

# A-Minor Tertiary Interactions in RNA Kink-Turns. Molecular Dynamics and Quantum Chemical Analysis

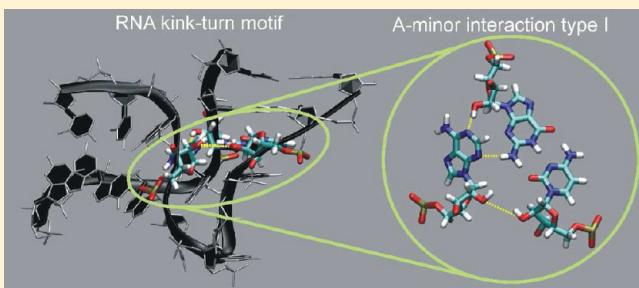
Kamila Réblová,<sup>\*,†</sup> Judit E. Šponer,<sup>†,‡</sup> Náda Špačková,<sup>†</sup> Ivana Beššeová,<sup>†</sup> and Jiří Šponer<sup>\*,†,‡</sup>

<sup>†</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 61265 Brno, Czech Republic

<sup>‡</sup>CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic

 Supporting Information

**ABSTRACT:** The RNA kink-turn is an important recurrent RNA motif, an internal loop with characteristic consensus sequence forming highly conserved three-dimensional structure. Functional arrangement of RNA kink-turns shows a sharp bend in the phosphodiester backbone. Among other signature interactions, kink-turns form A-minor interaction between their two stems. Most kink-turns possess extended A-minor I (A-I) interaction where adenine of the second A•G base pair of the NC-stem interacts with the first canonical pair of the C-stem (i.e., the receptor pair) via *trans*-sugar-edge/sugar-edge (tSS) and *cis*-sugar-edge/sugar-edge (cSS) interactions. The remaining kink-turns have less compact A-minor 0 (A-0) interaction with just one tSS contact. We show that kink-turns with A-I in ribosomal X-ray structures keep G=C receptor base pair during evolution while the inverted pair (C=G) is not realized. In contrast, kink-turns with A-0 in the observed structures alternate G=C and C=G base pairs in sequences. We carried out an extended set ( $\sim 5 \mu\text{s}$ ) of explicit-solvent molecular dynamics simulations of kink-turns to rationalize this structural/evolutionary pattern. The simulations were done using a net-neutral  $\text{Na}^+$  cation atmosphere (with  $\sim 0.25 \text{ M}$  cation concentration) supplemented by simulations with either excess salt  $\text{KCl}$  atmosphere or inclusion of  $\text{Mg}^{2+}$ . The results do not seem to depend on the treatment of ions. The simulations started with X-ray structures of several kink-turns while we tested the response of the simulated system to base substitutions, modest structural perturbations and constraints. The trends seen in the simulations reveal that the A-I/G=C arrangement is preferred over all three other structures. The A-I/C=G triple appears structurally entirely unstable, consistent with the covariation patterns seen during the evolution. The A-0 arrangements tend to shift toward the A-I pattern in simulations, which suggests that formation of the A-0 interaction is likely supported by the surrounding protein and RNA molecules. A-0 may also be stabilized by additional kink-turn nucleotides not belonging to the kink-turn consensus, as shown for the kink-turn from ribosomal Helix 15. Quantum-chemical calculations on all four A-minor triples suggest that there is a different balance of electrostatic and dispersion stabilization in the A-I/G=C and A-I/C=G triples, which may explain different behavior of these otherwise isosteric triples in the context of kink-turns.



## INTRODUCTION

Kink-turns are recurrent asymmetric RNA internal loops, occurring in the ribosome<sup>1,2</sup> and other RNAs.<sup>3–7</sup> The secondary structure of kink-turns contains a noncanonical (NC) stem composed of non-Watson–Crick (non-WC) base pairs, a canonical (C) stem with WC base pairs, and a bulge segment between the stems (Figure 1A).<sup>1</sup> These three parts comprise characteristic base pairs and tertiary interactions shaping up the kink-turn into a specific functional fold with a sharp bend in the phosphodiester backbone (Figure 1A).<sup>1,2</sup> Kink-turns play a key role in protein-assisted RNA folding<sup>8</sup> and likely contribute to functional dynamics of RNA.<sup>9–11</sup>

Kink-turns are stabilized by a set of signature interactions<sup>1,2</sup> (see ref 12 for the nomenclature of base pairing): (i) canonical (usually G=C) base pair at the first position in the C-stem, (ii) tandem *trans* Hoogsteen/sugar-edge (tHS) “sheared” base pairs

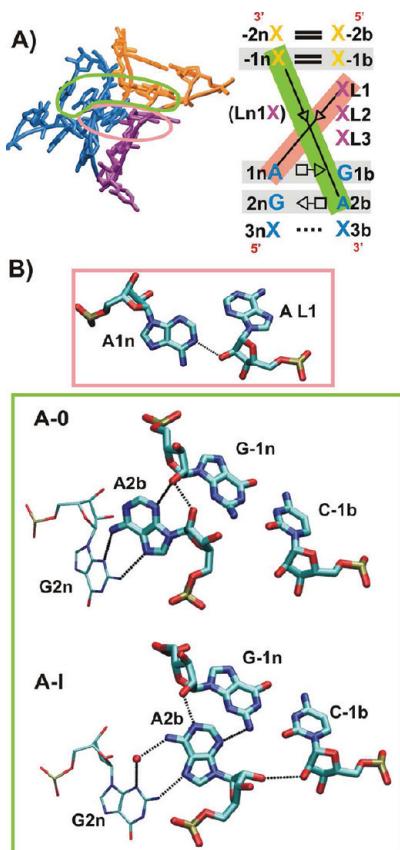
at the first and second positions of the NC-stem, and (iii) two tertiary *trans* sugar-edge/sugar-edge (tSS) interactions (Figure 1A). The first occurs between the 5'-end nucleotide of the bulge and the adenine of the first tHS pair of the NC-stem. It includes key 2'OH $\cdots$ A(N1) H-bond (Figure 1), which is essential for folding of kink-turns.<sup>13,14</sup> The second tertiary contact interconnects the NC and C-stems, as it occurs between adenine of the second NC-stem tHS pair and the first C-stem canonical pair (Figure 1A).

The second tertiary contact is represented by A-minor interaction. A-minor interactions are the most abundant tertiary contacts in large functional RNAs.<sup>15–17</sup> They in general involve adenine nucleotide inserted through its Sugar-Edge into the

**Received:** July 11, 2011

**Revised:** October 10, 2011

**Published:** October 17, 2011



**Figure 1.** (A) Right: Consensus secondary structure of kink-turn with unified numbering and description of nucleotides<sup>13</sup> (nucleotides of the C-stem, NC-stem, and bulge are in orange, blue, and violet, respectively. The Ln1X which is in parentheses is a deviation from the kink-turn signature). The three characteristic stem base pairs are highlighted by gray boxes, the first tSS contact essential for folding of the kink-turn is in the pink box, and the second tSS contact representing the A-minor interaction is in the green box. Left: Tertiary structure of the kink-turn colored identically as the secondary structure. The two tSS contacts are indicated by colored ovals. (B) Detailed views on the two tSS contacts; the first tertiary contact with signature H-bond marked by the dashed line is in the pink box, and the two A-minor interactions are in the green box. The A-minor arrangements also show a sheared A•G pair (with guanine in thin licorice representation). The nucleotides are labeled by unified numbering and description for kink-turns.<sup>13</sup> In the A-I arrangement, the sheared A•G pair has one H-bond direct and a second one mediated via a water molecule (the red ball). The dashed lines indicate H-bonds between adenine and guanine and between adenine and the receptor pair.

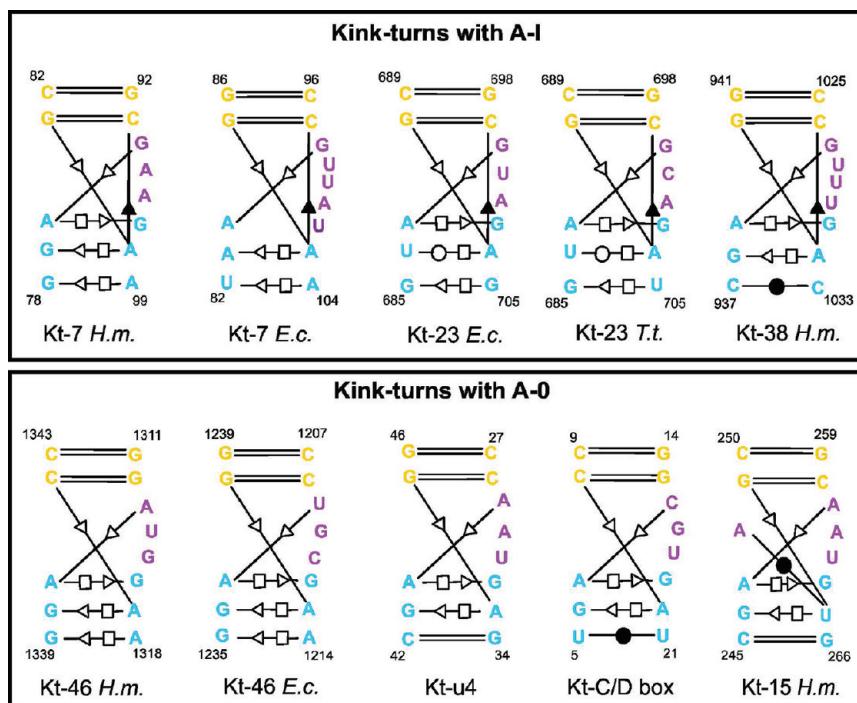
minor groove of A-RNA double helix which serves as the receptor. There are four basic substates of A-minor interactions, namely types I, II, III, and 0, differing in the position of the adenine nucleotide with respect to the receptor canonical pair.<sup>15</sup> The type I interaction is the most frequent (52%), while others occur much less frequently (type II, 31%; type 0, 10%; and type III, 7%).<sup>18</sup> Known kink-turn structures contain either A-minor I (A-I) or A-minor 0 (A-0) (Figure 1B).

Kink-turns have been extensively studied by experimental and theoretical methods. Biochemical studies have shown that formation of the sharply bent (folded) kink-turn structure requires the presence of divalent cations and binding proteins.<sup>19–23</sup>

Despite limitations, the explicit solvent molecular dynamics (MD) simulation technique provides useful insights into structural dynamics of known RNA structures, including impacts of base substitutions and modifications.<sup>24–39</sup> Previous MD simulations performed on a 10+ nanosecond time scale revealed considerable dynamics in kink-turn structures.<sup>9,11,40–42</sup> Simulations of kink-turns and larger molecular segments containing kink-turns showed that folded kink-turns exhibit significant anisotropic elbow-like fluctuations, which could play roles in large-scale functional movements within molecular assemblies.<sup>9,11,40</sup> The simulations investigated properties of isolated kink-turns folded in their native geometries taken from X-ray structures of the respective ribonucleoprotein systems. Although the kink-turns would probably unfold if the simulations are long enough (see above), the native and unfolded states are separated by a large enough energy barrier. Thus, unfolding of isolated kink-turns known from solution experiments was not seen on the present simulation time scales. Unfolding of kink-turns was investigated using enhanced sampling methods.<sup>43</sup> Although the solution simulations at first sight look to be related to the solution experiment, they in fact characterize the native kink-turn topology and are thus strictly relevant to the starting X-ray structures. The unfolded topology probably does not have any significant biochemical relevance and its existence does not affect the simulations sampling the conformational space of folded kink-turns. Simulations of protein/kink-turn complexes have been recently reported, too.<sup>44</sup>

The characteristic base pairs in ribosomal kink-turns are structurally conserved; that is, they are substituted exclusively by isosteric base pairs in the course of evolution.<sup>2</sup> However, frequency of occurrence of isosteric substitutions is not equal; that is, some isosteric substitutions occur frequently, whereas others are not observed at all. This finding motivated us to perform the present study. We first analyzed the structural and bioinformatics data. We compared sequence variability of the first C-stem base pair which is involved in the A-minor interaction as receptor (Figure 1) using data by Lescoute et al.<sup>2</sup> with the observed X-ray structures of ribosomal kink-turns. We found correlation between the A-minor interaction type and orientation of the receptor pair (i.e., either G=C or C=G). In particular, we noticed that kink-turns with A-I in the X-ray structures, i.e., 23S rRNA Kt-38, 16S rRNA Kt-23, 23S rRNA Kt-42, 23S rRNA Kt-7, 16S rRNA Kt-11, 23S rRNA Kt-94/99, and 23S rRNA Kt-4/5, keep primarily the G=C pair in the first C-stem position in the evolution. The inverted (C=G) base pair is basically not realized during the evolution. In contrast, ribosomal kink-turns with A-0 in the observed structures, i.e., 23S rRNA Kt-46, Kt-15, and Kt-78, alternate G=C and C=G base pairs according to the sequence analysis. We created a new isosteric matrix for the receptor pair in the 23S Kt-78 as it is not indicated in Lescoute et al.,<sup>2</sup> see the Supporting Information, Figure S1. The reason why kink-turns with A-I predominantly prefer the G=C pair is, however, not known. This observation motivated us to carry out a theoretical study which combines MD simulations and quantum chemical (QM) calculations, with the aim to better understand the role of the A-minor interaction type in the kink-turn structure and sequence evolution.

We investigate five kink-turns with A-I comprising the G=C pair and five kink-turns with A-0 comprising both G=C and C=G pairs (Figure 2 and Table 1, see also the Methods section). We analyze the simulation behavior of the A-0 and the A-I interactions and evaluate the influence of the orientation of the



**Figure 2.** Secondary structures of central parts of the studied kink-turns with base pairs marked according to the standard classification.<sup>12</sup> See the Methods section for comment on the tSS family nomenclature. Nucleotides are colored as in the consensus secondary structure in Figure 1. X-ray numbers are in black. Complete simulated sequences are in Table 1.

receptor pair on the structural stability of the A-minor interactions. The standard 100+ ns scale simulations are supplemented by simulations with base substitutions and restrained simulations, where we probe the effect of perturbations of the experimentally observed A-minor arrangements. QM calculations evaluate the intrinsic strength of the A-minor interactions. QM calculations have been recently applied to many RNA base pairs, all six “sugar-edge” base pair families,<sup>45,46</sup> consensus A-minor and P-tertiary interactions,<sup>46</sup> and some other RNA interactions.<sup>47</sup> The QM technique includes neither the dynamics (including the potentially important entropy effects) nor the context of the interactions. However, it provides a more accurate description of the intrinsic molecular interactions than the force fields. The force fields in fact try to approximate the QM electronic structure description. The application of standard QM calculations of interaction energies is in our case justified by the fact that we investigate a set of isosteric interactions (i.e., A-minor triplets) in basically identical contexts. In such situation it is plausible to expect that variations of the intrinsic interaction energies (intrinsic strength of the H-bonds) and their nature contribute substantially to the differences in the overall stability and dynamics of the studied interactions. The capability of QM calculations to complement structural bioinformatics by energy data has been demonstrated for RNA base-phosphate interactions,<sup>48</sup> see also the review in ref 49.

## METHODS

**Starting Structures.** We simulated noncomposite ribosomal kink-turns (23S rRNA *H.m.* and *E.c.* Kt-7; 16S rRNA *E.c.* and *T.t.* Kt-23; 23S rRNA *H.m.* Kt-38; 23S rRNA *H.m.* and *E.c.* Kt-46; and 23S rRNA *H.m.* Kt-15) and kink-turns from small noncoding RNAs (Kt-u4 and Kt-C/D box) (Figure 2 and Table 1). Most of

these structures are close to the consensus kink-turn sequence,<sup>1</sup> which comprises a three-nucleotide bulge segment, an NC-stem with two tHS “sheared” A•G base pairs, and a C-stem with several consecutive canonical base pairs (Figures 1 and 2). We use a general nucleotide numbering system for kink-turns<sup>13</sup> see Figure 1. Differences from the consensus sequence can be seen in the *E.c.* Kt-7, which has an incomplete first pair in the NC-stem followed by a tHS A•A pair, in the *H.m.* Kt-38 with four bases in the bulge segment, and in both *E.c.* and *T.t.* Kt-23, which have tHW A•U pair at the 2b•2n position (Figure 2). The largest difference from the consensus sequence shows *H.m.* Kt-15, which exhibits extra adenine at the L<sub>n</sub>1 position and tHS U•G pair at the 2b•2n position (Figure 2). The AL<sub>n</sub>1 makes additional contacts to the A-0 triplet; particularly, it forms a cWW pair with U2b (Figures 2 and S2), which is not seen in other kink-turn structures (Figure 2).

All studied kink-turns comprise two conserved tSS tertiary interactions.<sup>2</sup> The first one between L<sub>n</sub>1 and L<sub>1</sub> contains an important (signature) L<sub>1</sub>(O<sub>2'</sub>)...L<sub>n</sub>(N<sub>1</sub>) H-bond (Figure 1B).<sup>13</sup> The second one between -L<sub>n</sub>1 and 2b represents A-minor interaction, either A-I or A-0 (Figure 1B). A-I consists of the -L<sub>n</sub>1•2b tSS pair complemented by the 2b•1b cSS interaction (Figure 1B), which may be either direct or water-mediated.<sup>11</sup> The A-I can be seen in the *H.m.* and *E.c.* Kt-7, *E.c.* and *T.t.* Kt-23, and *H.m.* Kt-38 (Table 1). The A-0 comprises just the -L<sub>n</sub>1•2b tSS pair (different from the tSS pair in A-I type as the 2b base is shifted toward the -L<sub>n</sub>, see Figure 1B). It occurs in the *H.m.* and *E.c.* Kt-46, Kt-u4, Kt-C/D box and *H.m.* Kt-15 (Table 1). All studied kink-turns with A-I have receptor G=C pair (-L<sub>n</sub>1•1b position, see Figures 1 and 2), termed “A-I/G=C” triplet throughout the text. The kink-turns with A-0 have either the G=C receptor pair (“A-0/G=C” triplet) or the C=G receptor pair (A-0/C=G triplet; see Figure 2). For the tSS base pair,

**Table 1.** Survey of the Simulations

kink-turn	organism	simulated segment (original experimental numbering)	resolution (Å); PDB code	length of simulation (ns)
Free Simulations of Kink-Turns with A-I				
Kt-38	<i>H.m.</i>	931–952, 1015–1039	2.4; 1S72	200
Kt-38 control simulation	<i>H.m.</i>	934–945, 1022–1036	2.4; 1S72	200
Kt-23	<i>T.t.</i>	679–711	2.9; 2UUA	200
Kt-23	<i>E.c.</i>	679–711	3.5; 2AVY	150
Kt-7	<i>E.c.</i>	78–108	3.5; 2AW4	150
Kt-7	<i>H.m.</i>	76–83, 91–101	2.4; 1S72	200
Free Simulations of Kink-Turns with A-0				
Kt-46	<i>E.c.</i>	1196–1222, 1227–1250	3.5; 2AW4	100
Kt-46 control simulation	<i>E.c.</i>	1200–1245	3.5; 2AW4	150
Kt-u4	N/A <sup>a</sup>	26–35, 41–47	2.9; 1E7K <sup>b</sup>	200
Kt-u4	N/A <sup>a</sup>	26–35, 41–47	2.9; 1E7K <sup>c</sup>	200
Kt-46	<i>H.m.</i>	1305–1349	2.4; 1S72	200
Kt-46 control simulation	<i>H.m.</i>	1305–1349	2.4; 1S72	200
Kt-C/D box	N/A <sup>a</sup>	1–25	2.7; 1RLG <sup>d</sup>	100
Kt-C/D box control simulation	N/A <sup>a</sup>	1–25	2.7; 1RLG <sup>d</sup>	200
Kt-15	<i>H.m.</i>	244–267	2.4; 1S72	200
Kt-15 <sup>e</sup>	<i>H.m.</i>	244–267	2.4; 1S72	100
Simulations with Restraints				
Kt-38 <sup>f</sup>	<i>H.m.</i>	935–942, 1024–1035	2.4; 1S72	150
Kt-38 <sup>g</sup>	<i>H.m.</i>	935–942, 1024–1035	2.4; 1S72	150
Kt-u4 <sup>h</sup>	N/A <sup>a</sup>	26–35, 41–47	2.9; 1E7K <sup>b</sup>	150
Kt-u4 <sup>g</sup>	N/A <sup>a</sup>	26–35, 41–47	2.9; 1E7K <sup>b</sup>	150
Kt-46 <sup>i</sup>	<i>E.c.</i>	1200–1245	3.5; 2AW4	100
Simulations with Substitutions				
Kt-38 <sup>j</sup>	<i>H.m.</i>	931–952, 1015–1039	2.4; 1S72	100
Kt-38 <sup>k</sup>	<i>H.m.</i>	931–952, 1015–1039	2.4; 1S72	300
Simulation with 6 Mg <sup>2+</sup> Ions				
Kt-u4	N/A <sup>a</sup>	26–35, 41–47	2.9; 1E7K <sup>b</sup>	100
Simulations with 1 Mg <sup>2+</sup> Ion				
Kt-u4	N/A <sup>a</sup>	26–35, 41–47	2.9; 1E7K <sup>c</sup>	125
Kt-46	<i>E.c.</i>	1200–1245	3.5; 2AW4	125
Kt-46	<i>H.m.</i>	1305–1349	2.4; 1S72	125
Kt-C/D box	N/A <sup>a</sup>	1–25	2.7; 1RLG <sup>d</sup>	125

<sup>a</sup> Synthetic construct. <sup>b</sup> Chain C. <sup>c</sup> Chain D. <sup>d</sup> Chain C. <sup>e</sup> ALn1 was pulled out of the bulge segment. <sup>f</sup> Simulation starting from A-0 arrangement imposed by restraints (see the Methods section). <sup>g</sup> Simulation starting from geometry with increased distance between C and NC-stems and disrupted A-minor contact (see the Methods section). <sup>h</sup> Simulation starting from A-I arrangement imposed by restraints (see the Methods section). <sup>i</sup> Simulation, where both direct H-bonds in the second sheared A•G base pair were stabilized by restraints. <sup>j</sup> C1026G and G940C substitutions, parmbsc0 force field. <sup>k</sup> C1026G and G940C substitutions, parmb99 force field.

where the nomenclature is not always clear, we use the rule that the nucleotide bound via its 2'OH group to the partner nucleobase is listed first while the arrow of triangle points from the nucleotide interacting with its 2'OH with the paired base.

**MD Simulations.** The studied kink-turns were simulated on a 100+ ns time scale using the PMEMD module of AMBER-10.0<sup>50</sup> with the parmb99<sup>51</sup> force field<sup>52</sup> (Table 1). One control simulation was run with parmbsc0<sup>53</sup> variant designed to prevent degradation of B-DNA. For RNA molecules, parmb99 and parmbsc0 provide similar results.<sup>54</sup> Recently the  $\chi$  dihedral potential of the force field was completely reparametrized, to prevent degradation of the RNA molecules in long simulations to ladder-like structures.<sup>55,56</sup> Fortunately, on the present time scale the kink-

turns are not prone to the ladder-like degradation while our tests so far (unpublished data) reveal that the kink-turn simulations are not sensitive to the force field choice. As the project has been initiated a long time before the  $\chi$  modification has been available and tested, we decided to finish all calculations with the parmb99 force field version. The molecules were solvated by a periodic TIP3P water box extending 10–15 Å away from the solute in all directions. We mostly used an octahedral box extending 10 Å away from the solute. Few simulation runs at the beginning of the project were done with rectangular boxes with appropriate sizes in each dimension to allow rotation of the molecule within the box. The box choice is not assumed to have any detectable effect on the simulation behavior. The molecules were neutralized by

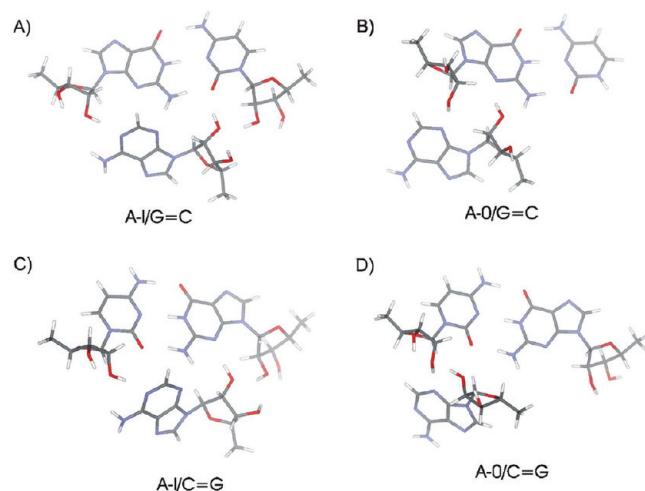
**Table 2.** Survey of the Simulations Carried out with KCl

kink-turn	organism	simulated segment (original experimental numbering)	resolution (Å); PDB code	length of simulation (ns)
Free Simulations of Kink-Turns with A-I				
Kt-38	<i>H.m.</i>	931–952, 1015–1039	2.4; 1S72	100
Kt-23	<i>E.c.</i>	679–711	3.5; 2AVY	100
Free Simulations of Kink-Turns with A-0				
Kt-46	<i>E.c.</i>	1200–1245	3.5; 2AW4	100
Kt-46	<i>H.m.</i>	1305–1349	2.4; 1S72	100
Kt-C/D box	N/A <sup>a</sup>	1–25	2.7; 1RLG <sup>b</sup>	100
Kt-15	<i>H.m.</i>	244–267	2.4; 1S72	100

<sup>a</sup> Synthetic construct. <sup>b</sup> Chain C.

the earlier standard AMBER Na<sup>+</sup> monovalent cations (radius 1.868 Å and well depth 0.00277 kcal/mol)<sup>57</sup> initially placed using the Xleap to the most negative sites around the solute, resulting in effective concentration of cations of ~0.2–0.25 M under the net-neutralization. The cation concentration can be obtained by considering the number of ions vs the box volume or number of ions vs number of water molecules, with similar result. We have extensively discussed the basic correctness of this ion treatment for example in;<sup>9,58</sup> for independent analysis of ion treatment in nucleic acid simulations see also.<sup>59</sup> For RNA molecules net-neutralizing ion condition and excess-salt simulations provide typically similar results.<sup>58</sup> The above used Na<sup>+</sup> parameters are appropriate if they are not mixed with imbalanced parametrizations of anions. Nevertheless, we also carried out tests with different ion conditions. (i) six control ~0.2 M excess salt KCl simulations using parameters for K<sup>+</sup> (radius 1.705 Å and well depth 0.1936829 kcal/mol) and Cl<sup>-</sup> (radius 2.513 Å and well depth 0.0355910 kcal/mol), which prevent salt crystallization at low to medium salt concentrations,<sup>60</sup> (ii) one simulation with 6 Mg<sup>2+</sup> ions (radius 0.7926 Å and well depth 0.8947 kcal/mol),<sup>57</sup> and (iii) four simulations with single Mg<sup>2+</sup> (Tables 1 and 2). The Mg<sup>2+</sup> ions were initially manually shifted ~5 Å away the RNA molecule to prevent inner-shell contacts during equilibration due to limited accuracy of force fields for divalent ions.<sup>28</sup> The Mg<sup>2+</sup> simulations were carried out with additional Na<sup>+</sup> ions used for neutralization. The crystal structure of Kt-u4 (1E7K.pdb) and in general also other structures do not contain Mg<sup>2+</sup> ions specifically bound to kink-turns. Mg<sup>2+</sup> ions used in the simulations were added when building the initial model in Xleap module using the ESP potential.

The particle mesh Ewald technique<sup>61</sup> was used while we applied standard protocols described elsewhere.<sup>62</sup> We carried out two simulations where we replaced the original receptor G=C pair in *H.m.* Kt-38 with C=G pair forbidden in evolution.<sup>2</sup> We also ran simulations with perturbed A-minor interaction. In particular, we gradually increased the distance between C and NC-stems using restraints in *H.m.* Kt-38 and Kt-u4, which resulted in disruption of A-minor interactions (the signature L1(O2')···In(N1) interaction and base pairs in the stems were not disrupted). After 5 ns, the restraints were released to run free simulations. Further, we ran simulations where the initial A-minor interactions were changed using restraints. In particular, we transformed A-I to A-0 in *H.m.* Kt-38 and A-0 to A-I in Kt-u4 during 5–6 ns. After that, the restraints were lifted. We ran also a restrained simulation of *E.c.* Kt-46, where we stabilized both direct H-bonds in the second sheared A•G base pair during the

**Figure 3.** B3LYP/6-31G\*\* optimized geometries of the A-minor interactions.

entire simulation. Restraints were applied on distances (see Supporting Information), IALTD parameter defining restraint function was set to 0 (for more details see Amber 10 manual part: Distances, angle and torsional restraints). Prior to each MD simulation with restraints, the model was minimized, thermalized, and equilibrated using standard protocol.<sup>62</sup>

**Quantum Chemical Calculations.** QM calculations followed the standard computational approach used in our previous studies.<sup>45,46,63–65</sup> Geometry optimizations were carried out at DFT (Density Functional Theory) level of theory using the Gaussian03 program package.<sup>66</sup> The density functional was built up by Becke's three-parameter exchange<sup>67</sup> and Lee–Yang–Parr's correlation functional (abbreviated as B3LYP).<sup>68</sup> The 6-31G\*\* basis set was used for geometry optimizations. Geometries of A-I/G=C and A-0/G=C were optimized with full relaxation of all structural parameters (Figure 3). Since the A-I/C=G triplet underwent modest changes upon full optimization (the adenine became perpendicular to the C=G pair) geometrical constraints had to be imposed to preserve the mutual position of the interacting nucleosides in an experiment-like arrangement (optimized geometry with constraints is in Figure 3). These involved the C5(G)-C6(G)-N3(A)-C6(A) and C5(G)-C6(G)-C4(A)-N1(A) torsions as well as the C5(G)-C6(G)-N3(A) valence angle. Similarly, the N2(G)-O4'(rA) distance had to be fixed at the value obtained from an equilibrated MD simulation run when optimizing A-0/C=G in order to avoid formation

of artificial H-bonds (optimized geometry with constraint is in Figure 3). The initial geometry of A-I/G=C was taken from the X-ray structure of *T.t.* Kt-11 (PDB code: 2J00), where both tSS G•A and cSS A•C interactions are complete. The A-I/C=G was built from A-I/G=C by rotating the G=C pair by 180° around the axis defined by the bisector of the line connecting the center of masses of G and C (lying in the plane of the G=C pair). A-0/G=C was taken from X-ray structure of *E.c.* Kt-46 (PDB code: 2AW4), while for its inverted variant we used structural analog from *H.m.* Kt-46 (PDB code: 1S72), which was equilibrated using MD procedure (see the Methods section).

Interaction energies were computed on the optimized structures using the RIMP2 approach combined with the aug-cc-pVDZ basis set of atomic orbitals.<sup>69–71</sup> The RIMP2 method for calculating interaction energies has been validated in refs 72 and 73.

Interaction energy of a dimer consisting of two subsystems A and B ( $\Delta E^{AB}$ ) is defined as

$$\Delta E^{AB} = E^{AB} - E^A - E^B \quad (1)$$

where  $E^{AB}$  stands for the electronic energy of the whole system and  $E^A$  and  $E^B$  are the electronic energies of the isolated subsystems A and B.

The interaction energy ( $\Delta E$ ) has two components: the Hartree–Fock (HF) term ( $\Delta E^{HF}$ ) and the electron correlation term ( $\Delta E^{corr}$ ). All interaction energy data have been corrected for the basis set superposition error.

$$\Delta E = \Delta E^{HF} + \Delta E^{corr} \quad (2)$$

The  $\Delta E^{HF}$  term includes mainly the electrostatic part of the interaction energy, short-range exchange repulsion, and polarization/charge transfer contributions. The  $\Delta E^{corr}$  term is dominated by the dispersion attraction and further includes the electron correlation correction to the electrostatic energy. The later term is mostly repulsive since the electron correlation reduces the dipole moments of the monomers.

The physical chemistry meaning of the interaction energies is as follows. They capture direct interaction between the interacting monomers in the absence of all the other forces (i.e., in complete isolation, intrinsic interaction energies). The interaction energy of course differs from experimental binding free energies which are determined by a very complex balance of all contributing forces and are substantially context-dependent. Nevertheless, the interaction energies characterize one of the most important contributions to free energies. In contrast to the other contributions, intrinsic interaction energies are not context-dependent, so they represent an important and invariant physical chemistry signature of the studied base pair arrangements.<sup>49</sup>

## RESULTS AND DISCUSSION

**Interpretation of the Results.** All simulations started from experimental structures of folded (tightly kinked) kink-turns or their straightforward modifications. In the course of the 100+ ns simulations, the structures sampled conformational space of fully folded kink-turns, which is the native structure. By definition, the simulations reflect 100+ ns scale structural stability of the simulated molecules within the approximation of the force field. This structural stability is the direct result of the simulations. Our multiple simulations provide rather good sampling of the conformational space of the folded kink-turns. For this particular system, the simulation time scale is sufficient to indirectly derive meaningful

conclusions about the relative preference for different A-minor arrangements in folded kink-turns.

**Structural Stability of A-Minor Interactions Based on MD Simulations.** *A-I Is a Stable Arrangement.* The A-I interaction was stable in the majority of simulations (i.e., in the simulation of *H.m.* Kt-38 and in the simulations of *E.c.* Kt-7, *E.c.* Kt-23, and *T.t.* Kt-23 (Figure 4, Table 1). Changes were seen in the control Kt-38 simulation, where A-I was reversibly disrupted and in the *H.m.* Kt-7 simulation where A-I irreversibly disrupted after 120 ns (Figure 4). The disruption in *H.m.* Kt-7 was followed by opening of the kink-turn (Figure S3 in the Supporting Information), which may indicate either imperfections of the force field or beginning of the genuine unfolding process (see the Introduction).

We carried out two supplemental simulations where we perturbed A-I of *H.m.* Kt-38 using restraints. First, we initially enforced the  $A-I \rightarrow A-0$  transition. The imposed A-0 was then stable up to  $\sim 115$  ns. After that, the A-0 converted back to the original A-I arrangement at 125 ns. In the second simulation, we initially increased the distance between the stems and disrupted the A-I interaction. Once the restraints were released, the kink-turn formed the A-0 (at 15 ns) and then it adopted the original A-I (at 72 ns), which was then stable. In both simulations the A-I interaction was restored spontaneously.

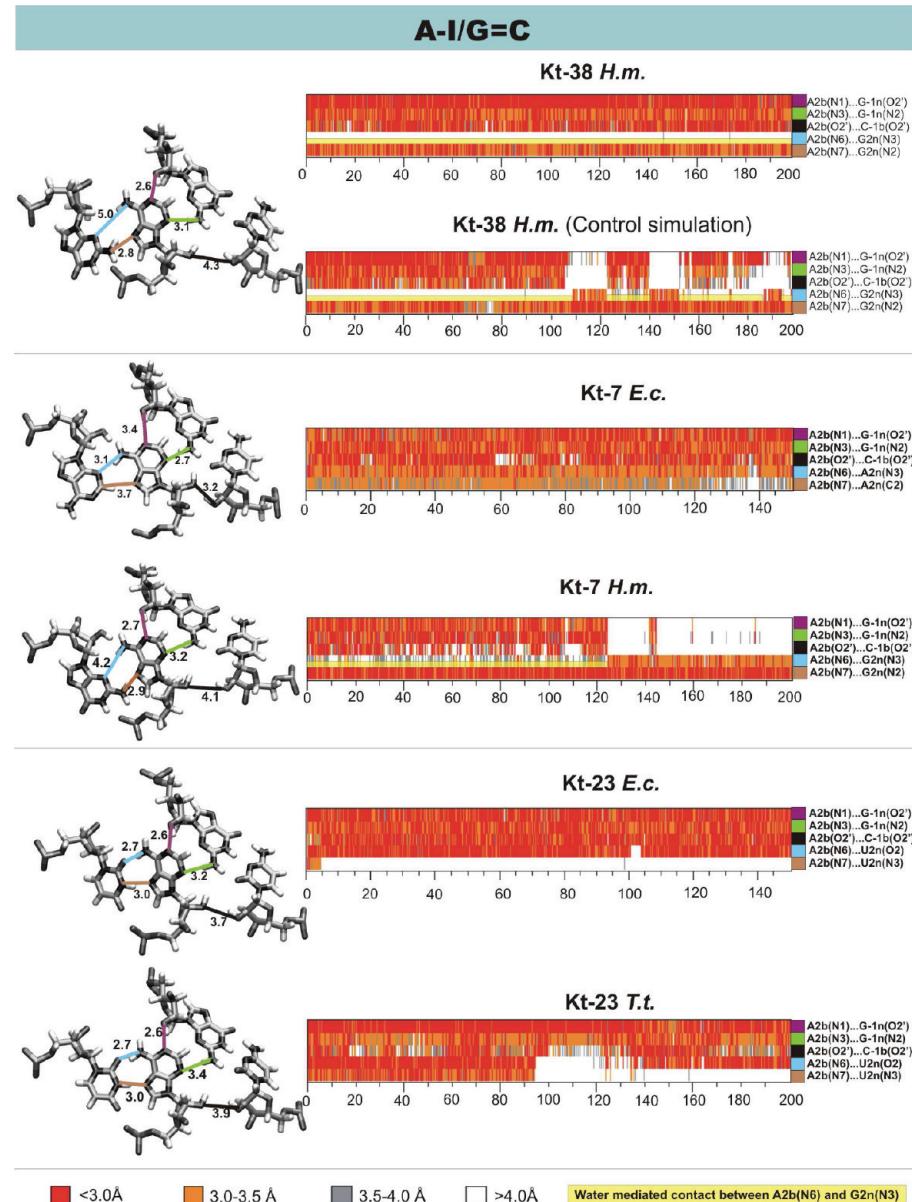
*Additional Interactions not Related to A-I.* Let us briefly summarize some additional structural-dynamics features of the simulated kink-turns which are not directly affecting the A-I dynamics. Base pairs of the C and NC-stems were essentially stable (not shown). There is, however, structural communication between the A-I interaction and the second sheared A•G 2b•2n pair whose adenine participates in A-I. In the crystal structures, the second A•G pair is stabilized by G2n(N2)···A2b(N7) direct H-bond and A2b(N6)···G2n(N3) water-mediated H-bond (Figures 1 and 4). The water mediated H-bond facilitates proper orientation of the interacting A2b (cf. Figure 1) with respect to the receptor pair. When the A-I interaction is disrupted direct A2b(N6)···G2n(N3) H-bond is formed (Figure 4). (See also below and ref 44.)

For the sake of completeness, in the crystal structures of *T.t.* and *E.c.* Kt-23, the 2b•2n tHW A•U pair (Figure 4) is additionally stabilized by 1b(O2')···2n(O4) H-bond. In the course of the *E.c.* and *T.t.* Kt-23 simulations, we observed disruption of this H-bond after 4 and 93 ns, respectively, which was followed by disruption of 2n(N3)···2b(N7) H-bond (Figure 4). These local changes may be related to relaxation of kink-turn structure outside the ribosomal context and do not affect stability of A-I (Figure 4).

The signature interaction between the bulge and the first NC-stem A•G base pair often fluctuates between direct and water-bridge H-bond in the simulations (Figure S4 in Supporting Information). This was analyzed elsewhere<sup>44</sup> and is not important for the scope of this paper.

The H-bonds in the bulge segment mediated by 2'-OH groups also fluctuate; however, details of this dynamics are unrelated to the A-minor dynamics.

*A-0 Interaction Is Less Stable.* The A-0 interaction is less stable in simulations. The structural changes depended on the orientation of the receptor pair. Kink-turns with the G=C pair (i.e., *E.c.* Kt-46 and Kt-u4, see Figure 2) showed spontaneous A-0  $\rightarrow$  A-I transition. In particular, in the simulation of *E.c.* Kt-46, the transition was seen during the first 5 ns (Figure 5). An identical behavior was observed in the control simulation of *E.c.* Kt-46 (Figure 5). Simulations of Kt-u4 also revealed the A-0  $\rightarrow$  A-I transition. We simulated both crystallographically independent



**Figure 4.** (Left) X-ray arrangements of A-I/G=C interactions and 2b•2n pairs in the studied kink-turns. Key H-bonds are highlighted by color lines. The numbers indicate X-ray distances in Å. (Right) Time developments of key H-bonds in the course of free simulations (Table 1). The Kt-38 H.m. trajectory shows the most typical behavior with stable dynamics of kink-turn having the 2b•2n A•G base pair and fully paired A-I interaction; note that there is a water insertion into the 2b•2n A•G base pair in this case (see the text and Figure 1), which was specifically monitored for this H-bond and marked by the yellow line along the time development. The other five examples illustrate simulations with either substituted 2b•2n base pairs (A•A or A•U) or with visible structural instabilities.

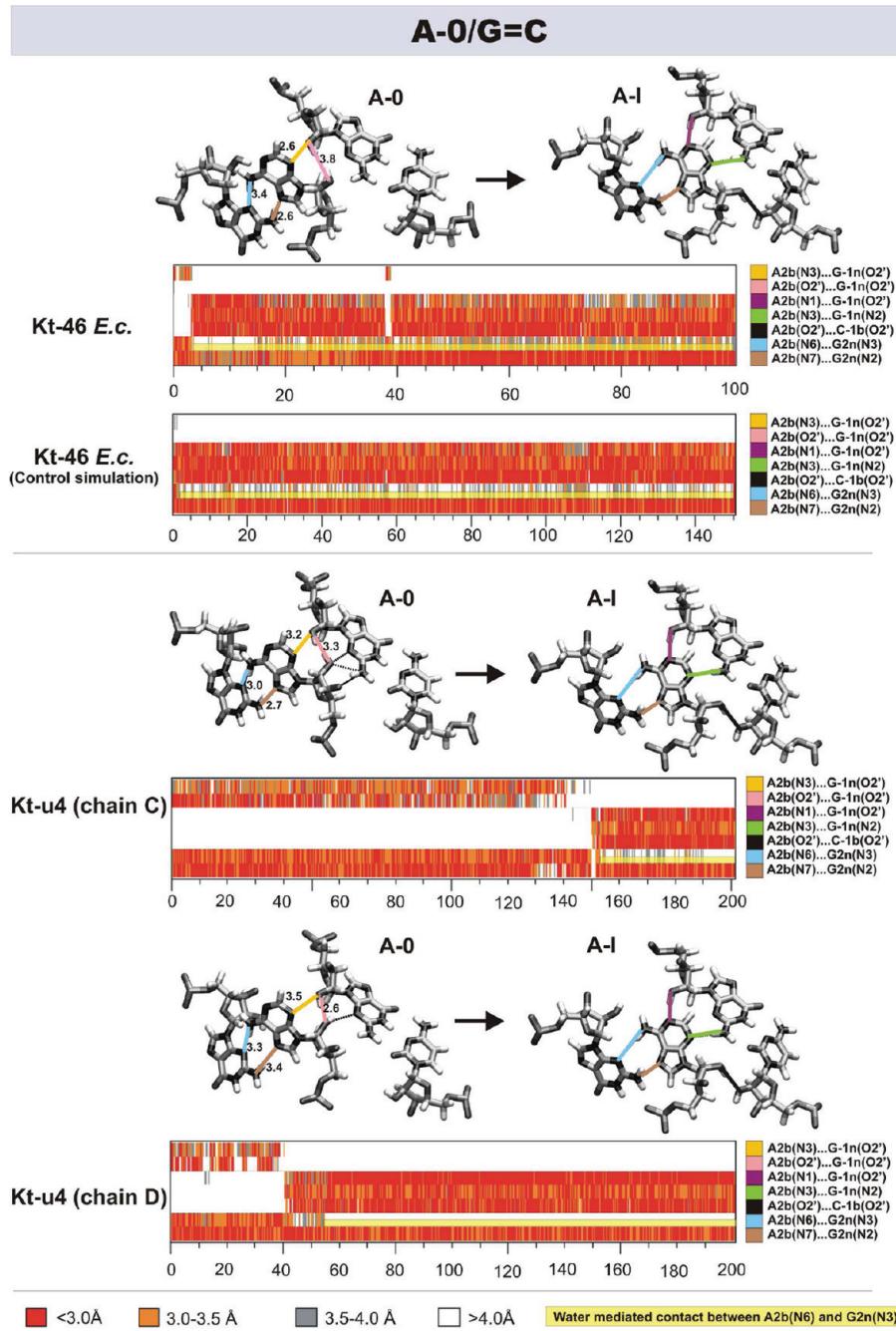
RNA molecules from the pdb 1E7K, termed as C and D chains (Table 1); the transition was detected at 150 and 40 ns, respectively (Figure 5). The transitions did not perturb the overall structure of these kink-turns, and the newly formed A-I arrangements were stable (Figure 5).

We carried out two simulations of the Kt-u4, where the initial A-0 was perturbed using restraints (see the Methods section and Table 1). First, the A-0 was changed to A-I. This arrangement was stable. In the second simulation, we initially increased the distance between the stems to disrupt the initial A-0 arrangement. The kink-turn formed stable A-I after 40 ns.

To understand the different speed of the A-0 → A-I transition in simulations of *E.c.* Kt-46 and Kt-u4, we compared the respective

crystal structures. In *E.c.* Kt-46, A-0 is stabilized by just one H-bond, whereas in the chains C and D of Kt-u4, it is stabilized by 4–5 and 2–3 H-bonds, respectively (Figure 5). Thus, the simulation behavior is consistent with the number of H-bonds in the starting structures (their quality may be affected by the resolution limits).

Kink-turns with the inverted receptor pair (i.e., *H.m.* Kt-46 and Kt-C/D box, see Figure 2), exhibited different behavior. In the simulation of *H.m.* Kt-46 (Table 1), the initial A-0 was disrupted (Figure 6). The interacting nucleotide A2b, however, did not move along the receptor pair toward the G-1b, because it was stabilized by G-1b(N2) ··· A2b(O2') H-bond, which hampered the formation of A-I interaction (Figure 6). The orientation of A2b and G-1b resembles incomplete cSS base pair.

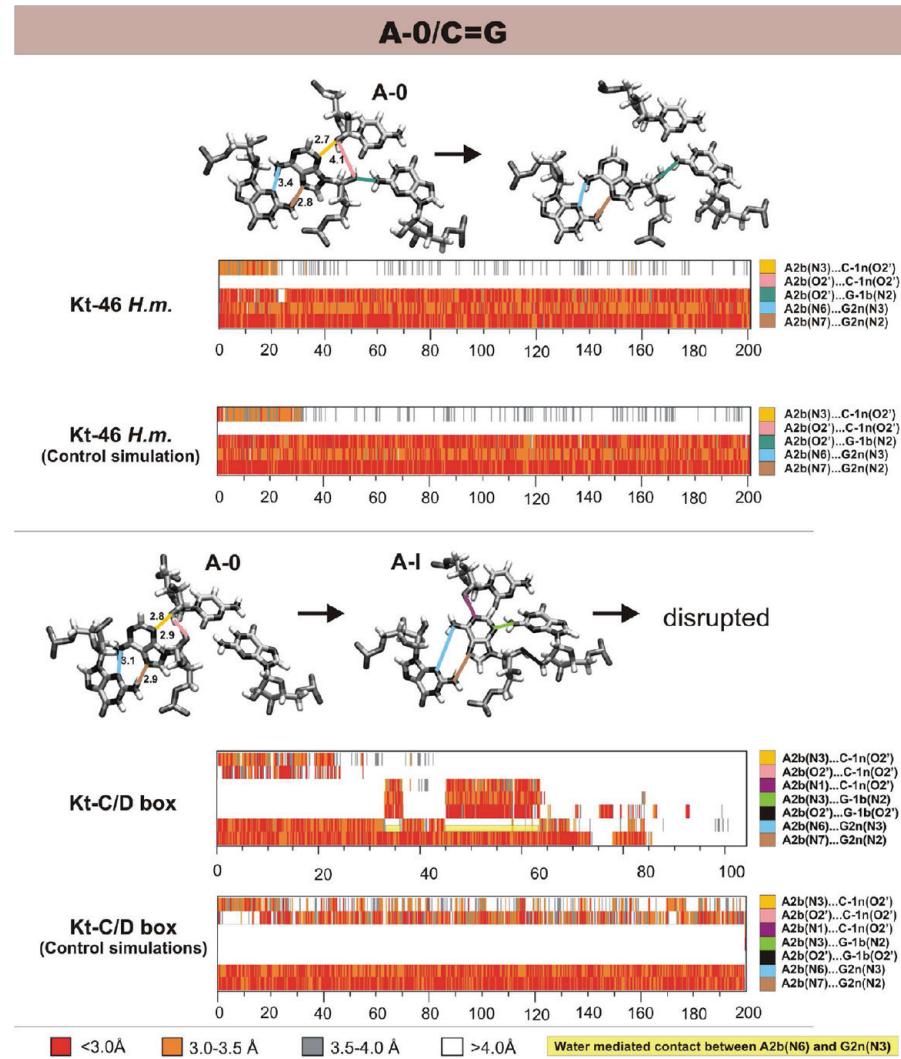


**Figure 5.** Schemes showing transitions of the initial X-ray arrangements of A-0/G=C triplet and 2b•2n base pair into the A-I arrangement in the course of the E.c. Kt-46 and Kt-u4 simulations. The key H-bonds are highlighted by colored lines, and their developments are shown in the corresponding plots. The numbers in A-0 arrangements indicate X-ray distances in Å. The depicted A-I arrangements represent MD snapshot structures. Water insertion into the 2b•2n A•G base pair monitored as in Figure 4.

Control simulation of *H.m.* Kt-46 (Table 1) revealed a similar behavior (Figure 6). In the simulation of Kt-C/D box, the original A-0 disrupted after ~20 ns and at 33 ns the kink-turn formed A-I (Figure 6). This arrangement was then quickly disrupted, formed again in the time period 45–62 ns and then it was irreversibly disrupted. This was followed by loss of the whole kink-turn structure. In second simulation of Kt-C/D box (Table 1), the initial A-0 fluctuated but it was not fully disrupted (Figure 6).

**Second Sheared A•G Base Pair in the NC-Stem Struggles with the A-I Triad.** As noted above, in all simulations we have

detected a tight structural relation between the configuration of the A2b•G2n base pair and the formation of the A-I interaction. When the A-I interaction is formed then the A2b•G2n base pair is water-mediated. In all other instances it is fully paired (Figures 4 and 5). This correlation was further tested for the E.c. Kt-46, of which A-0 quickly changes to the A-I in simulations (Figure 5). We run a simulation while stabilizing the two direct H-bonds in the 2b•2n sheared A•G base pair by restraints. After 10 ns, we observed a partial A-0 → A-I transition. Due to the restraints only  $-1n(N2)\cdots 2b(N3)$  and  $-1b(O2')\cdots 2b(O2')$  H-bonds formed.



**Figure 6.** Schemes showing changes of the initial X-ray arrangements of the A-0/C=G triplet and the 2b•2n pair in the course of the H.m. Kt-46 and Kt-C/D box simulations. The key H-bonds are highlighted by color lines and their developments are shown in the corresponding plots. The numbers in A-0 arrangements indicate X-ray distances in Å. In H.m. Kt-46 simulations, the A-0 was disrupted (the disrupted geometry is depicted by snapshot structure). In Kt-C/D box simulation, the A-0 transformed temporarily into A-I, which is depicted by snapshot structure, and then it was disrupted. Water insertion into the 2b•2n A•G base pair monitored as in Figure 4.

The third specific  $-1n(O2') \cdots 2b(N1)$  H-bond did not develop. This imperfect A-I was eventually disrupted. This confirms that a fully paired second sheared A•G base pair and complete A-I interaction are not mutually compatible within the specific context of folded kink-turns. This interplay modulates the balance between A-I and A-0 interactions in kink-turns (See below).

*Extra Bulge Base May Stabilize A-0. H.m. Kt-15 differs more visibly from the consensus kink-turns (Figure 2). There is extra adenine at the L<sub>n</sub>1 position (cf. Figures 1a and 2) which forms additional contacts to the adjacent uracil-based A-0 triple (Figure S2). In the first simulation, we made no changes in the starting crystal structure. The base pairing interactions and the initial A-0 were stable. In the second simulation, we pulled the ALn1 out of the bulge segment to disrupt its additional contacts to A-0, particularly, we disrupted the U2b•ALn1 cWW pair. After 5 ns, the A-0 interaction completely deteriorated. ALn1, however, quickly returned to its original orientation observed in the crystal structure and formed new triad with the 3n•3b C=G pair.*

This new interaction interconnected the bulge segment with the NC-stem, which probably limited the movement of bases within the kink-turn. We observed no changes until the end of the simulation; that is, the A-0 stayed disrupted. We nevertheless assume that the structure could fully repair on a longer time scale that is not affordable by contemporary simulations. Nevertheless, based on these simulations we suggest that additional “non-signature” nucleotides may tune the preference for A-0 and fine properties of kink-turns.

*A-I Unambiguously Prefers the Receptor G=C Pair.* Bioinformatics analyses of kink-turns with A-I reveal that G=C pair is superior receptor to other base pair combinations, whereas the isosteric C=G pair is almost not observed.<sup>2</sup> In contrast, isostericity matrices of the ribosomal kink-turns with A-0 show occurrence of both G=C and C=G receptor pairs<sup>2</sup> (see also Figure S1). We have carried out two simulations of H.m. Kt-38, where the original receptor G=C pair was inverted to C=G, which is not populated during evolution. In the first simulation, A-I was disrupted after 56 ns. No other changes were seen. In the

second simulation, A-I was disrupted after 230 ns. The interacting A2b slightly shifted toward the C-1n and at 240 ns the kink-turn formed the A-0/C=G pattern. This interaction was stable until the end of the simulation. These simulations further confirm that the inverted receptor pair is not compatible with A-I but may support A-0 interaction.

*Testing Different Ion Conditions.* In experiments, RNA structures are sensitive to ion condition. Thus, it is sometimes presumed that the RNA simulations might be sensitive to the ion treatment in a similar manner. Still, most NA simulations published so far were done with the net-neutralizing set of monovalent cations, usually  $\text{Na}^+$ . We are not aware of any report proving deficiency of this simulation condition or showing markedly different RNA structural dynamics caused by net-neutral and excess salt conditions. Experimentally detected ion binding sites are usually well reproduced; cf., e.g. the theoretically predicted<sup>74</sup> and subsequently experimentally confirmed<sup>75</sup> monovalent ion binding site in catalytic pocket of Hepatitis Delta Virus ribozyme. Recent systematic test of different ion treatments for RNA reverse-kink turns does not indicate any detectable effect of the ion treatment on the RNA dynamics, at least not on the present time scale and with the present simulation protocols.<sup>58</sup> The dynamics is clearly dominated by the RNA force field and modestly modulated by the water model. Perhaps, simulations do not need to be too strictly compared with experiments regarding the ion conditions, due to the variety of other and more significant differences between the experimental and simulation setups. Besides the genuine force field limitations (the force field ions are definitely not the same as the real ions), the present simulations are done with small solvent boxes and using periodic boundary conditions. Large boxes are not affordable since they would slow down the simulations and limit the solute dynamics sampling. It has several implications. With small box the ion distribution at the box border does not reach bulk properties.<sup>76</sup> Net-neutralization leads to cation concentrations of  $\sim 0.2 - 0.3 \text{ M}$  (see the Methods section) while adding excess salt further increases the cation concentration. In fact, also the formal solute concentration is high in simulations ( $\sim 0.01 \text{ M}$ ), but solute–solute interactions are prevented by the periodicity. The simulations are short and mostly stay within one conformational substate, so they are likely less sensitive to ion treatment than RNA folding experiments. We in no case claim that the description of ion conditions in contemporary simulations is perfect but we suggest that varying ion conditions does not substantially affect the RNA dynamics in these simulations.<sup>58</sup>

We carried out eleven simulations with different ion treatments. First we conducted six control simulations in KCl with  $\sim 0.2$  M excess salt concentration (Table 2). Similarly to the equivalent net-neutral  $\text{Na}^+$  simulations, the A-I/G=C interaction was stable in simulations of *H.m.* Kt-38 and *E.c.* Kt-23. The A-0/G=C in *E.c.* Kt-46 quickly transformed to A-I/G=C. The A-0/C=G in *H.m.* Kt-46 disrupted after 25 ns; however, the A-I interaction was not formed due to formation of G-1b(N2)  $\cdots$  A2b(O2') H-bond. Exactly the same behavior has been seen in  $\text{Na}^+$  simulations. In the simulation of Kt-C/D box, the A-0/C=G transformed to A-I/C=G after 60 ns and stayed stable until the end of the simulation. A-0/G=C in *H.m.* Kt-15, which contains an extra bulge base (see Figure 2), was unchanged in the course of the 100 ns simulation. In summary, the KCl simulations provided a close to identical picture as the  $\text{Na}^+$  simulations.

Then we ran one 100 ns simulation of Kt-u4 (chain C) with 6  $Mg^{2+}$  ions. The A-0  $\rightarrow$  A-I transition, which was seen in  $Na^+$

simulations was not achieved probably due to the short time scale. In the equivalent chain C Kt-u4 Na<sup>+</sup> simulation, it actually occurred at 150 ns (see above). The simulation revealed an identical behavior of other local interactions, such as fluctuations between direct and water-mediated interaction of the signature bulge: NC-stem L1(O2') ··· 1n(N1) H-bond as the Na<sup>+</sup> simulations (Figure S4). We did not see any specific Mg<sup>2+</sup> binding sites with high occupancy. Due to the generally lower reliability of the force field description of Mg<sup>2+</sup> cations and their poor sampling,<sup>28,77</sup> we decided to not make further such simulations. Note that 6 Mg<sup>2+</sup> ions mean a very high ~0.04 M Mg<sup>2+</sup> concentration, which is far from the experimental μM concentrations.<sup>21</sup> In addition, no binding of a structural (specifically bound) divalent ion to kink-turns has been revealed by experiments so far. Thus, although kink-turn folding is Mg<sup>2+</sup> dependent, kink-turns do not appear to be associated with some clear structural Mg<sup>2+</sup> binding sites, which would have to be included in simulations.

Then, we carried out four 125 ns simulations with just one  $Mg^{2+}$  (Table 1) localized initially near the kink-turn electrostatic potential minimum. All simulations were initiated with the A-0 arrangements (testing both orientations of the GC base pairs), however, due to the observed transitions we tested all four A-minor triples. The largest difference with respect to the  $Na^+$  and KCl simulations was seen for the *H.m.* Kt-46 (initially with A-0/C=G) which revealed spontaneous A-0 $\rightarrow$ A-I transition after 45 ns. This has not been seen in the equivalent  $Na^+$  and KCl simulations, where full formation of the A-I interaction was precluded by G-1b(N2) $\cdots$ A2b(O2') contact which prevented full shift of the A2b along the receptor pair toward the G-1b, although the direction of the movement was similar in all simulations. The newly formed A-I/C=G arrangement was stable until the end of the 125 ns simulation. This, however, can still be explained within the limits of conformational sampling, since also in monovalent ion simulations we have seen a case when the inverted A-I/C=G triple was disrupted after 230 ns. The  $Mg^{2+}$  ion formed several consecutive outer-shell contacts with RNA during the simulation.

Simulation of *E.c.* Kt-46 (with A-0/G=C) revealed spontaneous A-0 → A-I transition during the first two nanoseconds in agreement with the equivalent Na<sup>+</sup> and KCl simulations. In this simulation we could not prevent Mg<sup>2+</sup> inner-shell contact to phosphate group of the fourth base pair in the NC stem during the equilibration phase. Its position did not change in the course of the simulation. Although it most likely is a force field artifact we let the simulation go as we do not know the correct Mg<sup>2+</sup> binding pattern.

The A-0/C=G arrangement in the Kt-C/D box simulation fluctuated. In particular, A-0 was stabilized mainly by fluctuating A2b(O2') $\cdots$ C-1n(O2') H-bond. The second A2b(N3) $\cdots$ C-1n(O2') H-bond was observed for  $\sim$ 50% of the simulation time. This behavior matches one of the corresponding Na $^+$  simulations. The Mg $^{2+}$  ion formed several outer-shell contacts with RNA during the simulation.

In the simulation of Kt-u4 (with A-0/G=C) the A2b-(O2') $\cdots$ G-1n(O2') H-bond stabilizing A-0 disrupted after 15 ns and the structure moved in the same direction as in the Na<sup>+</sup> simulations. The A2b, however, did not complete the move along the receptor pair toward the C-1b, because it was stabilized by G-1n(N2) $\cdots$ A2b(O2') H-bond, which hampered formation of the A-I interaction. (Similar behavior has been seen in the Na<sup>+</sup> simulations of *H.m.* Kt-46.) Note that also in one of the Kt-u4

**Table 3.** Binding Energy of Adenine Nucleoside in Various A-Minor Interactions (kcal/mol) Computed at the RIMP2/aug-cc-pVDZ Level of Theory<sup>a</sup>

interaction	interaction energies <sup>b</sup>		
	$\Delta E^{\text{HF}}$	$\Delta E^{\text{corr}}$	$\Delta E$
A-I/C=G	-14.0	-19.0	-33.0
A-I/G=C	-13.1	-15.6	-28.7
A-0/C=G	-10.2	-11.3	-21.5
A-0/G=C	-6.7	-12.7	-19.4

<sup>a</sup> The interaction energy is calculated between the adenine nucleoside and the canonical base pair. <sup>b</sup>  $\Delta E^{\text{HF}}$ : HF-component,  $\Delta E^{\text{corr}}$ : correlation component,  $\Delta E$ : interaction energy at MP2 level.

$\text{Na}^+$  simulations the transition to A-I was completed after 150 ns. The  $\text{Mg}^{2+}$  ion formed stable inner-shell contact to phosphate group of the third base pair in the NC stem.

In summary, the  $\text{Mg}^{2+}$  simulations are mostly consistent with the  $\text{Na}^+$  and KCl simulations. Some modest differences could be reasonably explained by the sampling. We reiterate that the  $\text{Mg}^{2+}$  simulations face considerable force field and sampling uncertainties.

**Stability of the A-Minor Interactions Estimated by QM Calculations.** *Intrinsic Binding Strength of Adenine in Different A-Minor Patterns.* Table 3 summarizes the binding energy of adenine in the four A-minor models (see Figure 3) along with its HF and correlation components. The largest binding strength is found for A-I/C=G, which is followed by A-I/G=C and A-0/C=G. The least stable binding mode is found for A-0/G=C. At first sight, this finding contradicts the picture emerging from MD, where the intrinsically most stable A-I/C=G interaction either converts to A-0/C=G or breaks down in the course of the simulations (see above). In contrast, A-I/G=C is stable in a majority of the simulations. However, the QM calculations show that A-I/G=C and A-I/C=G exhibit a different (inverted) balance of the electrostatic and dispersion (or HF and electron correlation) interaction energy components (Table 4). We suggest that it influences the dynamical behavior of the kink-turn and rationalizes the simulation and bioinformatics data.

*tSS Base Pair Is Likely Critically Important for the Kink-Turn Structure.* In order to explain the behavior, we need to assume that the tSS contact is the more important for the kink-turn while the cSS contact is only auxiliary. There are two pieces of evidence that justify the assumption.

First, the kink-turns are quite stable in simulations when having the A-0 arrangement lacking the cSS interaction entirely. Further, we have shown in our earlier studies (in agreement with static experimental structures) that there is frequent water insertion into the cSS A•C base pair of the A-I/G=C triad,<sup>11,42</sup> see also Figure S6 in SI. Thus, the cSS interaction is dispensable.

Second, the overall kink-turn topology interplays with the conformational space of the A-minor interaction in an asymmetrical manner. While the overall kink-turn architecture tolerates shifts from A-I toward A-0, the opposite shifts (from the A-I arrangement toward A-minor II) are in conflict with the overall kink-turn structure. The experimental structures show no example of a shift in the direction toward A-II, whereas such motions are not sampled in the simulations. Indirect evidence that kink-turns face energy penalty in case of fluctuations in the A-II direction is the fact that even the fully paired A-I interaction can

**Table 4.** RIMP2/aug-cc-pVDZ Interaction Energies of the Component SS Contacts of the Two A-Minor I Interactions (kcal/mol), Decomposition of the Data from Table 3 into the Individual Pairwise Terms

interaction	component	interaction energies <sup>a</sup>		
		$\Delta E^{\text{HF}}$	$\Delta E^{\text{corr}}$	$\Delta E$
A-I/G=C	tSS G•A	-3.5	-10.3	-13.8
	cSS A•C	-11.4	-5.0	-16.4
A-I/C=G	tSS C•A	-5.4	-6.4	-11.8
	cSS A•G	-9.1	-12.7	-21.8

<sup>a</sup>  $\Delta E^{\text{HF}}$ : HF-component,  $\Delta E^{\text{corr}}$ : correlation component,  $\Delta E$ : interaction energy at MP2 level.

be formed only at the expense of perturbation of the NC-stem, i.e., the partial opening of its second sheared A•G base pair. Further shift toward the A-II would likely lead to further perturbation and eventual disruption of the A•G base pair to prevent steric clash between the NC and C stems.

*A-I/G=C and A-I/C=G Have an Inverted Balance of Electrostatic and Dispersion Stabilization Patterns.* The primary tSS G•A contact in the A-I/G=C triad is strongly dominated by the correlation component of the interaction energy (-10.3 kcal/mol; Table 4). On the contrary, the corresponding tSS C•A interaction in A-I/C=G has a significantly lower correlation component (-6.4 kcal/mol; see Table 4). That means that the tSS G•A interaction is shifted toward dispersion stabilization while the tSS C•A base pair is shifted toward electrostatic stabilization. Since the electrostatic stabilization competes with solvation energy contributions, the tSS G•A base pair is certainly more hydrophobic than the tSS C•A one and thus more suitable to stabilize tertiary contacts. In other words, the tSS interaction of the A-I/GC triad is much more potent stabilizer of the structure than the equivalent tSS interaction in the A-I/C=G triad.

Further (Table 4), the auxiliary cSS A•C contact of A-I/G=C is primarily electrostatic (hydrophilic), while the corresponding cSS A•G in A-I/C=G is controlled by the correlation component. Thus, the A-I/G=C and A-I/C=G interactions have inverted balance of dispersion and electrostatic contributions. Considering the balance, it is not surprising that the A-I/G=C triad often adopts the cSS water-assisted substates while keeping the tSS interaction intact.<sup>11,42</sup>

We also carried out electrostatic potential calculations of the whole kink-turn using CMIP program,<sup>78</sup> which solves the non-linear Poisson–Boltzmann equation. Probably due to the high concentration of the highly electronegative endocyclic N-sites, there is a profound negative potential minimum right above the cSS contact (Figure S5). This could support hydration in this region and is probably interplaying with the intrinsic molecular interactions analyzed above. We hypothesize that these profound electrostatic potential minimum may further reduce the capability of the cSS interaction to stabilize kink-turn topologies, although we do not have a suitable tool to quantify its role.

Regarding the A-0 interactions, they are weaker in the gas phase than the A-I arrangements. However, both A-0 arrangements have rather large electron correlation attraction indicating substantial dispersion energy stabilization, whereas basically all of the binding comes from the tSS contact. Further, as explained above, the A-0 interaction is more compatible with the overall kink-turn topology than A-I and thus it is more easily

accommodated in the overall kink-turn context. This reduces the genuine energy difference between the A-I and A-0 triads in the context of a kink-turn. In other words, although locally the A-minor interaction prefers to stay in the A-I arrangement, globally the overall kink-turn topology supports movements toward the A-0.

## CONCLUSIONS

We have investigated two A-minor interaction types occurring in kink-turns using MD simulations and QM calculations. The A-I interaction is primarily formed between adenine of the NC-stem and guanine from the C-stem via a tSS G•A interaction which is supplemented by cSS A•C interaction (Figure 1). The A-0 interaction consists of a deformed tSS-like G•A base pair. (The tSS pairs in A-0 patterns differ from the optimal tSS pair described by the standard classification<sup>12</sup> which can be found in A-I triads, cf. Figure 1.)

The most stable arrangement in simulations of kink-turns is the A-I interaction involving the G=C base pair as receptor (Figure 4). The inverted A-I/C=G interaction is unstable in simulations, not seen in experimental structures, and not supported by bioinformatics data. The A-0/G=C interaction tends to transform to A-I in simulations of isolated folded kink-turns (Figure 5). The sequence data for Kt-46, Kt-15, and Kt-78 show that, once the kink-turn, for whatever reason, prefers the A-0 interaction, G=C and C=G receptor pairs covary, in contrast to A-I kink-turns. Simulations with A-0/C=G often attempt to make A-0 → A-I transitions which, however, are not successful as the A-I/C=G pattern is disrupted in simulations. Taking all data together we conclude that among the four studied A-minor interaction types the A-I/G=C variant is the genuine one in kink-turns.

We suggest that the occurrence of the A-0 interaction may be induced by the context (the surrounding molecules) of the kink-turn, i.e., the bound proteins or other RNAs which may be specifically adapted to recognize and stabilize the A-0 pattern and kink-turn shape. We assume that Mg<sup>2+</sup> ions do not influence the balance between A-I and A-0 interactions, since it is not obvious how Mg<sup>2+</sup> could interplay with the subtly different shapes of A-I and A-0 kink-turns. Thus we suggest that, in solution Mg<sup>2+</sup> experiments done with isolated kink-turns, the kink-turns would tend to fold into A-I arrangement with the G=C pair and into the A-0 arrangement with the C=G receptor pair, irrespective of whether they possess A-I or A-0 in their functional contexts. It is likely that the G=C to C=G inversion would also weaken stability of the folded structure to a certain extent.

A stable A-0 binding pattern with the receptor G=C pair is seen in the *H.m.* Kt-15, where it is stabilized by extra bulge base (in the shorter strand) not present in other kink-turns (Figures 2 and S2). Therefore, the balance of forces can be shifted in favor of A-0 interstem contact also by modulation of the kink-turn sequence outside the kink-turn consensus sequence.

The occurrence of different variants of the A-minor interactions can be rationalized by QM calculations capturing the intrinsic interactions in the A-minor triples (Tables 3 and 4). By intrinsic stability we mean gas phase interaction energy which shows stability of the interaction per se, in absence of solvent and context effects.<sup>49</sup> The A-I arrangement is intrinsically more stable than the A-0 arrangement. The A-I pattern allows the tSS interaction to be maximized and to form the cSS interaction. The A-I/G=C and A-I/C=G triples possess similar overall interaction energies. However, there is a substantial difference

in the balance of interactions in the A-I/G=C and A-I/C=G triples. The key tSS G•A interaction of the A-I/G=C triple shows large electron correlation stabilization, which means a substantial London dispersion energy component. Such interactions are suitable to form tertiary contacts as they are less likely to be disrupted by polar water molecules. In contrast, the tSS C•A pair of the A-I/C=G arrangement is shifted more toward electrostatic binding, which increases capability of polar solvent to disrupt the structure, as seen in simulations. The A-0 interactions, although intrinsically weaker, show also rather large electron correlation stabilization, so they are suitable to form tertiary interaction.

The overall structure of kink-turn also specifically affects the balance of the A-I and A-0 interactions. There is structural struggle between the second sheared A•G pair in the NC-stem and the A-I interaction. In order to complete the A-I interaction, the A•G base pair is perturbed, having a water-bridge instead of the direct A(N6)···G(N3) H-bond. The A•G pair is completed when the A-I interaction is not formed (Figures 4–6). This structural struggle between two signature interactions which cannot be simultaneously optimized in the folded kink-turn contributes to flexibility of kink-turns and reduces the energy difference between the A-I and A-0 triples.

The present results show that magnitude and balance of stabilizing forces of molecular interactions, which is not considered in structural bioinformatics, may modulate the basic structural considerations.<sup>49</sup> The interaction energies cannot be straightforwardly used to predict free energies of RNA molecules.<sup>49</sup> However, for isosteric and near-isosteric interactions occurring in similar contexts, the interaction energy calculations provide valuable insights.<sup>49</sup>

In summary, the computations suggest the following picture. The most typical arrangement in kink-turns is the A-I interaction with G=C receptor pair. Inversion of the receptor pair is destabilizing. Some kink-turns, however, may adopt A-0 interaction, provided it gets some specific stabilization either due to the surrounding molecules or due to additional (nonconsensus) sequence features of the kink-turn. Once a given kink-turn prefers or tolerates the A-0 interaction, then both orientations of the receptor pair are possible. The stability of different variants of the A-minor interactions stems from the balance of stabilizing forces of molecular interactions that modulate the basic isostericity. We suggest the performance of solution experiments for close-to-consensus kink-turns with the C=G receptor pair. Our results indicate that inversion of the receptor pair could measurably destabilize the kink-turn.

## ASSOCIATED CONTENT

**S Supporting Information.** Figures S1–S7; listed restraints used in our simulations; and Cartesian coordinates of optimized geometries of the four A-minor patterns considered in this study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

(K.R.) Tel.: +420 541517250. Fax: +420 541212179. E-mail: kristina@physics.muni.cz. (J.Š.) Tel.: +420 541517133. Fax: +420 541212179. E-mail: sponer@ncbr.chemi.muni.cz.

## ■ ACKNOWLEDGMENT

This work was supported by the Grant Agency of the Academy of Sciences of the Czech Republic (CR) (KJB400040901 and IAA400040802), by the Grant Agency of CR (203/09/H046, 203/09/1476 and P208/10/2302), and by Ministry of Education of CR (LC06030), Academy of Sciences of CR (AV0Z50040507 and AV0Z50040702), and “CEITEC - Central European Institute of Technology” (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund.

## ■ REFERENCES

- (1) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 4214.
- (2) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. *Nucleic Acids Res.* **2005**, *33*, 2395.
- (3) Chao, J. A.; Williamson, J. R. *Structure* **2004**, *12*, 1165.
- (4) Montange, R. K.; Batey, R. T. *Nature* **2006**, *441*, 1172.
- (5) Moore, T.; Zhang, Y. M.; Fenley, M. O.; Li, H. *Structure* **2004**, *12*, 807.
- (6) Schroeder, K. T.; McPhee, S. A.; Ouellet, J.; Lilley, D. M. J. *RNA* **2010**, *16*, 1463.
- (7) Vidovic, I.; Nottrott, S.; Hartmuth, K.; Luhrmann, R.; Ficner, R. *Mol. Cell* **2000**, *6*, 1331.
- (8) Woolstenhulme, C. J.; Hill, W. E. *J. Mol. Biol.* **2009**, *392*, 645.
- (9) Reblova, K.; Razga, F.; Li, W.; Gao, H.; Frank, J.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 1325.
- (10) Besseova, I.; Reblova, K.; Leontis, N. B.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 6247.
- (11) Razga, F.; Koca, J.; Sponer, J.; Leontis, N. B. *Biophys. J.* **2005**, *88*, 3466.
- (12) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497.
- (13) Liu, J.; Lilley, D. M. J. *RNA* **2007**, *13*, 200.
- (14) Szewczak, L. B. W.; Gabrielsen, J. S.; DeGregorio, S. J.; Strobel, S. A.; Steitz, J. A. *RNA* **2005**, *11*, 1407.
- (15) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899.
- (16) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K. H.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678.
- (17) Doherty, E. A.; Batey, R. T.; Masquida, B.; Doudna, J. A. *Nat. Struct. Biol.* **2001**, *8*, 339.
- (18) Xin, Y. R.; Laing, C.; Leontis, N. B.; Schlick, T. *RNA* **2008**, *14*, 2465.
- (19) Goody, T. A.; Melcher, S. E.; Norman, D. G.; Lilley, D. M. J. *RNA* **2004**, *10*, 254.
- (20) Matsumura, S.; Ikawa, Y.; Inoue, T. *Nucleic Acids Res.* **2003**, *31*, 5544.
- (21) Schroeder, K. T.; Lilley, D. M. J. *Nucleic Acids Res.* **2009**, *37*, 7281.
- (22) Turner, B.; Melcher, S. E.; Wilson, T. J.; Norman, D. G.; Lilley, D. M. J. *RNA* **2005**, *11*, 1192.
- (23) Falb, M.; Amata, I.; Gabel, F.; Simon, B.; Carlomagno, T. *Nucleic Acids Res.* **2010**, *38*, 6274.
- (24) Cheatham, T. E., III *Curr. Opin. Struct. Biol.* **2004**, *14*, 360.
- (25) Fulle, S.; Gohlke, H. *J. Mol. Recognit.* **2010**, *23*, 220.
- (26) Hall, K. B. *Curr. Opin. Struct. Biol.* **2008**, *12*, 612.
- (27) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers* **2007**, *85*, 169.
- (28) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. *Acc. Chem. Res.* **2010**, *43*, 40.
- (29) Yildirim, I.; Stern, H. A.; Sponer, J.; Spackova, N.; Turner, D. H. *J. Chem. Theory Comput.* **2009**, *5*, 2088.
- (30) Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114*, 8701.
- (31) Koller, A. N.; Bozilovic, J.; Engels, J. W.; Gohlke, H. *Nucleic Acids Res.* **2010**, *38*, 3133.
- (32) Whitford, P. C.; Geggier, P.; Altman, R. B.; Blanchard, S. C.; Onuchic, J. N.; Sanbonmatsu, K. Y. *RNA* **2010**, *16*, 1196.
- (33) Trabuco, L. G.; Schreiner, E.; Eargle, J.; Cornish, P.; Ha, T.; Luthy-Schulten, Z.; Schulten, K. *J. Mol. Biol.* **2010**, *402*, 741.
- (34) Huang, W.; Kim, J.; Jha, S.; Aboul-Ela, F. *Nucleic Acids Res.* **2009**, *37*, 6528.
- (35) Romanowska, J.; Setny, P.; Trylska, J. *J. Phys. Chem. B* **2008**, *112*, 15227.
- (36) Veeraraghavan, N.; Ganguly, A.; Chen, J. H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, *50*, 2672.
- (37) Villa, A.; Wohnert, J.; Stock, G. *Nucleic Acids Res.* **2009**, *37*, 4774.
- (38) Xin, Y.; Hamelberg, D. *RNA* **2010**, *16*, 2455.
- (39) Singh, A.; Sethaphong, L.; Yingling, Y. G. *Biophys. J.* **2011**, *101*, 727.
- (40) Razga, F.; Spackova, N.; Reblova, K.; Koca, J.; Leontis, N. B.; Sponer, J. *J. Biomol. Struct. Dyn.* **2004**, *22*, 183.
- (41) Cojocaru, V.; Klement, R.; Jovin, T. M. *Nucleic Acids Res.* **2005**, *33*, 3435.
- (42) Razga, F.; Zacharias, M.; Reblova, K.; Koca, J.; Sponer, J. *Structure* **2006**, *14*, 825.
- (43) Curuksu, J.; Sponer, J.; Zacharias, M. *Biophys. J.* **2009**, *97*, 2004.
- (44) Spackova, N.; Reblova, K.; Sponer, J. *J. Phys. Chem. B* **2010**, *114*, 10581.
- (45) Sponer, J. E.; Leszczynski, J.; Sychrovsky, V.; Sponer, J. *J. Phys. Chem. B* **2005**, *109*, 18680.
- (46) Sponer, J. E.; Reblova, K.; Mokdad, A.; Sychrovsky, V.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2007**, *111*, 9153.
- (47) Sharma, P.; Chawla, M.; Sharma, S.; Mitra, A. *RNA* **2010**, *16*, 942.
- (48) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 4898.
- (49) Sponer, J.; Sponer, J.; Petrov, A. I.; Leontis, N. B. *J. Phys. Chem. B* **2010**, *114*, 15723.
- (50) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W. et al. *AMBER 10*; University of California: San Francisco, CA, 2008.
- (51) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
- (52) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (53) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
- (54) Besseova, I.; Otyepka, M.; Reblova, K.; Sponer, J. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701.
- (55) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E., III; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836.
- (56) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886.
- (57) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021.
- (58) Sklenovsky, P.; Florova, P.; Banas, P.; Reblova, K.; Lankas, F.; Otyepka, M.; Sponer, J. *J. Chem. Theory Comput.* **2011**, *7*, 2963.
- (59) Noy, A.; Soteras, I.; Luque, F. J.; Orozco, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596.
- (60) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020.
- (61) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (62) Reblova, K.; Fadrna, E.; Sarzynska, J.; Kulinski, T.; Kulhanek, P.; Ennifar, E.; Koca, J.; Sponer, J. *Biophys. J.* **2007**, *93*, 3932.
- (63) Mladek, A.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Sponer, J.; Sponer, J. E. *J. Phys. Chem. B* **2009**, *113*, 1743.
- (64) Sponer, J. E.; Spackova, N.; Kulhanek, P.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. A* **2005**, *109*, 2292.
- (65) Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399.

- (66) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C. et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (67) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (68) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (69) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652.
- (70) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
- (71) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.
- (72) Jurecka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
- (73) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.
- (74) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. *Biophys. J.* **2006**, *91*, 626.
- (75) Ke, A.; Ding, F.; Batchelor, J. D.; Doudna, J. A. *Structure* **2007**, *15*, 281.
- (76) Chen, A. A.; Draper, D. E.; Pappu, R. V. *J. Mol. Biol.* **2009**, *390*, 805.
- (77) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669.
- (78) Gelpi, J. L.; Kalko, S. G.; Barril, X.; Cirera, J.; de la Cruz, X.; Luque, F. J.; Orozco, M. *Proteins* **2001**, *45*, 428.