that (i) there is a large degree of nonequivalency between the QM and MM potential energy surfaces and (ii) the QM method appears to punish (relatively to the MM description) many of the structures that occur at the later stages of the simulations, when the simulated structure already deviates from the native-like geometry (represented rather by 2b/3a/4a than by X). Thus, the results might explain why the MD simulation deviates from the initial native arrangement. When taking the QM energies as reference, the MM approach is biased toward the non-native structures compared to the QM data. These energy differences are too large to be attributed to accuracy limits of the continuum solvent calculations. They likely reflect true differences between the solute force field and QM descriptions of the RNA molecules. We noticed above that results obtained with a single implicit solvation model should be interpreted with caution. Nevertheless, the reported differences between QM and MM relative energies are several times larger than the differences between the three different MM solvation models. This suggests that the main part of the QM vs MM difference can be primarily attributed to different description of the intrinsic energetics of the RNA by QM and MM levels of theory. In our earlier analogous study of smaller two-tetrad guanine quadruplex stems, we have been able to prove this suggestion explicitly by analysis of gas phase energies of the GpG dinucleotides.<sup>22a</sup> Unfortunately, such gas phase calculations are not feasible for the present RNA system as it carries a charge of -9.

As explained elsewhere, potential energies calculated for a series of single geometries cannot be used to predict the free energy order of different conformational states.<sup>22a</sup> This would require at least a set of temperature accessible conformers for each state and ideally explicit solvent MD simulations. However, visible differences between the QM and MM relative energies, such as those shown here, indicate that the MM description is intrinsically biased, and this bias should affect MM explicit solvent simulations. Thus, we speculate that a hypothetical DFT-D3 explicit solvent simulation would probably be able to keep the simulated structure in the region of the native-like 2b/3a/4a arrangement or even in the native structure which has not been sampled at all (and thus we could not evaluate its energy).

The calculations were repeated using MM geometries derived with the PB and APBS optimizations. Both the PB and APBS method solve the Poisson–Boltzmann equation with numeric integration, applying a less empiric model than the GB approach, but with a significantly higher computational cost (though still largely negligible for the studied systems). The

picture emerging from the PB and APBS calculations is qualitatively consistent with the data based on GB-optimized geometries. The observed differences are within the expected accuracy. The results for the PB energies vary across all structures a bit more than for the other solvation models. This does not necessarily indicate problems with the PB approach but rather shows differences in solvation models and their individual approximations and parametrizations. Note (see Computational Details) that the default settings for the numeric integration were increased compared to common standards for PB and APBS to ensure numeric stability as grid spacing can have a large influence on MM-PBSA free energy estimates.<sup>70</sup> There is no clear consensus for which solvation model is better. The PB method is considered more rigorous than GB, but the GB parameters are fitted to reproduce experimental data. Neither approximation is clearly favored.<sup>69a,71</sup>

**QM Optimizations Result in Sizable Geometrical Changes.** Finally, full QM optimizations have been carried out. Computation time is multiple weeks for a single structure. The results were at first sight surprising because the high energies of the structures 2a/3b/4b, 2b/3b/4b, 2a/3c/4c, 2a/3c/4d, and 2b/3c/4c disappeared. In addition, the energies obtained by the QM and MM methods are quite similar. However, the QM optimizations resulted in much larger conformational changes than the MM optimizations, preventing any consistent comparison of the QM and MM descriptions of the conformations observed in the simulations. As indicated in Table 1, five out of the eight QM optimizations resulted into structures very different from the intended structures.

The QM optimization restores the G8(N1/N2)-U4(O2P) base-phosphate (BPh) interaction in all conformations not having this interaction in the starting geometry, i.e., in the QM-optimized 2a/3b/4b, 2b/3b/4b, 2a/3c/4c, 2a/3c/4d, and 2b/3c/4c conformations. This leads to the transition of suite 4 in these structures to the native 4a (= family #4) substate in all cases. The root-mean-square deviation (RMSD) of these structures from the 2a/3c/4a QM conformation is smaller than the RMSD from their respective starting conformations. Thus, RMSD of the QM-optimized 2a/3c/4c conformation is 0.37 Å with respect to the 2a/3c/4a QM structure and 1.1 Å with respect to its initial conformation, confirming the rearrangement. The formal description is more complicated in the case of QM-optimized 2a/3b/4b and 2b/3b/4b structures. They resemble 2a/3c/4a but with significant differences in suites 2 and 3. Thus, RMSD of QM-minimized 2b/3b/4b from QM-minimized 2a/3c/4a is 0.8 Å, and the RMSD from its starting structure is 0.95 Å. Although the QM-optimized 2a/3c/4d structure shows the same substates for suites 2, 3, and 4 as 2a/3c/4a after 4d changes to 4a, its RMSD with respect to QM-optimized 2a/3c/4a is 0.49 Å, which is higher than for the QM-optimized 2a/3c/4c conformation described above. A similar RMSD to QM-optimized 2a/3c/4a shows QM-optimized 2b/3c/4c structure because its suite 2 adopts another substate.

**QM Optimizations Re-Establish a Native BPh Interaction.** The hallmark structural change in the QM optimizations is the formation of the G8(N1/N2)-U4(O2P) BPh interaction. This BPh interaction is native to the SR motif and is present in most experimental structures.<sup>18</sup> The anionic BPh interactions belong to the strongest H-bonds occurring in nucleic acids,<sup>11a</sup> and thus, it is not surprising that their formation during optimizations depends on the level of theory

**Table 2. QM versus MM Cross-Comparison of Conformational Energies (kcal/mol) between SR Motif Substates Based on the Simulation of the High-Resolution Structure 483D<sup>a</sup>**

state	MM GB opt.				MM PB opt.				MM APBS opt.				QM COSMO opt.			
	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$
X <sup>b</sup>	14.1	16.5	16.5	15.8	16.5	20.6	17.3	15.2	30.3	44.8	31.1	23.3	14.9	20.6	12.6	16.1
2b/3a/4a <sup>c</sup>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2a/3c/4c	1.7	−0.3	8.6	15.7	0.2	−0.5	2.4	17.2	−0.4	−3.5	1.3	15.6	−18.8 <sup>d</sup>	−13.4 <sup>d</sup>	−16.7 <sup>d</sup>	−2.0 <sup>d</sup>
													(−16.7) <sup>e</sup>	(−10.5) <sup>e</sup>	(−6.8) <sup>e</sup>	(1.7) <sup>e</sup>
2a/3c/4a	−3.5	−1.0	−2.1	−7.4	−2.5	0.4	−3.3	−1.3	−3.0	−0.2	−5.4	−2.6	−14.0	−10.9	−13.1	−27.3

<sup>a</sup>MM refers to the bsc0 $\chi_{OL3}$  force field with GB ( $\Delta E_{GB}$ ), PB ( $\Delta E_{PB}$ ), and APBS ( $\Delta E_{APBS}$ ) implicit solvation models, while QM refers to COSMO-TPSS-D3/def2-TZVP ( $\Delta E_{QM}$ ). See Figures 2 and 3 for the definition of the optimized 11-nt fragment and Figure 4 for the nomenclature of the conformations. <sup>b</sup>X-ray structure. <sup>c</sup>Reference structure. The geometry was taken shortly after the simulation start when the structure was already well equilibrated. With the exception of suite 2, the experimental substate is maintained. <sup>d</sup>QM optimization converges suite 4 to the 4a substate (i.e., #a family), thus changing the intended conformation. <sup>e</sup>Estimated energy contribution of the terminal 5'-end hydroxyl H-bond removed for this particular structure (in kcal/mol: QM = 3.7, GB = 2.1, PB = 2.9, and APBS = 9.9).

and may alter the energy picture significantly. Although interaction energy calculations on small model complexes indicated that the MM method describes acceptably the BPh H-bonding energy,<sup>11b</sup> it cannot be ruled out that inclusion of this BPh interaction within the full structural context of the SR motif may lead to nonequivalent MM and QM energy descriptions. The starting X, 2b/3a/4a and 2a/3c/4a structures contain this interaction and remain stable during the QM optimization.

#### Other Structural Changes in QM Optimizations.

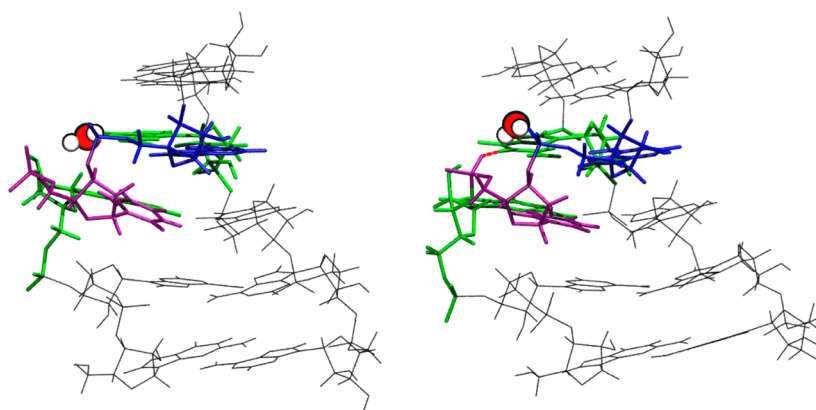
Further complication arises due to the formation of a terminal 5'-end H-bond in residue 7 between the 5'-terminal hydroxymethyl C5'-OS'H5 group and the residue 7 phosphate in most QM optimized conformations (this interaction is not formed in MM-optimized structures, see the overview in Table S7, Supporting Information). The possibility of this interaction results from the truncation of the system, and the only QM optimizations not showing this interaction are X and 2a/3c/4a. For 2a/3c/4a, this is of particular importance because the main structural difference between QM-optimized 2a/3c/4a and 2a/3c/4c structures is this particular 5'-end interaction. Besides rotation of the 5'-end hydroxymethyl group, the H-bond also includes a significant change in the  $\gamma$  angle of residue 7. The energy difference of this H-bond is estimated by calculating the relative energies of the G7pG8 dinucleotide in the 2a/3c/4a and 2a/3c/4c QM-optimized conformations (the 3'-end hydrogen was added using the x-leap template library). The 2a/3c/4a structure having this H-bond is 5.2 kcal/mol more stable (QM level). We have used this energy difference as a correction to the QM energies of the X and 2a/3c/4a QM structures in Table 1 (data in parentheses; we have also calculated similar corrections for the MM single points on the QM-optimized geometries; see Table 1). The correction brings the QM energies of QM conformations of 2a/3c/4a and 2a/3c/4c very close in energy. The other QM structures whose suite 4 is also optimized into the 4a conformations, i.e., 2a/3b/4b, 2b/3b/4b, 2a/3c/4d, and 2b/3c/4c, are structurally less close to 2a/3c/4a, so that we naturally see somewhat larger energy differences. It is somewhat surprising that QM-optimized 2a/3c/4d, with suite 4 changed to 4a, shows a difference of ~5 kcal/mol with respect to the 2a/3c/4a QM-optimized structure, even though it also shows the above-mentioned 5'-end rotation. However, as mentioned above, its RMSD to the 2a/3c/4a QM conformation is a bit higher than for the 2a/3c/4c QM conformation, indicating some additional structural differences.

There are yet additional significant structural changes in the QM-optimized structures. The terminal base pairs are much more distorted during QM optimization compared to the MM optimizations. This results into the formation of further (non-native) BPh interactions, namely, G1(N2)-A11(O2P), G1(N1)-A10(O2P), and C6(N4)-A5(O2P) BPh interactions. While the formation of the C6(N4)-A5(O2P) interaction can be prevented by insertion of an explicit water molecule (see below), the other non-native BPh interactions are present in most QM-optimized structures. Although their impact on the energy calculation is non-negligible, they affect all QM-optimized structures to a similar extent. We suggest that the formation of these interactions is facilitated by the truncation of the system, allowing large movements in this region.

The results mean that the QM-optimized structures significantly deviate from the starting (intended) conformations and are not consistent with the geometries populated in explicit solvent simulations. In other words, the MM- and QM-optimized structures are not equivalent, which should be considered when evaluating the data from Table 1. The QM-optimized structures are not suitable to monitor the energies of different substates occurring during the simulation.

The results show that the QM and MM optimizations, despite starting from the same structure, sometimes lead to quite different geometries of optimized nucleic acids building blocks. This is consistent with the data for two-quartet G-DNA stems.<sup>22a</sup> The QM optimization procedure is more flexible and allows more conformational changes in the NA building block than the MM force field. Another obvious factor is the truncation of the system that artificially increases the flexibility in both terminal regions, allowing the QM optimization to achieve conformational changes that would otherwise be restricted by enclosing base pairs. However, this does not explain the difference between the QM and MM data. We assume that the electronic structure description gives more flexibility to the RNA molecule than the pairwise additive nonpolarizable force field, which is consistent with increased biopolymer flexibility seen in simulations of proteins using polarizable force field.<sup>72</sup> QM optimizations are more prone to form various intramolecular H-bonds than MM, often moving the structure away from the intended geometry region, especially for terminal regions of truncated systems. We suggest that one of the reasons for this behavior is paradoxically the better quality of the QM description, its capability to realistically include various electronic structure redistributions, and a more physical description of the (truncated) molecules.





**Figure 5.** Molecular representation of the *2a/3b/4b* (left) and *2b/3a/4a* (right) substates (MD structures) including the explicit water molecule. The residues G3, A5, U4, and G8 are highlighted in stick representation with the following color code: G = green, A = blue, and U = purple. The explicit water molecule is highlighted in vdW sphere representation. (Left) The explicit water bridge prevents (replaces) the direct G8/U4 BPh interaction in *2a/3b/4b*. (Right) The direct BPh interaction is present in this structure (indicated by the red line), and the water molecule is placed at the alternative nearby position as described in the text.

**Table 3. QM versus MM Cross-Comparison of Conformational Energies (kcal/mol) between SR Motif Substates of the Structure 1S72 Optimized with Inclusion of One Explicit Water Molecule<sup>a</sup>**

state	GB opt.				PB opt.				APBS opt.				QM opt.			
	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$	$\Delta E_{GB}$	$\Delta E_{PB}$	$\Delta E_{APBS}$	$\Delta E_{QM}$
<i>X</i> <sup>b</sup>	18.9	16.6	17.9	20.8	21.1	20.7	20.2	19.9	47.3	66.0	57.9	38.4	34.7 (33.5) <sup>c</sup>	43.8 (42.0) <sup>c</sup>	39.0 (30.5) <sup>c</sup>	22.6 (17.9) <sup>c</sup>
<i>2b/3a/4a</i> <sup>c</sup>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>2a/3b/4b</i>	1.5	−9.4	4.4	15.9	0.5	−2.9	6.3	20.7	5.6	9.8	13.6	27.4	8.8	2.6	9.0	6.5
<i>2b/3b/4b</i>	−0.1	−11.9	−1.0	16.8	2.0	−2.8	5.2	12.8	5.6	4.0	10.6	20.8	6.0	0.6	7.0	3.7
<i>2a/3c/4c</i>	−1.9	−8.2	−2.5	13.6	−2.9	−2.9	−1.3	4.7	−0.4	0.2	3.1	12.4	−0.5 <sup>d</sup>	1.6 <sup>d</sup>	−0.1 <sup>d</sup>	−5.8 <sup>d</sup>
<i>2a/3c/4a</i>	1.6	−9.7	−2.4	−0.9	−1.2	−2.5	−2.0	−5.8	−2.9	−4.5	−5.4	−7.8	−3.8 (−5.0) <sup>e</sup>	−2.0 (−3.8) <sup>e</sup>	−2.8 <sup>e</sup> (−11.3)	−2.3 (−7.0) <sup>e</sup>
<i>2a/3c/4d</i>	−0.9	−12.7	−2.6	6.8	−3.3	−5.9	−1.3	5.8	−11.1	−30.1	−21.5	6.7	1.5	−1.4	2.4	0.4
<i>2b/3c/4c</i>	0.0	−9.3	0.2	15.8	0.4	0.2	5.7	12.5	0.0	−1.9	3.1	13.6	4.4	2.8	8.1	6.9

<sup>a</sup>MM refers to the bsc0 $\chi_{OL3}$  force field with GB ( $\Delta E_{GB}$ ), PB ( $\Delta E_{PB}$ ), and APBS ( $\Delta E_{APBS}$ ) implicit solvation models, while QM refers to COSMO-TPSS-D3/def2-TZVP ( $\Delta E_{QM}$ ). See Figures 2 and 3 for the definition of the optimized 11-nt fragment and Figure 4 for the nomenclature of the conformations. <sup>b</sup>X-ray structure. <sup>c</sup>Reference structure. The geometry was taken shortly after the simulation start when the structure was already well equilibrated. With exception of suite 2, the experimental substate is maintained. <sup>d</sup>QM optimization leads to the formation of direct BPh interaction during QM optimization even with inclusion of the water, and leads to the 4a substate in suite 4. <sup>e</sup>Energy corrected by the estimate (in kcal/mol: QM = 4.7, GB = 1.2, PB = 1.8, APBS = 8.5), for the terminal 5'-end hydroxyl H-bond absent in these optimized structures.

Thus, although the QM description is more accurate than the MM description per se, the QM optimizations are more likely to form non-native interactions in the continuum solvent environment and are more sensitive to the truncation. In our earlier study on G-DNA stems, we were able to prevent non-native H-bonds occurring in QM but not MM optimizations by inclusion of a single X-ray water molecule.<sup>22a</sup> We emphasize that although the MM optimizations may sometimes result into better structures than QM optimizations (in continuum solvent), it does not indicate better intrinsic accuracy of the MM description. It rather reflects a specific compensation of errors of the MM and continuum solvent approximations. Profound differences between QM and MM relative energies (Table 1) indicate that the MM force field is likely to introduce bias when used in explicit solvent simulations.

**Computations on the 438D SRL Structure Confirm the Results.** We repeated the computations for the high-resolution 438D X-ray SRL structure. Because the two SRL motifs differ a little in their sequence, their direct energy comparison is not possible. For the 438D structure, we identified three conformations in the simulation equivalent to the *2b/3a/4a*,

*2a/3c/4c*, and *2a/3c/4a* conformations from the 1S72 simulation (Figure 4). The simulation (as all SRL simulations)<sup>18</sup> departs from the native geometry, and going through the *2b/3a/4a* conformation stabilizes in the region with *2a/3c/4c* structural features (including population of the flipped state *2a/3c/4a*).

The computations (Table 2) are consistent with the data from Table 1. The optimized X-ray structure is again much higher in energy than the *2b/3a/4a* conformation, despite the fact that the *X* optimization now started from a high-resolution structure. The energy difference between the *X* and *2b/3a/4a* states is nevertheless somewhat reduced. As for the 1S72 structure, the high energy is eliminated when optimizations are done using the starting structure processed via the MD equilibration protocol (see the Supporting Information for a short note on the *5z* family for suite 2 in the equilibrated structure).

There is a similar variability between the three types of continuum solvent computations as in Table 1. Most importantly, the MM and QM descriptions again give very different relative energies for the MM-optimized *2a/3c/4c*

**Table 4.** Survey of Changes of Backbone and Glycosidic Torsional Angles for Optimized Structures Across All Substates Listed in Table 1 Covering All Occurrences of Listed Angles<sup>a</sup>

	MM GB			MM PB			MM APBS			QM COSMO		
	range(+)	range(−)	MAD	range(+)	range(−)	MAD	range(+)	range(−)	MAD	range(+)	range(−)	MAD
$\alpha$	8.5	−19.0	4.1	5.7	−16.1	2.7	4.5	−12.4	1.5	84.4	−20.5	10.3
$\beta$	35.6	−17.6	4.6	34.4	−7.5	2.9	15.8	−5.5	1.4	68.0	−41.7	15.3
$\gamma$	18.4	−12.1	4.0	14.0	−8.9	2.5	9.4	−5.0	1.1	12.4	−27.6	6.3
$\delta$	13.9	−25.1	3.5	6.8	−14.7	1.8	3.4	−9.2	1.1	21.1	−17.2	4.6
$\epsilon$	37.5	−11.5	4.9	14.6	−16.0	2.6	6.4	−7.0	1.1	43.1	−75.8	13.5
$\zeta$	15.1	−10.0	3.8	16.5	−8.0	2.5	13.7	−7.5	1.1	54.8	−59.3	12.5
$\chi$	16.4	−14.6	4.8	7.1	−4.5	1.3	4.6	−3.1	0.9	30.9	−41.0	12.3

<sup>a</sup>The changes are calculated against the X-ray structure for state X and against the respective MD-based starting structures for all the other states (angle(initial) − angle(optimized). Range(+) and range(−) show the maximal deviations; negative deviations mean the angle becomes larger; MAD is the mean absolute deviation (MAD).

conformations, explaining why the MD simulation does not stay in the native-like  $2b/3a/4a$  region and moves to the  $2a/3c/4c$  region.

The QM optimizations lead to even larger geometrical changes than for the 1S72 conformations, making their further analysis difficult. More details are given in Supporting Information.

**Inclusion of an Explicit Water Molecule.** The major limitation of QM computations on nucleic acids building blocks is the continuum solvent approximation. Inclusion of specific water molecules based on X-ray data or explicit solvent simulations prevented formation of non-native interactions in the optimized structures of guanine quadruplexes.<sup>22a</sup> Thus, we have carried out similar calculation also for the SRL motif.

The direct BPh interactions in the SR motif are replaced by water-bridged BPh interactions in some simulation substates.<sup>18</sup> In the simulation substates  $2a/3b/4b$ ,  $2b/3b/4b$ ,  $2a/3c/4c$ ,  $2a/3c/4d$ , and  $2b/3c/4c$ , the space between U4(O2P) and the Watson–Crick edge of G8 is highly populated with water molecules, often bridging the BPh interaction. Another proximal position with a high population of water molecules is between U4(O2P) and C6(N4) (Figure 5). Selection of water positions for the optimizations is not straightforward because including the explicit water molecules via structural averaging of MD snapshots is not a viable option. In the MD simulation, the water molecules dynamically rattle and exchange with the bulk. Selection of the explicit water was made by carefully analyzing the trajectory and manually placing the water, using exactly the same averaged structures as used for the original starting structures lacking the water (structures used in Table 1). The optimizations then included this explicit water molecule.

Water molecules were included in the structures lacking the direct G8(N1/N2)–U4(O2P) interaction, i.e., in the  $2a/3b/4b$ ,  $2b/3b/4b$ ,  $2a/3c/4c$ ,  $2a/3c/4d$ , and  $2b/3c/4c$  conformations. The water bridge aims to prevent the formation of direct BPh interaction in the QM optimizations. To maintain energy comparability with the conformations having the direct BPh interaction ( $X$ ,  $2b/3a/4a$ , and  $2a/3c/4a$ ), we needed to add a water molecule also for these structures. We used the second water position described above.

The energies of the structures that include one explicit water molecule (Table 3) mostly show similar results as without inclusion of water (Table 1). The calculations confirm the high relative energy of the optimized X-ray structure ( $X$ ). The calculations (on MM geometries) also confirm the over-stabilization of the non-native  $2a/3c/4c$  structure by MM. For

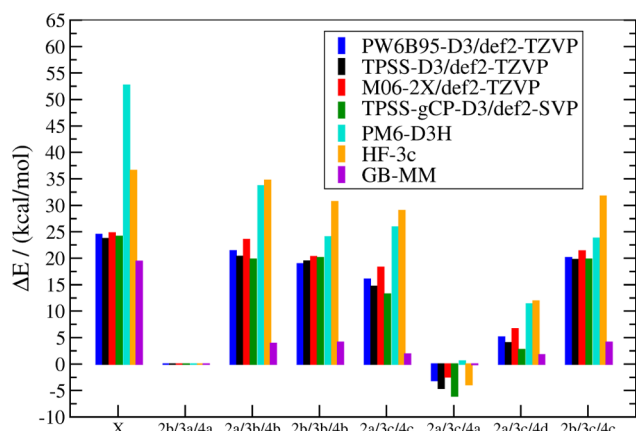
the GB structures, QM data still show  $2a/3c/4a$  as the most stable substate, while PB and APBS predict  $2a/3c/4d$  to be the most stable. Conformational energies calculated on the PB and APBS geometries show slight energy reordering. However, given the small energy values and the uncertainties in the solvation models, such reordering is insignificant.

Structural changes in the SR motif during the MM optimizations are rather small. However, the choice of MM solvation model has some influence on the optimized water position. In structures in which the water molecule is not bridging the G8/U4 BPh interaction (the second binding mode), the water modestly rearranges compared to the starting geometry. This is not surprising, as this particular water binding mode is less well defined compared to the BPh-bridging binding mode. The inclusion of a water molecule leads to slightly larger differences in the MM structures than in optimizations without a water molecule (Table 4 and Table S6, Supporting Information). In other words, inclusion of explicit water molecules somewhat increases diversity of the optimized MM structures because the water brings additional flexible degrees of freedom. Thus, unless explicit water molecules are used to prevent undesired interactions,<sup>22a</sup> their inclusion can make the energy data slightly more noisy. Because the MM structures do not form the BPh interactions during optimizations, we consider the data from Table 1 as more representative.

The water molecules bridging the BPh interactions were added in order to prevent the changes of the initial conformations in the QM optimizations (see discussion about Table 1). This goal was achieved for all conformations except for the  $2a/3c/4c$  conformation. For this structure, the QM optimization displaces the water molecule into the second hydration position and forms the undesired direct BPh interaction. As a result, the  $2a/3c/4c$  and  $2a/3c/4a$  conformations are quite identical except for the 5'-end H-bond interaction. In summary, although explicit inclusion of water molecules may help to stabilize the QM optimizations near the target structure, the procedure is not always successful.

**Comparison of DFT-D3, HF-3c, and PM6-D3H Methods.** As common benchmark computations of CCSD(T)/CBS quality<sup>30b,73</sup> are limited to small model systems, which may insufficiently represent the properties of large biomolecules, we propose that DFT-D3 computations on large biomolecular systems can serve as an alternative benchmark for MM and more approximate QM methods.

Figure 6 compares relative energies obtained by different QM methods using the GBMM-optimized structures (numer-



**Figure 6.** Single-point energies (kcal/mol) based on GB-MM optimized structures. All methods except GB-MM use the COSMO solvation model.

ical data can be found in Table S12, Supporting Information). First, comparison with the global hybrid functional PW6B95-D3<sup>49</sup> shows that the TPSS-D3 functional yields reliable results. PW6B95-D3 proved to be exceptionally robust for many properties in extended benchmark calculations.<sup>49b</sup> The TPSS-D3 results are very close to the PW6B95-D3 results with differences of only about 2–3 kcal/mol. Such differences are expected for systems of this size. Thus, we suggest that the TPSS-D3 method can be considered to reach semi-quantitative accuracy.

We additionally tested the widely used M06-2X functional, which was also used in a recent investigation on three base pair B-DNA duplexes.<sup>22b</sup> The performance of M06-2X is on par with TPSS-D3 and PW6B95-D3 as shown in Figure 6. M06-2X is known to be less accurate for London dispersion interactions in the long-range regime.<sup>28</sup> A correct long-range dispersion description is considered crucial for large systems. On the other hand, calculations of conformational energies are often rather robust due to error cancellation effects, especially if all conformations are rather similar, which can explain the good performance of M06-2X.

Another goal was to see if low-scaling methods like the recent HF-3c<sup>31</sup> or a dispersion- and H-bonding corrected PM6<sup>33</sup> can describe the RNA building blocks. The semi-empirical PM6-D3H method is at least 2 orders of magnitude faster than a typical DFT calculation with a large basis set and is much less memory demanding. PM6 is known to have difficulties with H-bond interactions, and thus several H-bond corrections have been proposed in recent years.<sup>74</sup> Herein, the D3H correction (zero-damping D3-dispersion correction plus a simple H-bond correction)<sup>33b</sup> is used. HF-3c employs HF with an (almost) minimal basis set and scales formally equal to a hybrid-functional DFT calculation. However, the very small basis set enables quite efficient calculations of large systems, which are often faster than a corresponding (pure) DFT calculation with a reasonable basis set. Both HF-3c and PM6-D3H qualitatively reproduce the major differences between QM and MM for the 2a/3b/4b, 2b/3b/4b, and 2a/3c/4c structures. However, although the results are encouraging, quantitative agreement with DFT-D3 was not reached. We suggest that both methods could be used to make preliminary QM calculations on nucleic acids building blocks, which can then be refined using DFT-D3.

The geometrical counterpoise correction<sup>32</sup> (gCP, also applied in the HF-3c scheme) enables computations with small basis sets by eliminating major parts of the inter- and intramolecular basis set superposition error (BSSE). We applied the correction to the TPSS-D3/def2-SVP computations (TPSS-gCP-D3) and obtained results very close to the computations using the much larger def2-TZVP basis set, saving significant computer resources. For the computations with def2-SVP only, around 4000 basis functions are assigned, while for def2-TZVP the number of basis functions exceeds 8000.

**Geometry Analyses.** In the Supporting Information, we provide the coordinates of all optimized structures reported in this paper. In the subsequent paragraphs, we provide a few geometrical analyses; however, for space reasons, it is not possible to present a complete analyses in the paper. Readers can use the coordinates to make their own assessments.

Table 4 gives preliminary insight into the sensitivity of the backbone angles to the optimization protocols. It shows the (maximum) deviation range and mean absolute deviation (MAD) for all angles, calculated for all eight optimized structures from Table 1, and including all nucleotides. The deviation range in the positive direction, i.e., range(+), indicates that the angle becomes larger in the optimization. The deviation for each conformation is calculated with respect to its initial starting geometry, i.e., the X-ray structure for X and the MD-averaged structures for all others.

These statistics are not intended to evaluate the quality of the methods for structure optimization, as this would assume the initial (X-ray and MD) structures are of ideal quality. The table reflects magnitudes of changes of the structures upon optimization and thus reveals mainly differences in the potential energy surface and optimization protocols of the various methods, taking into account all substates. It shows which backbone torsional angles are more flexible in the optimizations than the others.

Among the MM methods, the GB optimization yields large deviations in both directions. The largest deviations are found for  $\beta$  with 35.6° and  $\epsilon$  with 37.5°. The large maximal (negative) deviation of  $\delta$  with −25.1° is also significant. The MADs of the GB angles lie mostly between 4° and 5°. The  $\delta$  angle yields the smallest MAD with 3.5°, while  $\epsilon$  shows the largest with 4.9°. PB structures show smaller MADs than GB with values between 1.3° and 2.9°. Large deviations are again observed for the  $\beta$  and  $\epsilon$  angles. However, the  $\epsilon$  angle is statistically closer to the starting structures (14.6° maximal deviation) than the respective GB angle (37.5° maximal deviation). The APBS method yields overall the lowest deviations from the initial structure, where only the  $\beta$  and  $\zeta$  angles deviate more than 10° in the positive direction, and  $\alpha$  is the only one changing in the negative direction by more than 10°. The MADs reflect these findings with values between 0.9° and 1.5°. Thus, among the MM methods, the MM APBS optimization shows on average the smallest changes from the starting geometries, while the MM GB optimizations show the largest ones. The reason for the different flexibility of the MM methods is unknown and may reflect properties of both the continuum solvent models and the respective optimization protocols. A similar difference between PB and GB optimizations has been noticed earlier.<sup>22a</sup>

MADs of the QM structures are significantly larger than for the MM optimization methods, for reasons explained earlier in this paper. The formation of the G8/U4 BPh interaction, non-native BPh interactions, and accompanied changes in the



backbone topology have obviously a huge impact on the calculated deviations. The MADs for  $\gamma$  ( $4.6^\circ$ ) and  $\epsilon$  ( $6.3^\circ$ ) are still quite small, while the other angles have MADs above  $10^\circ$ . Largest deviations are found in the  $\beta$  angle with a MAD of  $15.3^\circ$ . All these numbers reflect the substantial changes of geometries during the QM optimizations. As mentioned above, we think that the increased flexibility of the QM-optimized systems compared to the MM optimizations reflects the more realistic description of the RNA molecule by the DFT-D3 method, making the continuum solvent approximation less efficient in replacing the explicit solvent.

The backbone changes for structures containing the explicit water molecule are slightly larger, probably due to the additional degrees of freedom stemming from the water molecule (Table S6, Supporting Information).

## CONCLUSIONS

The SR motif is an intricate folded RNA molecule with complex backbone topology that is not well reproduced by the MM-based explicit solvent simulations. In this study, we address this issue using large-scale QM computations on the complete SR motif consisting of 11 nucleotides (Figure 2).

A series of conformations for the Sarcin–Ricin loop motif were derived from explicit solvent MD simulations. The simulations populate a set of non-native conformations of the S-turn backbone region of the SR motif. The obtained conformations were used as starting structures for investigations using the TPSS-D3/def2-TZVP QM method and parm99bsc0+ $\chi_{ol3}$  AMBER MM. The QM calculations were coupled with the COSMO continuum solvent model, while the MM calculations were done with three continuum solvent models, namely, GB, PB, and APBS. All four methods were used for gradient optimizations, and for each optimized structure, all four methods were used to calculate the conformational energy.

MM optimizations keep the intended starting structures and show similar qualitative structural features with all three solvent models. However, there are non-negligible quantitative differences (5–10 kcal/mol) in the MM relative energies (Tables 1 and 2), illustrating uncertainty in implicit solvent modeling of nucleic acids. Still, all MM computations indicate that all structures occurring along the MD simulation trajectories belong to the low-energy conformational region.

In contrast, QM energy calculations on MM-optimized geometries show qualitatively different energy order and penalize most geometries populated in the later stages of simulations compared to the MM (Tables 1 and 2). We suggest that this result reflects the intrinsic inaccuracy of the MM description, which is biased in favor of the non-native (incorrect) structures of the SR motif that are dominant in the simulations. This is consistent with the fact that MD simulations of the SR motif do not reproduce its native conformation. The calculations show the potential of large-scale QM calculations in studies of RNA motifs and their capability to refine the picture obtained by MM methods.

QM optimizations of the SR motif conformations resulted in considerable geometrical changes of the optimized molecules. The results are consistent with our recent study on G-DNA stems<sup>22a</sup> and illustrate challenges in using continuum solvent QM optimizations in studies of larger nucleic acids fragments. Note that another group recently reported QM optimization of three base pair B-DNA double helices.<sup>22b</sup> Although the authors suggested that optimized structures remained in good agree-

ment with experimental B-DNA structures, closer inspection of the actual data reveals sizable deviations from the experimentally known ranges of B-DNA conformations. Thus, although the QM description is undoubtedly more accurate than the MM description per se, the QM optimizations are more likely to form non-native interactions in a continuum solvent environment and are more sensitive to truncation of the system. The continuum solvent models do not fully mimic the explicit solvent environment of real nucleic acids and miss specific solute–water H-bonding. Thus, high-quality QM structure–energy descriptions often overpower the mean field solvent screening of the continuum solvent and establish gas-phase-like molecular interactions that would not form in explicit solvent. To conduct large-scale QM studies on nucleic acids molecules, it will be necessary to find efficient means to keep the studied structures in desired geometries. One of the options is an inclusion of selected specific water molecules. This, however, was only partially successful in our case, and in addition, the explicit water molecules make the computations noisier due to the additional degrees of freedom in the optimization (Table 3 and Table S6, Supporting Information). Another complication may arise from insufficiently relaxed experimental structures (even high-resolution ones) during the gradient optimization. This problem can be eliminated by a short MD or even by the MD equilibration protocol. Both issues are discussed in detail in our study.

Disregarding the above difficulties to keep QM optimizations of nucleic acids building blocks under control, DFT-D3 computations on complete nucleic acids building blocks can be used as a benchmark for fast QM methods and various MM approaches. Such calculations can complement common CCSD(T)/CBS benchmarks on small model systems. To demonstrate this possibility, we compared energy predictions of different SR motif conformations using several QM methods (Figure 6). Results obtained by TPSS-D3/def2-TZVP, PW6B95-D3/def2-TZVP, and M06-2X calculations are very consistent. Also the TPSS-gCP-D3/def2-SVP method neatly reproduces the TPSS-D3/def2-TZVP data. However, other low scaling methods, namely, PM6-D3H and HF-3c, lead to large energy differences for the relative energies of different SR motif conformations.

In summary, the computations reveal large nonequivalency between the conformational energies of the SR RNA motif as described by MM and QM methods. This result illustrates limitations of MM-based MD simulations of nucleic acids and is consistent with MD simulations of the SR motif. Therefore, high-quality large-scale DFT-D3 QM computations represent a promising complement of MM-based methods to study various aspects of nucleic acid structures, which can refine the picture obtained by MM-based methods. However, we also underline that the QM calculations need to be performed in conjunction with the MM computations, as demonstrated here and elsewhere.<sup>22a</sup> The large-scale QM calculations can also serve as benchmarks to assess the quality of fast QM and MM methods. However, such QM computations are far from being straightforward, and their routine use will require overcoming the above highlighted methodological challenges.

## ASSOCIATED CONTENT

### Supporting Information

PDB coordinates of the initial and optimized geometries; detailed information about the MD simulation (technical details, RMSD development, backbone topology development

for the key suites, MM-PBSA energies along the trajectory, averaging windows for the snapshots); tables with the backbone families and their suitability for each conformation; an overview of key structural features in the MM and QM optimized conformations; backbone topology for 1S72 and 483D X-ray structures for the three key suites; further discussion about the 483D system; torsional angles statistics for the 1S72 structure with explicit water; numerical data for Figure 6; and minimal AMBER input for GB, PB, and APBS computations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [sponer@ncbr.muni.cz](mailto:sponer@ncbr.muni.cz).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was funded by Grant P305/12/G034 of the Czech Science Foundation, by the project "CEITEC - Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund, and by the EU Seventh Framework Programme under the "Capacities" specific program (Contract 286154).

## REFERENCES

- (1) Mathews, D. H.; Turner, D. H. Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **2006**, *16* (3), 270–278.
- (2) (a) Leontis, N. B.; Stombaugh, J.; Westhof, E. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002**, *30* (16), 3497–3531. (b) Spomer, J.; Spomer, 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.
- (3) Leontis, N. B.; Lescoute, A.; Westhof, E. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **2006**, *16* (3), 279–287.
- (4) (a) Lu, X. J.; Olson, W. K.; Bussemaker, H. J. The RNA backbone plays a crucial role in mediating the intrinsic stability of the GpU dinucleotide platform and the GpUpA/GpA miniduplex. *Nucleic Acids Res.* **2010**, *38* (14), 4868–4876. (b) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Herschkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M.; Consortium, R. N. A. O. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **2008**, *14* (3), 465–481.
- (5) Bida, J. P.; Das, R. Squaring theory with practice in RNA design. *Curr. Opin. Struct. Biol.* **2012**, *22* (4), 457–466.
- (6) (a) Ditzler, M. A.; Otyepka, M.; Spomer, J.; Walter, N. G. Molecular dynamics and quantum mechanics of RNA: Conformational and chemical change we can believe. *Acc. Chem. Res.* **2010**, *43* (1), 40–47. (b) Chawla, M.; Abdel-Azeim, S.; Oliva, R.; Cavallo, L. Higher order structural effects stabilizing the reverse Watson–Crick guanine–cytosine base pair in functional RNAs. *Nucleic Acids Res.* **2014**, *42* (2), 714–726.
- (7) Oliva, R.; Cavallo, L.; Tramontano, A. Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Res.* **2006**, *34* (3), 865–879.
- (8) Spomer, J.; Mladek, A.; Spomer, J. E.; Svozil, D.; Zgarbova, M.; Banas, P.; Jurecka, P.; Otyepka, M. The DNA and RNA sugar–phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies. *Phys. Chem. Chem. Phys.* **2012**, *14* (44), 15257–15277.
- (9) (a) Endo, Y.; Mitsui, K.; Motizuki, M.; Tsurugi, K. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes –the site and the characteristics of the modification in 28-s ribosomal-RNA caused by the toxins. *J. Biol. Chem.* **1987**, *262* (12), 5908–5912. (b) Qin, S. B.; Zhou, H. X. Dissection of the high rate constant for the binding of a ribotoxin to the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (17), 6974–6979. (c) Lacadena, J.; Alvarez-Garcia, E.; Carreras-Sangra, N.; Herrero-Galan, E.; Alegre-Cebollada, J.; Garcia-Ortega, L.; Onaderra, M.; Gavilanes, J. G.; del Pozo, A. M. Fungal ribotoxins: Molecular dissection of a family of natural killers. *FEMS Microbiol. Rev.* **2007**, *31* (2), 212–237.
- (10) (a) Klein, D. J.; Moore, P. B.; Steitz, T. A. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **2004**, *340* (1), 141–177. (b) Correll, C. C.; Wool, I. G.; Munishkin, A. The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: The structure at 1.11 Å resolution. *J. Mol. Biol.* **1999**, *292* (2), 275–287. (c) Leontis, N. B.; Stombaugh, J.; Westhof, E. Motif prediction in ribosomal RNAs – Lessons and prospects for automated motif prediction in homologous. *Biochimie* **2002**, *84* (9), 961–973.
- (11) (a) Zirbel, C. L.; Spomer, J. E.; Spomer, J.; Stombaugh, J.; Leontis, N. B. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* **2009**, *37* (15), 4898–4918. (b) Zgarbova, M.; Jurecka, P.; Banas, P.; Otyepka, M.; Spomer, J. E.; Leontis, N. B.; Zirbel, C. L.; Spomer, J. Noncanonical hydrogen bonding in nucleic acids. Benchmark evaluation of key base-phosphate interactions in folded RNA molecules using quantum-chemical calculations and molecular dynamics simulations. *J. Phys. Chem. A* **2011**, *115* (41), 11277–11292.
- (12) Spackova, N.; Spomer, J. Molecular dynamics simulations of sarcin–ricin rRNA motif. *Nucleic Acids Res.* **2006**, *34* (2), 697–708.
- (13) Mladek, A.; Spomer, J. E.; Kulhanek, P.; Lu, X. J.; Olson, W. K.; Spomer, J. Understanding the sequence preference of recurrent RNA building blocks using quantum chemistry: The intrastand RNA dinucleotide platform. *J. Chem. Theory Comput.* **2012**, *8* (1), 335–347.
- (14) (a) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Spomer, J. Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: Quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8* (7), 2506–2520. (b) Spomer, J.; Cang, X. H.; Cheatham, T. E. Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods* **2012**, *57* (1), 25–39.
- (15) 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 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
- (16) (a) Zgarbova, M.; Otyepka, M.; Spomer, J.; Mladek, A.; Banas, P.; Cheatham, 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. (b) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Spomer, J.; Otyepka, M. Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836–3849.
- (17) Perez, A.; Marchan, I.; Svozil, D.; Spomer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829.
- (18) Havrila, M.; Řeblová, K.; Zirbel, C. L.; Leontis, N. B.; Šponer, J. Isosteric and nonisosteric base pairs in RNA motifs: Molecular dynamics and bioinformatics study of the sarcin–ricin internal loop. *J. Phys. Chem. B* **2013**, *117* (46), 14302–14319.
- (19) (a) Spomer, J.; Spomer, J. E.; Mladek, A.; Banas, P.; Jurecka, P.; Otyepka, M. How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists. *Methods* **2013**, *64* (1), 3–11. (b) Spomer, J.; Jurecka, P.; Hobza, P. Accurate

interaction energies of hydrogen-bonded nucleic acid base pairs. *J. Am. Chem. Soc.* **2004**, *126* (32), 10142–10151. (c) Hobza, P. Calculations on noncovalent interactions and databases of benchmark interaction energies. *Acc. Chem. Res.* **2012**, *45* (4), 663–672.

(20) (a) Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (2), 211–228. (b) Riley, K. E.; Hobza, P. Noncovalent interactions in biochemistry. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (1), 3–17. (c) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2* (2), 304–326.

(21) (a) Guerra, C. F.; Zijlstra, H.; Paragi, G.; Bickelhaupt, F. M. Telomere structure and stability: Covalency in hydrogen bonds, not resonance assistance, causes cooperativity in guanine quartets. *Chem.—Eur. J.* **2011**, *17* (45), 12612–12622. (b) Kellie, J. L.; Wetmore, S. D. Selecting DFT methods for use in optimizations of enzyme active sites: applications to ONIOM treatments of DNA glycosylases. *Can. J. Chem.* **2013**, *91* (7), 559–572. (c) Barone, G.; Guerra, C. F.; Bickelhaupt, F. M. B-DNA structure and stability as function of nucleic acid composition: Dispersion-corrected DFT study of dinucleoside monophosphate single and double strands. *ChemistryOpen* **2013**, *2* (5–6), 186–193. (d) Wick, C. R.; Lanig, H.; Jager, C. M.; Burzlaff, N.; Clark, T. Structural insight into the prolyl hydroxylase PHD2: A molecular dynamics and DFT study. *Eur. J. Inorg. Chem.* **2012**, *31*, 4973–4985.

(22) (a) Šponer, J.; Mládek, A.; Špačková, N.; Cang, X.; Cheatham, T. E.; Grimme, S. Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations. *J. Am. Chem. Soc.* **2013**, *135* (26), 9785–9796. (b) Zubatiuk, T. A.; Shishkin, O. V.; Gorb, L.; Hovorun, D. M.; Leszczynski, J. B-DNA characteristics are preserved in double stranded d(A)(3)·d(T)(3) and d(G)(3)·d(C)(3) mini-helices: Conclusions from DFT/M06-2X study. *Phys. Chem. Chem. Phys.* **2013**, *15* (41), 18155–18166.

(23) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. Hydrogen bonding in DNA base pairs: Reconciliation of theory and experiment. *J. Am. Chem. Soc.* **2000**, *122* (17), 4117–4128.

(24) (a) Waller, M. P.; Kruse, H.; Mueck-Lichtenfeld, C.; Grimme, S. Investigating inclusion complexes using quantum chemical methods. *Chem. Soc. Rev.* **2012**, *41* (8), 3119–3128. (b) Risthaus, T.; Grimme, S. Benchmarking of London dispersion-accounting density functional theory methods on very large molecular complexes. *J. Chem. Theory Comput.* **2013**, *9* (3), 1580–1591.

(25) (a) Kristyan, S.; Pulay, P. Can (semi)local density-functional theory account for the London dispersion forces. *Chem. Phys. Lett.* **1994**, *229* (3), 175–180. (b) Perezjorda, J. M.; Becke, A. D. A density-functional study of van-der-Waals forces – Rare-gas diatomics. *Chem. Phys. Lett.* **1995**, *233* (1–2), 134–137. (c) Hobza, P.; Sponer, J.; Reschel, T. Density-functional theory and molecular clusters. *J. Comput. Chem.* **1995**, *16* (11), 1315–1325.

(26) (a) 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. (b) 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.

(27) Goerigk, L.; Grimme, S. Efficient and accurate double-hybrid-meta-GGA density functionals—Evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2011**, *7* (2), 291–309.

(28) Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking density functional methods against the S66 and S66 × 8 datasets for non-covalent interactions. *ChemPhysChem* **2011**, *12* (17), 3421–3433.

(29) Cang, X. H.; Sponer, J.; Cheatham, T. E. Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.* **2011**, *39* (10), 4499–4512.

(30) (a) Mládek, A.; Krepl, M.; Svozil, D.; Cech, P.; Otyepka, M.; Banas, P.; Zgarbova, M.; Jurecka, P.; Sponer, J. Benchmark quantum-chemical calculations on a complete set of rotameric families of the DNA sugar-phosphate backbone and their comparison with modern density functional theory. *Phys. Chem. Chem. Phys.* **2013**, *15* (19), 7295–7310. (b) Mládek, A.; Banáš, P.; Jurečka, P.; Otyepka, M.; Zgarbová, M.; Šponer, J. Energies and 2'-hydroxyl group orientations of RNA backbone conformations. Benchmark CCSD(T)/CBS database, electronic analysis, and assessment of DFT methods and MD simulations. *J. Chem. Theory Comput.* **2014**, *10* (1), 463–480.

(31) Sure, R.; Grimme, S. Corrected small basis set Hartree–Fock method for large systems. *J. Comput. Chem.* **2013**, *34* (19), 1672–1685.

(32) Kruse, H.; Grimme, S. A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree–Fock and density functional theory calculations for large systems. *J. Chem. Phys.* **2012**, *136* (15), 154101.

(33) (a) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13* (12), 1173–1213. (b) Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem.—Eur. J.* **2012**, *18* (32), 9955–9964.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(35) (a) Sklenovsky, P.; Florova, P.; Banas, P.; Reblova, K.; Lankas, F.; Otyepka, M.; Sponer, J. Understanding RNA flexibility using explicit solvent simulations: The ribosomal and group I intron reverse kink-turn motifs. *J. Chem. Theory Comput.* **2011**, *7* (9), 2963–2980. (b) Besseova, I.; Banas, P.; Kuhrova, P.; Kosinova, P.; Otyepka, M.; Sponer, J. Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration. *J. Phys. Chem. B* **2012**, *116* (33), 9899–9916.

(36) Case, D.A.; T. A. D, Cheatham, T.E.; , III, Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Walker, R.C.; Zhang, W.; Merz, K.M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; A. W. Götz, I. Kolossváry, Wong, K.F.; Paesani, F.; Vanicek, J.; Wolf, R.M.; Liu, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.R.; Mathews, D.H.; Seetin, M.G.; R. Salomon-Ferrer, Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; , and Kollman, P.A. AMBER 12, University of California: San Francisco, 2012.

(37) Reblova, K.; Fadrna, E.; Sarzynska, J.; Kulinski, T.; Kulhanek, P.; Ennifar, E.; Koca, J.; Sponer, J. Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics. *Biophys. J.* **2007**, *93* (11), 3932–3949.

(38) (a) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh ewald method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. (b) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald – An N·Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(39) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of Cartesian equations of motion of a system with constraints – Molecular-dynamics of N-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(40) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(41) (a) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100* (S1), 19824–19839. (b) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246* (1–2), 122–129.

(42) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular Simulations. *Biopolymers* **2001**, *56* (4), 275–291.



- (43) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson–Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **2002**, *23* (13), 1244–1253.
- (44) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (18), 10037–10041.
- (45) Konecny, R.; Baker, N. A.; McCammon, J. A. iAPBS: A programming interface to adaptive Poisson–Boltzmann solver. *Comput. Sci. Discovery* **2012**, *5* (1), 015005–015013.
- (46) AMBER mailing list archive. <http://lists.ambermd.org/mailman/listinfo/amber> (accessed November 2013).
- (47) (a) TURBOMOLE, V6.5: A development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007; TURBOMOLE GmbH: Karlsruhe, Germany, since 2007, 2013. (b) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic-structure calculations on workstation computers – The program system Turbomole. *Chem. Phys. Lett.* **1989**, *162* (3), 165–169.
- (48) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91* (14), 146401.
- (49) (a) Zhao, Y.; Truhlar, D. G. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A* **2005**, *109* (25), 5656–5667. (b) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13* (14), 6670–6688.
- (50) Zhao, Y.; Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06 functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *119* (5–6), 525–525.
- (51) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2* (1), 73–78.
- (52) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8* (9), 1057–1065.
- (53) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4* (18), 4285–4291.
- (54) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356* (1–3), 98–109.
- (55) Sierka, M.; Hogeekamp, A.; Ahlrichs, R. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.* **2003**, *118* (20), 9136–9148.
- (56) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* **1997**, *97* (1–4), 119–124.
- (57) Eckert, F.; Pulay, P.; Werner, H.-J. Ab initio geometry optimization for large molecules. *J. Comput. Chem.* **1997**, *18* (12), 1473–1483.
- (58) Klamt, A.; Schuurmann, G. Cosmo – A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, No. 5, 799–805.
- (59) Richardson, D. C. *Suitename v. 0.3.070628*. <http://kinemage.biochem.duke.edu/software/suitename.php>.
- (60) O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
- (61) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera – A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
- (62) 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.
- (63) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.
- (64) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (65) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B.  $\alpha/\gamma$  Transitions in the B-DNA backbone. *Nucleic Acids Res.* **2002**, *30* (24), 5398–5406.
- (66) Ho, J. M.; Klamt, A.; Coote, M. L. Comment on the correct use of continuum solvent models. *J. Phys. Chem. A* **2010**, *114* (S1), 13442–13444.
- (67) (a) Spomer, J.; Kypr, J. Theoretical-analysis of the base stacking in DNA – Choice of the force-field and a comparison with the oligonucleotide crystal-structures. *J. Biomol. Struct. Dyn.* **1993**, *11* (2), 277–292. (b) Spomer, J.; Riley, K. E.; Hobza, P. Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.* **2008**, *10* (19), 2595–2610.
- (68) Muddana, H. S.; Sapra, N. V.; Fenley, A. T.; Gilson, M. K. The electrostatic response of water to neutral polar solutes: Implications for continuum solvent modeling. *J. Chem. Phys.* **2013**, *138* (22), 224504.
- (69) (a) Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, *51* (1), 69–82. (b) Gong, Z.; Zhao, Y. J.; Xiao, Y. RNA stability under different combinations of Amber force fields and solvation models. *J. Biomol. Struct. Dyn.* **2010**, *28* (3), 431–441.
- (70) Harris, R. C.; Boschitsch, A. H.; Fenley, M. O. Influence of grid spacing in Poisson–Boltzmann equation binding energy estimation. *J. Chem. Theory Comput.* **2013**, *9* (8), 3677–3685.
- (71) Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of protein–ligand binding free energies. *Proteins: Struct., Funct., Bioinf.* **2012**, *80* (5), 1326–1342.
- (72) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable force field for peptides and proteins based on the classical drude oscillator. *J. Chem. Theory Comput.* **2013**, *9* (12), 5430–5449.
- (73) (a) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction,  $\delta(\text{CCSD(T)})(\text{MP2})$ : Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 19. (b) Korth, M.; Grimme, S. “Mindless” DFT benchmarking. *J. Chem. Theory Comput.* **2009**, *5* (4), 993–1003.
- (74) Korth, M. Empirical hydrogen-bond potential functions—An old hat reconditioned. *ChemPhysChem* **2011**, *12* (17), 3131–3142.