Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C:G-A triple

Lukasz Popenda, Lukasz Bielecki, Zofia Gdaniec, and Ryszard W. Adamiak*

Laboratory of Structural Chemistry of Nucleic Acids, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland
E-mail: adamiakr@ibch.poznan.pl

Dedicated to Professor Harri Lönnberg on the occasion of his 60th birthday

Abstract

The RNA single bulge motif is an unpaired residue within a strand of several complementary base pairs. Structure and dynamics of RNA duplexes containing single adenosine (5′-GUCGAGCUG-3′/5′-CAGCCGAC-3′) and uridine bulge (5′-GUCGUGCUG-3′/5′-CAGCCGAC-3′) have been studied using NMR spectroscopy and *in aqua* simulation of molecular dynamics. In both NMR structures, the bulged residues are oriented toward the minor groove. The bulged adenosine participates in the base triple formation being coplanar with the 5′-neighboring residue G4 which pairs with cytidine C14 from the opposite strand. The uridine bulge adopts a looped-out conformation pointing toward the residue G15 and bridges the minor groove to 5′ side of the helix. NMR-derived structures have been used to formulate the initial RNA duplex models for the MD simulation. Results of molecular dynamics simulations in explicit solvent support the conclusion concerning the formation of a dinucleotide platform within the C:G-A triple. As suggested by NMR results, the uridine bulge shows considerable dynamics.

Keywords: RNA bulges, dinucleotide platform, base triple, NMR spectroscopy, molecular dynamics

Introduction

Bulges are among the most common non-Watson-Crick features in RNA. Single and multiple base bulges are important for the tertiary folding process and are known to be specific sites for RNA-RNA and RNA-protein recognition.[1-6] The RNA single bulge motif is an unpaired residue within a strand of several complementary base pairs. Determination of the preferred conformation of a single-base bulge may be important for the prediction and understanding of the three-dimensional structure of RNAs. The structure of DNA and RNA bulges has been

studied by X-ray[7-12] and NMR methods,[13-21] with the use of algorithms searching the conformational space for energetically favorable structures[22] and molecular simulations.[23,24] The X-ray crystallography and NMR spectroscopy are the principal methods for the evaluation of structural details of RNAs and their interactions. RNA structures determined with the use of X-ray crystallography and NMR show that single base bulges can adopt both stacked-in and looped-out conformations. In the high-resolution NMR studies, a bulged adenosine is generally found stacked into the helix while the X-ray structures, usually obtained under elevated salt conditions, reveal adenine residue mostly in the looped-out conformation. On the contrary, single uridine bulges show a preference for the looped-out structure in crystal and in solution.[12,13,15,25,26]

In the stacked-in conformation a bulged nucleotide residue can be positioned between the adjacent base pairs or can form a dinucleotide platform being arranged coplanar to the base, either to the 5′ or 3′ side of the bulge. In the looped-out conformation the bulged residue can be oriented towards one of the helix grooves or it can protrude away from the helix. Whether the bulged nucleobase adopts a looped-out or stacked-in conformation, depends on the nature of the bulged base itself, the identity of the neighboring residues, the sequence not adjacent to the bulge and the length of the helix. The conformational transitions in RNA single adenosine and uridine bulged duplexes have been investigated using the umbrella sampling method.[23] The calculations indicated that flipping out of bulge bases toward the minor groove appears to be more favorable than that toward the major groove. The latter led to significant structural changes including a disruption and opening of base pairs flanking the bulge.

In this work we have studied the conformation of single adenosine and uridine bulges located within the same sequence using the NMR spectroscopy and *in aqua* simulation of molecular dynamics. The data presented extend our former results concerning thermodynamics, spectrofluorimetry and dynamics[27,28] of the 2-aminopurine modified adenosine bulges. To our best knowledge this is the first report on the structure of RNA bulges studied by a combination of the methods of NMR spectroscopy and molecular dynamics, an approach currently used for other RNA structures.[29,30]

Results and Discussion

The sequences of both molecules, together with the numbering scheme used in this work, are shown in Figure 1.

Figure 1. The sequence and numbering scheme for the adenine (**I**) and uridine bulged duplexes (**II**).

NMR spectroscopy

Assignments of protons in the duplex I

Standard methods were employed to assign the nonexchangeable protons.[31-33] The sequential assignment of this duplex followed that of a right-handed RNA helix. The conventional sequential H8/H6-H1′ connectivities are continuous through all nucleotides in both strands in the 400 ms NOESY spectrum (Fig. 2).

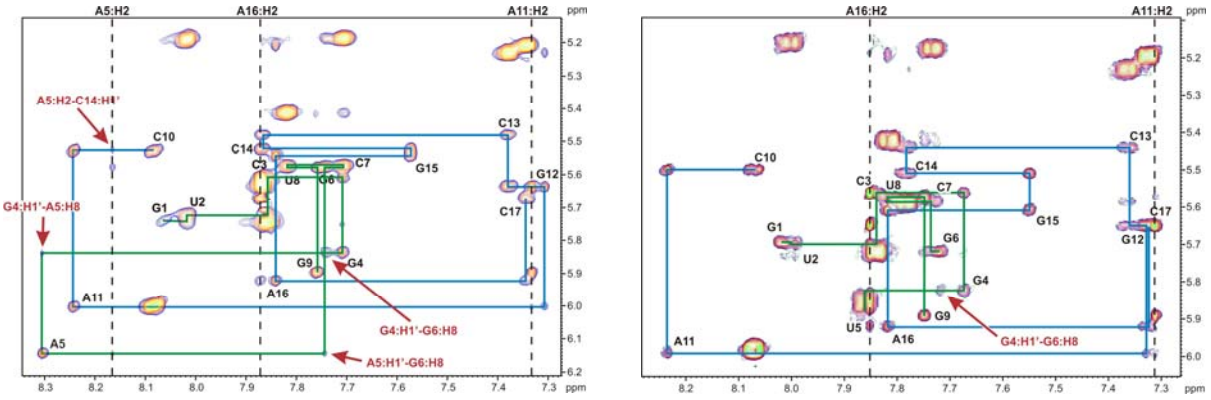

Figure 2. Aromatic H6/H8-H1′ regions of the 400 ms NOESY spectra. Sequential connectivities are shown for the top (*green line*) and bottom (*blue line*) strands for the adenosine (*left*) and uridine (*right*) bulge duplexes. The cross-peaks discussed in the text are marked with arrows.

However, the interresidue connectivities between G4, A5 and G6 in the bulge region are weak in the 400 ms and absent in the 150 ms spectrum. In the upper strand, sequential aromatic-aromatic cross-peaks appear for residues G1 through G9, including the bulged adenine A5 for which the weak NOEs from G4-H8 to A5-H8 and from A5-H8 to G6-H8 are observed. Sequential aromatic-aromatic cross-peaks are also observed for all the residues in the lower strand. The A5-H2 resonance exhibits NOEs suggesting a stacked-in conformation of that residue. Most significantly, except NOEs to G6-H1′, the H2 proton of A5 residue shows a weak intensity NOE with C14-H1′ on the opposite strand. Additionally, apart from the presence of uninterrupted sequential base-H1′ connectivities, cross-peaks of moderate intensity between G4-H2′ and A5-H8 as well as G6-H8 and A5-H2′ indicated at least partial stacking and partial intercalation of the adenosine residue. However, except for the NOE cross peaks characteristic of the stacked-in conformation of the bulged residue, a NOE between G6-H8 and G4-H1′ was observed implying that adenosine A5 was stacked, though in a different manner than that observed for A-RNA. This finding was further confirmed by the presence of a weak NOE between G6-H8 and G4-H2′ which would not be possible if A5 were sandwiched between neighboring guanosines G4 and G6.

Analysis of the DQF-COSY spectrum shows that most of the helix residues have N-puckered sugars, and that deviations occur in the bulge region of the duplex. A total of five H1′-H2′ distinct cross peaks are observed indicative of $^{3}J_{H1',H2'}$ couplings larger than 2 Hz. Moderate

signal intensities for A5, G6 and C17 indicate that the ribose residues at these positions can be described as a mixture of C3′-*endo* and C2′-*endo* conformations. The guanosine G4 flanking the bulge residue from 5′ side, as well as the 5′-terminal residue G9 exhibits large H1′ to H2′ coupling constants revealing that conformational equilibria of these ribose rings are shifted toward the C2′-*endo* conformation.

The imino proton spectrum of the bulged duplex **I** is well resolved, and its resonances were readily assigned using NOESY spectra recorded in 90% H_2O / 90% D_2O at 10 °C. The imino proton resonances were identified for all expected Watson-Crick base pairs except the terminal GC base pairs. The imino proton resonances for G4 and G6 residues were unambiguously assigned. The presence of a characteristic pattern of cross-peaks between these imino protons and their respective cytosine partners provides evidence that G4-C14 and G6-C13 base pairs flanking the bulge are classical Watson-Crick base-pairs. Additionally, sequential NOE was observed between the imino protons of G4 and G6. This observation strongly suggested that the bulged residue A5 cannot be fully stacked-in between its flanking residues, but rather rotated away from the regular stacking, a conclusion reinforced by the NOE between G6-H8 and G4-H1′.

Most ^{31}P resonances appear in a narrow region between -3.8 and -4.5 ppm. Only the ^{31}P signal of A5 at -3.4 ppm is shifted downfield from other resonances and falls just outside the regular helix range providing further evidence of the unusual conformation of the bulge region.

Assignments of protons in the duplex II

Comparison of chemical shifts of both duplexes, reveal that most resonances of all residues, excluding A5 (**I**) and U5 (**II**), differ only slightly (BMRB, codes *15780* and *15781*). The largest chemical shift difference, not exceeding 0.2 ppm, was observed for the residues adjacent to the bulge site i.e. for G6-H1′, G6-H2′ and G4-H4′.The same trend was observed for ^{13}C chemical shifts. For guanine rings of the G4 and G6 the difference was smaller than 0.3 ppm. This result suggested that the presence of the bulged residue had no great impact on the structure of helical stems flanking the bulge.

The aromatic-anomeric part of the NOESY spectrum of the duplex **II** is shown in Fig. 2. A continuous set of sequential H8/H6-H1′ NOE connectivities can be traced for the lower strand, although a break occurs between the bulged uridine U5 and the G6 step. Two interresidue cross-peaks were observed from uridine U5. The U5-H6 showed NOEs only to G4-H2′ and G4-H3′ protons. The absence of stacking interactions between the residue U5 and its 3′-neighboring G6 residue, along with the presence of a NOE between G6-H8 and G4-H1′ implied that in duplex **II** the bulged residue U5 was completely bulged-out.

In the H1′-H2′ region of the DQF-COSY spectrum, only G4 has strong cross-peaks corresponding to $^3J_{H1',H2'} = 6 - 8$ Hz. Residues G6 and U5 exhibit only weak H1′-H2′ COSY peaks. These data indicate that the ribose ring of G4 adopts predominantly the C2′-*endo* conformation. G6 and U5 ribose residues probably exist in equilibrium between C3′-*endo* and C2′-*endo* conformations. The absence of COSY H1′-H2′ cross-peaks for the rest of the residues except the terminal ones suggests that their riboses are predominantly C3′-*endo*.

A continuous set of the sequential imino-imino connectivities, including G4-C13 and G6-C13 base pairs, could be seen in the NOESY spectrum recorded in 90% H_2O / 90% D_2O at 10 °C. The imino resonance corresponding to U5-H3 was not observed in any of the exchangeable proton spectra due to the rapid exchange with bulk water. The sequential NOE connectivities observed for imino proton resonances of G15, G4, G6 and G12 indicated that G15-C3, G4-C14, G6-C13 and G12-C7 formed continuous base pairs surrounding the bulge region.

All ^{31}P shifts fall into the range from -3.7 to -4.5 ppm, typical of helices, including the ^{31}P signal of U5 at -3.7 ppm.

NMR structures of the adenosine (I) and uridine (II) bulged duplexes

Table 1 summarizes the statistics on the distance and dihedral constraints used in the structure calculations of the two bulged duplexes. Both structures were calculated using the restrained molecular dynamics protocol described in the Experimental Section.

Table 1. NMR structure determination statistics[a]

	duplex **I**	duplex **II**
distance restraints	288	274
intranucleotide NOEs	180	171
internucleotide NOEs	108	103
negative NOE	8	17
hydrogen bonds	50	48
planarity restraints	10	8
dihedral angle restraints	171	165
backbone	69	69
ribose pucker	85	80
glycosidic	17	16
NOEs per residue	16.94	16.12
NOEs and dihedrals per residue	27.00	25.82
average r.m.s.d. from ideal covalent geometry		
bond length (Å)	0.0021 (0.0001)	0.0019 (0.0001)
angle (°)	0.68 (0.01)	0.67 (0.01)
impropers (°)	0.37 (0.03)	0.38 (0.03)
number of NOE violations (>0.3 Å)	0	0
number of dihedral violations (>5°)	0	0
avg. non-hydrogen atom pairwise r.m.s.d.	0.71 (0.17)	0.68 (0.19)
avg. non-hydrogen atom pairwise r.m.s.d. for the bulge region (residues 4, 5, 6 and 13, 14)	0.39 (0.11)	0.82 (0.29)

[a]Standard deviations from the family of structures calculated are given in parentheses

The calculations used a total of 288 distance constraints and 171 dihedral angle constraints for the duplex containing bulged adenosine (**I**) and 274 distance constraints and 165 dihedral angle constraints for the duplex with bulged uridine (**II**). Out of 50 starting structures with completely random backbone dihedral angles, 38 and 36 structures converged to low NOE and dihedral angle violation energies for duplexes **I** and **II**, respectively. The structures were classified as converged if they were consistent with the NMR data and maintained correct stereochemistry. All converged structures had no NOE constraints violated by more than 0.3 Å and no dihedral angle violations larger than 5°. The superposition of ten lowest-energy structures of both molecules is shown in Fig. 3. The average pairwise r.m.s.d. calculated for all heavy atoms for these ten structures is 0.71 (±0.17) Å and 0.68 (±0.19) Å, for duplex **I** and duplex **II**, respectively.

The structure of duplex **I** comprises two A-type helical Watson-Crick segments as evidenced by characteristic NOEs cross peaks between G4 imino proton at 12.52 ppm and amino and H5 protons of C14 typical of canonical Watson-Crick interactions.

The location of the bulged nucleotide was determined by several NOEs including A5-H2 NOE to C14-H1′ and G6-H1′, A5-H8 to G4-H2′ and G4-H3′ as well as NOE between A5-H2′ and G6-H8. In all calculated structures, G4 and A5 residues form a dinucleotide platform being part of the C14:G4-A5 base triple. The position of adenosine A5 in the minor groove is probably stabilized by the hydrogen bonds with G4 and C14 on their sugar edges: N7(A5)…H22(G4), H61(A5)…O2(C14) and H62(A5)...O2′(C14).

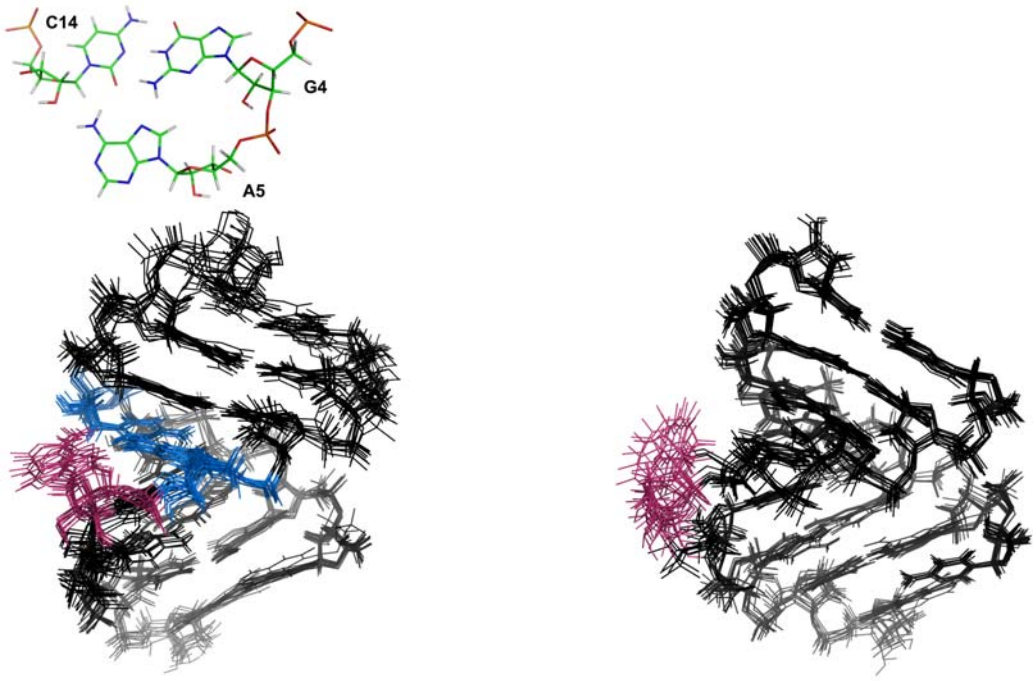

Figure 3. Superimposition of ten lowest energy structures of adenosine (*left*) and uridine (*right*) bulged duplexes. Bulged residues A5 and U5 are shown in magenta. For the duplex **I**, the C14-G4 base pair is colored blue and the insert shows the C14:G4-A5 base triple in detail.

However, at this point, we could not unambiguously confirm the presence of those hydrogen bonds. Although base stacking interaction with guanosine G6 is likely to contribute to the stability of the base triple, as revealed by a downfield shift of the A5-H2 and A5-H8 resonances with increasing temperature, hydrogen bonding interactions must play an additional significant role.

The helical stems of duplex **II** with the uridine bulge form well-defined A-type structures. The two helical segments stack one on the other with U5 positioned outside the helix. There were much fewer NOEs available to determine the conformation of the bulged residue U5 than for A5 in the duplex **I**. The key NOEs observed for uridine U5 include the interactions of U5-H6 with G4-H3′ and G4-H2′. Although the analysis of the final ensemble of duplex **II** structures reveals a dynamic behavior of U5, in all structures the bulged uridine is looped-out and points towards the residue G15, thus bridging the minor groove to the 5′ side of the helix (Fig. 3).

The NMR calculations resulted in two families of structures with a characteristic pattern of α/γ backbone torsion angles at the bulge site. In one family, consisting of six structures, the α torsion angle ($111 \pm 2°$) of guanosine G6 corresponds with the γ value of $-128 \pm 2°$. In the second family of four structures, the values of $\alpha = 29 \pm 11°$ of guanosine G6 correlate with $\gamma = 16 \pm 7°$. The ribose ring of G4 adopts a C2′-*endo* conformation. For some conformers, a shift of the sugar pucker at the bulge nucleotide towards C2′-*endo* was observed, which was in agreement with the NMR observation that the sugar of U5 existed in equilibrium between the C3′-*endo* and C2′-*endo* conformations.

Molecular dynamics

In order to gain more insight into the dynamics of adenosine and uridine bulges in RNA duplexes, the molecular dynamics simulations were performed. For both duplexes, the structure coordinates delivered by the NMR data were used to build the initial models for simulation, but without any experimental restraints. The trajectories acquired from the molecular dynamics study (Fig. 4) show that the conformation of the model with bulged adenosine (trajectories labeled A1 and A2) is more stable than that of the model with bulged uridine (trajectory U).

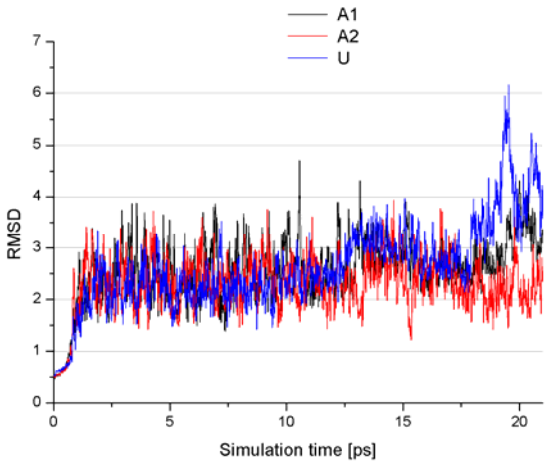

Figure 4. Molecular dynamics simulation for trajectories A1, A2 and U. The pairwise r.m.s.d. of heavy atoms positions referred to the respective initial structure obtained from NMR experiment.

In A1 trajectory, the dinucleotide platform in the C:G-A triple is retained for about 14 ns but then the nucleobases G4 and A5 switch to a stacked alignment which remains stable up to the end of the trajectory (Fig. 4). In A1 trajectory, the terminal G9-C10 base pair was broken at the end of the simulation (not presented).

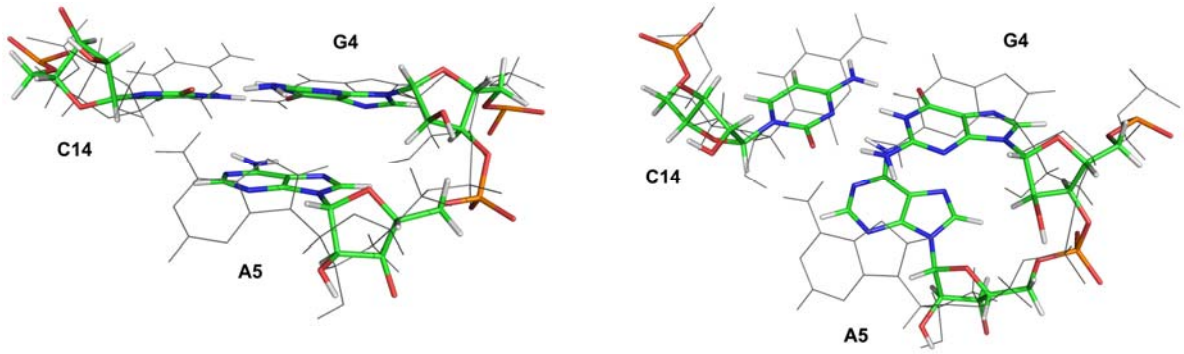

Figure 5. The final frame of trajectory A1. Two views showing the alignment of the C4, A5 and G14 residues (*color*) with respect to the NMR-derived structure (*gray)*.

On the other hand, the platform is stable over the entire A2 trajectory which correlates with the generally lower r.m.s.d. values for that simulation. The presence of a local network of strong hydrogen bonds, anticipated by NMR spectroscopy, stabilizes the three nucleobases which form the platform (Fig. 6). In the same time, the stacking interactions between G4, A5, C14 and their neighboring residues, keep the planar alignment of the platform.

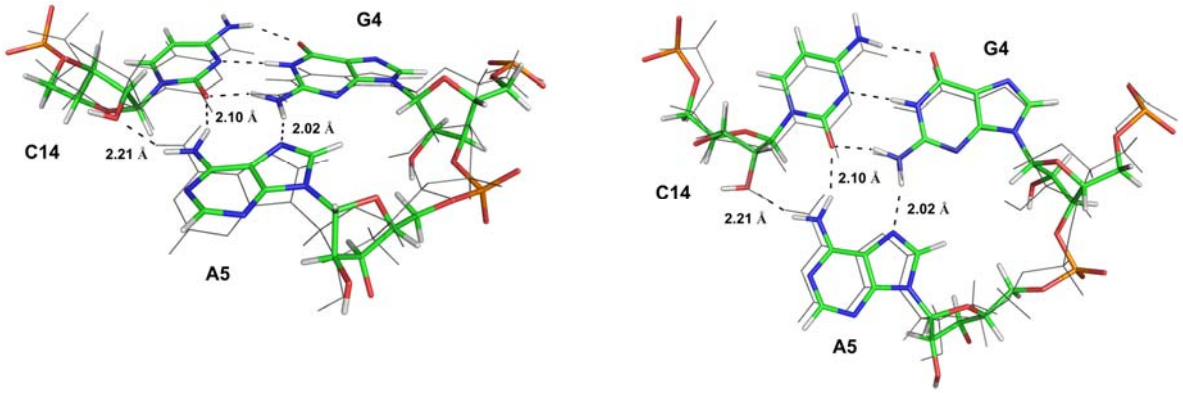

Figure 6. The final frame of trajectory A2. Two views showing the alignment of the dinucleotide platform in the C:G-A triple (*color*) with respect to the NMR-derived structure (*gray)*.

The trajectory obtained for the bulged uridine model (trajectory U, Figure 4) shows clearly higher r.m.s.d. than trajectories A1 and A2, which corresponds to the observed greater conformational flexibility. Interestingly, the initial orientation of the unpaired uridine, which was bulged out, has changed into a stacked-in conformation in about the simulation time of 12 ns and the resultant alignment remained stable until the end of the simulation (Fig. 7). On the other hand, the terminal base pairs (U8-A11 and G9-C10) were disrupted over the second half of the trajectory. This does not seem to affect the stability of the other parts of the duplex.

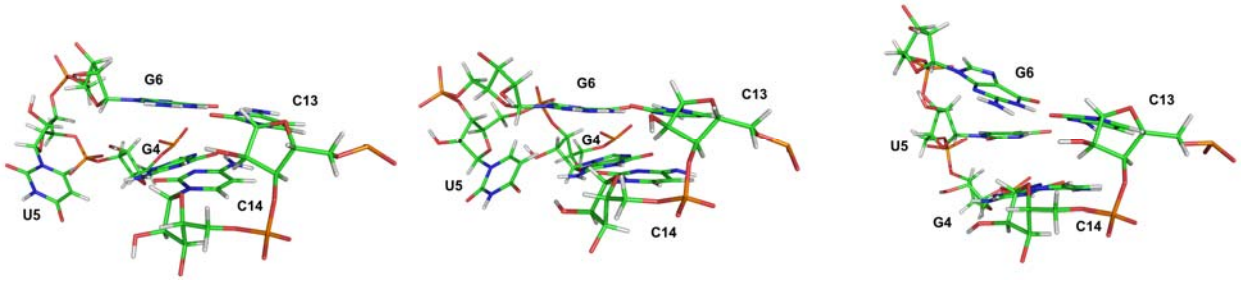

Figure 7. Selected frames of the trajectory U. Conformations of residue U5 observed at the beginning (*left*), in the middle (*centre*) and at the end (*right*) of trajectory U.

Previous NMR studies of the RNAs containing single nucleotide bulges show that adenine bulges tend to adopt the stacked-in conformation in solution,[18,19,21,34-36] in contrast to uridine bulges that prefer the looped-out structure.[13,15,26,37,38] The NMR structure of the duplex with the adenosine bulge (**I**) determined in the present study implies a bulge region conformation different from those described above. The bulged residue A5 is positioned in the minor groove and forms a base triple with the 5′-flanking G4-C14 pair (C:G-A triple). This motif has been previously predicted by the molecular dynamics simulation studies.[23,24] Results of our molecular dynamics simulations confirm the formation of a stable C14-G4:A5 base triple. Although in one

of the trajectories (A1), the bulge residue A5 switches to the stacked-in conformation after 14 ns, the C14-G4:A5 base triple remains stable over the entire trajectory A2. The network of hydrogen bonds suggested by NMR calculations, contributes to the stability of the base triple (Fig. 6). The analysis of NMR structures reveals that stacking interactions of residue A5 with its 3′-neighboring guanosine G6 additionally stabilize the conformation of the triple.

This study is concerned with the adenosine bulge placed between two GC base pairs in the 5′-GAG-3′/5′-CC-3′ context. In the previously reported structures of adenine bulges in the same context,[21,34] the bulge residue intercalates between the neighboring base pairs. Our study clearly suggests that the conformation of the bulged base might depend not only on the type of the flanking residues but also on the sequence not directly adjacent to the bulge.

To our knowledge, there are only two examples of NMR structures with a single bulged residue being involved in a base triple formation. In one of those structures, namely the stem-loop SL2 of the HIV-1 Ψ RNA packaging signal, the bulged adenosine residue, within the 5′-UAC-3′/5′-GA-3′ context, forms the A-U-A base triple.[39] The notation A-U-A is used because the base triple formation has led to a disruption of the 5′-neighboring Watson-Crick AU base pair. This conformation is stabilized by several non-canonical hydrogen bonds. The second example was found in the structure of the P4 stem of RNaseP in which the uracil base was mutated to adenine.[40] In this structure both Watson-Crick base pairs flanking the adenosine bulge are preserved and the bulged adenosine in the 5′-GAC-3′/5′-GC-3′ context forms a base triple with its 5′-adjacent residue.

In the uridine bulged duplex **II**, our studies indicate a dynamic character of bulge U5. In all NMR structures the bulged uridine is looped-out and points at the minor groove to side 5′ of the helix (Fig. 3). Conformational flexibility of uridine U5 was also observed in our molecular dynamics simulation (trajectory U) in which the initial looped-out orientation of the unpaired uridine changed into a stacked-in conformation.

Experimental Section

NMR samples. Oligoribonucleotides 5′-GUCGAGCUG-3′ (A), 5′-GUCGUGCUG-3′ (B) and 5′-CAGCCGAC-3′ (C) were synthesized using the standard phosphoramidite chemistry approach.[41] NMR samples were prepared by mixing equimolar amounts of strands A+C or B+C. Molar extinction coefficients were used to calculate the 1:1 stoichiometric ratio of these strands in the sample. Final concentrations of RNAs were ~1 mM. Each duplex was dissolved in a buffer containing 50 mM NaCl, 10 mM sodium phosphate (pH 6.8), 0.1 mM EDTA and placed in a Shigemi tube. In the next step, they were heated to 80 ^{o}C and allowed to cool slowly to the room temperature. For the experiments carried out in D_2O, the duplexes were evaporated from D_2O three times and redissolved in 99.96% D_2O. A mixture of 90% H_2O and 10% D_2O was used for experiments undertaken to study exchangeable protons

NMR spectroscopy. All NMR experiments were performed on a Bruker Avance 600 MHz spectrometer, processed with TopSpin (Bruker Inc.) and analyzed with the FELIX (Accelrys) software. NOESY spectra in D_2O were acquired at 25 °C with mixing times of 50, 150 and 400 ms using 2048 complex points in t_2 and a spectral width of 5000 Hz. A total of 512 t_1 experiments were recorded. The residual water signal in D_2O samples was suppressed by the low-power presaturation during the relaxation delay. Exchangeable proton resonances were assigned using the 150 ms mixing time NOESY spectra obtained at 10 °C. Solvent suppression for samples in 90% H_2O /10% D_2O was accomplished using the 3-9-19 WATERGATE[42] pulse sequence. The spectra were acquired with a sweep width of 14000 Hz in both dimensions. Typically, 512 FIDs of 2048 complex points were collected.

In the high resolution DQF- COSY spectra, broadband phosphorus decoupling was achieved by GARP[43], and a narrow spectral width (2800 Hz) was used for 4096 complex data points. A total of 512 FIDs were collected. Proton detected ^{31}P-^{1}H COSY experiments were acquired with 1200 Hz spectral width in the ^{31}P dimension and 1500 Hz spectral width in the ^{1}H dimension. Ninety-six scans of 2048 complex points and 256 FIDs were acquired. ^{1}H-^{13}C HMQC experiments were run with a sweep width of 20000 Hz in the ^{13}C dimension and the ^{13}C carrier at 100 ppm and 256 FIDs of 2048 complex points were acquired

Structure calculations. Nonexchangeable interproton restraints were obtained from the 150 ms NOESY (D_2O) spectrum recorded at 25 °C. The average pyrimidine H5-H6 distance (2.45 Å) was used as a reference. The lower and upper bounds for the restraints were set to -15% and +30%, respectively. The NOE distances for all exchangeable protons were restrained to 1.8 – 6.0 Å. Dihedral constraints on the sugar and backbone dihedral angles were restrained on the basis of the estimates of $^{3}J_{H1',H2'}$, $^{3}J_{H4',H5'}$, $^{3}J_{H4',H5''}$, $^{3}J_{P,H3'}$, $^{3}J_{P,H5'}$, $^{3}J_{P,H5''}$ as described previously.[19]

Torsion angle restraints χ were constrained to *anti* for all nucleotides except uridine U5 of duplex **II**, on the basis of analysis of the H1′-H8/H6 cross-peaks intensities in a short mixing time NOESY spectrum (50 ms).[33] For residue U5 of duplex **II** no torsion angle restraints were applied for the χ angle. The α and ζ torsion angles were restrained to 0 ± 120° for all nucleotides with phosphorus chemical shift between -3.7 and -4.5 ppm. The torsion angle constraints on β were restrained in order to achieve the *trans* conformation (178 ± 30°) for all residues in the stems of both **I** and **II** duplexes. For duplexes **I** and **II** no constraints on α, β and ζ were used for the residues G4 (ζ), A5 and U5 (α, β, ζ), and G6 (α, β) in the bulge region. The backbone dihedral angle γ was restrained to 54 ± 30° for all residues except the bulged residue A5 of duplex **I**, which was loosely restrained to exclude the *gauche*$^{-}$ conformation. For residue U5 of duplex **II**, and G6 in both duplexes, γ was left unconstrained. For the residues in the bulge region, the backbone dihedral angle ε was restrained to -153 ± 55°, and for the remaining residues it was given a tighter range of -153 ± 30°.

The sugar puckers of the majority of riboses were restrained to C3′-*endo* on the basis of the weak H1′ to H2′ scalar couplings. Residues A5, G6 and terminal G9 and C17 of duplex **I** and G6, G9, C10, C17 of duplex **II** were restrained to the intermediate C2′-*endo*/C3′-*endo* sugar puckering

conformation. For the residues G4 of both duplexes, sugar conformations were restrained to C2′-*endo*. No restraints were used for the ribose torsion angles of U5 in duplex **II**. Hydrogen bonding distance restraints used for the duplex with the adenosine bulge, were either consistent with experimental data or inferred from the initial structure calculated without explicit hydrogen bond restraints.

The structures were calculated using the TAMD[44] algorithm implemented in the XPLOR NIH package.[45,46] A total of 50 starting structures were generated using randomized coordinates and subjected to the four-step global fold as described previously.[19] The high-temperature TAMD stage consisted of 8000 steps with a time step of 0.008 ps and a bath temperature of 20000 K. It was followed by 10000 cycles of TAMD annealing from 20000 K to 300 K, with a time step of 0.007 ps. In the second cooling phase which consisted of 6000 steps in Cartesian space, the temperature decreased from 3000 K to 300 K. Afterwards, 1200 cycles of restrained Powell minimization were applied. Finally, the structures were subjected to the refinement procedure using the torsion angle and base-base positional database potentials of the mean force protocol.[47] The DELPHIC potentials were turned on for all residues except the bulged A5 and U5 residues. The parameter and topology sets used (dna-rna-allatom.top and dna-rna-allatom.param) were taken from the CNS distribution.

Finally, the 10 lowest energy structures with no violations to the NOE distances (0.3 Å), and dihedral angles (5°) were selected.

Structure coordinates have been deposited in the Protein Data bank as entries *2k3z* and *2k41*. Chemical shift data for both molecules have been deposited in BioMagResBank (BMRB) under entry codes *15780* and *15781*.

Molecular dynamics. Three 21 nanosecond MD simulations were performed in total: two were based on the adenosine bulge sequence: A1 and A2, and one with the uridine bulge sequence. The coordinates for the starting RNA duplexes were derived from the NMR experiment. Each model was neutralized with 15 Na^+ ions and placed into a PBC box filled with TIP3P water[48] layer of 12 Å. This procedure yielded 3612 water molecules in a 57 Å x 54 Å x 49 Å box for the (A) model and 3931 water molecules in a 56 Å x 53 Å x 54 Å box for the (U) model (the box dimensions slightly shrink during the subsequent equilibration). The simulations were run at a constant temperature (300 K) and pressure (1 atm) by using the AMBER 8 simulation package[49] with the Amber forcefield.[50] The particle mesh Ewald summation method was used for the treatment of long-range electrostatic interactions.[51] Each trajectory was obtained with a 2 fs time step using the SHAKE bond constraints scheme. The equilibration procedure followed the protocol used in an earlier paper[52] but the timing of all stages was elongated so that the whole equilibration phase reached 1 ns. The production stage comprised 20 nanoseconds of unrestrained MD simulation. The trajectories were subjected to conformational analysis with AMBER package tools and our own procedures.

Acknowledgements

This work was supported by a research grant from the Polish Ministry of Science and Higher Education (No. 2 P04 A 033 30). The generous support of the Poznan Supercomputing and Networking Centre is acknowledged.

References and Notes

1. Egli, M. *Curr. Opin. Chem. Biol.* **2004**, *8*, 580.
2. Hermann, T.; Patel, D. J. *Structure* **2000**, *8*, R47.
3. Turner, D. H. *Curr. Opin. Struct. Biol.* **1992**, *2*, 334.
4. Peattie, D. A.; Douthwaite, S.; Garret, R. A.;Noller, H. F. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 7331.
5. Valegard, K.; Murray, J. B.; Stonehouse, N. J.; van den Worm, S.; Stockley, P. G.;Lijias, L. *J. Mol. Biol.* **1997**, *270*, 724.
6. Vicens, Q.; Westhof, E. *J. Mol. Biol.* **2003**, *326*, 1175.
7. Correll, C. C.; Beneken, J.; Plantinga, M. J.; Lubbers, M.;Chan, Y. L. *J. Mol. Biol.* **2003**, *31*, 6806.
8. Ennifar, E.; Yusupov, M.; Walter, P.; Marquet, R.; Ehresmann, B.; Ehresmann, C.;Dumas, P. *Structure* **1999**, *7*, 1439.
9. Joshua-Tor, L.; Rabinovich, D.; Hope, H.; Frolow, F.; Appela, E.;Sussman, J. L. *Nature* **1988**, *334*, 82.
10. Portmann, S.; Grimm, S.; Workman, C.; Usman, N.;Egli, M. *Chem. Biol.* **1996**, *3*, 173.
11. Sudarsanakumar, Ch.; Xiong, M.;Sundaralingam, M. *J. Mol. Biol.* **2000**, *299*, 103.
12. Xiong, Y.; Deng, J.; Sudarsanakumar, C.;Sundaralingam, M. *J. Mol. Biol.* **2001**, *313*, 573.
13. Cabello-Villegas, J.; Giles, K.E.; Soto, A. M.; Yu, P.; Mougin, A.; Beemon, K. L.;Wang, Y. X. *RNA* **2004**, *10*, 1388.
14. Dornberger, U.; Hillisch, A.; Gollmick, F. A.; Fritzsche, H.;Diekmann, S. *Biochemistry* **1999**, *38*, 12860.
15. Finger, L. D.; Trantirek, L.; Johansson, C.;Feigon, J. *Nucl. Acids Res.* **2003**, *31*, 6461.
16. Gollmick, F. A.; Lorenz, M.; Dornberger, U.; von Langen, J.; Diekmann, S.; Fritzsche, H. *Nucleic Acids Res.* **2002**, *30*, 2669.
17. Greenbaum, N. L.; Radhakrishnan, I.; Patel, D. J.;Hirsh, D. *Structure* **1996**, *4*, 725.
18. Newby, M. I.; Greenbaum, N. L. *Nature Struct. Biol.* **2002**, *9*, 958.
19. Popenda, L.; Adamiak, R. W.;Gdaniec, Z. *Biochemistry* **2008**, *47*, 5059.
20. Rosen, M. A.; Live, D.;Patel, D. J. *Biochemistry* **1992**, *31*, 4004.
21. Thiviyanathan, V.; Guliaev, A. B.; Leontis, N. B.;Gorenstein, D. G. *J. Mol. Biol.* **2000**, *300*, 1143.
22. Zacharias, M.; Sklenar, H. *J. Mol. Biol.* **1999**, *289*, 261.

23. Barthel, A.; Zacharias, M. *Biophys. J.* **2006**, *90*, 2450.
24. Hastings, W. A.; Yingling, Y. G.; Chirikjian, G. S.;Shapiro, B. A. *J. Comput. Theor. Nanosci.* **2006**, *3*, 63.
25. Kazantsev, A. V.; Krivenko, A. A.; Harrington, D. J.; Holbrook, S. R.; Adams, P. D.;Pace, N. R. *Proc. Natl Acad. Sci. USA* **2005**, *102*, 13392.
26. Leeper, T. C.; Martin, M. B.; Kim, H.; Cox, S.; Semenchenko, V.; Schmidt, F. J.;Van Doren, S. R. *Nature Struct. Biol.* **2002**, *9*, 397.
27. Zagorowska, I.; Adamiak, R. W. *Biochimie* **1996**, *78*, 123.
28. Kulinski, T.; Bielecki, L.;Adamiak, R. W. *Nucleic Acids Res. Supplement No 1* **2001**, 139.
29. Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wohnert, J.; Stock, G.;Schwalbe, H. *Nucleic Acids Res.* **2008**, *36*, 1928.
30. Villa, A.; Stock, G. *J. Chem. Theory Comput.* **2006**, *2*, 1228.
31. Varani, G.; Tinoco, I.Jr. *Q. Rev. Biophys.* **1991**, *24*, 479.
32. Varani, G.; Aboulela, F.;Allain, F. H. T. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51.
33. Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287.
34. Sashital, D. G.; Allmann, A. M.; Van Doren, S. R.;Butcher, S. E. *Biochemistry* **2003**, *42*, 1470.
35. Smith, J. S.; Nikonowicz, E. P. *Biochemistry* **1998**, *37*, 13486.
36. Smith, J. S.; Nikonowicz, E. P. *Biochemistry* **2000**, *39*, 5642.
37. D'Souza, V.; Dey, A.; Habib, D.;Summers, M. F. *J. Mol. Biol.* **2004**, *337*, 427.
38. Lukavsky, P. J.; Kim, I.; Otto, G. A.;Puglisi, J. D. *Nature Struct. Biol.* **2003**, *10*, 1033.
39. Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.;Summers, M. F. *J. Mol. Biol.* **2000**, *299*, 145.
40. Schmitz, M. *Nucleic Acids Res.* **2004**, *32*, 6358.
41. Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859.
42. Piotto, M.; Saudek, V.;Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661.
43. Shaka, A. J.; Barker, P. B.;Freeman, R. *J. Magn. Reson.* **1985**, *64*, 547.
44. Stein, E. G.; Rice, L. M.;Brunger, A. T. *J. Magn. Reson.* **1997**, *124*, 154.
45. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.
46. Schwieters, C. D.; Kuszewski, J. J.;Clore, G. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47.
47. Clore, G. M.; Kuszewski, J. *J. Am. Chem. Soc.* **2003**, *125*, 1518.
48. Jorgensen, W.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.;Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
49. Case, D. A.; Darden, T. A.; Cheatham, I. T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.;Kollman, P. A. *AMBER 8* **2004**, University of California, San Francisco.
50. Duan, Y.; Chowdhury, S.; Wu, C.; Xiong, G. M.; Zhang, W.; Yang, R.; Lee, M.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A. *J. Comp. Chem.* **2003**, *24*, 1999.

51. Darden, T.; Perera, L.; Li, L.P.;Pedersen, L. *Structure* **1999**, *7*, R55-R60.
52. Auffinger, P.; Bielecki, L.;Westhof, E. *Chemistry & Biology* **2003**, *10*, 551.