Published in final edited form as:

J Phys Chem B. 2019 January 10; 123(1): 57–65. doi:10.1021/acs.jpcb.8b09139.

Computational Investigation of RNA A-Bulges Related to the Microtubule-Associated Protein Tau Causing Frontotemporal Dementia and Parkinsonism

David J. Wales[†], Matthew D. Disney[‡], and Ilyas Yildirim^{*,‡,§}

[†]Department of Chemistry, University of Cambridge, Cambridge, Cambridgeshire CB2 1EW, U.K.

[‡]Department of Chemistry, Scripps Research Institute, Jupiter, Florida 33458, United States

[§]Department of Chemistry and Biochemistry, Florida Atlantic University, Jupiter, Florida 33458, United States

Abstract

Mutations in the human *tau* gene result in alternative splicing of the *tau* protein, which causes frontotemporal dementia and Parkinsonism. One disease mechanism is linked to the stability of a hairpin within the microtubule-associated protein tau (MAPT) mRNA, which contains an A-bulge. Here we employ computational methods to investigate the structural and thermodynamic properties of several A-bulge RNAs with different closing base-pairs. We find that the current amber RNA force field has a preference to overstabilize base-triple over stacked states, even though some of the A-bulges are known to prefer stacked states according to NMR studies. We further determined that if the neighboring base-pairs of A-bulges are AU, this situation can lead to base slippage. However, when the 3'-side of the A-bulge has an UA base-pair, the stacked state is stabilized by an extra interaction that is not observed in the other sequences. We suggest that these A-bulge RNA systems could be used as benchmarks to improve the current RNA force fields.

Graphical Abstract

*Corresponding Author Phone: (561) 799-8325; iyildirim@fau.edu.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jpcb.8b09139](https://doi.org/10.1021/acs.jpcb.8b09139).

Movie descriptions, sample input AMBER/OPTIM/PATHSAMPLE input files, RMSD analyses, distributions of backbone torsions, and distribution analyses for umbrella sampling calculation (PDF)

Movie of MD simulation of 5'-GC^AUU/5'-AAGC (MPG)

Movie of MD simulation of 5'-GC^AGU/5'-ACGU (AVI)

The authors declare no competing financial interest.

INTRODUCTION

Along with translating genetic information in protein synthesis, different RNA molecules have crucial roles in cellular function, including retroviruses,^{1,2} riboswitches,^{3,4} retrotransposons,⁵ RNA interference,^{6,7} ribozymes,^{8,9} RNA aptamers,¹⁰ anti-microRNAs,¹¹ and CRISPR.¹² One important RNA system is the microtubule-associated protein tau (MAPT) mature mRNA, which is associated with inherited frontotemporal dementia and Parkinsonism, caused by inclusion of exon 10 in MAPT mRNA. The over inclusion of exon 10 creates too much of the 4R (microtubule-binding domain) version of tau, and this inclusion is linked to the stability of a hairpin structure within MAPT pre-mRNA, which is controlled by an adenosine bulge loop (A-bulge) (Figure 1). Because the energetics of RNA loops make them biologically significant, investigation of their structural and dynamic properties can provide information, which can be used to identify and study lead medicines to target these structures.

Several NMR studies have been reported for RNA systems suggesting that A-bulges prefer stacked, unstacked, or base-triple states (Table 1 and Figure 2).^{13–24} The multiconformational nature of A-bulges implies that the closing base-pairs at the bulge site might have direct roles in the conformational preference of the bulged adenosine. For example, NMR structures reveal that 5'-UCACC/5'-GGGA,²² 5'-GCAGU/5'-ACGU,^{16,23} 5'-AGAGU/5'-ACCU,¹⁵ 5'-UGACG/5'-CGCA,^{13,14} 5'-UAAGG/5'-CCUA,²⁵ and 5'-AGAAG/5'-CUCU¹⁸ mainly prefer the stacked adenosine in the A-bulge sites (Figure 2A). On the other hand, NMR structures also reveal that 5'-AGACC/5'-GGCU,¹⁷ 5'-CGAGC/5'-GCCG,¹⁹ 5'-CUACC/5'-GGGG,²¹ and 5'-GGAUG/5'-UACC²⁴ mainly prefer the adenosine in base-triple states (Figure 2B), while 5'-GUAGC/5'-GCGC²⁰ have a fully unstacked A-bulge state (Figure 2C). An in-depth understanding of structural preferences of A-bulges could explain why hairpin structures in MAPT mRNA are destabilized through specific mutations that cause frontotemporal dementia and Parkinsonism.

RNA structures in solution are dynamic and will sample all the possible conformations allowed by energetics and environment. If the structure prefers only one conformation, NMR data is useful to get a 3D model for the structure. In the case where the structure has two competing conformations with a fast exchange rate, the NOE data, which are the ensemble averages of $1/r^6$ and are used to determine the NMR distance restraints, will not be successful enough to determine the conformations. In the slow exchange limit, however, it

might be possible to determine the two conformations because two separate peaks each representing individual conformations will be observed in the NMR spectrum. For example, the 4×4 internal loop in $(5' \text{GACGAGUGUCA})_2$ has been shown to prefer two conformations in equilibrium by Turner and co-workers using NMR spectroscopy.²⁶ A similar result was observed in the NMR studies of single-stranded RNA GACC, which displayed two structures representing the major and minor conformations.²⁷ As a result, one has to be careful when utilizing the NMR results.

One of the aims of computational chemistry is to develop physically meaningful models to study RNA. However, this is a big challenge, because of the dynamic nature of RNA molecules that have complex architecture including loops and/or pseudoknots. A systematic study of RNA molecules starting with the simplest forms should help to create better RNA models that can improve the accuracy of RNA force fields. Previously, Barthel and Zacharias studied the conformational preference of single uridine and adenosine bulges in RNA systems using an obsolete AMBER force field, which has been shown to have serious artifacts.²⁸ Recent advancements in RNA force fields, however, made it possible to study challenging RNA systems.²⁹ We previously showed that RNA mononucleosides and single-stranded RNA tetramers provide ideal benchmarks to optimize the parameters of the amber force field.^{27,30,31} In this article, we propose that RNA adenosine bulges (A-bulge) can be used as another benchmark to improve current RNA force fields.

In this contribution, we investigate the properties of model RNA systems mimicking A-bulges. Specifically, two A-bulge systems, $5' \text{-CCGGCAGUGUG/5' -CACACGUCGG}$ and $5' \text{-CCGCGAGCGUG/5' -CACGCCGCGG}$, that are known to prefer stacked and base-triple states, respectively, by NMR are studied using 2D umbrella sampling (US) calculations, discrete path sampling (DPS), and standard molecular dynamics (MD) simulations. Moreover, 16 unique A-bulges of $5' \text{-CCGGXAYU-GUG-3' / 5' -CACAY*X*CCGG-3'}$ (X , Y , X^* , and $\text{Y}^* = \text{A}, \text{C}, \text{G}, \text{and U}$, and XX^* and YY^* are forming Watson—Crick base-pairs) were studied by MD simulations. We discovered that the current amber RNA force field tends to prefer the base-triple over the stacked state for A-bulges, which is stabilized by hydrogen-bond interactions between the A-bulge and one of the closing base-pairs. Furthermore, we discovered that A-bulge systems $5' \text{-XAU/5' -AX}^*$ have higher tendencies to form a stacked state, compared to other systems we studied with the amber force field. We also discovered that A-bulges closed by AU base-pairs (such as $5' \text{-AAX/5' -X}^*\text{U}$ and $5' \text{-XAA/5' -UX}^*$) often exhibit base slippage that might affect the stability of the RNA hairpin observed in MAPT pre-RNA at the exon 10-intron junction. The results indicate that the current amber RNA force fields require revision to correct the imbalance between stacked and base-triple states observed in A-bulges.

METHODS

Preparation of Initial Structures.

For explicit solvent MD simulations, 18 A-bulge systems, shown in Table 2, were prepared for MD simulations. The NMR structure of $5' \text{-CCGGCAGUGUG/5' -CACACGUCGG}$ (GCAGU), which displays the A-bulge in a stacked state (Figure 3A), is used as a homology model for the A-bulge RNA systems studied (Table 2). When building the homology

models, the backbone conformation of GCAGU was kept intact while mutating the residues around the A-bulge site. Minimization was performed using Watson–Crick and torsional restraints to constrain the initial structures for MD simulations to be in stacked states. For the 2D US and DPS calculations, two model systems were utilized, (i) 5′-GCAGU-3′/5′-ACGU-3′ and (ii) 5′-CGAGC-3′/5′-GCCG-3′, that are known to prefer the stacked and base-triple states by NMR spectroscopy, respectively.^{19,23} The leap module of the AMBER MD package³² was used to prepare the files required for these simulations. For the explicit solvent MD runs, each system was first neutralized with Na⁺ ions.³³ We then added five Na⁺ and Cl[−] ions³³ to each system. Then, 6676 TIP3P³⁴ water molecules were used to solvate each system in a truncated octahedral box. We employed implicit solvent models (GB^{OBC})³⁵ in the 2D US and DPS calculations. In all of the simulations and calculations, the revised χ ³¹ and α/γ ³⁶ torsional parameters, which have been shown to improve predictions,^{27,30,31,37–42} were used with the amber force field⁴³ to describe the RNA molecules.

MD Simulations.

Each system was first minimized using Watson–Crick base-pairing restraints. We minimized the structures using both the steepest-descent and conjugate-gradient minimization methods. Nonbonded interactions were included by an 8.0 Å cutoff during minimization (see Table S1 for sample input files). After minimization, Watson–Crick base-pairing restraints were utilized in the first step of equilibration, where temperature was increased from 0 to 300 K over 2 ns under constant volume dynamics (*NVT*). Another 2 ns of MD at 300 K under constant pressure dynamics (*NPT*) was run on each system, while still imposing Watson–Crick base-pairing restraints (see Table S1 for sample inputs).

After equilibration, a similar protocol to that previously described was followed for the production runs.^{36,38,44,45} In each case, we utilized the constant pressure dynamics (*NPT*) with uniform scaling (isotropic position scaling), where the reference pressure was set to be 1 atm with 2 ps pressure relaxation time. SHAKE⁴⁶ was used to constrain the bonds involving hydrogen atoms. Similar to minimization, the atom-based long-range cutoff of 8.0 Å was used in all of the MD simulations. Over 500 ns of MD was run with 2 fs time steps for each system. The GPU version of pmemd (pmemd.cuda)³² was employed for all of the MD simulations.

US.

There are two important motions involving adenosine bulges, (i) the base rotation with respect to sugar, which can be mimicked with χ torsion, and (ii) base stacking $\leftrightarrow$ unstacking, which can be mimicked with a pseudotorision (θ). We previously built the 2D PMF surfaces for RNA CAG and CUG repeats successfully using such pseudotorisions.^{38,44} In this study, we again used the χ and θ torsions (Figure 4) to mimic these two motions, and constructed the 2D PMF surfaces for (i) 5′-GCAGU-3′/5′-ACGU-3′ and (ii) 5′-CGAGC-3′/5′-GCCG-3′. The initial conformations for US calculations were created by rotating the χ and θ angles by 10°, yielding $36 \times 36 = 1296$ structures. Each US window was simulated for 100 ns using implicit solvent models (GB^{OBC}),³⁵ producing ~130 μ s MD simulation to build each 2D PMF surface. We decided to use the continuum solvent model for two reasons. (1) We previously built the 2D PMF surfaces of 1 $\times$ 1 UU RNA internal

loops in explicit solvent that overlapped well with the DPS results, which utilized implicit solvent models.⁴⁴ (2) Explicit solvent MD simulations are much slower than implicit solvent ones due to explicit inclusions of water molecules and ions. The frictional force between solute and solvent (viscosity) hinders the conformational change of the solute. As a result, utilization of implicit solvent models provides better conformational sampling.

DPS.

To efficiently scan the configurational space, we utilized the DPS^{47,48} method, which we successfully applied in the studies of 1×1 internal loops in RNA CAG and CUG repeat expansions, and single-stranded RNA tetramers.^{36,44,49} In the present work, DPS was applied to 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3' A-bulge RNAs to compare the predictions with the NMR structures, which are known to prefer stacked and base-triple states, respectively.

The initial conformations for DPS calculations were created by rotating the pseudotorsion (Figure 4B) in steps of 10° , yielding 36 structures. We then built the initial database by first minimizing the starting conformations with a modified version of the LBFGS algorithm⁵⁰ and then attempting to make a connection between the minima. For the root-mean-square gradient, we used a convergence criterion of 10^{-6} kcal mol⁻¹. Connection attempts were made between the minima. The UNTRAP scheme⁵¹ was performed to refine the database further. Disconnectivity graphs⁵²⁻⁵⁴ were created to highlight the stable conformational states. The OPTIM and PATHSAMPLE programs were used for all the DPS calculations (see Table S2 for sample inputs). For the 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-375'-GCCG-3' A-bulge RNAs, the final stationary point databases contain ~56 K minima and ~81 K transition states, and ~124 K minima and ~173 K transition states, respectively. The harmonic superposition approximation was employed on the database to estimate the free energies.⁵⁵

Analysis.

For the dihedral and root-mean-square deviation (rmsd) analyses, we utilized the ptraj module of AMBER16.³² In-house code was written to perform cluster analyses. Disconnectivity graphs were plotted to visualize the free energy landscapes.⁵⁶ The disconnectionDPS program⁵⁷ was utilized on the stationary point databases to construct the disconnectivity graphs.

RESULTS AND DISCUSSION

MD Simulations of 16 Unique A-Bulges Indicate that A-Bulges Prefer the Base-Triple Over the Stacked-A State.

Sixteen model A-bulge systems (Table 2, systems # 1–16), 5'-CCGGXAYUGUG-3'/5'-CACAYX*CCGG-3', where X, Y, X*, and Y*=A, C, G, and U, and XX* and YY* form Watson–Crick base-pairs, were created in stacked A-bulge orientations (Figure 3A) to investigate their structural preferences. Almost all of them immediately transformed to a base-triple state (Figures 3B and S1–S16). Except for GGAUU, GGAAU, GCAUU, GCAAU, and GAAUU, it was observed that over 90% of the snapshots were in base-triple

states in the MD simulations (Table 2). MD simulations display stacked $\leftrightarrow$ base-triple transformations, but the stacked states are not long-lived in most cases (Table 2 and Figures S1–S16). It is important to note that the simulations might not have sampled all the conformational space, as each MD simulation is run for 500 ns. Nevertheless, there is a consistent behavior in all of the A-bulges. Even though the initial conformations were designed to be in stacked states, the A-bulges transform to base-triple states (Figures S1–S16).

A-Bulges Closed by AU Base-Pairs, Such as 5'-AAX/5'-X*U and 5'-XAA/5'-UX*, Often Exhibit Base Slippage.

Base slippage occurs if the closing base-pairs are AU, such as 5'-AAX/5'-X*U and 5'-XAA/5'-UX*, where XX* is a Watson–Crick base-pair (Table 2 and Figures S1–S5 and S9). It is noteworthy, that no base slippage was observed in GUAU (Figure S13 and Table 2, system # 13). Such slippages can affect the stabilities of the systems, as the A-bulge can have more than one conformational choice. As an example, system # 1 (Table 2), which is GAAU, can have three alternative A-bulge sequences due to base slippage (GGAU, GAAU, and AAU) (Figure S1). Similar base slippages are observed in systems # 2, 3, 4, 5, and 9. These types of base slippage were observed by others using NMR spectroscopy in A-bulge systems closed by AU base-pairs where uridine can base-pair with the A-bulge to cause the slippage.^{25,58}

A-Bulge Systems of 5'-XAU/5'-AX* Have Higher Tendencies To Form a Stacked State Compared to the Other Systems Studied.

MD simulations of A-bulge systems of 5'-XAU/5'-AX*, except GUAU (system # 16 in Table 2), tend to stabilize the stacked state by an electrostatic interaction between the A-bulge and AU closing base-pair (Figure 5 and Table 2). As an example, it was observed that the GCAU A-bulge (system # 8 in Table 2) prefers 49.4 and 50.6% of the stacked and base-triple states, respectively (Figures 5 and S8). Similar trends have been observed in A-bulge systems of 1, 4, 5, 9, and 12 (Table 2, and Figures S1, S4–5, S9, and S12). The stabilizing interaction between A6-NH₂ and A16-N1 is the main reason why the stacked state in these A-bulge systems is stabilized, as depicted in Figure 5 (see Movie S1).

Current Amber RNA Force Field Does Not Support the A-Bulge Structure of 5'-GCAGU/5'-ACGU but Is Effective for 5'-CGAGC/5'-GCCG.

Varani and co-workers solved the NMR structure of an A-bulge, 5'-GCAGU/5'-ACGU,^{16,23} which exhibited an A-bulge in a stacked state (Table 1 and Figure 3A). Adamiak and co-workers solved the NMR structure of an A-bulge, 5'-CGAGC-3'/5'-GCCG-3',¹⁹ which displayed an A-bulge in a base-triple state (Figure 3C). MD simulations of model A-bulge systems mimicking these two structures (system # 17–18 in Table 2), which are started from initial states similar to the NMR structures, suggest that both systems strongly prefer the base-triple states (see Figures 3B,D and 6). 5'-GCAGU/5'-ACGU, which has an A-bulge in the stacked state according to NMR, transforms to a base-triple state within 10 ns and stays in this conformation for the rest of the MD simulation (Figure 6B and Movie S2). 5'-CGAGC/5'-GCCG, which has an A-bulge in a base-triple state according to NMR, transforms to a stacked state for a short amount of time (time ~10 ns in Figure 6A) but stays

in the base-triple state for the rest of the MD simulation, as expected (Figure 6A). Compared to experimental results, analyses of the backbone torsions near the A-bulge site display sampling of A-form-like dihedral angles (Figure S17).⁵⁹

Similar to the MD Simulations, the DPS Method Predicts the Base-Triple States To Be the Global Minimum in Both 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3'.

The free energy landscapes of GCAGU/ACGU and CGAGC/GCCG were calculated using the DPS approach (Figure 7). These results reveal several A-bulge states in the disconnectivity graphs, but the global minimum structures in both systems are the base-triple states (black conformations in Figures 7 and 8A), which are similar to the results observed in the MD simulations (Figure 3B,D). The stacked A-bulge states (Figure 8B), highlighted in red in Figure 7, are ~3.5 and 4.5 kcal/mol less stable than the base-triple states in GCAGU/ACGU and CGAGC/GCCG, respectively (Figure 7). Another state revealed by the DPS approach is a conformation unstacked via the major groove (green conformations in Figures 7 and 8C). Fully unstacked states (Figure 8D) are highlighted in blue in Figure 7.

US Calculations also Predict the Base-Triple State To Be the Global Minimum for A-Bulges in 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3'.

We built the 2D free energy landscapes (θ , χ) for A-bulges in 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3' using the χ torsion and a pseudotorsion (θ) representing the base flipping of the A-bulge as the reaction coordinates in US calculations (Figure 9). Similar analyses were performed previously to study 1 × 1 AA and UU internal loops in RNA CAG and CUG repeat expansions.^{38,44} Analysis of the distributions for each US window overlap well (Figure S18). The 2D PMF surfaces predicted for the A-bulges in 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3' and have very similar characteristics, except for a couple of slight differences. In both 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3', the global minimum structure is the base-triple state ($\theta \sim -60^\circ$, $\chi \sim 180^\circ$) (state b in Figure 9) similar to the one predicted by the MD simulations and DPS calculations (Figure 8A). The A-bulge state at ($\theta \sim 50^\circ$, $\chi \sim 200^\circ$) (state c in Figure 9) represents the stacked state (Figure 8B), which is a local minimum in both systems. The G of the base-triple (state b in Figure 9) $\rightarrow$ stacked (state c in Figure 9) transformation of A-bulges in 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3' are 1.1 and 2.2 kcal/mol, respectively (Figure 9). US calculations identify other local minima states for the A-bulges in both RNA systems as well, such as states d, e, f, and g in Figure 9, which have A-bulges in syn orientations ($\chi \sim 60^\circ$), and these are at least 1.7 and 2.8 kcal/mol less stable than global minimum (b in Figure 9) in 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3', respectively. Furthermore, state a in Figure 9 represents the fully unstacked A-bulge state (Figure 8D), which is 2.5 and 3.6 kcal/mol less stable than the global minimum (b in Figure 9) in 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3', respectively. Furthermore, US calculations in 5'-CGAGC-3'/5'-GCCG-3' display a state at ($\theta \sim 90^\circ$, $\chi \sim 240^\circ$) (h in Figure 9B), which is the major-groove state displayed in Figure 8C that is 3.1 kcal/mol less stable than the global minimum b (Figure 9B).

SUMMARY AND CONCLUSIONS

The mature mRNA of MAPT is associated with inherited frontotemporal dementia and Parkinsonism. This disease is linked to the stability of a hairpin structure within MAPT pre-mRNA with an A-bulge. Such systems can prefer different conformations that are determined by neighboring base-pairs. The atomistic details of the structure and dynamics of A-bulges yield valuable information, including the conditions that determine the stability of A-bulges and, thus, provide a mechanism that explains how the disease arises.

In the present contribution, we utilized MD simulations for model RNA A-bulges to describe the roles of neighboring base-pairs. Furthermore, we utilized US and DPS calculations to test the quality of current amber RNA force fields by studying two known RNA A-bulges observed in NMR spectroscopy. The results indicate that the current amber RNA force field prefers the base-triple over the stacked state in A-bulges, which is one of the A-bulge conformations observed in NMR spectroscopy. This conclusion was verified in the studies of 5'-GCAGU-3'/5'-ACGU-3' and 5'-CGAGC-3'/5'-GCCG-3' A-bulges, which are known to prefer stacked and base-triple states, respectively, using US and DPS calculations and analysis of free energy landscapes. The base-triple state of A-bulges is overstabilized by three attractive interactions between the A-bulge and the neighboring base-pair at the 5'-side, which explains why A-bulges known to prefer the stacked state cannot be represented properly. Furthermore, MD simulations of A-bulges with neighboring AU base-pairs display base-slippage, where the system transforms to another type of A-bulge that can affect the stability. Moreover, A-bulges with neighboring UA base-pairs at the 3'-side exhibit relatively higher populations of stacked states compared to the other A-bulges studied here, due to the attractive interaction observed between the A-bulge and the adenosine residue of the UA base-pair, which keeps the A-bulge in a stacked orientation.

Overall, the results indicate that the forces stabilizing the basetriples and stacked states in A-bulges are not described properly by the current amber RNA force field. One might suppose that the improper description of π - π interactions, which can stabilize the stacked state in A-bulges, is responsible for the inaccurate predictions. Another reason could be the description of the backbone parameters, which might also require revision. Overall, RNA A-bulges are relatively small compared to other RNA internal loops, which makes them ideal systems to benchmark and hence improve RNA force fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Computations were performed using the high-performance computing (HPC) cluster, KoKo, at the Florida Atlantic University. This work was supported by the Department of Chemistry and Biochemistry, Florida Atlantic University (I.Y.); the Engineering and Physical Sciences Research Council (EPSRC EP/N035003/1) (D.J.W.); and the National Institute of Health (R01 GM97455).

REFERENCES

- (1). Bishop JM Cellular Oncogenes and Retroviruses. *Annu. Rev. Biochem.* 1983, 52, 301–354. [PubMed: 6351725]
- (2). Gifford R; Tristem M The Evolution, Distribution and Diversity of Endogenous Retroviruses. *Virus Genes* 2003, 26, 291–316. [PubMed: 12876457]
- (3). Haller A; Souliere MF; Micura R The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms. *Acc. Chem. Res.* 2011, 44, 1339–1348. [PubMed: 21678902]
- (4). Smith AM; Fuchs RT; Grundy FJ; Henkin TM Riboswitch RNAs Regulation of Gene Expression by Direct Monitoring of a Physiological Signal. *RNA Biol.* 2010, 7, 104–110. [PubMed: 20061810]
- (5). Sabot F; Schulman AH Parasitism and the Retrotransposon Life Cycle in Plants: A Hitchhiker's Guide to the Genome. *Heredity* 2006, 97, 381–388. [PubMed: 16985508]
- (6). Hannon GJ RNA Interference. *Nature* 2002, 418, 244–251. [PubMed: 12110901]
- (7). Kole R; Krainer AR; Altman S RNA Therapeutics: Beyond RNA Interference and Antisense Oligonucleotides. *Nat. Rev. Drug Discovery* 2012, 11, 125–140. [PubMed: 22262036]
- (8). Birikh KR; Heaton PA; Eckstein F The Structure, Function and Application of the Hammerhead Ribozyme. *Eur. J. Biochem.* 1997, 245, 1–16. [PubMed: 9128718]
- (9). Frank DN; Pace NR RIBONUCLEASE P: Unity and Diversity in a Trna Processing Ribozyme. *Annu. Rev. Biochem.* 1998, 67, 153–180. [PubMed: 9759486]
- (10). Patel DJ; Suri AK; Jiang F; Jiang LC; Fan P; Kumar RA; Nonin S Structure, Recognition and Adaptive Binding in RNA Aptamer Complexes. *J. Mol. Biol.* 1997, 272, 645–664. [PubMed: 9368648]
- (11). Lu YJ; Xiao JN; Lin HX; Bai YL; Luo XB; Wang ZG; Yang BF A Single Anti-Microrna Antisense Oligodeoxyribo-Nucleotide (Amo) Targeting Multiple Micrornas Offers an Improved Approach for Microrna Interference. *Nucleic Acids Res.* 2009, 37, e24. [PubMed: 19136465]
- (12). Cong L; Ran FA; Cox D; Lin SL; Barretto R; Habib N; Hsu PD; Wu XB; Jiang WY; Marraffini LA; et al. Multiplex Genome Engineering Using Crispr/Cas Systems. *Science* 2013, 339, 819–823. [PubMed: 23287718]
- (13). Borer PN; Lin Y; Wang S; Roggenbuck MW; Gott JM; Uhlenbeck OC; Pelczar I Proton NMR and Structural Features of a 24-Nucleotide RNA Hairpin. *Biochemistry* 1995, 34, 6488–6503. [PubMed: 7756280]
- (14). Kerwood DJ; Borer PN Structure Refinement for a 24-Nucleotide RNA Hairpin. *Magn. Reson. Chem.* 1996, 34, S136–S146.
- (15). Thivyanathan V; Guliaev AB; Leontis NB; Gorenstein DG Solution Conformation of a Bulged Adenosine Base in an RNA Duplex by Relaxation Matrix Refinement. *J. Mol. Biol.* 2000, 300, 1143–1154. [PubMed: 10903860]
- (16). Varani L; Spillanti MG; Goedert M; Varani G Structural Basis for Recognition of the RNA Major Groove in the Tau Exon 10 Splicing Regulatory Element by Aminoglycoside Antibiotics. *Nucleic Acids Res.* 2000, 28, 710–719. [PubMed: 10637322]
- (17). Schmitz M Change of Rnase P RNA Function by Single Base Mutation Correlates with Perturbation of Metal Ion Binding in P4 as Determined by NMR Spectroscopy. *NucleicAcids Res.* 2004, 32, 6358–6366.
- (18). Popena L; Adamiak RW; Gdaniec Z Bulged Adenosine Influence on the RNA Duplex Conformation in Solution. *Biochemistry* 2008, 47, 5059–5067. [PubMed: 18399645]
- (19). Popena L; Bielecki L; Gdaniec Z; Adamiak RW Structure and Dynamics of Adenosine Bulged RNA Duplex Reveals Formation of the Dinucleotide Platform in the C:G-a Triple. *Arkivoc* 2009, 2009, 130–144.
- (20). Falb M; Amata I; Gabel F; Simon B; Carlomagno T Structure of the K-Turn U4 RNA: A Combined NMR and Sans Study. *Nucleic Acids Res.* 2010, 38, 6274–6285. [PubMed: 20466811]
- (21). Miller SB; Yildiz FZ; Lo JA; Wang B; D'Souza VM A Structure-Based Mechanism for Trna and Retroviral RNA Remodelling During Primer Annealing. *Nature* 2014, 515, 591–595. [PubMed: 25209668]

- (22). Smith JS; Nikonowicz EP Phosphorothioate Substitution Can Substantially Alter RNA Conformation. *Biochemistry* 2000, 39, 5642–5652. [PubMed: 10801314]
- (23). Varani L; Hasegawa M; Spillantini MG; Smith MJ; Murrell JR; Ghetti B; Klug A; Goedert M; Varani G Structure of Tau Exon 10 Splicing Regulatory Element RNA and Destabilization by Mutations of Frontotemporal Dementia and Parkinsonism Linked to Chromosome 17. *Proc. Natl. Acad. Sci U. S. A.* 1999, 96, 8229–8234. [PubMed: 10393977]
- (24). Chen JL; Kennedy SD; Turner DH Structural Features of a 3' Splice Site in Influenza A. *Biochemistry* 2015, 54, 3269–3285. [PubMed: 25909229]
- (25). Smith JS; Nikonowicz EP NMR Structure and Dynamics of an RNA Motif Common to the Spliceosome Branch-Point Helix and the RNA-Binding Site for Phage GA Coat Protein. *Biochemistry* 1998, 37, 13486–13498. [PubMed: 9753434]
- (26). Kennedy SD; Kierzek R; Turner DH Novel Conformation of an RNA Structural Switch. *Biochemistry* 2012, 51, 9257–9259. [PubMed: 23134175]
- (27). Yildirim I; Stern HA; Tubbs JD; Kennedy SD; Turner DH Benchmarking Amber Force Fields for RNA: Comparisons to NMR Spectra for Single-Stranded r(GACC) Are Improved by Revised X Torsions. *J. Phys. Chem. B* 2011, 115, 9261–9270. [PubMed: 21721539]
- (28). Barthel A; Zacharias M Conformational Transitions in RNA Single Uridine and Adenosine Bulge Structures: A Molecular Dynamics Free Energy Simulation Study. *Biophys. J.* 2006, 90, 2450–2462. [PubMed: 16399833]
- (29). Sponer J; Bussi G; Krepl M; Banas P; Bottaro S; Cunha RA; Gil-Ley A; Pinamonti G; Poblete S; Jurecka P; et al. RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* 2018, 118, 4177–4338. [PubMed: 29297679]
- (30). Yildirim I; Kennedy SD; Stern HA; Hart JM; Kierzek R; Turner DH Revision of Amber Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iG/C Base Pairs. *J. Chem. Theory Comput.* 2012, 8, 172–181. [PubMed: 22249447]
- (31). Yildirim I; Stern HA; Kennedy SD; Tubbs JD; Turner DH Reparameterization of RNA X Torsion Parameters for the Amber Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* 2010, 6, 1520–1531. [PubMed: 20463845]
- (32). Case DA; Betz RM; Cerutti DS; Cheatham TE; Darden TA; Duke RE; Giese TJ; Gohlke H; Goetz AW; Homeyer N, et al. *Amber 16*; University of California: San Francisco, CA, 2016.
- (33). Joung IS; Cheatham TE Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* 2008, 112, 9020–9041. [PubMed: 18593145]
- (34). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983, 79, 926–935.
- (35). Onufriev A; Bashford D; Case DA Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Genet.* 2004, 55, 383–394. [PubMed: 15048829]
- (36). Wales DJ; Yildirim I Improving Computational Predictions of Single-Stranded RNA Tetramers with Revised A/T Torsional Parameters for the Amber Force Field. *J. Phys. Chem. B* 2017, 121, 2989–2999. [PubMed: 28319659]
- (37). Condon DE; Kennedy SD; Mort BC; Kierzek R; Yildirim I; Turner DH Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* 2015, 11, 2729–2742. [PubMed: 26082675]
- (38). Yildirim I; Park H; Disney MD; Schatz GC A Dynamic Structural Model of Expanded RNA CAG Repeats: A Refined X-Ray Structure and Computational Investigations Using Molecular Dynamics and Umbrella Sampling Simulations. *J. Am. Chem. Soc.* 2013, 135, 3528–3538. [PubMed: 23441937]
- (39). Banas P; Hollas D; Zgarbova M; Jurecka P; Orozco M; Cheatham TE; Sponer J; Otyepka M Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* 2010, 6, 3836–3849.
- (40). Deb I; Sarzynska J; Nilsson L; Lahiri A Conformational Preferences of Modified Uridines: Comparison of Amber Derived Force Fields. *J. Chem. Inf. Model.* 2014, 54, 1129–1142. [PubMed: 24697757]

- (41). Tubbs JD; Condon DE; Kennedy SD; Hauser M; Bevilacqua PC; Turner DH The Nuclear Magnetic Resonance of CCCC RNA Reveals a Right-Handed Helix, and Revised Parameters for Amber Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry* 2013, 52, 996–1010. [PubMed: 23286901]
- (42). Condon DE; Yildirim I; Kennedy SD; Mort BC; Kierzek R; Turner DH Optimization of an Amber Force Field for the Artificial Nucleic Acid, Lna, and Benchmarking with NMR of L(CAAU). *J. Phys. Chem. B* 2014, 118, 1216–1228. [PubMed: 24377321]
- (43). Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 1995, 117, 5179–5197.
- (44). Yildirim I; Chakraborty D; Disney MD; Wales DJ; Schatz GC Computational Investigation of RNA CUG Repeats Responsible for Myotonic Dystrophy 1. *J. Chem. Theory Comput.* 2015, 11, 4943–4958. [PubMed: 26500461]
- (45). Yildirim I; Stern HA; Sponer J; Spackova N; Turner DH Effects of Restrained Sampling Space and Non-Planar Amino Groups on Free Energy Predictions for RNA with Imino and Sheared Tandem GA Base Pairs Flanked by GC, CG, iGiC or iCiG Base Pairs. *J. Chem. Theory Comput.* 2009, 5, 2088–2100. [PubMed: 20090924]
- (46). Ryckaert JP; Ciccotti G; Berendsen HJC Numerical-Integration of Cartesian Equations of Motion of a System with Constraints: Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* 1977, 23, 327–341.
- (47). Wales DJ Discrete Path Sampling. *Mol Phys.* 2002, 100, 3285–3305.
- (48). Wales DJ Some Further Applications of Discrete Path Sampling to Cluster Isomerization. *Mol. Phys.* 2004, 102, 891–908.
- (49). Chen JL; VanEtten DM; Fountain MA; Yildirim I; Disney MD Structure and Dynamics of RNA Repeat Expansions That Cause Huntington's Disease and Myotonic Dystrophy Type 1. *Biochemistry* 2017, 56, 3463–3474. [PubMed: 28617590]
- (50). Liu DC; Nocedal J On the Limited Memory Bfgs Method for Large-Scale Optimization. *Math. Program.* 1989, 45, 503–528.
- (51). Strodel B; Whittleston CS; Wales DJ Thermodynamics and Kinetics of Aggregation for the Gnnqqny Peptide. *J. Am. Chem. Soc.* 2007, 129, 16005–16014. [PubMed: 18052168]
- (52). Wales DJ; Miller MA; Walsh TR Archetypal Energy Landscapes. *Nature* 1998, 394, 758–760.
- (53). Becker OM; Karplus M The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics. *J. Chem. Phys.* 1997, 106, 1495–1517.
- (54). Krivov SV; Karplus M Free Energy Disconnectivity Graphs: Application to Peptide Models. *J. Chem. Phys.* 2002, 117, 10894–10903.
- (55). Strodel B; Wales DJ Free Energy Surfaces from an Extended Harmonic Superposition Approach and Kinetics for Alanine Dipeptide. *Chem. Phys. Lett.* 2008, 466, 105–115.
- (56). Evans DA; Wales DJ Free Energy Landscapes of Model Peptides and Proteins. *J. Chem. Phys.* 2003, 118, 3891–3897.
- (57). Miller M; Wales DJ; de Souza V DisconnectionDPS, <http://www-wales.ch.cam.ac.uk/software.html> (accessed May 30, 2018).
- (58). Newby MI; Greenbaum NL Sculpting of the Spliceosomal Branch Site Recognition Motif by a Conserved Pseudouridine. *Nat. Struct. Biol.* 2002, 9, 958–965. [PubMed: 12426583]
- (59). Schneider B; Moravsek Z; Berman HM RNA Conformational Classes. *Nucleic Acids Res.* 2004, 32, 1666–1677. [PubMed: 15016910]

Figure 1.
Sequence of the RNA hairpin observed in MAPT pre-mRNA at the exon 10-intron junction.
Mutations destabilizing this RNA hairpin are linked to splice site selection.

Figure 2. Overlap of stacked (A), base-triple (B), and unstacked states of A-bulges observed in RNA structures determined by NMR spectroscopy (Table 1). The highlighted gray residues represent adenosine bulges.

Figure 3. NMR modeled and computationally predicted structures of A-bulges in 5'-GCAGU/5'-ACGU (A and B) and 5'-CGAGC/5'-GCCG (C and D). Blue and black colored residues are the closing base-pairs, while A-bulges are highlighted in red. Dashed red and blue lines represent Watson-Crick base-pairing hydrogen bonds and stabilizing electrostatic interactions between the A-bulge and one of the closing base-pairs, respectively. Note, that the NMR structure of A-bulge in 5'-GCAGU/5'-ACGU (A), which is in a stacked state, is not represented correctly with current computational methods (B). In contrast, the NMR structure of 5'-CGAGC/5'-GCCG (C) is correctly predicted (D).

Figure 4.

Coordinates used in 2D US calculations for 5'-GCAGU-3'/5'-ACGU-3'. Similar torsions were used to study 5'-CGAGC-3'/5'-GCCG-3'. Torsions highlighted with red in A and B, respectively, are the chi (χ) and pseudotorion (θ), which are defined as @O4'-@C1'-@N9-@C4 and :G8@C1'-:C2@C1'-:G4@C1'-:A3@C5, where ":" and "@" denote the residue and atom names, respectively.

Figure 5.

Average structure of the stacked states observed in the MD simulations of the GCAUU A-bulge (system # 8 in Table 2). The bulge region is highlighted in a ball and stick model. Note the stabilizing electrostatic interaction between A6-NH₂ and A16-N1, highlighted by a dashed blue line (see Movie S1).

Figure 6.

Pseudorotation angle, representing base flipping, versus time extracted from MD simulations of (A) 5'-CGAGC/5'-GCCG and (B) 5'-GCAGU/5'-ACGU. Pseudorotation around 30 and -50° represent stacked and base-triple states, respectively. Empty parts in the plots (e.g., the time between ~30–50 ns in (B)) represent conformations not resembling stacked or base-triple states due to distortions observed in the closing base-pairs.

Figure 7.

Disconnectivity graphs of GCAGU/ACGU (A) and CGAGC/GCCG (B) calculated using the DPS approach. Coloring is used to emphasize the conformational preferences of A-bulges. Structures highlighted with black, red, green, and blue represent base-triple, stacked, major-groove, and “other” states, respectively, which have pseudoangles between $[-100:0]$, $[0:70]$, $[70:120]$, and $[120:260]$ (Figure 8). “Other” states mostly correspond to fully unstacked A-bulges. The lowest 10 000 minima in the DPS databases were used to build the disconnectivity graphs. Note that the base-triple state is the global minimum in both A-bulge systems.

Figure 8. Base-triple (black) (A), stacked (red) (B), major-groove (green) (C), and fully unstacked (blue) (D) A-bulge states observed in DPS calculations (Figure 7). Hydrogen-bonds are shown as dashed blue lines, while the closing CG/GC base pairs are highlighted in orange.

Figure 9.

2D (θ , χ) PMF surfaces (in kcal/mol) of A-bulges in 5'-GCAGU/5'-ACGU (A) and 5'-CGAGC/5'-GCCG (B) predicted by umbrella sampling calculations. θ around -180 (a), -60 (b), 50 (c), and 70° with $\chi \sim 200^{\circ}$ represent the fully unstacked, base-triple, stacked, and major-groove states, respectively (Figure 8). Syn orientation of A is represented at $\chi \sim 60^{\circ}$ (d, e, f, and g). Note that, in both systems, the global minimum ($\theta \sim -60^{\circ}$, $\chi \sim 200^{\circ}$) conformation represents the base-triple state.

Table 1.

Conformations of Adenosine in A-Bulges Solved by NMR Spectroscopy

PDB ID	A-bulge sequence	conformation
1D0U ²²	5'-UC <u>A</u> CC-3'	stacked
	3'-AG.GG-5'	
1QC8 ²³ and 1EI2 ¹⁶	5'-GC <u>A</u> GU-3'	stacked
	3'-UG.CA-5'	
1K8S ¹⁵	5'-AG <u>A</u> GU-3'	stacked
	3'-UC.CA-5'	
1TFN ¹⁴ and 1RHT ¹³	5'-UG <u>A</u> CG-3'	stacked
	3'-AC.GC-5'	
2JXS ¹⁸	5'-AG <u>A</u> AG-3'	stacked
	3'-UC.UC-5'	
17RA ²⁵	5'-UA <u>A</u> GG-3'	stacked
	3'-AU.CC-5'	
1XSG ¹⁷	5'-AG <u>A</u> CC-3'	base-triple
	3'-UC.GG-5'	
2K3Z ¹⁹	5'-CG <u>A</u> GC-3'	base-triple
	3'-GC.CG-5'	
2MQT ²¹ and 2MQV ²¹	5'-CU <u>A</u> CC-3'	base-triple
	3'-GG.GG-5'	
2MXL ²⁴	5'-GGA <u>A</u> UG-3'	base-triple
	3'-CC.AU-5'	
2XEB ²⁰	5'-GU <u>A</u> GC-3'	unstacked
	3'-CG.CG-5'	

Table 2.Conformational Analyses of the A-Bulge Systems (Base-Triple vs Stacked)^a

system	model A-bulge	observed ^b A-bulges	base-triple (%)	stacked (%)
1	5'-CCGGA <u>A</u> AUGUG-3'	GG <u>A</u> AA	3.5	0.2
	3'-GGCCU.UACAC-5'	GA <u>A</u> AU	8.7	0.7
		AA <u>A</u> UG	79.0	8.0
2	5'-CCGGA <u>A</u> CUGUG-3'	GA <u>A</u> CU	98.5	0.1
	3'-GGCCU.GACAC-5'	GA <u>A</u> CU	0.8	0.6
3	5'-CCGGA <u>A</u> GUGUG-3'	GG <u>A</u> AG	1.3	0.5
	3'-GGCCU.CACAC-5'	GA <u>A</u> GU	96.5	1.7
4	5'-CCGGA <u>A</u> UUGUG-3'	GG <u>A</u> AU	7.3	0.6
	3'-GGCCU.AACAC-5'	GA <u>A</u> UU	65.6	26.5
5	5'-CCGGC <u>A</u> AUGUG-3'	GC <u>A</u> AU	0.2	1.2
	3'-GGCCG.UACAC-5'	CA <u>A</u> UG	81.9	16.8
6	5'-CCGGC <u>A</u> CUGUG-3'	GC <u>A</u> CU	92.9	7.1
	3'-GGCCG.GACAC-5'			
7	5'-CCGGC <u>A</u> GUGUG-3'	GC <u>A</u> GU	98.3	1.7
	3'-GGCCG.CACAC-5'			
8	5'-CCGGC <u>A</u> UUGUG-3'	GC <u>A</u> UU	49.4	50.6
	3'-GGCCG.AACAC-5'			
9	5'-CCGGG <u>A</u> AUGUG-3'	GG <u>A</u> AU	8.4	2.6
	3'-GGCCC.UACAC-5'	GA <u>A</u> UG	62.4	26.7
10	5'-CCGGG <u>A</u> CUGUG-3'	GG <u>A</u> CU	100.0	0.0
	3'-GGCCC.GACAC-5'			
11	5'-CCGGG <u>A</u> GUGUG-3'	GG <u>A</u> GU	96.1	3.9
	3'-GGCCC.CACAC-5'			
12	5'-CCGGG <u>A</u> UUGUG-3'	GG <u>A</u> UU	69.9	30.1
	3'-GGCCC.AACAC-5'			
13	5'-CCGGU <u>A</u> AUGUG-3'	GU <u>A</u> AU	99.8	0.2
	3'-GGCCA.UACAC-5'	UA <u>A</u> UG	0.0	0.0
14	5'-CCGGU <u>A</u> CUGUG-3'	GU <u>A</u> CU	100.0	0.0
	3'-GGCCA.GACAC-5'			
15	5'-CCGGU <u>A</u> GUGUG-3'	GU <u>A</u> GU	100.0	0.0
	3'-GGCCA.CACAC-5'			
16	5'-CCGGU <u>A</u> UUGUG-3'	GU <u>A</u> UU	99.4	0.6
	3'-GGCCA.AACAC-5'			
17	5'-CCGGC <u>A</u> GUGUG-3'	GC <u>A</u> GU	100.0	0.0
	3'-GGCUG.CACAC-5'			
18	5'-CCGCG <u>A</u> GCGUG-3'	CG <u>A</u> GC	99.5	0.5
	3'-GGCGC.CGCAC-5'			

^aIf an A-bulge is closed by AU base-pairs, base-slipping can occur, creating a modified version of the A-bulge (systems # 3–7, 11, and 15). Except for systems 1 and 2, the sequence of the model RNA system used to study the effect of closing base-pairs in A-bulges is 5'-CCGGXAYUGUG-375'-CACAYX*CCGG-3', where X, Y, X*, and Y* = A, C, G, and U, and XX* and YY* form Watson–Crick base-pairs (see also Figures 3 and S1–S16).

^bWhen there are adenosine residues neighboring an A-bulge, base-slipping can occur, which will create a different A-bulge sequence, as noted in the table.