ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

RNA as a Major-Groove Ligand: RNA−RNA and RNA−DNA Triplexes Formed by GAA and UUC or TTC Sequences

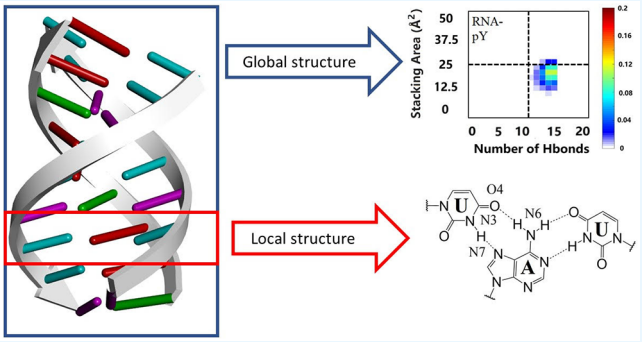
Jiahui Zhang, Ashkan Fakharzadeh, Christopher Roland, and Celeste Sagui*

Cite This: *ACS Omega* 2022, 7, 38728−38743

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Friedreich's ataxia is associated with noncanonical nucleic acid structures that emerge when GAA:TTC repeats in the first intron of the FXN gene expand beyond a critical number of repeats. Specifically, the noncanonical repeats are associated with both triplexes and R-loops. Here, we present an *in silico* investigation of all possible triplexes that form by attaching a third RNA strand to an RNA:RNA or DNA:DNA duplex, complementing previous DNA-based triplex studies. For both new triplexes results are similar. For a pyridimine UUC$^+$ third strand, the parallel orientation is stable while its antiparallel counterpart is unstable. For a neutral GAA third strand, the parallel conformation is stable. A protonated GA$^+$A third strand is stable in both parallel and antiparallel orientations. We have also investigated Na^+ and Mg^{2+} ion distributions around the triplexes. The presence of Mg^{2+} ions helps stabilize neutral, antiparallel GAA triplexes. These results (along with previous DNA-based studies) allow for the emergence of a complete picture of the stability and structural characteristics of triplexes based on the GAA and TTC/UUC sequences, thereby contributing to the field of trinucleotide repeats and the associated unusual structures that trigger expansion.



INTRODUCTION

RNA plays a crucial role in fundamental biological process such as coding, decoding, regulation, and gene expression.[1−4] Consequently, there are many types of RNAs: messenger RNA (mRNA) is used by cellular organisms to convey genetic information that directs protein synthesis;[5] transfer RNA (tRNA) delivers amino acids to ribosomes for protein synthesis;[6] ribosomal RNA (rRNA) links amino acids together to form coded proteins;[5] transfer-messenger RNA(tmRNA) tags proteins encoded by mRNAs that lack stop codons for degradation and prevents the ribosome from stalling;[7] small nuclear RNA (snRNA) is another type of RNA, whose primary function is the processing of premessenger RNA in the nucleus.[8] Given the number of different types of RNA and their flexibility, it is not surprising that there exists a variety of RNA secondary structures.[9] Examples of rather simple motifs that can form include the following: an RNA single strand can fold back on itself and form an RNA hairpin;[10] two complementary RNA strands can come together and form an A-RNA[11] or a Z-RNA helix;[12] three RNA strands with the proper complementary sequence can form an RNA triplex;[13] while four stranded C- and G-rich RNA can form either an i-motif[14] or a G-quadruplex RNA.[15] RNA can bind to proteins,[16,16,17] DNA,[18,19] and drug molecules.[20,21] It is believed that the complex biological functions undertaken by RNA are related to its binding, and so many investigations of complex biological functions and pathological biological mechanisms[22−26] focus on this important aspect.

In this work, we present results about RNA as a major-groove ligand in DNA and RNA helical duplexes, which results in pure RNA triplexes or in RNA−DNA hybrid triplexes. In particular, we study triplexes formed by the GAA/TTC(UUC) trinucleotide repeats whose expansion is behind Friedreich's ataxia (FRDA) disease. FRDA is caused by a GAA expansion in the first intron of the frataxin (FXN) gene, and is a member of the so-called trinucleotide repeat expansion diseases (TREDs).[27−32] TREDS are caused by the expansion of trinucleotide repeats (TRs), which gives rise to pathological disorders after the expansion goes beyond a critical threshold in the length of the repeat. TREDs are typically inherited neurological disorders that exhibit a phenomenon known as "anticipation", where the age of the onset of the disease typically decreases and the severity of the disease phenotype increases with each successive generation.[30,33−36] Longer repeat tracts become progressively more deleterious and

Received: July 10, 2022
Accepted: October 4, 2022
Published: October 24, 2022

constitute the dominant molecular determinant of anticipation in a significant number of diseases, with other genetic modifiers and environmental factors accounting for the remainder of the effects.[36,37] The critical step in all models of repeat instabilities is the transient formation of atypical, stable secondary structures in the expandable repeats.[38−42] Expandable repeats can lead to atypical structures such as single-stranded hairpins, Z-DNA/RNA, triple helices, G-quadruplexes, i-motifs, slipped-stranded duplexes, and R-loops. As may be anticipated, the quest to understand the mechanisms behind these diseases has also led to a better understanding of the lesser known nucleic acid structures, thereby contributing to the basic science of nucleic acid research.

A wide variety of triple helices may be formed by combining RNA and DNA strands. The triplex itself consists of an antiparallel helical duplex, where the bases of the two antiparallel strands form traditional Watson−Crick base pairs, and a third strand in the major groove of the duplex attached via Hoogsteen (or Hoogsteen-like) hydrogen bonds to the purine strand of the duplex. In principle, using the notation ":" for a double helix and "·" for the third strand attached to the double helix, a first classification of these triplexes includes DNA·DNA:DNA and RNA·RNA:RNA (pure DNA and RNA triplexes); RNA·DNA:DNA and DNA·RNA:RNA (where the duplex is formed by the same nucleic acid type and the third strand belongs to the other nucleic acid type); and DNA·DNA:RNA and RNA·DNA:RNA (where the duplex itself is a hybrid, formed by Watson−Crick base pairs between DNA and RNA). In turn, each of these classes comprises a family of different, *potential* triplexes (not all of which are stable, as we showed in our previous work[43]). Following standard notation,[44] the double helix in the triplex can be written as R:Y, where R represents the purine-rich strand (not to be confused with the "R" in "R-loops") and Y the pyrimidine-rich strand, respectively. For the chosen sequences, R = GAA and Y = TTC (DNA) or Y = UUC (RNA). Thus, in each family, the duplex is always R:Y and the third strand can be either R or Y. The next step for building the potential triplex is the choice of orientation of the third strand. By convention, in a "parallel" ("antiparallel") triplex, the third strand is parallel (antiparallel) to the R strand in the duplex. Other important decisions for building the triplex involve the protonation state of bases in the third strand (for instance, unprotonated C in the third Y strand results in an unstable triplex, while one of the As in the R = GAA, third strand can be protonated or unprotonated), and the alignment of the third strand with respect to the duplex, since shifting it results in different sets of steps.

In our previous work,[43] we reported on a systematic characterization of eight possible pure DNA triplexes that could be assembled with GAA and TTC strands. We also presented results about the two hybrid duplexes [r(GAA):d(TTC) and d(GAA):r(UUC)] that could form in an R-loop with bidirectional transcription. An R-loop is a three-stranded nucleic acid structure consisting of a hybrid RNA:DNA duplex formed by the template DNA and the RNA strands, along with a displaced, nontemplate, single-strand DNA. Finally, we also studied the three hybrid triplexes that could form during bidirectional transcription when the nontemplate DNA strand bonds with the hybrid duplex (a collapsed R-loop, DNA·DNA:RNA, where the two DNA strands are antiparallel). We found that for both Y·R:Y and R·R:Y DNA triplexes, the parallel third strand orientation is more stable, while both parallel and antiparallel protonated d(GA^+A)·d(GAA):d(TTC) triplexes are stable. Apparent contradictions in the literature about the R·R:Y triplex stability are probably due to the lack of experimental molecular resolution, since shifting the third strand by one or two nucleotides alters the stability of the triplex. Among the collapsed R-loops, antiparallel d(TTC^+)·d(GAA):r(UUC) is unstable, while parallel d(GAA)·r(GAA):d(TTC) and d(GA^+A)·r(GAA):d(TTC) are stable.

In this work, we present results for pure RNA triplexes and for hybrid triplexes of the form RNA·DNA:DNA. Naturally occurring RNA triplexes play important roles in many biological functions.[45−49] For example, RNA triplexes have been detected in telomerase RNAs,[50−54] metabolite-sensing riboswitches,[55−62] −1 ribosomal frameshift-inducing mRNA pseudoknots,[63−67] long noncoding RNAs,[68−71] group I introns,[72−74] group II introns,[75] and rRNAs.[76−78] Since it is known that the minor-groove RNA triplexes are usually not stable in isolation, we do not consider such structures in this work, and focus only on the more stable major-groove RNA triplexes.[79]

The most common RNA−DNA hybrid triplex is RNA·DNA:DNA, with standard Watson−Crick (WC) base pairing in the DNA duplex and the RNA strand binding to the major-groove of the DNA duplex via Hoogsteen (or Hoogsteen-like) hydrogen bonds.[80] It has been reported that this RNA−DNA hybrid triplex can be formed by long noncoding RNA in vivo.[80] Recent studies show that this kind of RNA−DNA hybrid triplex is related to the TGF-β pathway gene[81] and participates in the regulation of the β-globin locus.[82] Although a few pure RNA and RNA·DNA:DNA hybrid triplexes have been determined experimentally, the molecular structures for the GAA and TTC(UUC) sequences are largely unknown. The aim of our work is to provide a systematic characterization of these structures and their relative stability. To that end, we employ classical molecular dynamics (MD) simulations as our main investigative tool. MD simulations have proven themselves to be extremely valuable as they possess the ability to probe molecular structures, dynamics, and mechanisms at the atomic level, which are often beyond the resolution of current experiment. Our recent characterization of homoduplexes, quadruplexes and hairpins conformations corresponding to the most common TRs and to several hexanucleotide repeats is an example of this,[43,83−90] as these simulations sample both DNA and RNA sequences of different lengths, different nonequivalent nucleotide arrangements (such as $(GCC)_n$ and $(CCG)_n$ homoduplexes, with CpG and GpC steps between the C−C mismatches); provide free energies, dynamics of conformational transitions, etc. These conformation studies do not address the formation of the given atypical structures: In order for these atypical structures to nucleate, it is necessary to cross a free energy barrier that is both sequence and repeat-length dependent. One of the advantages of MD simulations is one can start the structure in any minimum of the free energy landscape. Thus, one can simply study the resulting structure after its nucleation has taken place (similarly to studying a folded protein after the folding has taken place), and the length-dependence for the nucleation becomes irrelevant. We expect the present study of RNA triplexes to contribute to the widening knowledge of RNA-based atypical structures.

METHODS

Here, we briefly discuss three aspects of this work: the initial construction of the RNA-based triplexes, the subsequent MD

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

simulations, and the analysis. The initial structures for these RNA triplexes are based on well-equilibrated DNA triplexes as reported in our previous work.[43] We then mutated the sugar rings and bases to obtain the RNA sequences used in this study. Although the initial duplex generated in this form is in B-DNA form, crossover to the A-RNA form occurs relatively fast. Indeed, in our previous work, we showed that full convergence between different initial ideal B- and A-forms in hybrid duplexes and triplexes is achieved quite early in the simulations. The initial structures are further described in the next section.

The MD simulations of the triplexes were carried out with the *Amber18* package[91] with force field BSC1[92] for the DNA and BSC0[93] + OL3[94] RNA parts, respectively. Protonation as needed was accomplished with the protonated *AMBER* force field[95] using tleap.[91] The TIP3P water model[96] was used for the explicit solvent. Periodic boundary conditions were implemented in a truncated octahedron water box and the appropriate number of Na^+ ions (parameters in ref 97) were added in order to neutralize the nucleic acid charges. Electrostatics were handled by the Particle Mesh Ewald Method[98] with a direct space cutoff of 9 Å; the same cutoff was used for the van der Waals interaction. We used Langevin dynamics to control the temperature with a coupling parameter of 1.0 ps^{-1}. The *SHAKE* algorithm[99] was applied to all bonds involving hydrogen atoms.

In addition, we ran simulations in the presence of Mg^{2+} ions, motivated by their physiological importance and experiments showing that these ions help promote triplex formation.[100,101] We therefore added 14 Mg^{2+} ions (concentration of 280 mM with parameters given in ref 102) randomly to the water box, but not closer to 3.5 Å to the nucleic acids. Quickly, the cations became hexahydrated forming complex $Mg[(H_2O)_6]^{2+}$ ions and remained in that hydration state during the length of the simulations. The necessary Cl^- ions[97] were also added and carefully equilibrated.

The initial conformations for the MD simulations were obtained as follows. After mutating the DNA to RNA, MD runs with a constraint of 20 kcal/mol on the triplex hydrogen bonds were performed in order to ensure initial structural stability up to 20 ns. These triplexes were then considered as starting conformations and we carried out a full equilibration procedure on them, starting with energy minimization. Subsequently, the temperature of the system was gradually raised using conditions of constant volume from zero to 300 K over 50 ps runs with a 1 fs time step. Then, 100 ps runs at constant volume were employed to gradually reduce the harmonic restraining force for the nucleic acids and ions. During the equilibration runs, a weak constraint of 1 kcal/mol was placed on the hydrogen bonds of the end bases in order to reduce any artificial fraying effects. The MD production runs were performed over 1 μs with a 2 fs time step under conditions of 1 atm constant pressure. For analysis, conformations were saved every 20 ps. Typically, a few hundred nanoseconds into the MD simulations, some structures maintained a stable hydrogen-bond pattern, while others evolved to form different stable structures, or simply became unstable.

Turning to the analysis, we note that from a structural point of view, the two most important quantities stabilizing the third nucleic acid strand are the hydrogen bonds with the R strand of the antiparallel double helix and the $\pi-\pi$ stacking interaction between the bases of the third strand. In the analysis, hydrogen bonds were identified using Cptraj as supplied by *Ambertools18*,[91] with a distance cutoff of 3.5 Å and an angular cutoff of 140°. Base-pair stacking was conveniently analyzed using a heuristic method described in our previous publication on DNA triplexes[43] and summarized in the Supporting Information .

Additionally, we found it convenient to introduce two other structural parameters—distance d and angle ϕ—which help reveal geometrical aspects of the triplexes. These are shown in Figure 1.

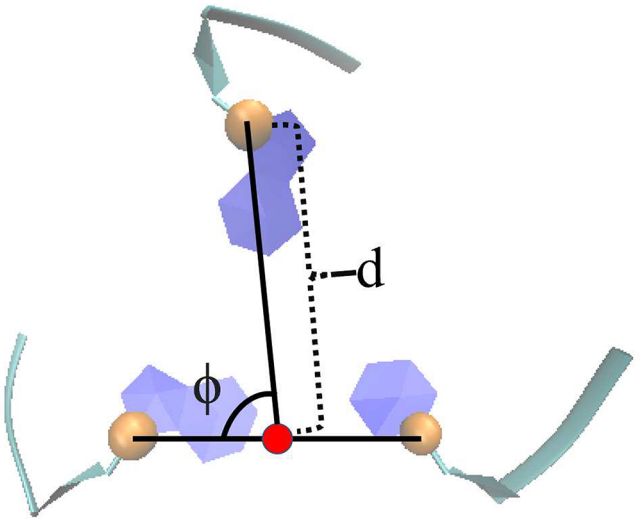


Figure 1. Schematic illustrating the definition of parameters ϕ and d.

The parameter d is defined as the distance between the N9 atom of purine or the N1 atom of pyrimidine on the third strand and the average position of the corresponding N atoms of the WC base pairs in the duplex. The angle ϕ is defined as the left angle between the line connecting the N9 or N1 atoms of the third strand and the line connecting the average position of the corresponding N atoms of the WC base pairs. These structural parameters characterize the distance from the third strand to the WC base pair in the duplex. A positive correlation with $\phi > 90°$ indicates that the third strand is closer to the pyrimidine base, most likely with enhanced hydrogen bonding, while $\phi < 90°$ indicates that the third strand is closer to the purine base. The presence of more than one cluster on the $\phi-d$ diagram signals the presence of more than one hydrogen-bond pattern for the third strand.

RESULTS

Structure of the Initial RNA·RNA:RNA and RNA·DNA:DNA Triplexes. The triple helices investigated consist of a pure DNA or pure RNA antiparallel double helix with WC base pairs, with a single RNA strand binding to the major groove of the duplex. The third RNA strand may be either a pyrimidine or a purine, and these may be either parallel or antiparallel to the purine strand of the duplex. For the case of a pyrimidine third strand, we only considered the case of protonated cytosines, as supported by experimental evidence.[103−107] This results in two cases where the third strand can form hydrogen bonds with the duplex: the protonated pyrimidine third strand in either a parallel or antiparallel direction, with the latter shifted by a single base. For the purine third strand, we considered the strand in either parallel and

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

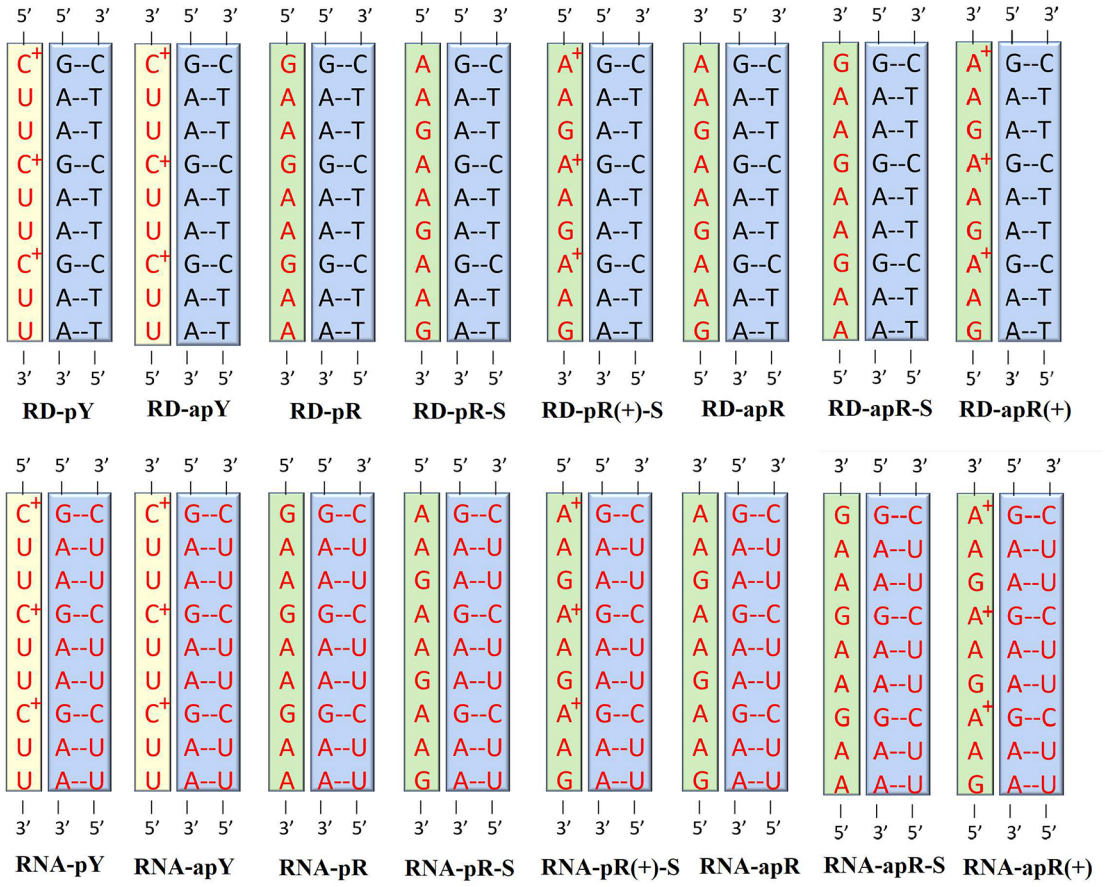


Figure 2. Schematic illustrating all the nonequivalent triplexes investigated. WC bonded structures are encased in blue boxes; the pyrimidine and purine third strands are in yellow and green boxes, respectively. Top row indicates the eight RNA−DNA triplexes, while the bottom row indicates the pure RNA triplexes. RNA and DNA bases are written in red and black, respectively. We note that we only consider the third strand nucleic bases to be in their *anti* conformation, as the *syn* is always higher energy and less stable.[43]

antiparallel orientation. We also consider the case where the trinucleotides of the third strand are perfectly aligned with those of the duplex and where the third strand is shifted with respect to the duplex. We also included two protonated cases, since protonated adenines can form stable hydrogen bonds with G:C base pairs.[108] This leads to 16 different structures, which are shown in Figure 2.

For discussion purposes, it is convenient to introduce the following notation for the different triplexes. When the third strand is UUC or GAA, we use the notation Y or R, respectively. The C's in the third UUC are always protonated, so this need not be indicated by extra notation. On the other hand, the first A in the GAA third strand can be either neutral or protonated in order to form A^+·G-C triple base planes with G-C WC pairs of the helical duplex. In this case, we use the R(+) notation. In order to indicate the parallel or antiparallel nature of the third strand, we precede the Y or R by p or ap. If the strand is shifted, we end the notation with "-S". If we are dealing with a pure RNA (DNA) triplex, we place an "RNA-" ("DNA-") before the structural notation; likewise we use the prefix "RD-" to indicate a mixed triplex in which an RNA third strand is combined with a B-DNA double helix. For example, "RD-pR(+)-S" refers to a hybrid RNA·DNA:DNA triplex with a parallel purine third strand formed by GA^+A TRs shifted with respect to the trinucleotides in the helical duplex. Finally, we note that structures with a DNA third strand placed in the major groove of an RNA double helix (DNA·RNA:RNA) have not been reported in the literature, and it is not clear whether they are biologically relevant.

The initial hydrogen-bond pattern of these structures is similar to those of the DNA triplexes previously considered.[43] The hydrogen-bond patterns include traditional Hoogsteen (H) and reverse Hoogsteen (RH) patterns, as well as new patterns. These hydrogen bonds are shown in Figure 3, while Table 1 gives the notation of each structure along with the corresponding hydrogen bond type.

MD Results for RNA·RNA:RNA and RNA·DNA:DNA Triplexes. In this section, we present the MD results for pure RNA and hybrid triplexes in the presence of neutralizing Na^+ ions. Figure 4 shows the final pure RNA triplex structures obtained after 1 μs MD runs. The top view allows for a visual appreciation of the third-strand attachment (or lack thereof), while the side view gives an idea of the base stacking. Clearly, not all configurations are stable as evidenced by the detachment of the third strand from the triplex structure. Visually, only RNA-pY (pyrimidine third strand) and RNA-pR and RNA-apR(+) (purine cases) appear to be stable, while RNA-pR(+)-S and RNA-apR-S appear to be marginally stable. In order to obtain a quantitative, statistical assessment of the structural stability, we computed the effective stacking area versus the number of hydrogen bonds, as defined in the previous section (Figure 5), over the last 800 ns of the simulations. The most stable triplexes are characterized by more hydrogen bonds and larger stacking areas, which

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

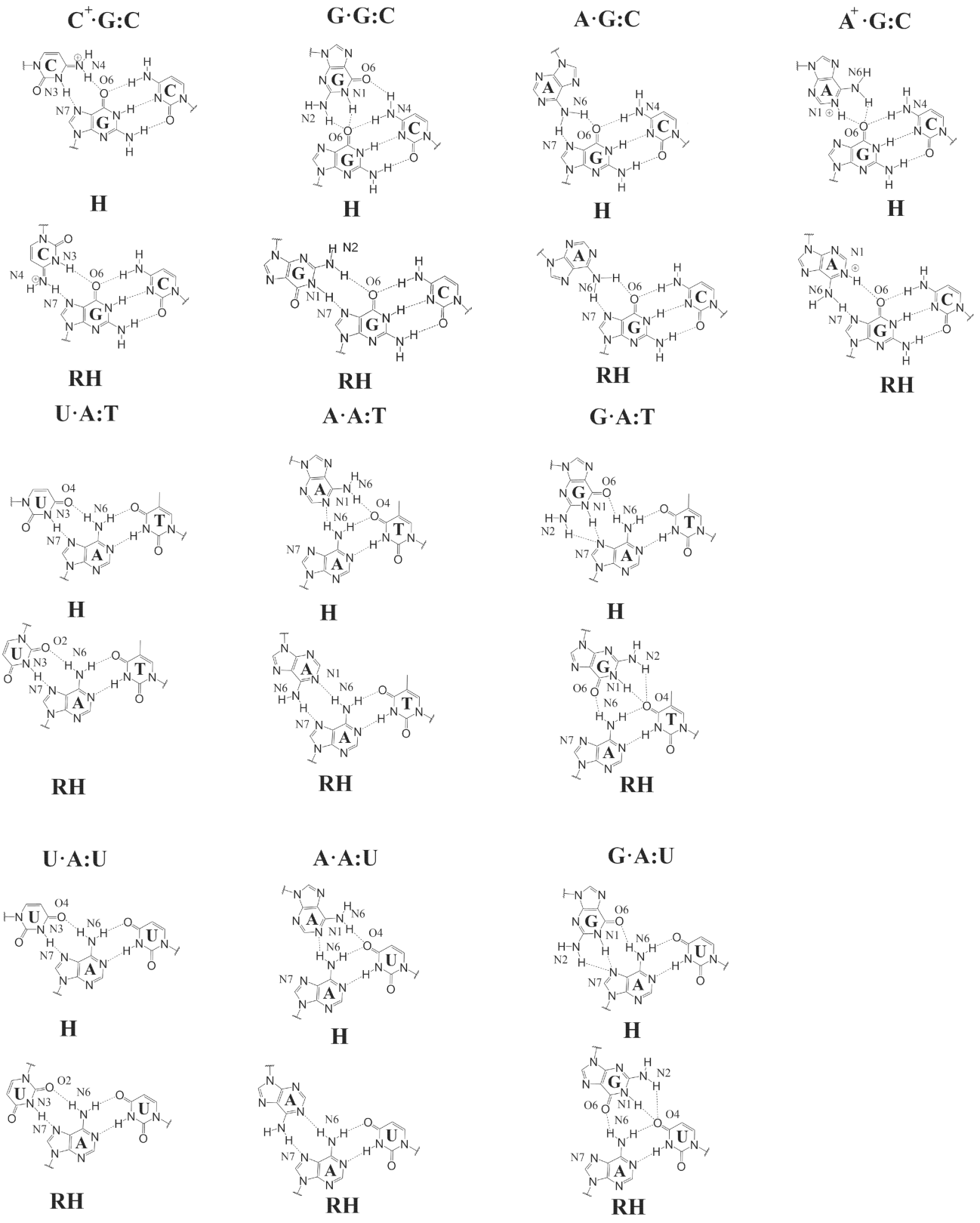


Figure 3. Initial hydrogen-bond pattern for all triplexes.

corresponds to distributions in the upper right quadrant of these plots.

In terms of stability, the graphs indicate RNA-apR(+) > RNA-pY > RNA-pR ≃ RNA-pR(+)-S. Thus, RNA-pY is the only stable pyrimidine structure and RNA-apR(+) is the most stable purine case. Note that RNA structures apY, pR-S, and apR are quite unstable, registering close to zero stacking and hydrogen bond number; the third strand peels off from the duplex during the first 300 ns of the simulation. Figures 6 and 7 show the same analysis for the RNA·DNA:DNA hybrid

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

Table 1. Summary of Nucleotide Triplets for Both Pure and Hybrid RNA Triplexes

triple step type	corresponding triple sequence
H-type C^+·G:C	pY
RH-type C^+·G:C	apY
H-type G·G:C	pR
RH-type G·G:C	apR-S
H-type A·G:C	pR-S
RH-type A·G:C	apR
H-type A^+·G:C	pR(+)-S
RH-type A^+·G:C	apR(+)
H-type U·A:T(U)	pY
RH-type U·A:T(U)	apY
H-type G·A:T(U)	pR-S
	pR(+)-S
RH-type G·A:T(U)	apR
	apR(+)
H-type A·A:T(U)	pR
	pR-S
	pR(+)-S
RH-type A·A:T(U)	apR
	apR-S
	apR(+)-S

triplexes. Combining information from the final configurations in Figure 6 and the statistical analysis in Figure 7, the stability ranking is RD-apR(+) > RD-pR(+)-S > RD-pY > RD-pR ≃ RD-apR-S.

We now turn to an examination of the dominant hydrogen-bond patterns associated with these stable structures. For descriptive purposes, we use the following notation to describe a hydrogen bond: A(N)-B(M), where A and B represent the donor and acceptor atoms, and N and M are the bases containing atoms A and B. The initial hydrogen-bond pattern evolves with time. Figure 8 shows the dominant hydrogen bonds for the inner three steps only (to avoid edge effects).

Hydrogen bond patterns are as follows. (i) RNA-pY: Two hydrogen bonds on the fourth plane, N4(C^+)-O6(G) and N3(C^+)-N7(G), and two hydrogen bonds on the fifth and sixth planes, N6(A)-O(U) and N3(U)-N7(A), are very stable throughout the simulation. (ii) RNA-pR: A pattern of three hydrogen bonds on the fourth plane, N4(C)-O6(G), N1(G)-O6(G), and N2(G)-O6(G), coexists with a pattern of only one bond, N2(G)-O6(G), or none. On the fifth and sixth planes, the dominant hydrogen-bond pattern, N6(A)-O4(U) and N6(A)-N1(A), fluctuates with only one of these bonds, and the other is absent. On the sixth plane, occasionally there are no hydrogen bonds present. (iii) RNA-pR(+)-S: On the fourth plane, the dominant hydrogen bonds, N1(A^+)-O6(G) and N6(A^+)-O6(G), coexist with a pattern in which N1(A^+)-O6(G) is replaced by N1(A^+)-N7(G). On the fifth plane, the hydrogen bonds are N6(A)-N1(A) and N6(A)-O4(U); while on the sixth plane they are N2(G)-N7(A), N1(G)-N7(A) and N6(A)-O6(G). (iv) RNA-apR-S: On the fourth plane, the dominant hydrogen bonds, N1(G)-N7(G) and N2(G)-O6(G), coexist with a similar pattern, N1(G)-N7(G) and N2(G)-N7(G), and with only one-bond pattern, N2(G)-O6(G). On the fifth and sixth planes, the dominant hydrogen-bond pattern is N6(A)-N1(A) and N6(A)-N1(A). On the fifth plane, hydrogen bonds can also disappear, and different stacking effects are observed as the position and orientation of the third-chain base changes. On the sixth plane, an alternative pattern, only N2(A)-O4(U), was observed. (v) RNA-apR(+): On the fourth plane, there is only one hydrogen-bond pattern, N1(A^+)-O6(G) and N6(A^+)-G(N7). On the fifth plane, the dominant hydrogen-bond pattern, N6(A)-N1(A) and N6(A)-N7(A), alternates with a single N6(A)-O4(U) bond. On the sixth plane, the dominant pattern, N2(G)-O4(U), N1(G)-O4(U) and N6(A)-O6(G), alternates with the hydrogen-bond pattern N2(G)-N7(A) and N1(G)-N7(A).

Similar hydrogen-bond patterns were observed for the mixed RNA·DNA:DNA hybrid triplexes (with T replacing U in DNA, of course). These are discussed in the SI (see Figure S2). All the major hydrogen bonds for the stable cases are exactly the same as the initial hydrogen-bond patterns; in turn, these are the same as the hydrogen-bond patterns of the corresponding DNA triplexes covered in our previous work.[43]

To further characterize the geometry of the RNA triple helices, we plotted the ϕ-d diagrams (Figure 9) for the three innermost planes of the triplexes.

There is little variation of the distributions from layer to layer. For all the three planes, $\phi < 90°$, indicating that the third strand is closer to the purine strand of the RNA duplex. For the pY triplex, the three planes have an average ϕ around 50° and an average d around 9 Å. For the purine third strand, both ϕ

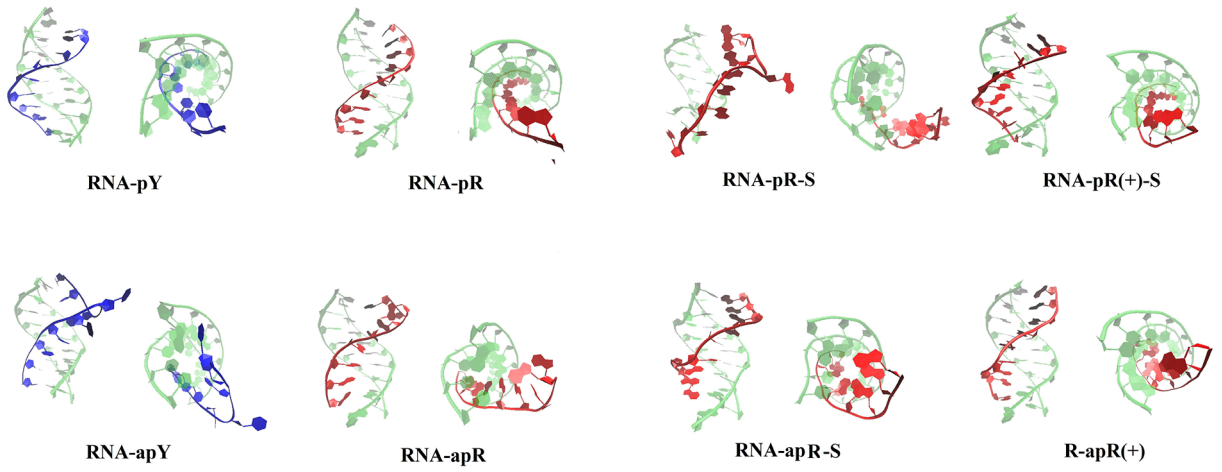


Figure 4. Final conformations of the RNA triplexes after 1μs MD simulations. The RNA WC antiparallel helix is colored light green, while the third RNA strand is colored blue (pyrimidine) or red (purine).

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

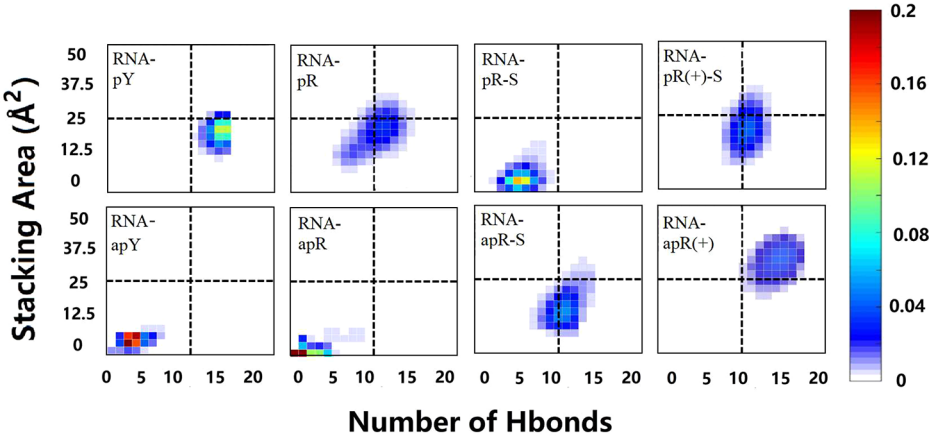


Figure 5. Effective stacking area versus effective hydrogen bond number of RNA triplexes.

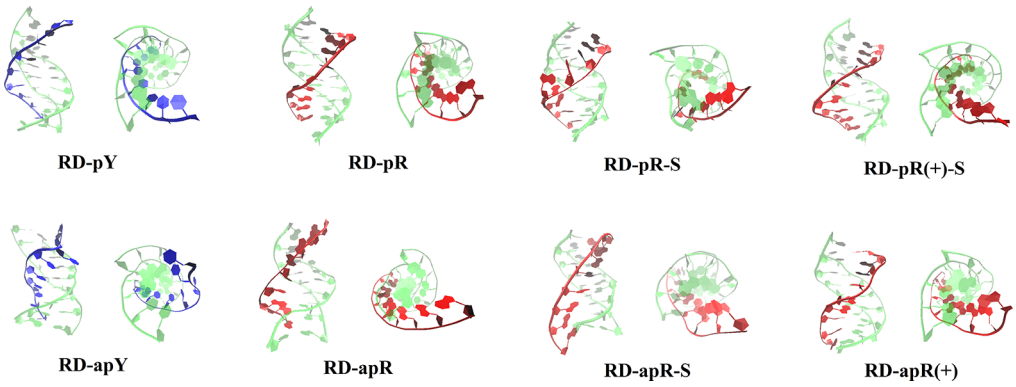


Figure 6. Final conformations of the RNA·DNA:DNA hybrid triplexes after 1 μs MD simulations. The DNA WC antiparallel helix is colored light green, while the third RNA strand is colored blue (pyrimidine) or red (purine).

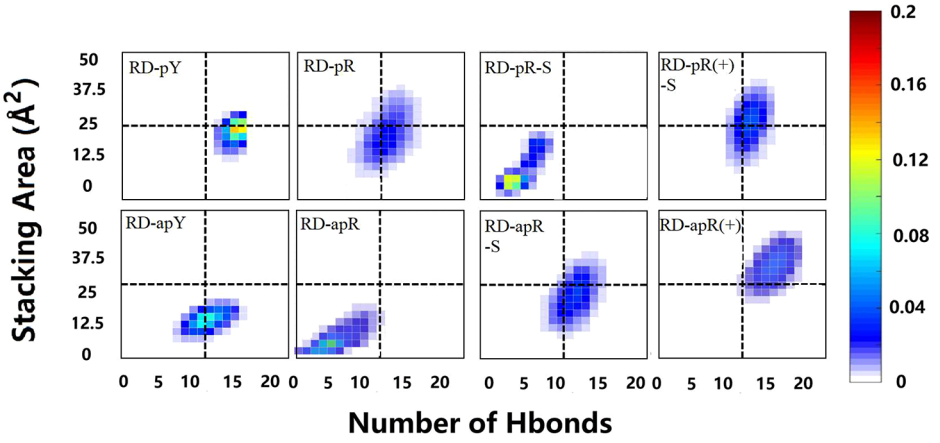


Figure 7. Effective stacking area versus effective hydrogen bond number for the RNA−DNA hybrid triplex structures.

and d tend to have larger values than for the pyrimidine strand, which still indicates a closer proximity to the purine strand of the duplex, but less pronounced than for the pyrimidine case. For pR, all three planes have a value ϕ around 80° and a d around 10 Å. For the pR(+)-S case, the fourth and fifth planes have a ϕ near 70° and a d around 10 Å. The distribution of the sixth plane is shifted a bit toward the left, with an average ϕ around 60° and d around 10.5 Å. For apR-S, the fourth plane has $d \simeq 10$ Å and $\phi \simeq 65°$, while the fifth plane has an average $\phi \simeq 75°$ and $d \simeq 9.5$ Å. While the fourth and the fifth planes' distribution indicate a third strand preference for the purine RNA strand, the sixth plane is characterized by two clusters. The major one is very similar to that of the fifth plane with a slight shift to the right, while the minor distribution shows a positive correlation. Finally for apR(+), the average ϕ for the fourth plane is 80° and the average d is slightly above 9 Å. The major distribution on the fifth plane has an average $\phi \simeq 75°$ and an average $d \simeq 9.5$ Å. The sixth plane has $\phi \simeq 90°$ and $d \simeq 10$ Å. The results for the hybrid triplex closely resemble those for the pure RNA triplex and are shown in Figure S3.

We looked at the conformations of the Watson−Crick helical duplexes that form part of the triplex, and compared

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

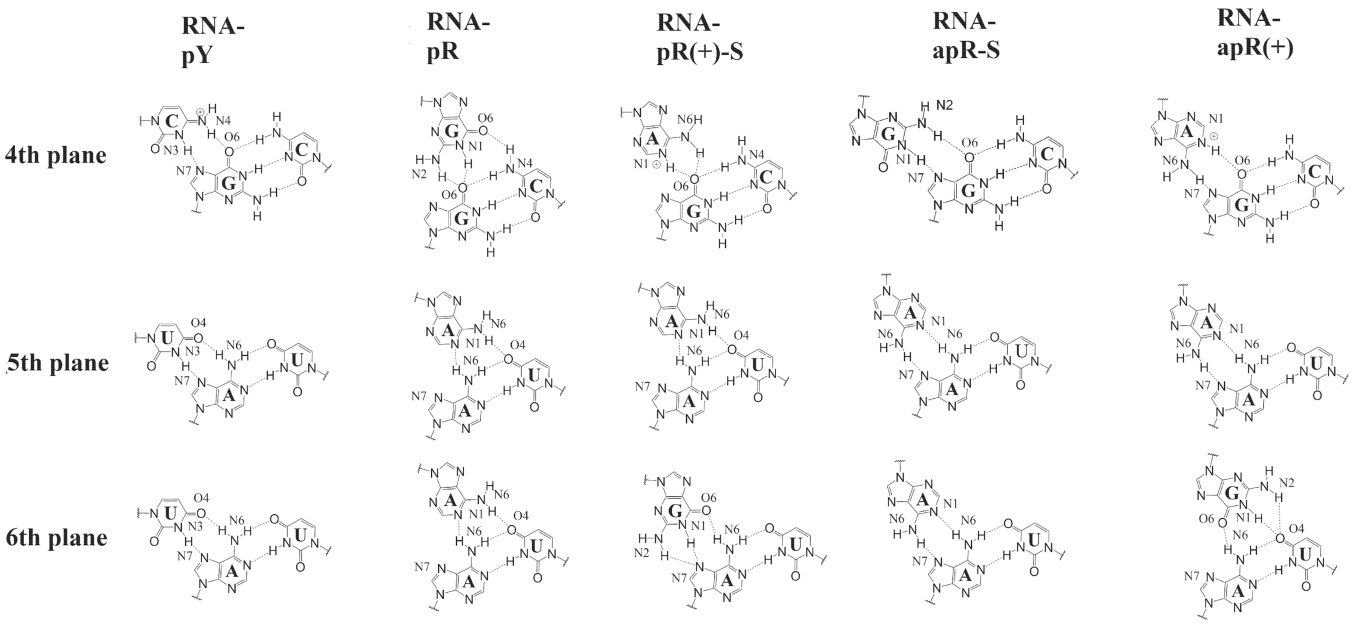


Figure 8. Dominant hydrogen-bond pattern for RNA triple helices during the last 800 ns of the MD simulation.

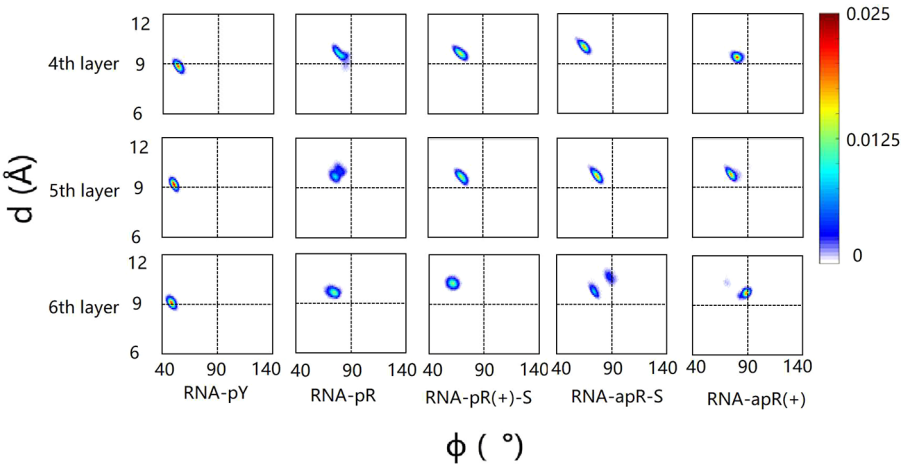


Figure 9. (ϕ, d) diagram for the RNA triplexes.

them to the corresponding free-standing B-DNA and A-RNA helices of the same sequence. First, we consider the DNA helical duplexes that belong to pure DNA triplexes (Figure S12) and to the hybrid RNA·DNA:DNA triplexes (Figure S13). The attachment of the third strand changes the conformations of these DNA duplexes with respect to the corresponding B-DNA free helix. For the helical twist, the values of both DNA and RD triplex cases are distributed between 25° and 40°, spanning values between B-DNA and A-RNA. Inclination angles are distributed between −5° and 10°, closer to B-DNA values. Roll angles are distributed between −5° and 5°, and helical rise varies between 3.0 Å to 3.7 Å; both parameters are also closer to B-DNA values. Slide varies between −2 and 0 Å spanning values between B-DNA and A-RNA. Of the different parameters, Z_p is considered to have the most discriminating power between A- and B- forms.[109] Z_p is given by the mean z-coordinates of the backbone phosphorus atoms with respect to individual dimer reference frames. Typically, Z_p is greater than 1.5 Å for A-RNA and less than 0.5 Å for B-DNA. For the considered DNA duplexes, -0.5 Å $\leq Z_p \leq 0.5$ Å, which is quite close to B-DNA.

Results for the RNA helical duplex belonging to a pure RNA triplex are shown in S14. As with the DNA counterparts, twist angles vary between 25 and 40°, spanning values between B-DNA and A-RNA. Inclination angles are distributed between 0 and 20°, roll angles are distributed between 0 and 10°, and helical rise varies between 2.5 Å to 3.7 Å; the three parameters span values between B-DNA and A-RNA. As for slide, its values run from −2 Å to −1 Å, the pyrimidine third strand case is closer to B-DNA while the purine third strand cases are closer to A-RNA. Finally, 0 Å $\leq Z_p \leq$ 2 Å; the values of the pyrimidine third strand case are closer to the B-DNA, while the purine third strand cases are closer to the A-RNA. Clearly, the presence of the RNA third strand affects the structure of the RNA duplex part, which exhibits features with large variations between B-DNA and A-RNA. Finally, the large peaks observed for some steps in some of the parameters are due to the presence of close Na^+ ions. For the RNA-pY case, the large population of Na^+ ions around the OP2 and O4 atoms of A5 and U13 causes larger roll and inclination values in the AA/UU steps. For the RNA apR(+) case, the presence of Na^+ ions

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

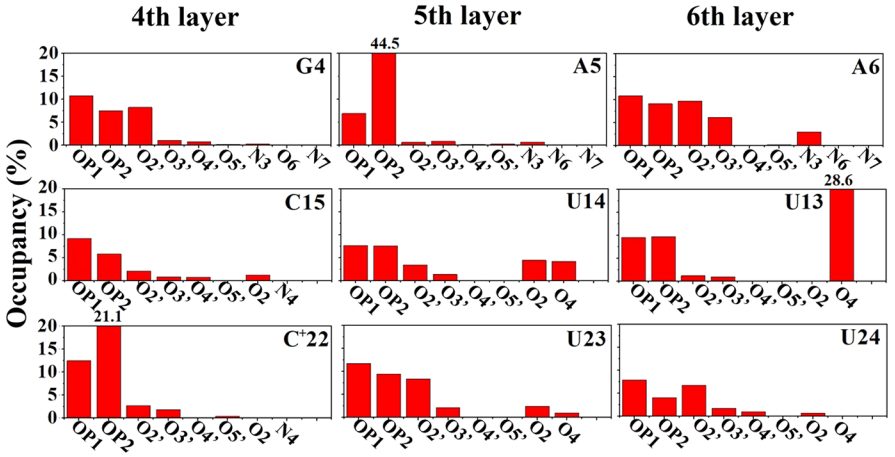


Figure 10. Na^+ occupancy for the middle three planes of the RNA-pY triplex.

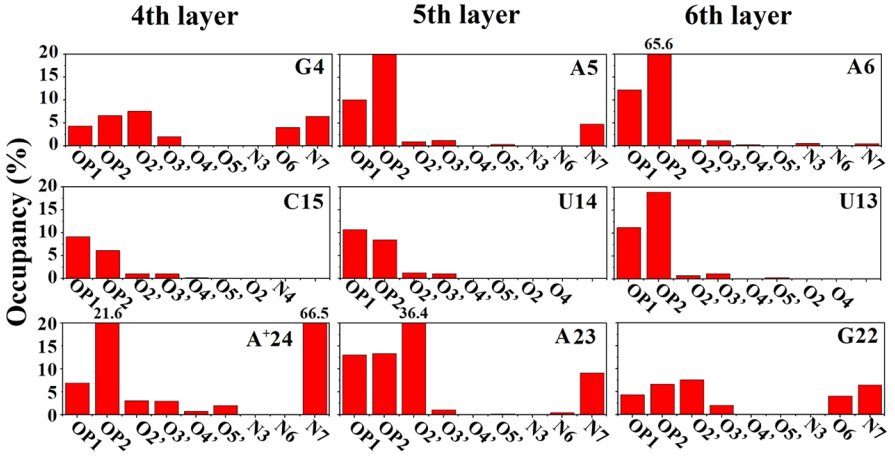


Figure 11. Na^+ occupancy for the middle three planes of the RNA-apR(+) triplex.

around the N7 atom of A(+)24 on the fourth layer contributes to a high helical rise in the GC/AU step.

Ion Distributions and Interactions. We have characterized the ion distribution around the triple helices, with a focus on the neutralizing Na^+ and Mg^{2+} ions. For the former, we focus primarily on the distributions around the most stable structures: RNA-pY and RNA-apR(+), and the corresponding mixed RD counterparts. The labeling of the atomic structures is given in Figure S11. The ion occupancy for the middle three planes of these structures is shown in Figures 10 and 11.

Naturally, the ions interact significantly with the OP1 and OP2 oxygen atoms in the phosphate backbone, where the ion occupancy is consistent with other studies on mismatched nucleic acids.[83,110,111] Certain bases display very large ion occupancy in OP2, such as C^+22 and A5 in RNA-pY; and A^+22, A5, A6, and U13 in RNA-apR(+). Other significant sites include the O4 in U13 in RNA-pY; and N7 in A^+24, and O2′ in A23. The results for the mixed RNA−DNA triple helices show similar trends, and are shown as Figures S6 and S7. Important differences are the occupation of OP2 only reaches large values for A5 both in RD-pY and RD-apR(+), otherwise OP1 has equal or larger occupation values; and T13 lacks a peak on O4. A visual representation of the ion cloud density around selected triplex structures is shown in Figures S8 and S9. Some typical binding sites are shown in Figure 12a for RD-pY; and in Figure 12b for RD-apR(+), which shows the binding of Na^+ to O2′ and N7 of A^+24.

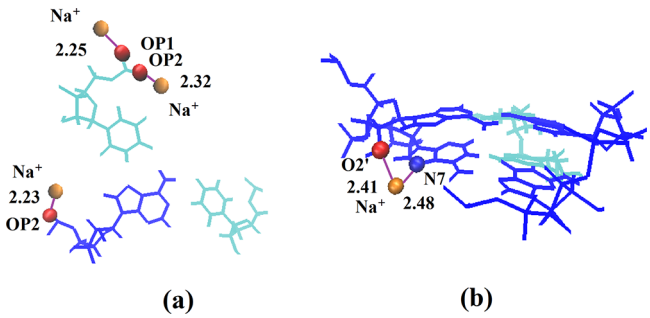


Figure 12. Schematic illustrating some typical Na^+ binding sites for (a) RD-pY, and (b) RD-apR(+). Ion binding distances are indicated in angstroms.

The importance of the interactions between Mg^{2+} ions and nucleic acids is well documented,[112−119] and experiments show that these divalent ions helps the formation of RNA-based triplexes.[100,101] To address this issue theoretically, we carried out 1 μs simulations of each of the triplex structures in the presence of hydrated Mg^{2+} ions. The Mg cations retained their solvation shell in the form of $Mg[(H_2O)_6]^{2+}$ ions, as expected.[110] Figures S10 and S11 show ion cloud densities

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

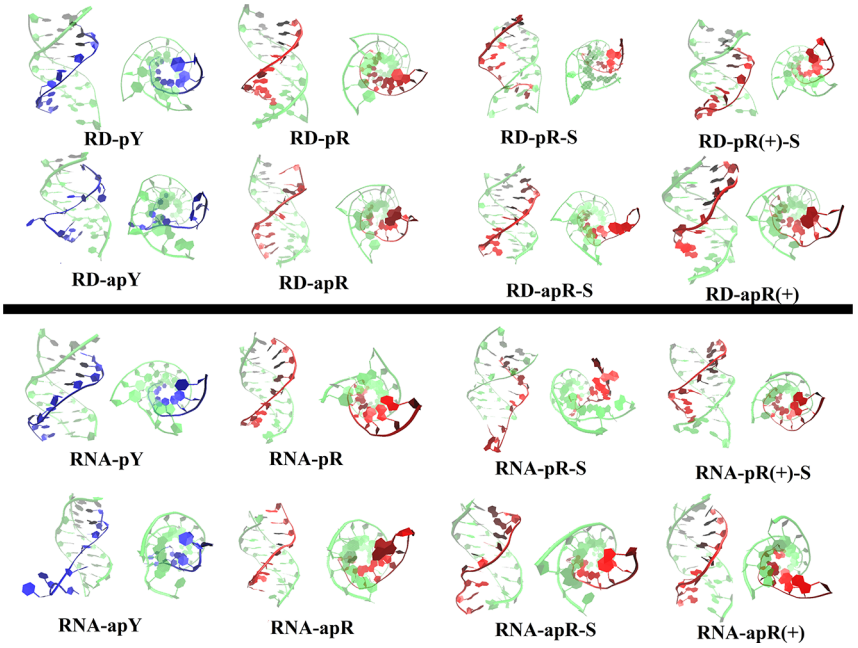


Figure 13. Final conformations of RNA·DNA:DNA hybrid triplexes and RNA·RNA:RNA triplexes as obtained after 1 μs simulations with hydrated Mg^{2+} ions.

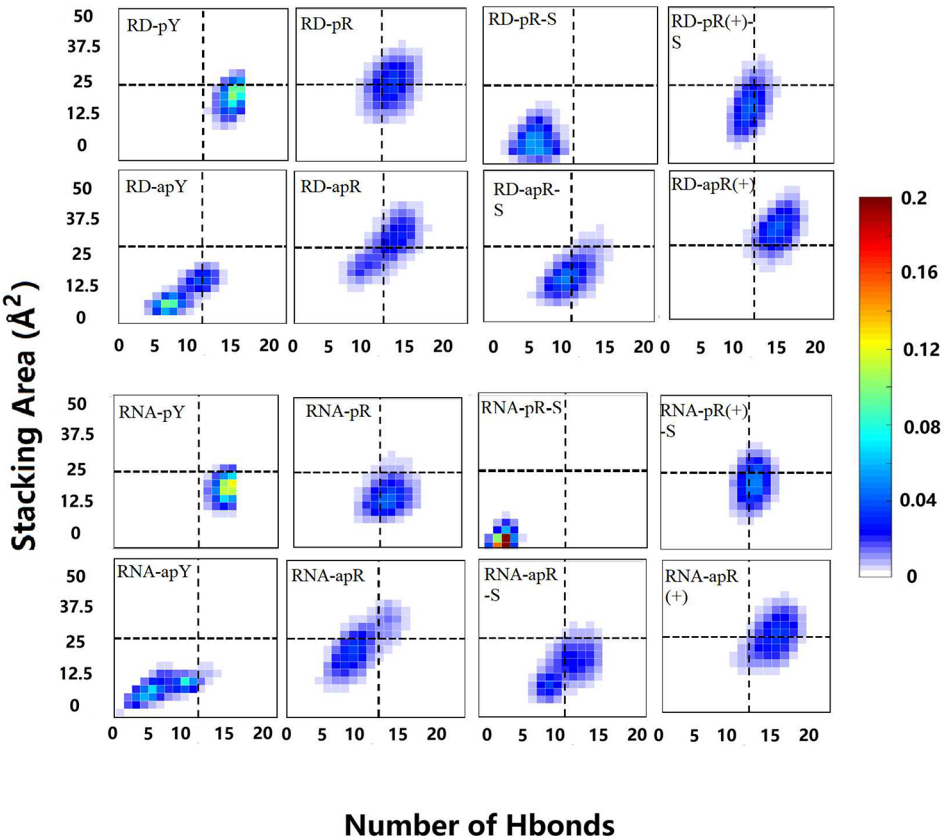


Figure 14. Effective stacking area versus effective hydrogen bond number for the RNA−DNA hybrid and pure RNA triplexes in the presence of hydrated Mg^{2+}.

around the different triplex structures. These figures show that in contrast to the Na^+ ions, the majority of the hydrated Mg ions are localized toward the outside of the triplex structures, with the majority to be found near the groove formed by the third strand and the purine strand of the duplex. In contrast to the Na^+ results, the hydrated Mg ions do not appear to bind or associate with any specific atom on the nucleic acid structure.

Figure 13 shows the final configurations of the different RNA based triplexes. From the figure, it is clear that structures apY and pR-S are unstable for both kinds of triplexes; all the

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

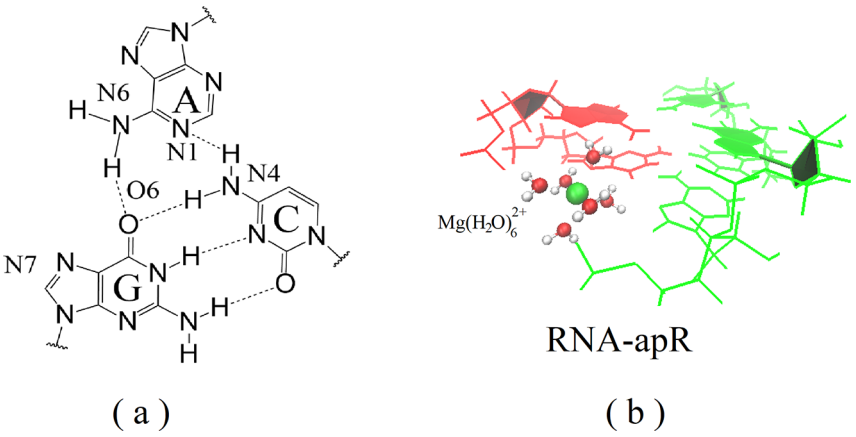


Figure 15. Stabilization by Mg^{2+} ions. (a) Rearrangement of hydrogen bonds on the A·G:C plane. (b) Binding of the hydrated Mg^{2+} ion to the first groove of the RNA-apR triplex.

rest of the structures are stable over the duration of the simulation.

To quantify these results, we show in Figure 14 histograms of the effective stacking area and hydrogen bond number for the triplexes. For the most part, these results are quite similar to results obtained in the absence of Mg^{2+} ions (see Figures 7 and 9).

The main difference with respect to the simulations with Na^+ ions is that the previously very unstable apR case gains stability in the presence of the $Mg[(H_2O)_6]^{2+}$ ions, which rearrange the hydrogen bond pattern. Figure 15 shows that the presence of a hydrated Mg ion exchanges the N3−N7 bond in the RH A·G:C pattern in Figure 3 with an N1−N4 bond that stabilizes the apR triplex, making it similar to previous results for pure DNA-based triplexes.[43]

DISCUSSION

Many Nonequivalent RNA·RNA:RNA and RNA·DNA:DNA Triplexes Can Be Assembled from GAA and UUC or TTC Strands. Schemes of 16 triplexes are shown in Figure 2. In contrast to our previous DNA triplex investigations, we focus only on structures with glycosidic torsion angles in *anti* conformation, since the *syn* conformations are energetically unfavorable.[43] These triplexes consist of antiparallel, helical RNA:RNA and DNA:DNA duplexes with WC base pairing with a single RNA strand binding to the major groove of the duplex. The third RNA strand may be either a pyrimidine or a purine, and these may be either parallel or antiparallel to the purine strand of the duplex. For the case of a pyrimidine third strand, we only considered the case of protonated cytosines, as supported by experimental evidence.[103−107] This results in two cases where the third strand can form hydrogen bonds with the duplex: the protonated pyrimidine third strand in either a parallel or antiparallel direction, with the latter shifted by a single base. For the purine third strand, we considered the strand in either parallel and antiparallel orientation. We considered both cases where the trinucleotides of the third strand are perfectly aligned with those of the duplex and where the third strand is shifted with respect to the duplex. We also included two protonated cases, since protonated adenines can form stable hydrogen bonds with G:C base pairs.[108]

Stability of Pure RNA and Hybrid RNA−DNA Triplexes Is Determined by the Stacking of the Bases of the Third Strand and by the Hydrogen Bonds between the Third Strand Base and the Duplex WC Bases on the Same Plane. To characterize the hydrogen bonds, we have extended the definitions of Hoogsteen (H) and reverse Hoogsteen (RH) hydrogen-bond patterns as shown in Figure 3 and Table 1. These definitions are similar to those for DNA triplexes, and are simply taken over here. We previously noted that standard ways of characterizing hydrogen bonds and base stacking fail to give a good measure when applied to the third strand. Hence, we extended the standard protocols explicitly taking into account the third strand, and successfully used this approach to characterize DNA triple helices. Details of this approach are provided in the Supporting Information. The same analysis applies to the RNA triplexes, so that two-dimensional histograms of the effective base stacking versus hydrogen bond number provide for a good measure of the stability of a given triplex. The histogram results are completely in line with the visual observations of the structures: for all unstable triplexes, the third strand peels off, often over a 300 ns time frame, while leaving the duplex structure intact.

We have also found it convenient to examine the (ϕ, d) histograms for each three-base plane, which serves to quickly reveal the position and orientation of the third base with respect to the WC bases. Values of ϕ less than (greater than) 90° indicate a preference of the third strand to be closer to the duplex purine (pyrimidine) strand. Almost universally, the third strand is closer to the purine strand of the duplex. In addition, these histograms also provide for a measure of the fluctuations that the third strand undergoes, given by the spread associated with each cluster. For the most part, the data for stable or nearly stable triplexes point to single, relatively compact clusters. In a small number of cases (RNA-apR-S, sixth plane), the data is split into two clusters. These clusters are associated with two different hydrogen bond conformations for that particular plane.

For Both Pure RNA and Mixed RNA−DNA Triple Helices, the Pyrimidine UUC^+ Third Strand in Its Parallel Alignment, RNA-pY and RD-pY, Is Stable, while Its apY Antiparallel Counterpart is Unstable. Both parallel and antiparallel configurations for r(UUC^+)·r(GAA):r(UUC) and r(UUC^+)·d(GAA):d(TTC) triplexes were studied, and it was found that only the parallel pY triplexes are stable. The antiparallel UUC^+ third strand cannot form stable hydrogen bonds with either the RNA−RNA or the RNA−DNA helix because of steric hindrance. On the other hand, the parallel UUC^+ third strand forms two hydrogen bonds with the WC

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

base pair of the duplex on each plane, U·A:T or U·A:U and C^+·G:C, in a similar pattern as the pY DNA triplex reported previously.[43] For pure DNA triplexes, the pY conformation is the *most* stable; however, the apY conformation is also stable at room temperature, unlike the pure RNA triplex and the hybrid triplex r(UUC$^+$)·d(GAA):d(TTC) studied here, where the apY conformation is totally unstable. We also note that the other hybrid triplex studied previously, the antiparallel d(TTC$^+$)·d(GAA):r(UUC) collapsed R-loop, is completely unstable. This instability could be traced to the reduced stability of the hybrid d(GAA):r(UUC) duplex, that although stable on its own, is not stable enough to carry the third strand.[43] When comparing the DNA third strand with the RNA third strand, we notice that the absence of the methyl group in the U base of the third strand effectively reduces its area compared to the T base, which in turns makes the stacking of the bases of the third strand slightly weaker, and the U bases more likely to rotate and turn away from the duplex. In the parallel case, the presence of the two hydrogen bonds anchors the third strand, but the reduced number of hydrogen bonds already manifested in the antiparallel DNA triplex makes the antiparallel RNA third strand completely unstable.

For Both Pure RNA and RNA−DNA Triplexes, When the Purine GAA Third Strand is Unprotonated, the Most Stable Conformations Are Parallel, Mainly RNA-pR and RD-pR; When It Is Protonated, the Most Stable Conformations Are Antiparallel, Mainly RNA-apR(+) and RD-apR(+), Which Are the Most Stable Triplexes for the Purine Third Strand. The stacking area and hydrogen bond analysis presented in Figures 5 and 7 gives a good idea of the triplex stability. The distributions pick the apR(+) conformations as the most stable for both pure RNA and RNA−DNA triplexes. For the parallel cases, the pR and pR(+)-S conformations are more stable−although slightly less than the apR(+) conformations. These results agree with those obtained for the pure DNA triplexes.[43] A comparison of the distributions for the two types of triplexes, pure RNA and RNA−DNA, shows that the distributions are more shifted to the upper right panels in the hybrid RNA−DNA cases, suggesting that the RNA·DNA:DNA triplexes are more stable than the RNA·RNA:RNA triplexes. Indeed, the RD-apR-S triplex is relatively stable, while the RNA-apR-S one is unstable.

Distribution of Na^+ and Mg^{2+} Ions around the Triplexes. Neutralizing Na^+ ions, not surprisingly, concentrate around the backbone OP1 and OP2 sites, and O2′ in some bases. There are three grooves associated with a given triplex: the groove formed between the third strand and purine strand of the helical duplex (first groove); the original minor groove of the duplex (second groove); and the groove formed between the third strand and pyrimidine strand of the helical duplex (third groove). For the RNA-pY triplex, most of the Na^+ ions are associated with the backbone and found in the first and third groove, with considerably fewer ions in the second groove. Notable binding sites include the O4 site in a U base in the third groove of the RNA-pY triplex. For RNA-apR(+), Na^+ ions are mainly found in the first and third grooves. Notable binding sites include the N7 site in A bases in the first groove. Turning to the RNA−DNA triplexes, the Na^+ ions around RD-pY triplexes are primarily found in the first and second grooves (the third groove is almost empty), while for the RD-apR(+) case the three grooves exhibit a roughly equal ion distribution. Na^+ ions found in the first groove are associated with OP1, OP2 and N7 on the protonated A^+. These ion distributions are illustrated in the Supporting Information. We note that in previous work,[43] we examined the Na^+ distribution associated with pure DNA triplex structures. For DNA-pY, most of the ions were found in the first groove, with fewer ions in the second groove, and almost no ions in the third groove; for DNA-apR(+), there is considerable ion presence in all of the grooves.

The Mg^{2+} ions are found in their hexahydrated state and do not appear to be directly associated with any specific atoms on the nucleic acids; rather they stay outside of the triplex structure and are found primarily in the first groove. The biggest change with respect to the Na^+ ions is that the hexahydrated Mg^{2+} ions help stabilize the apR triplexes (both pure RNA and RNA−DNA) through a change in the hydrogen bond pattern on the A·G:C planes.

Watson−Crick Helical Duplex in the Triplex Undergoes More Conformational Variations than the Corresponding Free-Standing Helix of the Same Sequence; DNA Duplexes Stay Closer to the Free B-DNA Helix, While the RNA Duplexes Tend to Stray Further from the Free A-RNA Helix. The structure of the helical duplex in a triplex is modified by the third strand. Some helical base step parameters for the different duplexes are shown in Figures S12−S14, which indicate that the duplexes are neither B-form or A-form, but display features of both forms. We first consider results for the DNA duplexes from stable pure DNA triplexes (Figures S12) and from mixed RNA−DNA triplexes (Figures S13). In both cases, the duplexes clearly stay closer to B-DNA except for twist and slide, whose values fall evenly between those of the B-DNA and A-RNA. The Z_p parameter is considered to have the most discriminating power between A- and B-forms.[109] Typically, Z_p is greater than 1.5 Å for A-RNA and less than 0.5 Å for B-DNA. Measured values for Z_p for these duplexes are in the range $-0.5\ \text{Å} \leq Z_p \leq 0.5\ \text{Å}$, which is quite close to B-DNA. Next, we consider the pure RNA triplexes, shown in Figure S14. The corresponding duplexes present larger fluctuations, with a wider distribution of values, spanning both A- and B-forms. The measured values for Z_p of the RNA duplex are mostly between 1.0 to 2.0 Å, characteristic of A-RNA values, especially for the purine third strand triplexes.

SUMMARY

In this work, we have studied the RNA and hybrid RNA−DNA triplexes that can be constructed from GAA and UUC or TTC sequences. This investigation is motivated by the fact that these noncanonical nucleic acid structures are associated with Friedreich's ataxia when the first intron of the FXN gene expands beyond a critical number of repeats; the study also complements previous investigations of DNA-based triplexes for the GAA/TTC sequence, thereby allowing a more complete picture of these triplexes to emerge. All in all, 16 nonequivalent RNA·RNA:RNA and RNA·DNA:DNA triplexes were considered; for the most part, results for the pure RNA triplexes and hybrid RNA−DNA triplexes are similar. Based on an analysis of the hydrogen-bond patterns and stacking of bases in the third strand, we found that the pyrimidine UUC$^+$ third strand in its parallel arrangement is stable, while (in contrast to DNA results) its antiparallel counterpart is unstable. For the GAA purine third strand, the parallel arrangements pR and pR(+)-S and the antiparallel apR(+) are stable, with the latter being the most stable structure. We have

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

also characterized the distribution of neutralizing Na^+ ions, as well as hydrated Mg^{2+} ions, around the triplexes. The results reported here provide c meticulous insight into the conformations and stability of the triplexes that can be assembled from GAA and TTC or UUC sequences, thereby contributing to further understanding of trinucleotide repeats and the associated unusual structures that trigger expansion.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c04358.

Supplementary details about the simulations, analysis, most of the RD triplex data, and Figures S1−S14 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Celeste Sagui − *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202, United States;* orcid.org/0000-0002-2750-5186; Email: sagui@ncsu.edu

Authors

Jiahui Zhang − *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202, United States*

Ashkan Fakharzadeh − *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202, United States*

Christopher Roland − *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c04358

Author Contributions

Designed research: all authors. Performed research: J.Z. Analyzed data: J.Z. with help from A.F. Wrote paper: all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the NC State HPC Center for computational support. This work was supported by the National Institutes for Health (NIH) via R01GM118508 and the National Science Foundation (NSF).

■ REFERENCES

(1) Prives, C.; Gilboa, E.; Revel, M.; Winocour, E. Cell-free translation of simian virus 40 early messenger RNA coding for viral T-antigen. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 457−461.

(2) François, B.; Szychowski, J.; Adhikari, S. S.; Pachamuthu, K.; Swayze, E. E.; Griffey, R. H.; Migawa, M. T.; Westhof, E.; Hanessian, S. Antibacterial aminoglycosides with a modified mode of binding to the ribosomal-RNA decoding site. *Angew. Chem., Int. Ed.* **2004**, *43*, 6735−6738.

(3) Mattick, J. S.; Amaral, P. P.; Dinger, M. E.; Mercer, T. R.; Mehler, M. F. RNA regulation of epigenetic processes. *Bioessays* **2009**, *31*, 51−59.

(4) Massé, E.; Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 4620−4625.

(5) Cooper, G. M.; Hausman, R. E. *The Cell: Molecular Approach*; Medicinska Naklada, 2004.

(6) Hatfield, D.; Lee, B. J.; Pirtle, R. M. *Transfer RNA in Protein Synthesis*; CRC Press, 1992.

(7) Gueneau de Novoa, P.; Williams, K. P. The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts. *Nucleic Acids Res.* **2004**, *32*, 104D−108D.

(8) Setyono, B.; Pederson, T. Ribonucleoprotein organization of eukaryotic RNA: XXX. Evidence that U1 small nuclear RNA is a ribonucleoprotein when base-paired with pre-messenger RNA in vivo. *Journal of molecular biology* **1984**, *174*, 285−295.

(9) Jones, C. P.; Ferré-D'Amaré, A. R. RNA quaternary structure and global symmetry. *Trends in biochemical sciences* **2015**, *40*, 211−220.

(10) Cheong, C.; Varani, G.; Tinoco, I. Solution structure of an unusually stable RNA hairpin, 5GGAC (UUCG) GUCC. *Nature* **1990**, *346*, 680−682.

(11) Arnott, S.; Hukins, D.; Dover, S.; Fuller, W.; Hodgson, A. Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations: X-ray diffraction analyses of the molecular conformations of polyadenylic acid· polyuridylic acid and polyinosinic acid· polycytidylic acid. *Journal of molecular biology* **1973**, *81*, 107−122.

(12) Davis, P. W.; Adamiak, R. W.; Tinoco, I., Jr Z-RNA: The solution NMR structure of r (CGCGCG). *Biopolymers: Original Research on Biomolecules* **1990**, *29*, 109−122.

(13) Sinha, R.; Kumar, G. S. Interaction of isoquinoline alkaloids with an RNA triplex: structural and thermodynamic studies of berberine, palmatine, and coralyne binding to poly (U). poly (A)* poly (U). *J. Phys. Chem. B* **2009**, *113*, 13410−13420.

(14) Snoussi, K.; Nonin-Lecomte, S.; Leroy, J.-L. The RNA i-motif. *Journal of molecular biology* **2001**, *309*, 139−153.

(15) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **2007**, *3*, 218−221.

(16) Moallem, E.; Kilav, R.; Silver, J.; Naveh-Many, T. RNA-protein binding and post-transcriptional regulation of parathyroid hormone gene expression by calcium and phosphate. *J. Biol. Chem.* **1998**, *273*, 5253−5259.

(17) Bley, C. J.; Qi, X.; Rand, D. P.; Borges, C. R.; Nelson, R. W.; Chen, J. J.-L. RNA−protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 20333−20338.

(18) Qi, X.; Xie, M.; Brown, A. F.; Bley, C. J.; Podlevsky, J. D.; Chen, J. J.-L. RNA/DNA hybrid binding affinity determines telomerase template-translocation efficiency. *EMBO journal* **2012**, *31*, 150−161.

(19) Nygaard, A.; Hall, B. Formation and properties of RNA-DNA complexes. *Journal of molecular biology* **1964**, *9*, 125−142.

(20) Thompson, J.; Cundliffe, E. The binding of thiostrepton to 23S ribosomal RNA. *Biochimie* **1991**, *73*, 1131−1135.

(21) Garvin, R. T.; Biswas, D. K.; Gorini, L. The effects of streptomycin or dihydrostreptomycin binding to 16S RNA or to 30S ribosomal subunits. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 3814−3818.

(22) Conne, B.; Stutz, A.; Vassalli, J.-D. The 3′ untranslated region of messenger RNA: a molecular 'hotspot'for pathology? *Nature medicine* **2000**, *6*, 637−641.

(23) Ito, D.; Suzuki, N. Conjoint pathologic cascades mediated by ALS/FTLD-U linked RNA-binding proteins TDP-43 and FUS. *Neurology* **2011**, *77*, 1636−1643.

(24) Ito, D.; Hatano, M.; Suzuki, N. RNA binding proteins and the pathological cascade in ALS/FTD neurodegeneration. *Science Translational Medicine* **2017**, *9*, eaah5436.

(25) Yiakouvaki, A.; Dimitriou, M.; Karakasiliotis, I.; Eftychi, C.; Theocharis, S.; Kontoyiannis, D. L. Myeloid cell expression of the RNA-binding protein HuR protects mice from pathologic inflammation and colorectal carcinogenesis. *J. Clin. Invest.* **2012**, *122*, 48−61.

(26) Findeis-Hosey, J. J.; Xu, H. The use of insulin like-growth factor II messenger RNA binding protein−3 in diagnostic pathology. *Human pathology* **2011**, *42*, 303−314.
(27) Verkerk, A. J.; Pieretti, M.; Sutcliffe, J. S.; Fu, Y.-H.; Kuhl, D. P.; Pizzuti, A.; Reiner, O.; Richards, S.; Victoria, M. F.; Zhang, F.; et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **1991**, *65*, 905−914.
(28) Yu, S.; Pritchard, M.; Kremer, E.; Lynch, M.; Nancarrow, J.; Baker, E.; Holman, K.; Mulley, J.; Warren, S.; Schlessinger, D.; et al. Fragile X genotype characterized by an unstable region of DNA. *Science* **1991**, *252*, 1179−1181.
(29) La Spada, A. R.; Wilson, E. M.; Lubahn, D. B.; Harding, A.; Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **1991**, *352*, 77.
(30) Giunti, P.; Sweeney, M. G.; Spadaro, M.; Jodice, C.; Novelletto, A.; Malaspina, P.; Frontali, M.; Harding, A. E. The Trinucleotide Repeat Expansion on Chromosome 6p (SCA1) in Autosomal Dominant Cerebellar Ataxias. *Brain* **1994**, *117*, 645−649.
(31) Wells, R. D.; Warren, S. *Genetic Instabilities and Neurological Diseases*; Academic Press, 1998.
(32) Orr, H.; Zoghbi, H. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **2007**, *30*, 575.
(33) Oberle, I.; Rouseau, F.; Heitz, D.; Devys, D.; Zengerling, S.; Mandel, J. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **1991**, *252*, 1097−1102.
(34) Campuzano, V.; et al. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **1996**, *271*, 1423−1427.
(35) Mirkin, S. M. DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. in Struct. Biol.* **2006**, *16*, 351−358.
(36) Khristich, A. N.; Mirkin, S. M. On the Wrong DNA Track: Molecular Mechanisms of Repeat-mediated Genome Instability. *J. Biol. Chem.* **2020**, *295*, 4134−4170.
(37) Paulson, H. Repeat expansion diseases. *Handb. Clin. Neurol.* **2018**, *147*, 105.
(38) Moore, H.; Greenwell, P. W.; Liu, C.-P.; Arnheim, N.; Petes, T. D. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1504.
(39) McMurray, C. DNA secondary structure: A common and causative factor for expansion in human disease. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1823−1825.
(40) Wells, R.; Dere, R.; Hebert, M.; Napierala, M.; Son, L. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **2005**, *33*, 3785−3798.
(41) Polak, U.; McIvor, E.; Dent, S. Y.; Wells, R. D.; Napierala, M. Expanded complexity of unstable repeat diseases. *Biofactors* **2013**, *39*, 164−175.
(42) Usdin, K.; House, N. C.; Freudenreich, C. H. Repeat instability during DNA repair: Insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* **2015**, *50*, 142−167.
(43) Zhang, J.; Fakharzadeh, A.; Pan, F.; Roland, C.; Sagui, C. Atypical structures of GAA/TTC trinucleotide repeats underlying Friedreich's ataxia: DNA triplexes and RNA/DNA hybrids. *Nucl. Acids Res.* **2020**, *48*, 9899.
(44) Liebecq, E. C. *Biochemical Nomenclature and Related Documents*, 2nd ed.; Portland Press, 1992.
(45) Vuyisich, M.; Beal, P. A. Regulation of the RNA-dependent protein kinase by triple helix formation. *Nucleic acids research* **2000**, *28*, 2369−2374.
(46) Ageeli, A. A.; McGovern-Gooch, K. R.; Kaminska, M. M.; Baird, N. J. Finely tuned conformational dynamics regulate the protective function of the lncRNA MALAT1 triple helix. *Nucleic acids research* **2019**, *47*, 1468−1481.
(47) Butcher, S. E.; Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of chemical research* **2011**, *44*, 1302−1311.
(48) Buske, F. A.; Mattick, J. S.; Bailey, T. L. Potential in vivo roles of nucleic acid triple-helices. *RNA biology* **2011**, *8*, 427−439.
(49) Conrad, N. K. The emerging role of triple helices in RNA biology. *Wiley Interdisciplinary Reviews: RNA* **2014**, *5*, 15−29.
(50) Cash, D. D.; Cohen-Zontag, O.; Kim, N.-K.; Shefer, K.; Brown, Y.; Ulyanov, N. B.; Tzfati, Y.; Feigon, J. Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10970−10975.
(51) Qiao, F.; Cech, T. R. Triple-helix structure in telomerase RNA contributes to catalysis. *Nature structural & molecular biology* **2008**, *15*, 634.
(52) Kim, N.-K.; Zhang, Q.; Zhou, J.; Theimer, C. A.; Peterson, R. D.; Feigon, J. Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *Journal of molecular biology* **2008**, *384*, 1249−1261.
(53) Shefer, K.; Brown, Y.; Gorkovoy, V.; Nussbaum, T.; Ulyanov, N. B.; Tzfati, Y. A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Molecular and cellular biology* **2007**, *27*, 2130−2143.
(54) Ulyanov, N. B.; Shefer, K.; James, T. L.; Tzfati, Y. Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates. *Nucleic acids research* **2007**, *35*, 6150−6160.
(55) Klein, D. J.; Ferré-D'Amaré, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **2006**, *313*, 1752−1756.
(56) Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chemistry & biology* **2007**, *14*, 97−105.
(57) Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature structural & molecular biology* **2008**, *15*, 177−182.
(58) Kang, M.; Peterson, R.; Feigon, J. Structural insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Molecular cell* **2009**, *33*, 784−790.
(59) Klein, D. J.; Edwards, T. E.; Ferré-D'Amaré, A. R. Cocrystal structure of a class I preQ 1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nature structural & molecular biology* **2009**, *16*, 343−344.
(60) Spitale, R. C.; Torelli, A. T.; Krucinska, J.; Bandarian, V.; Wedekind, J. E. The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. *J. Biol. Chem.* **2009**, *284*, 11012−11016.
(61) Batey, R. T. Structure and mechanism of purine binding riboswitches. *Q. Rev. Biophys.* **2012**, *45*, 345.
(62) Liberman, J. A.; Salim, M.; Krucinska, J.; Wedekind, J. E. Structure of a class II preQ 1 riboswitch reveals ligand recognition by a new fold. *Nat. Chem. Biol.* **2013**, *9*, 353−355.
(63) Michiels, P. J.; Versleijen, A. A.; Verlaan, P. W.; Pleij, C. W.; Hilbers, C. W.; Heus, H. A. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *Journal of molecular biology* **2001**, *310*, 1109−1123.
(64) Nixon, P. L.; Rangan, A.; Kim, Y.-G.; Rich, A.; Hoffman, D. W.; Hennig, M.; Giedroc, D. P. Solution structure of a luteoviral P1−P2 frameshifting mRNA pseudoknot. *Journal of molecular biology* **2002**, *322*, 621−633.
(65) Pallan, P. S.; Marshall, W. S.; Harp, J.; Jewett, F. C.; Wawrzak, Z.; Brown, B. A.; Rich, A.; Egli, M. Crystal structure of a luteoviral RNA pseudoknot and model for a minimal ribosomal frameshifting motif. *Biochemistry* **2005**, *44*, 11315−11322.
(66) Giedroc, D. P.; Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus research* **2009**, *139*, 193−208.
(67) Olsthoorn, R. C.; Reumerman, R.; Hilbers, C. W.; Pleij, C. W.; Heus, H. A. Functional analysis of the SRV-1 RNA frameshifting pseudoknot. *Nucleic acids research* **2010**, *38*, 7665−7672.
(68) Mitton-Fry, R. M.; DeGregorio, S. J.; Wang, J.; Steitz, T. A.; Steitz, J. A. Poly (A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **2010**, *330*, 1244−1247.
(69) Tycowski, K. T.; Shu, M.-D.; Borah, S.; Shi, M.; Steitz, J. A. Conservation of a triple-helix-forming RNA stability element in

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

noncoding and genomic RNAs of diverse viruses. *Cell reports* **2012**, *2*, 26−32.
(70) Brown, J. A.; Valenstein, M. L.; Yario, T. A.; Tycowski, K. T.; Steitz, J. A. Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19202−19207.
(71) Wilusz, J. E.; JnBaptiste, C. K.; Lu, L. Y.; Kuhn, C.-D.; Joshua-Tor, L.; Sharp, P. A. A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly (A) tails. *Genes & development* **2012**, *26*, 2392−2407.
(72) Adams, P. L.; Stahley, M. R.; Kosek, A. B.; Wang, J.; Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **2004**, *430*, 45−50.
(73) Golden, B. L.; Kim, H.; Chase, E. Crystal structure of a phage Twort group I ribozyme−product complex. *Nature structural & molecular biology* **2005**, *12*, 82−89.
(74) Guo, F.; Gooding, A. R.; Cech, T. R. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Molecular cell* **2004**, *16*, 351−362.
(75) Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **2008**, *320*, 77−82.
(76) Klein, D.; Schmeing, T.; Moore, P.; Steitz, T. The kink-turn: a new RNA secondary structure motif. *EMBO journal* **2001**, *20*, 4214−4221.
(77) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4899−4903.
(78) Noller, H. F. RNA structure: reading the ribosome. *Science* **2005**, *309*, 1508−1514.
(79) Devi, G.; Zhou, Y.; Zhong, Z.; Toh, D.-F. K.; Chen, G. RNA triplexes: from structural principles to biological and biotech applications. *Wiley Interdisciplinary Reviews: RNA* **2015**, *6*, 111−128.
(80) Li, Y.; Syed, J.; Sugiyama, H. RNA-DNA triplex formation by long noncoding RNAs. *Cell chemical biology* **2016**, *23*, 1325−1333.
(81) Mondal, T.; et al. MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA−DNA triplex structures. *Nat. Commun.* **2015**, *6*, 7743.
(82) Zhou, Z.; Giles, K. E.; Felsenfeld, G. DNA· RNA triple helix formation can function as a cis-acting regulatory mechanism at the human β-globin locus. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 6130−6139.
(83) Pan, F.; Man, V. H.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices of CAG and GAC Trinucleotide Repeats. *Biophys. J.* **2017**, *113*, 19−36.
(84) Zhang, Y.; Roland, C.; Sagui, C. Structure and dynamics of DNA and RNA double helices obtained from the GGGGCC and CCCCGG hexanucleotide repeats that are the hallmark of C9FTD/ALS diseases. *ACS Chem. Neurosci.* **2017**, *8*, 578−591.
(85) Zhang, Y.; Roland, C.; Sagui, C. Structural and Dynamical Characterization of DNA and RNA Quadruplexes Obtained from the GGGGCC and GGGCCT Hexanucleotide Repeats Associated with C9FTD/ALS and SCA36 Diseases. *ACS Chem. Neurosci.* **2018**, *9*, 1104−1117.
(86) Pan, F.; Zhang, Y.; Man, V. H.; Roland, C.; Sagui, C. E-motif Formed by Extrahelical Cytosine Bases in DNA Homoduplexes of Trinucleotide and Hexanucleotide Repeats. *Nucleic Acids Res.* **2018**, *46*, 942−955.
(87) Pan, F.; Man, V. H.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices Obtained from the CCG and GGC Trinucleotide Repeats. *J. Phys. Chem. B* **2018**, *122*, 4491−4512.
(88) Xu, P.; Pan, F.; Roland, C.; Sagui, C.; Weninger, K. Dynamics of strand slippage in DNA hairpins formed by CAG repeats: role of sequence parity and trinucleotide interrupts. *Nucleic Acids Res.* **2020**, *48*, 2232.
(89) Pan, F.; Zhang, Y.; Xu, P.; Man, V.-H.; Roland, C.; Weninger, K.; Sagui, C. Molecular conformations and dynamics of nucleotide repeats associated with neurodegenerate diseases: double helices and AG hairpin loops. *Comput. Struct. Biotech J.* **2021**, *19*, 2819.
(90) Fakharzadeh, A.; Zhang, J.; Roland, C.; Sagui, C. Novel eGZ-motif formed by regularly extruded guanine bases in a left-handed Z-DNA helix as a major motif behind CGG trinucleotide repeats. *Nucleic Acids Res.* **2022**, *50*, 4860.
(91) Case, D.; Walker, R. C.; Cheatham, T. E.; Simmerling, C.; Roitberg, A.; Merz, K. M.; Luo, R.; Darden, T.; Wang, J.; Duke, R. E.; et al. *Amber 18*; University of California: San Francisco, 2018.
(92) Ivani, I.; et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55.
(93) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophysical journal* **2007**, *92*, 3817−3829.
(94) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E., III; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(95) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *Journal of computational chemistry* **1986**, *7*, 230−252.
(96) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(97) Joung, I. S.; Cheatham, T. E., III Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(98) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(99) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(100) Kankia, B. I. Mg2+-induced triplex formation of an equimolar mixture of poly (rA) and poly (rU). *Nucleic acids research* **2003**, *31*, 5101−5107.
(101) Klener, J.; Štěpánek, J. UV resonance Raman study of PolyA and PolyU complexes: Mg2+-induced formation of PolyU· PolyA· PolyU triplexes. *Vib. Spectrosc.* **2015**, *81*, 32−39.
(102) Panteva, M. T.; Giambasu, G. M.; York, D. M. Force field for Mg2+, Mn2+, Zn2+, and Cd2+ ions that have balanced interactions with nucleic acids. *J. Phys. Chem. B* **2015**, *119*, 15460−15470.
(103) Misra, G. *Introduction to Biomolecular Structure and Biophysics: Basics of Biophysics*; Springer, 2017; p 86.
(104) Roberts, R. W.; Crothers, D. M. Prediction of the stability of DNA triplexes. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 4320−4325.
(105) Radhakrishnan, I.; Patel, D. J. Solution structure of a pyrimidine· purine· pyrimidine DNA triplex containing T· AT, C+· GC and G· TA triples. *Structure* **1994**, *2*, 17−32.
(106) Mariappan, S. V.; Catasti, P.; Silks, L. A.; Bradbury, E. M.; Gupta, G. The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J. Mol. Biol.* **1999**, *285*, 2035−2052.
(107) Soto, A. M.; Marky, L. A. Thermodynamic contributions for the incorporation of GTA triplets within canonical TAT/TAT and C + GC/C+ GC base-triplet stacks of DNA triplexes. *Biochemistry* **2002**, *41*, 12475−12482.
(108) Mirkin, S. M.; Frank-Kamenetskii, M. D. H-DNA and related structures. *Annual review of biophysics and biomolecular structure* **1994**, *23*, 541−576.
(109) Lu, X.-J.; Shakked, Z.; Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *Journal of molecular biology* **2000**, *300*, 819−840.
(110) Pan, F.; Roland, C.; Sagui, C. Ion distribution around left- and right-handed DNA and RNA duplexes: a comparative study. *Nucleic Acids Res.* **2014**, *42*, 13981−96.
(111) Pan, F.; Man, V. H.; Roland, C.; Sagui, C. Structure and dynamics of DNA and RNA double helices obtained from the CCG

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743

and GGC trinucleotide repeats. *J. Phys. Chem. B* **2018**, *122*, 4491−4512.
(112) Zhao, H.; Lin, Z.; Lynn, A. Y.; Varnado, B.; Beutler, J. A.; Murelli, R. P.; Le Grice, S. F.; Tang, L. Two distinct modes of metal ion binding in the nuclease active site of a viral DNA-packaging terminase: insight into the two-metal-ion catalytic mechanism. *Nucleic acids research* **2015**, *43*, 11003−11016.
(113) Leonarski, F.; d'Ascenzo, L.; Auffinger, P. Mg2+ ions: do they bind to nucleobase nitrogens? *Nucleic acids research* **2017**, *45*, 987−1004.
(114) Sun, L.-Z.; Zhang, D.; Chen, S.-J. Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annual review of biophysics* **2017**, *46*, 227−246.
(115) Kolev, S. K.; Petkov, P. S.; Rangelov, M. A.; Trifonov, D. V.; Milenov, T. I.; Vayssilov, G. N. Interaction of Na+, K+, Mg2+ and Ca2+ counter cations with RNA. *Metallomics* **2018**, *10*, 659−678.
(116) Bao, L.; Wang, J.; Xiao, Y. Dynamics of metal ions around an RNA molecule. *Phys. Rev. E* **2019**, *99*, 012420.
(117) Xi, K.; Wang, F.-H.; Xiong, G.; Zhang, Z.-L.; Tan, Z.-J. Competitive binding of Mg2+ and Na+ ions to nucleic acids: from helices to tertiary structures. *Biophysical journal* **2018**, *114*, 1776−1790.
(118) Fischer, N. M.; Polêto, M. D.; Steuer, J.; van der Spoel, D. Influence of Na+ and Mg2+ ions on RNA structures studied with molecular dynamics simulations. *Nucleic acids research* **2018**, *46*, 4872−4882.
(119) Zhang, Z.-L.; Wu, Y.-Y.; Xi, K.; Sang, J.-P.; Tan, Z.-J. Divalent ion-mediated DNA-DNA interactions: a comparative study of triplex and duplex. *Biophysical journal* **2017**, *113*, 517−528.

https://doi.org/10.1021/acsomega.2c04358
ACS Omega 2022, 7, 38728−38743