



Published in final edited form as:

*Biochemistry*. 2019 May 21; 58(20): 2474–2487. doi:10.1021/acs.biochem.9b00122.

## Effects of Noncanonical Base Pairing on RNA Folding: Structural Context and Spatial Arrangements of G-A Pairs

Wilma K. Olson<sup>1</sup>, Shuxiang Li<sup>1</sup>, Thomas Kaukonen<sup>1</sup>, Andrew V. Colasanti<sup>1</sup>, Yurong Xin<sup>1</sup>, Xiang-Jun Lu<sup>2</sup>

<sup>1</sup>Department of Chemistry & Chemical Biology and Center for Quantitative Biology, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, USA

<sup>2</sup>Department of Biological Sciences, Columbia University, New York, New York 10027, USA

### Abstract

Noncanonical base pairs play important roles in assembling the three-dimensional structures critical to the diverse functions of RNA. These associations contribute to the looped segments that intersperse the canonical double-helical elements within folded, globular RNA molecules. They stitch together various structural elements, serve as recognition elements for other molecules, and act as sites of intrinsic stiffness or deformability. This work takes advantage of new software (DSSR) designed to streamline the analysis and annotation of RNA three-dimensional structures. The multi-scale structural information gathered for individual molecules, combined with the growing number of unique, well-resolved RNA structures, makes it possible to examine the collective features deeply and to uncover previously unrecognized patterns of chain organization. Here we focus on a subset of noncanonical base pairs involving guanine and adenine and the links between their modes of association, secondary structural context, and contributions to tertiary folding. The rigorous descriptions of base-pair geometry that we employ facilitate characterization of recurrent geometric motifs and the structural settings in which these arrangements occur. Moreover, the numerical parameters hint of the natural motions of the interacting bases and the pathways likely to connect different spatial forms. We draw attention to higher-order multiplexes involving two or more G-A pairs and the roles these associations appear to play in bridging different secondary structural units. The collective data reveal pairing propensities in base organization, secondary structural context, and deformability and serve as a starting point for further multi-scale investigations and/or simulations of RNA folding.

---

**Corresponding Authors** Wilma K. Olson – wilma.olson@rutgers.edu. Phone: +1 848-445-3993., Xiang-Jun Lu – xiangjun@3dna.org. Phone: +1 732-447-7806.

#### Supporting Information

Three tables respectively listing the high-resolution structures that contain the G-A base pairs described herein (Table S1), the identities and relevant structural features of the selected pairs (Table S2), and the precise identities of illustrated G-A-paired structures (Table S3); ten figures respectively illustrating the rigid-body parameters and reference frames used to characterize the spatial arrangements of base pairs (Figure S1), the secondary structures of selected G-A pairs (Figure S2), the observed non-planarity of G-A base associations (Figure S3), the hydrogen bonds linking different modes of G-A pairing (Figure S4), the hydrogen-bonding features of G-A pairs in different spatial forms (Figure S5), the secondary structural contexts of the G-A pairs (Figure S6), the unusual structural setting of an m–M sheared G-A pair (Figure S7), the secondary structures of selected G-A-mediated multiplets (Figure S8), the relative deformabilities of the major G-A pairing schemes (Figure S9), and the influence of base-pairing on secondary structural pathways (Figure S10).

The authors declare no competing financial interest.

## INTRODUCTION

The structural and functional versatility of RNA arises in large part from the numerous modes of interaction of the constituent nucleotides. Notable in this regard are the various edge-to-edge associations of the planar, nitrogenous bases, such as the canonical Watson-Crick and wobble pairings found in short double-helical stretches of folded RNA molecules and the noncanonical arrangements of bases that may mediate long-range RNA interactions, <sup>1-3</sup> serve as recognition elements for proteins and other molecules, <sup>4-6</sup> and act as sites of intrinsic stiffness or deformability. <sup>7,8</sup> Among the most common noncanonical base pairs are those involving guanine (G) and adenine (A). The hydrogen-bonding capabilities of the two purine moieties make it possible to arrange a G and an A in multiple ways (Figure 1) and to incorporate the bases in different secondary structural contexts (Figure 2). For example, there are many anecdotal examples of Watson-Crick-like imino configurations of G and A interrupting stretches of canonical base pairs and forming so-called internal loops<sup>9</sup> and numerous reported instances of sheared G·A pairs lying within an internal loop or at the end of a short hairpin loop, most notably in the GNRA tetraloop (where the N refers to any base and the R to G or A).<sup>9-12</sup> The adenines of sheared G·A pairs may also form stabilizing A-minor interactions with canonical Watson-Crick pairs in various structural contexts, such as at kink-turns,<sup>13-15</sup> and may contribute to larger, conserved multi-base-pair juxtapositions, such as the loop E/bulged-G,<sup>16-18</sup> sarcin-ricin,<sup>19</sup> and 3RRs<sup>20</sup> motifs. While there are fewer documented examples of G·A pairing within the loops at the junctions of three or more covalently linked stretches of canonical base pairs,<sup>21,22</sup> recent studies point to the potential contribution of these noncanonical associations to the global flexing of ribosomal and transfer RNA molecules.<sup>8</sup> Knowledge of the intrinsic deformabilities, the preferred structural contexts, and the likely transition pathways among the many modes of G·A association is accordingly important to understanding the global folds and motions of RNA and is potentially useful in the prediction of RNA three-dimensional (3D) structure from nucleotide sequence.

One of the simplest ways to assess the deformations and structural context of RNA base pairing is through the spatial arrangements of the interacting bases with respect to one another and with respect to other bases and base pairs in the surrounding chemical environment. The relative orientation and displacement of the paired bases are easily understood in terms of six rigid-body parameters. The parameters — three translations (Shear, Stretch, Stagger) and three rotations (Buckle, Propeller, Opening), originally devised to characterize the arrangements of bases in double-helical DNA<sup>23</sup> — can be conceived without resort to detailed molecular models (Figure S1). Whereas the values of Shear, Stretch, and Opening distinguish different modes of base pairing, the values of Stagger, Buckle, and Propeller characterize the non-planarity of base association.<sup>24-27</sup> The precise numbers reflect the relative location and orientation of a pair of coordinate frames, one embedded in each of the interacting bases.<sup>28,29</sup> The frames are conventionally defined such that all six parameters are null in an ideal, planar, antiparallel Watson-Crick base pair,<sup>30</sup> where the faces of the two bases <sup>31,32</sup> and the axes normal to the base planes point in opposing directions. One of the bases, however, may flip in other pairing schemes (through reversal of chain direction or syn/anti rotation about the glycosidic linkage between the

sugar and base) so that the normal vectors point in the same direction.<sup>33</sup> The two directions of association (antiparallel and parallel) are respectively differentiated here by minus and plus signs.<sup>24</sup>

The coordinate axes in the planes of the paired bases capture the hydrogen-bonding and groove characteristics of a Watson-Crick pair and thus provide an automatic description of the interacting edges of the bases (see Materials and Methods for details). The minor- and major-groove edges of a base, here respectively abbreviated as *m* and *M*, correspond approximately to the so-called sugar and Hoogsteen/C-H edges used in earlier classifications of RNA base pairing.<sup>9,41</sup> The minor-groove edge includes the atoms closest to the site of base attachment to the sugar-phosphate backbone; the major-groove edge, on the opposite side of the base, lies far from the attachment point (see Figure S1). The Watson-Crick or *W* edge includes the purine atoms involved in canonical associations of G and A with their pyrimidine complements, i.e., guanine N1, N2, O6 and adenine N1, N6 respectively paired in such arrangements to cytosine N3, O2, N4 and uracil/thymine N3, O4 (Figure 1).

The orientation and displacement of a base pair with respect to the coordinate frames on nearby bases determine whether that pair occurs in the context of a multiplet, i.e., a higher-order, hydrogen-bonded coplanar association of three or more bases, or an array of stacked residues.<sup>27</sup> The paired bases typically lie near one another in roughly the same plane and form one or more hydrogen bonds, while the bases in stacked arrays lie one above the other in closely spaced, overlapping arrangements. The locations of the two nearest neighbors in a stacked array determine whether a base pair lies at the end or in the middle of the array. That is, the two nearest neighbors lie on one side of a terminal pair but on opposing sides of an internal pair. The ends of a stacked array of covalently connected canonical base pairs, in turn, determine the locations and types of adjoining loops. For example, a hairpin loop abuts a single stacked array, a bulge or internal loop two arrays, and a junction (multi-branched loop) three or more arrays (Figures 2 and S2). Thus, knowledge of the spatial arrangements of the RNA bases leads to direct characterization of secondary structure and makes it possible to examine the effects of base pairing in a higher-order structural context.

Interest in deciphering the complexities of RNA folding and dynamics has prompted the development of assorted methods to detect,<sup>25,27,42-46</sup> characterize,<sup>26,47,48</sup> and simulate<sup>48-51</sup> the spatial properties of the base-pair components. The composite data reveal six major categories of G-A association in high-resolution structures (Figures 1, 2). The interactions are stabilized by two or more hydrogen bonds, albeit in some cases through atoms on the sugar-phosphate backbone. In addition to the sheared and imino patterns noted above, the guanine frequently associates through its minor-groove edge with the minor-groove or Watson-Crick edge of adenine in either a parallel or antiparallel orientation. These four combinations, which have no common names, are denoted in the illustrated examples in terms of the interacting base edges and the signs of association (Figures 1, 2), i.e.,  $m\pm m$ ,  $m\pm W$ , where the first letter refers to the hydrogen-bonded edge of guanine, the second to that of adenine, and the + or - to the parallel or antiparallel direction of the two base normals. The antiparallel, Watson-Crick-like associations of G and A in imino pairs are thus labeled

W–W and the sheared interactions between the minor-groove edge of guanine and the major-groove edge of adenine m–M.

Current knowledge of the motions of noncanonical base pairs derives in large part from molecular simulations and/or computationally guided interpretations of solution measurements, e.g., models derived from atomic-level characterization of RNA energy landscapes<sup>48-54</sup> or consistent with observed experimental data.<sup>55</sup> Formation of a sheared G·A pair is predicted to stiffen an internal loop compared to an imino pair in terms of the simulated degree of bending of short double-helical molecules<sup>49</sup> and to lower the range of normal-mode deformations of rigid-body parameters.<sup>48</sup> On the other hand, the spread of rigid-body parameters of sheared base pairs is slightly wider compared to those of imino G·A pairs in some analyses of high-resolution RNA structures<sup>47</sup> and thus suggestive of greater deformability. The narrower range of parameters for sheared G·A base pairs reported here and in earlier studies from this lab<sup>26</sup> mirrors the lesser deformability detected in solution. Deciphering the effects of the two types of noncanonical pairs on the solution properties of RNA internal loops, however, is complicated by the very different sequence contexts in which the interactions have been examined, e.g., tandem imino G·A pairs closed by relatively stiff G·C pairs vs. tandem sheared G·A pairs closed by more deformable G·U pairs.<sup>56,57</sup> The spread of angular variables used to guide simulated transitions between sheared and imino states hints of a potentially wider range of accessible configurations for the imino pair but a deeper free-energy minimum for the sheared pair.<sup>50</sup>

Recent analyses of RNA base interactions in terms of the observed spatial density distribution of a representative point on one base with respect to a point on another base provide an interesting basis for the prediction of RNA secondary structure and a novel way to visualize the deformations of paired bases.<sup>51</sup> The distributions reveal characteristic zones of interaction on the edges and above a given base, disclosing the dominant sites of in-plane pairing and out-of-plane stacking. Treatment of G·A pairing in terms of the relative positions of the centers of the six-membered rings of the two bases, however, does not capture the diverse modes of their association. Guanine pairs through its minor-groove edge with adenine in at least five distinct ways (Figure 1), which are lost in representations of single points. Moreover, as reported below, the modes of base-pair association depend upon higher-order structural context. Predictions of RNA folding may benefit from this information.

Here we take advantage of new capabilities in the analysis of nucleic acid structures from the DSSR software program<sup>27</sup> to characterize the geometries, classes, and structural context of the base-pair interactions found in a set of representative, well-resolved RNA crystal structures. We focus on the abundant noncanonical pairings of guanine and adenine and find connections between the modes of association, the local deformations of the pair, and the secondary structural contexts in which the bases occur. The precise numerical descriptions of base-pair geometry that we employ facilitate classification of recurrent geometric motifs and the settings in which these arrangements occur, as well as hint of the natural motions of the interacting bases and the pathways likely to connect different spatial forms. The same approach can be applied to other base pairs, e.g., the diverse arrangements of guanine and uracil detected in recent solution studies,<sup>58</sup> and the 3D settings in which these interactions occur. We also draw attention to higher-order multiplexes involving two or more G·A pairs

and the roles these associations appear to play in bridging different secondary structural units. The collective data illustrate what can potentially be learned about RNA spatial organization from the wide variety of information that can be extracted with DSSR and serve as a starting point for further multi-scale investigations of RNA folding.

## MATERIALS AND METHODS

### RNA Databases.

The modes of G·A association reported herein are taken from the RNA-containing structures found in the November 9, 2018 release (3.47) of the dataset of non-redundant high-resolution crystal structures curated by Leontis and Zirbel.<sup>59</sup> The sample includes 3256 G·A pairs in representative integrated function elements (IFEs) from 282 Protein Data Bank entries of 3.0 Å or better resolution. The RNA molecules comprise assorted ribosomes, riboswitches, ribozymes, transfer RNAs, messenger RNAs, guide RNAs, RNA aptamers, isolated double- and multi-stranded helices, and protein/DNA/drug-RNA assemblies (see complete listing in Table S1).

### Base-pair Identification/Characterization.

The modes of G·A association are based on the rigid-body parameters and hydrogen-bonding patterns derived from DSSR version 1.8.5 with a hydrogen-bond distance cutoff of 3.4 Å.<sup>27</sup> The pairs identified in the analysis (see complete listing in Table S2) do not include any of the close associations of G and A detected by the software with proton-donor interactions suggestive of base isomerization or protonation, e.g., the 53 examples of G·A pairs potentially held in place through G(N2)···A(N6) amino-imino hydrogen bonding.

The pairing between G and A can be in the form of A·G or G·A associations, depending upon the context or sequential order. At the base-pair level, however, the two forms are geometrically equivalent. Here, all the rigid-body parameters and hydrogen-bonding patterns are expressed in terms of G·A pairs. Thus, the values of parameters characterizing the 1360 occurrences of A·G pairing in the dataset are converted so that the buckle and shear of the A–G pairs are negated to describe an antiparallel interaction with respect to guanine as the leading residue, and all six parameters of A+G pairs are negated to describe a parallel interaction. The 3256 consolidated G·A associations include a total of 6873 hydrogen bonds, with a very small number of pairs (~1%) held in place by four or more donor-acceptor couples. Roughly a third of the hydrogen bonds involve the interaction of a base atom with a donor-acceptor atom on the sugar-phosphate backbone, most frequently with the guanine 2'-hydroxyl (~21%) and the adenine 5'-phosphate (~7%) groups but rarely with the adenine 2'-hydroxyl and the guanine 5'-phosphate. Interactions of the bases with the 3'-phosphate groups of G and A, i.e., the 5'-phosphate groups on succeeding residues, are also rare.

The rigid-body components that contribute to base-pair non-planarity — Stagger, Buckle, and Propeller — are distributed fairly uniformly about null values, with respective averages and standard deviations of  $0.15 \pm 1.00$  Å,  $-0.3 \pm 23.2^\circ$ , and  $-0.5 \pm 20.2^\circ$  (Figure S3). The three remaining parameters — Shear, Stretch, and Opening — cluster in several discrete ranges (Figure 3), which are useful in grouping the G·A pairs into distinct categories.<sup>24,27</sup> The

qualitative categories of G-A pairing are named on the basis of the locations of the atoms on one base with respect to the coordinate frame of the partner (reference) base. For example, given the design of the standard base frame with the positive  $x$ -axis pointing in the direction of the major groove and the positive  $y$ -axis away from the Watson-Crick edge<sup>30</sup> (Figure S1), base atoms with positive  $x$  and  $y$  coordinates interact with the reference base through the major groove M, those with negative  $x$  and positive  $y$  coordinates through the minor groove m, and those with small  $x$  and negative  $y$  values through the Watson-Crick edge W. In cases where the base position is ambiguous, the atoms of the sugar and 5'-phosphate are added to the categorization, and if the interaction is still unclear, the category is denoted by a dot. Outlying states where the arrangements of G and A differ significantly from the values of Shear, Stretch, Opening characteristic of the initially assigned grouping are reclassified through stepwise comparison with the parameters from other groupings until no parameter deviates from the averages of the new grouping by more than two standard deviations, a process that accounts for all but three of the G-A pairs in the current dataset and assures that each grouping of base pairs falls in a well-behaved region of configuration space. The scalar product between the base normals, i.e., the  $z$ -axes of the base reference frames, determines the sign of the interaction (either '+' for parallel or '-' for antiparallel). See the DSSR Manual at <http://docs.x3dna.org/dssr-manual.pdf> for further details.

Notably, the DSSR software returns the secondary structural context in which each of the bases occurs, including information on whether the bases lie within the same or a different structural unit and an expanded dot-bracket description of canonical pairing that serves as input for the simple secondary structural diagrams<sup>60</sup> that we report. Here we focus attention on whether the guanine and adenine bases occur within the four common RNA loop types — hairpin loops, internal loops (or bulges), and junctions — or within a canonical double-helical stem, where one of the bases participates in a Watson-Crick or wobble (G-U) pairing with a third base and the G-A pair forms part of a triplet. The loops identified by the software form 'closed' circles with sequential nucleotides connected by either a phosphodiester linkage or a canonical base pair. The number of stems determines the classification, with a hairpin loop closed by one stem, an internal loop or bulge by two stems, and a junction (multi-branched) loop by three or more stems. For simplicity, pseudoknots, which are also identified with the software, are not considered. Account is taken, however, of both higher-order multiplets, in which a G-A pair associates with additional bases, and single-stranded nucleotides outside of loops and double-helical stems. This knowledge of secondary and higher-order structure in combination with a quantitative depiction of the local geometry (rigid-body parameters) and hydrogen-bonding patterns of the G-A pairs makes it possible to see how the modes of base interaction and the range of local deformations depend upon structural context.

The range of conformational states associated with a specific pairing mode and/or structural context is described in terms of the volume of conformation space  $V_{bp}$  within one standard deviation of the average rigid-body parameters of the collected examples. Values of  $V_{bp}$  are obtained from the dispersion of parameters, given by the product of the eigenvalues of a matrix with elements  $\langle \Delta\theta_{ij}^2 \rangle$  that incorporates the covariance of all pairwise combinations of parameters,  $\langle \Delta\theta_{ij}^2 \rangle = \langle \theta_i \theta_j \rangle - \langle \theta_i \rangle \langle \theta_j \rangle$ .<sup>61</sup> The  $\theta_i$  ( $i=1-6$ ) correspond respectively to the values

of Shear, Stretch, Stagger, Buckle, Propeller, and Opening for each sample entry. The influence of specific variables and/or structural context on deformability is estimated from subsets of the data, e.g., out-of-plane parameters  $\theta_i$  ( $i=3-5$ ). Values are based on the full range of examples within a given structural category.

## RESULTS

### G-A Pairing Motifs and Transitions.

The rigid-body parameters used to describe the spatial arrangements of guanine and adenine in the chosen set of RNA structures reveal potential pathways connecting the many documented examples of G-A base pairing, including the six dominant edge-to-edge motifs<sup>9,62</sup> highlighted in Figures 1 and 2. Although the numerical values of the parameters corresponding to the various pairing schemes tend to cluster tightly in distinct domains, close examination of the data suggests possible links between the different 3D forms (Table 1, Figure 3). For example, a subset of the m-W associations of G and A that bring atoms along the minor-groove edge of guanine in hydrogen-bonding contact with atoms along the Watson-Crick edge of adenine (configurations labeled m-W<sub>II</sub> in Table 1) differ only slightly from many of the sheared m-M arrangements of the two bases in terms of the six rigid-body parameters. Other configurations of the m-W pairs (m-W<sub>I</sub> states in Table 1) adopt configurations more closely related to the hydrogen-bonded m-m arrangements of G and A, where the two bases interact via their minor-groove edges. The bimodality of the m-W associations stands out in the scatter and distributions of the Shear, Stretch, and Opening values of these pairs (Figure 3). The plotted points and histograms, color-coded to match the shades used in the preceding illustrations of G-A pairing modes and secondary structural occurrences, include two clusters of m-W configurations, one abutting the corresponding features of the sheared m-M states and the other in the vicinity of the m-M values. The images also disclose a subset of sheared G-A pairs in a secondary m-M<sub>II</sub> geometric arrangement, with slightly more positive Shear and more negative Opening than the dominant m-M<sub>I</sub> state (note the small secondary peaks in the distributions of individual parameters along the sides of the scatter plots).

Thus one can deduce a continuum of movements, i.e., the in-plane displacements and the relative rotation of A with respect to G, that convert one pairing mode to another. Specifically, the composite data show how decreases in Stretch and Opening in combination with the increase in Shear successively transform the pairing of G and A from an m-m arrangement of antiparallel bases to m-W and m-M forms (Figure 4a and Table 1). The bimodal distribution of m-W configurations also points to a ‘smooth’ pathway linking the arrangements of imino and sheared G-A pairs. The rigid-body parameters of some of the W-W imino associations lie close to those of a few of the m-W pairs and, as noted above, the parameters of certain m-W forms abut those of the sheared base pairs. The interconversion between imino and sheared states seemingly occurs in two stages, first via the coupled increase of Stretch and Opening from the W-W state to a subset of m-W arrangements with unique hydrogen bonding (see below) followed by the correlated changes in Shear, Stretch, and Opening (noted above) that lead to the sheared m-M state. Reductions in the values of Stretch and Opening similarly transform the pairing of G and A in parallel orientations from

m+W to m+m arrangements, albeit in combination with a decrease in Shear (Figure 4b). Changes in Shear, Stretch, and Opening of the opposite sense convert the m+W state to a low-populated m+M form (Figure 4b and Table 1).

### Hydrogen-bond Rearrangements.

The relative movements of G and A naturally change the hydrogen-bonding patterns that stabilize the various associations of the two bases. The interactions found in one pairing motif, i.e., the polar atoms that share a hydrogen atom, however, may persist in a closely related form. For example, the N3 atom of guanine accepts a hydrogen atom from the exocyclic N6 amino group of adenine in both m-W<sub>II</sub> and m-M<sub>I</sub> base-paired arrangements, although with differently oriented hydrogens from the N6 amino group and with the change in base orientation attended by different hydrogen bonds (Figure 5a). The W-W and m-W<sub>I</sub> states similarly include a common G(N2)⋯A(N1) hydrogen bond that occurs in concert with G(N1)⋯A(N1) and G(O6)⋯A(N6) interactions in the former arrangement and a G(O2')⋯A(N3) interaction in the latter (Figure 5b). By contrast, the hydrogen bonding schemes in W-M pairs are identical to those of the small subset of sheared m-M<sub>II</sub> pairs. The N2 amino group of guanine associates exclusively with the adenine 5'-phosphate and N7 atom in the two forms, but in slightly different rotational settings of the sugar-phosphate backbone (Figure 5c). The 59 examples of W-M pairing occur in 29 different RNA structures, including some riboswitches and ribozymes, the ribonuclease P specificity domain, and assorted ribosomal fragments and intact ribosomes (see Tables S1 and S2 for details).

Hydrogen bonding contributions from the ribose similarly persist in closely related modes of G·A pairing. For example, the paired association of 2'-hydroxyl groups, one from G and the other from A, sometimes components of a ribose zipper,<sup>63</sup> occurs in both m-m and m-W<sub>I</sub> pairs (Figure 5d). Indeed, the guanine 2'-hydroxyl appears to serve as a pivot point for the rearrangement of G·A pairs, forming hydrogen bonds with different atoms on adenine as G and A convert from m-m to m-M<sub>I</sub> arrangements. The m-m states are anchored by interactions of guanine O2' with adenine O3' and O2', the m-W states by interactions with adenine O2', N3, and N1, and the m-M<sub>I</sub> states by interactions with adenine N6 (Figure S4). These and other patterns of hydrogen-bond association — e.g., the persistence of G(O6)⋯A(N6) associations in W-W, m-W, m-M, and various intermediate forms — are at once evident from comparison of scatter plots of the rigid-body parameters color-coded by G·A pairing mode (Figure 3) and by specific hydrogen-bond components (Figure S5).

Like their antiparallel m-W counterparts, the parallel m+W arrangements of guanine and adenine fall into two distinct classes with different hydrogen-bonding motifs (Figures 3, 4). The majority of m+W<sub>I</sub> examples — held in place by G(O2')⋯A(N6), G(N2)⋯A(N1) and G(N3)⋯A(N6) hydrogen bonds — adopt more positive values of Shear, Stretch, and Opening than their minor m+W<sub>II</sub> counterparts (Table 1). The m+W<sub>II</sub> population, stabilized by G(N2)⋯A(N3) and G(O2')⋯A(N1) associations, differs only slightly from the m+m arrangements of the two bases. Some of the hydrogen bonds persist in lesser-populated parallel forms, revealing a potential route between the 20 W+W, 69 W+M, 26 M+M, 29 M



+W, and 53 m+M collected pairings of guanine and adenine via the minor m+W<sub>II</sub> state (see Table 1 and Figure S4).

### Secondary Structural Context and Long-range Interactions.

The capability to detect the secondary structural context of the G·A pairs provides insights into how the modes of base association influence the higher-order spatial organization of RNA. For example, in nearly half the observed cases (627/1338) the sheared m–M pairs occur within internal loops (Table 2). Moreover, all but five of the bases in these examples lie within the same loop (Figure S6). The sheared G·A pairs occur in lesser numbers in hairpin and junction loops, with 351 and 261 examples, respectively, in the current dataset. The vast majority of the base associations in the latter loops also occur in the same secondary structural element (Figure S6) although there are eight examples, all in the structures of riboswitches, where a sheared G·A pair entails bases in different hairpin loops, e.g., G19 in loop L2 and A90 in loop L6 of the flavin mononucleotide riboswitch (Figure S7).<sup>70</sup> Interestingly, the subset of G·A pairs with more positive Shear and more negative Opening values, the minor m–M<sub>II</sub> states noted above, occur exclusively in hairpin loops, all of length four and all involving a G(N1)···A(OP2) hydrogen bond unique to these arrangements (e.g., Figures 2a and S5). These base pairs constitute about a third of the GNRA tetraloops within the dataset and as noted above are further stabilized by hydrogen bonding of G(N2) with A(N7) and A(OP2).

As evident from Table 2, none of the adenines and only a small number of guanines in the set of sheared G·A pairs fall within a double-helical stem and thereby also participate in a canonical Watson-Crick or wobble base pair. The simultaneous involvement of guanine in both a canonical and a sheared base pair leads to the formation of a triplet, e.g., a C–G–A hydrogen-bonded association with the G contacting the A via the sheared G–A arrangement and the C through a G–C Watson-Crick pair. Surprisingly, despite the few examples of either G or A in a double-helical stem, nearly a quarter of the m–M arrangements of the two bases (296/1338) occur in multiplets. Moreover, over half of these multiplets (160/296) involve a different type of G·A pairing, including 78 m+m, 32 m+W, and 21 m–m associations. These multiplets bring different pieces of RNA secondary structure in contact with one another and also contribute to the spatial organization of RNA junctions (Figures 6 and S8). The connections more frequently involve an adenine shared by two guanines, one from the m–M interaction and the other from a different mode of G·A pairing, rather than a guanine shared by two adenines (145 examples of noncanonical G–A–G or G–A+G multiplets vs. 15 examples of A–G–A or A+G–A interactions).

Despite the different mode of interaction, the imino W–W associations of guanine and adenine resemble the sheared G·A pairs in several respects. All but one of the imino interactions occur within the same secondary structural unit, and nearly a fifth (62/344) occur within a multiplet. Like the multiplets containing sheared G·A pairs, a large proportion of the multiplets containing imino pairs (40/62) entail a different mode of G·A pairing (including 12 m+W, 2 m–m, 6 m+m associations) and only a few incorporate a base in a canonical pair (five with G, two with A). The imino pairs also occur in roughly half the cases (169/344) within internal loops (e.g., Figure 2b) but rarely in hairpin loops (Table 2

and Figure S6a). The proportion of W–W pairs involving unstructured regions, such as the single-stranded fragments found at the ends of RNA chains, exceeds that of sheared m–m G·A base pairs by nearly a factor of two (30/344 vs. 66/1338). The W–W pairs frequently ‘extend’ a canonical double-helical stem, such as the G88·A10 association that abuts helix P2 in the tetrahydrofolate riboswitch<sup>71</sup> and respectively links the single-stranded 3′- and 5′- tails of the molecule.

In contrast to the sheared and imino arrangements of G·A pairs, there are relatively few cases where m±m associations of the base minor-groove edges occur within the same secondary structural element (Table 2). Moreover, most of these interactions form within multiplets (372/395 m+m, 136/141 m–m) and a large proportion of the collected pairs include a guanine canonically hydrogen-bonded to C or U (312/395 m+m, 102/141 m–m; see Figures 2c,d). The adenine, almost always free of canonical A·U pairing (394/395 m+m, 138/141 m–m), typically lies outside of a double-helical stem and in over a third of the examples within a junction loop regardless of base orientation (164/395 m+m, 52/141 m–m). The sign of base association influences the relative numbers of adenines in hairpin vs. internal loops (107:143 m+m vs. 54:57 m–m) and the likelihood, albeit small, that minor-groove edge associations of G and A occur within the same secondary structural unit (27 of 44 m+m vs. 0 of 10 m–m internal loops). The m±m G·A pairs thus tend to ‘glue’ the guanine in a double-helical stem to different looped segments of an RNA structure (see examples in Figure 6).

Given that the loops include the base pairs found at the ends of double-helical stems and in isolated Watson-Crick pairs (see Materials and Methods), the counts of secondary structural motifs exceed the number of pairs of a given type, e.g., the 395 examples of m+m pairing occur in 496 secondary structural settings. Notable in this regard are (i) arrangements of G and A counted among both the m+m G·A pairs in internal loops and the m+m examples with G in a stem and A in an internal loop and (ii) arrangements counted among both the m+m G·A pairs in junctions and the m+m examples with G in a stem and A in a junction. The double counting of guanines within looped secondary structural units similarly accounts for differences between the number of observed m–m G·A pairs and the number of secondary structural settings in which they occur (141 m–m pairs vs. 197 settings).

The m±W associations of G and A resemble the m±m pairs in that they too occur in multiplets and serve as links between different RNA secondary structural units (Table 2). Although comparable in number to their m±m counterparts, the m±W pairs within multiplets (309/388 m+W and 118/153 m–W) occur with slightly lower frequency, i.e., roughly 80% of the m±W pairs vs. over 90% of the m±m pairs. These differences arise in part from the greater likelihood of the G in the m±W pairs to occur within a hairpin loop as opposed to a double-helical stem (e.g., Figures 2e,f). As a consequence, the m±W pairs are more apt to mediate interactions between different loops (Figure 6).

### Deformability/Context.

The distributions of rigid-body parameters describing the various modes of G·A association provide insight into the relative flexibility of the different arrangements. The volumes of rigid-body space  $V_{bp}$ , obtained by diagonalizing the matrix of parameter covariances (see

Materials and Methods),<sup>61</sup> offer a measure of the deformability seen in superimposed images of the bases comprised in the different pairing schemes (Figure S9). When combined with the average base-pair parameters, the set of volumes complete the ‘fingerprint’ of each G·A pair, making it possible to identify potentially tighter and looser pairing modes. At once apparent from the volumes of the major pairing arrangements (Figure 7) is the marked stiffness of the m+m associations compared to all other G·A pairs. The small conformational volume reflects the much narrower ranges of the pattern-specific variables — Shear, Stretch, Opening, i.e.,  $\theta_1, \theta_2, \theta_6$  — in the documented examples. That is, the very small conformational volume of m+m states determined from Shear, Stretch, Opening has a greater influence on overall base-pair deformability than the out-of-plane parameters — Stagger, Buckle, and Propeller, i.e.,  $\theta_3, \theta_4, \theta_5$  (Figure 7a). If the analysis were confined to the out-of-plane parameters, the imino W–W arrangements of G and A would appear to be stiffer than most pairing modes. If all six rigid-body parameters are considered, the imino pairs are more flexible than both the m+m and sheared m–M pairs (also see Figure S9). The m–m and m±W associations of G and A are even more deformable in terms of occupied conformational space. The larger conformational volumes of the m±W arrangements over the m–m forms reflect the wider ranges of Stagger, Buckle, and Propeller in the former pairs (Figure 7a).

The deformability of the various G·A pairing schemes further depends upon structural context. For example, the sheared m–M pairs span a 1.3-fold wider range of rigid-body conformation space when located in hairpin loops as opposed to internal loops but a comparable proportion of the space adopted within junctions (Figure 7b). The imino W–W pairs show an even more pronounced (4.3-10.1-fold) enhancement in deformability when formed within junctions as opposed to internal loops and hairpins. The many m+m pairs that link a G within a double-helical stem to an A in a loop also span a wider range of conformation space than the smaller number of examples of the same pairing mode within a junction but a narrower range of space than the even fewer examples of the pair within an internal loop. The bridging m+m pairs, however, are more than an order of magnitude (over 14-fold) stiffer than their m+W counterparts in terms of rigid-body parameters. The bridging m+W linkers occupy larger conformational volumes than G·A pairs of the same type within hairpin and internal loops but roughly the same volume as those in junctions. The limited number of examples of m–m and m–W pairs in secondary structural settings other than stem-loop connectors rules out comparisons of deformability of such arrangements in different secondary structural contexts.

## DISCUSSION

Conventional understanding of RNA folding derives in large part from the qualitative descriptions of macromolecular organization collected over time through the determination of individual, high-resolution structures, e.g., the noncanonical base pairs and the organized hairpin folds first observed in the yeast tRNA<sup>Phe</sup> crystal,<sup>1</sup> the A-minor motifs that accompany the associations of secondary structural units in large ribosomes and ribozymes,<sup>13,72</sup> the novel base triples found in the *Tetrahymena* telomerase RNA pseudoknot,<sup>73</sup> etc. Despite the vast accumulated knowledge, understanding how the molecular components fit together remains an open question. While computations now capture the 3D folds of

assorted small RNA molecules from combinations of local structural, chemical, and genetic information,<sup>74,75</sup> viable predictions of large ribozyme and ribosomal structures are still beyond reach. Success in this regard requires a higher-level perspective, incorporating information beyond the spatial arrangements and composition of isolated base pairs and secondary structural units and the sites of chemical modification or mutation. This article illustrates how one can begin to attack the problem by taking advantage of the multi-scale structural information collected with DSSR, including the identities, spatial arrangements, and secondary structure context of associated bases, as well as the nucleotide configurations and global arrangements of secondary structural units.<sup>27</sup>

Here we focus on a subset of noncanonical base pairs involving guanine and adenine and the links between their modes of association, secondary structural context, and contributions to tertiary folding. The precise numerical descriptions of base-pair geometry that we employ not only facilitate classification of recurrent geometric motifs and the settings in which these arrangements occur but also hint of the natural motions of the interacting bases and the pathways likely to connect different spatial forms. The scatter of rigid-body parameters relating coordinate frame on the two bases point to ways by which the sheared and imino G·A pairs typically found within RNA hairpins, junctions, and internal loops might convert into spatially related forms that bridge different pieces of RNA secondary structure. The data also hint of a two-step mechanism linking sheared and imino G·A pairs, a result of potential utility in guiding computer-simulations of transitions between the two forms.<sup>50</sup> The reorientation of adenine with respect to guanine during these processes has a clear effect on RNA secondary structure. For example, the bulges incorporating sheared G·A pairs in the dataset of representative structures differ in length and assume quite different folds from those found in bulges with closely related (m–W<sub>II</sub>) pairing. The former bulges tend to be shorter with simple base flip-outs compared to the latter, which adopt more circuitous pathways (Figure S10).

The current study also draws attention to higher-order multiplexes of two or more G·A pairs bridging different secondary structural units. Although included in earlier collections of higher-order base pairings,<sup>76</sup> most structural work has focused on the three-base, i.e., triplet, interactions formed by an adenine within a loop and a guanine hydrogen-bonded to C or U in a double-helical fragment. The states termed m+m here correspond to the type I A-minor motif first identified by Nissen *et al.* in the *Haloarcula marismortui* large ribosomal subunit<sup>13</sup> and the m–W states to the type II triples found by Doherty *et al.* to contribute to helix packing in the *Tetrahymena thermophila* group I ribozyme.<sup>72</sup> The present survey uncovers even more extensive base associations involving different modes and numbers of G·A pairs, e.g., tetraplexes and pentaplexes linking hairpins and internal loops to double-helical stems and junctions. The collective data further reveal pairing propensities in base organization and secondary structural context. For example, the vast majority of sheared and imino G·A base pairs (m–M and W–W) occur within a looped region, and almost always in the same loop. The guanine and adenine pairs that bridge different secondary structural units preferentially adopt parallel over antiparallel forms (m+m, m+W vs. m–m, m–W). Not surprisingly, the bridging G·A pairs tend to be more deformable than their sheared and imino counterparts within loops in terms of the observed distributions of spatial states.

The rigid-body parameters used to characterize the G-A pairs differ in some studies from those reported here owing to a different choice of base reference frame. Rather than expressing the geometric arrangement of two bases in terms of the standard reference frame,<sup>30</sup> other groups prefer to describe the interactions in terms of base-edge specific reference frames. That is, the location and orientation of the base frames depend upon the type of association.<sup>47</sup> The latter approach masks the pattern-specific parameters used here to distinguish modes of pairing and instead draws attention to the differences in base-pair non-planarity that accompany the various modes of interaction. This different point of view may underlie the greater deformability of sheared compared to imino G-A pairs deduced from studies based on this vantage point.<sup>47</sup> The treatment of deformability based solely on the out-of-plane rigid-body parameters collected in this survey would lead to a similar conclusion, given that the conformational volume occupied by Stagger, Buckle, and Propeller is slightly (1.05 ×) greater for sheared m-M compared to imino W-W pairs in the current dataset (Figure 7a). As demonstrated here, the complete set of pattern-specific parameters obtained with the standard reference frame is needed to understand and characterize the relative deformations and transitions between different modes of G-A pairing.

Qualitative descriptions of non-canonical RNA base pairing, pioneered by Leontis and Westhof<sup>9,41</sup> and linked in this work to the rigid-body parameters of interacting bases, have proven valuable in deciphering the connections between RNA primary, secondary, and tertiary structures. The present categorization is based on the positions of the hydrogen-bonded atoms with respect to a standard, embedded base reference frame<sup>30</sup> defined in terms of an idealized Watson-Crick base pair. The major- and minor-groove base edges used here correspond in most cases to what are termed the Hoogsteen and sugar edges in the Leontis-Westhof scheme (one can compare the two classification schemes in Table S2). The + and – symbols introduced in 3DNA<sup>24</sup> and DSSR<sup>27</sup> unambiguously distinguish the relative orientations of the two bases. The *trans* and *cis* designations used in the earlier literature, however, are qualitative in nature and often uncertain. There are many ‘nc’ (near *cis*, as in ncWW) and ‘nt’ (near *trans*, as in ntSH) annotations listed in the RNA Structure Atlas; see, for example, the base pair interactions in the sarcin-ricin domain of *E. coli* 23S rRNA found by entering PDB entry 1msy at URL:<http://rna.bgsu.edu/rna3dhub/pdb>. The assignment of qualitative descriptors of RNA associations on the basis of atomic identity alone is generally not clear-cut. Numerical differences in the rigid-body parameters are critical to differentiating pairing schemes that share a common hydrogen bond, e.g., the G(N3)···A(N6) interaction found in m-W<sub>II</sub> and m-M<sub>I</sub> arrangements of G and A (Table 1, Figures 4 and S3). The numerical data also provide a basis for following conformational transitions and may potentially be of value in making functional and other meaningful distinctions among RNA base pairs.

Finally, our findings illustrate what can potentially be learned about RNA spatial organization by taking advantage of the variety of information that can now be collected with software, i.e., DSSR, designed to streamline the analysis and annotation of RNA 3D structures.<sup>27</sup> The results serve as a starting point for further multi-scale investigations of RNA folding. In addition to examining the linkages between noncanonical base pairing and secondary structures reported here, one can immediately address classic biochemical

questions with the accumulated data, such as the locations of noncanonical pairs relative to the ends of double-helical stems in RNA molecules<sup>21</sup> or the sites of noncanonical pairing within and between different types of loops.<sup>12,77</sup> It is also possible to explore connections between the modes of base pairing and a number of 3D structural motifs linked to RNA tertiary interactions and automatically identified by the DSSR software, e.g., kissing-loop, A-minor, ribose-zipper, kink-turn, U-turn, pseudoknot, G-quadruplex, i motif. Finally, one can investigate the noncanonical pairs in the context of the global axes of the double-helical stems and/or coaxially stacked helices comprised within a structure and introduce these components in simplified depictions of overall chain folding.<sup>78</sup>

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### Funding

The authors are grateful to the National Institute of General Medical Sciences for financial support (GM034809 to wko and GM096889 to xjl).

## REFERENCES

- (1). Rich A, and RajBhandary UL (1976) Transfer RNA: molecular structure, sequence, and properties, *Annu Rev Biochem* 45, 805–860. [PubMed: 60910]
- (2). Ferré-D'Amaré AR, and Doudna JA (1999) RNA folds: insights from recent crystal structures, *Annu Rev Biophys Biomol Struct* 28, 57–73. [PubMed: 10410795]
- (3). Hermann T, and Patel DJ (1999) Stitching together RNA tertiary architectures, *J Mol Biol* 294, 829–849. [PubMed: 10588890]
- (4). Draper DE (1995) Protein-RNA recognition, *Annu Rev Biochem* 64, 593–620. [PubMed: 7574494]
- (5). Hermann T, and Westhof E (1999) Non-Watson-Crick base pairs in RNA-protein recognition, *Chem Biol* 6, R335–R343. [PubMed: 10631510]
- (6). Rupert PB, and Ferré-D'Amaré AR (2000) SRPrises in RNA-protein recognition, *Structure* 8, R99–R104. [PubMed: 10896472]
- (7). Schlatterer JC, Kwok LW, Lamb JS, Park HY, Andresen K, Brenowitz M, and Pollack L (2008) Hinge stiffness is a barrier to RNA folding, *J Mol Biol* 379, 859–870. [PubMed: 18471829]
- (8). Mohan S, and Noller HF (2017) Recurring RNA structural motifs underlie the mechanics of L1 stalk movement., *Nat Commun* 8, 14285. [PubMed: 28176782]
- (9). Leontis NB, and Westhof E (1998) Conserved geometrical base-pairing patterns in RNA, *Q Rev Biophys* 31, 399–455. [PubMed: 10709244]
- (10). Woese CR, Winker S, and Gutell RR (1990) Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops", *Proc Natl Acad Sci, USA* 87, 8467–8471. [PubMed: 2236056]
- (11). Shen LX, Cai Z, and Tinoco I Jr (1995) RNA structure at high resolution., *FASEB J* 9, 1023–1033. [PubMed: 7544309]
- (12). Chen G, Znosko BM, Kennedy SD, Krugh TR, and Turner DH (2005) Solution structure of an RNA internal loop with three consecutive sheared GA pairs, *Biochemistry* 44, 2845–2856. [PubMed: 15723528]
- (13). Nissen P, Ippolito JA, Ban N, Moore PB, and Steitz TA (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif, *Proc Natl Acad Sci, USA* 98, 4899–4903. [PubMed: 11296253]

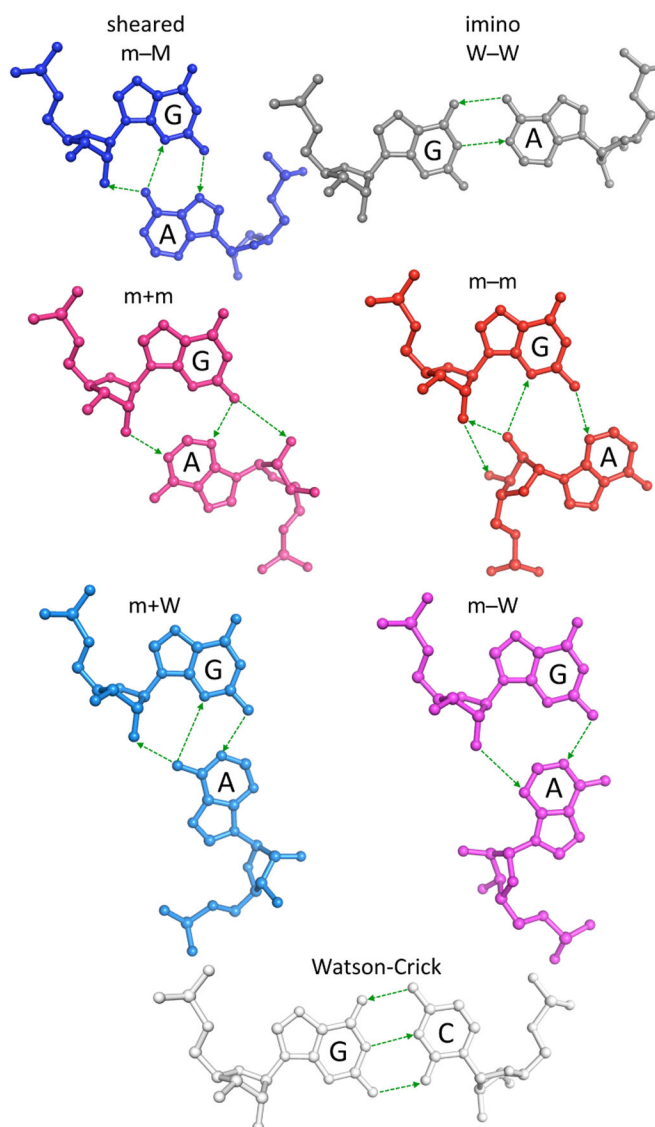
- (14). Klein DJ, Schmeing TM, Moore PB, and Steitz TA (2001) The kink-turn: a new RNA secondary structure motif, *EMBO J* 20, 4214–4221. [PubMed: 11483524]
- (15). Lilley DM (2012) The structure and folding of kink turns in RNA, *Wiley Interdiscip Rev RNA* 3, 797–805. [PubMed: 22976946]
- (16). Wimberly B, Varani G, and Tinoco I Jr (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA, *Biochemistry* 32, 1078–1087. [PubMed: 8424938]
- (17). Correll CC, Beneken J, Plantinga MJ, Lubbers M, and Chan Y-L (2003) The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure, *Nucleic Acids Res* 31, 6806–6818. [PubMed: 14627814]
- (18). Lu X-J, Olson WK, and Bussemaker HJ (2010) The RNA backbone plays a crucial role in mediating the intrinsic stability of the GpU dinucleotide platform and the GpUpA/GpA miniduplex, *Nucleic Acids Res* 38, 4868–4876. [PubMed: 20223772]
- (19). Szewczak AA, Moore PB, Chang Y-L, and Wool IG (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA, *Proc Natl Acad Sci, USA* 90, 9581–9585. [PubMed: 8415744]
- (20). Lerman YV, Kennedy SD, Shankar N, Parisien M, Major F, and Turner DH (2011) NMR structure of a 4 × 4 nucleotide RNA internal loop from an R2 retrotransposon: identification of a three purine-purine sheared pair motif and comparison to MC-SYM predictions, *RNA* 17, 1664–1677. [PubMed: 21778280]
- (21). Elgavish T, Cannone JJ, Lee JC, Harvey SC, and Gutell RR (2001) AA.AG@helix.ends: A:A and A:G base-pairs at the ends of 16 S and 23 S rRNA helices, *J Mol Biol* 310, 735–753. [PubMed: 11453684]
- (22). Laing C, and Schlick T (2009) Analysis of four-way junctions in RNA structures, *J Mol Biol* 390, 547–559. [PubMed: 19445952]
- (23). Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, Kennard O, von Kitzing E, Lavery R, Nelson HCM, Olson WK, Saenger W, Shakked Z, Sklenar H, Soumpasis DM, Tung C-S, Wang AH-J, and Zhurkin VB (1989) Definitions and nomenclature of nucleic acid structure parameters, *Nucleic Acids Res* 17, 1797–1803. [PubMed: 2928107]
- (24). Lu X-J, and Olson WK (2003) 3DNA: a software package for the analysis, rebuilding, and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res* 31, 5108–5121. [PubMed: 12930962]
- (25). Xin Y, and Olson WK (2009) BPS: a database of RNA base-pair structures, *Nucleic Acids Res* 37, D83–D88. [PubMed: 18845572]
- (26). Olson WK, Esguerra M, Xin Y, and Lu X-J (2009) New information content in RNA base pairing deduced from quantitative analysis of high-resolution structures, *Methods* 47, 117–186.
- (27). Lu X-J, Bussemaker HJ, and Olson WK (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA, *Nucleic Acids Res* 43, e142. [PubMed: 26184874]
- (28). Lu X-J, and Olson WK (1999) Resolving the discrepancies among nucleic acid conformational analyses, *J Mol Biol* 285, 1563–1575. [PubMed: 9917397]
- (29). Lu X-J, Babcock MS, and Olson WK (1999) Overview of nucleic acid analysis programs, *J Biomol Struct Dyn* 16, 833–843. [PubMed: 10217453]
- (30). Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu X-J, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung C-S, Westhof E, Wolberger C, and Berman HM (2001) A standard reference frame for the description of nucleic acid base-pair geometry, *J Mol Biol* 313, 229–237. [PubMed: 11601858]
- (31). Rose IA, Hanson KR, Wilkinson KD, and Wimme MJ (1980) A suggestion for naming faces of ring compounds, *Proc Natl Acad Sci, USA* 77, 2439–2441. [PubMed: 16592816]
- (32). Lavery R, Zakrzewska K, Sun J-S, and Harvey SC (1992) A comprehensive classification of nucleic acid structural families based on strand direction and base pairing, *Nucleic Acids Res* 20, 5011–5016. [PubMed: 1383936]
- (33). Burkard ME, Turner DH, and Tinoco I Jr (1999) Structures of base pairs involving at least two hydrogen bonds, In *The RNA World.*, Second Edition (Gesteland R, Cech T, and Atkins J, Eds.), pp 675–680, Cold Spring Harbor Laboratory Press, New York.

- (34). Chandrasekaran R, and Arnott S (1989) The structures of DNA and RNA helices in oriented fibers, In Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology, Group VII/1b, Nucleic Acids (Saenger W, Ed.), pp 31–170, Springer-Verlag, Berlin.
- (35). Klein DJ, Wilkinson SR, Been MD, and Ferré-D'Amaré AR (2007) Requirement of helix P2.2 and nucleotide G1 for positioning the cleavage site and cofactor of the glmS ribozyme, *J Mol Biol* 373, 178–189. [PubMed: 17804015]
- (36). Polikanov YS, Melnikov SV, Söll D, and Steitz TA (2015) Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly, *Nat Struct Mol Biol* 22, 342–344. [PubMed: 25775268]
- (37). Tishchenko S, Nikonova E, Nikulin A, Nevskaya N, Volchkov S, Piendl W, Garbera M, and Nikonova S (2006) Structure of the ribosomal protein L1-mRNA complex at 2.1 Å resolution: common features of crystal packing of L1-RNA complexes, *Acta Crystallogr D Biol Crystallogr* 62, 1545–1554. [PubMed: 17139090]
- (38). Ren A, and Patel DJ (2014) c-di-AMP binds the ydaO riboswitch in two pseudo-symmetry-related pockets, *Nat Chem Biol* 10, 780–786. [PubMed: 25086509]
- (39). Basavappa R, and Sigler PB (1991) The 3 Å crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination, *EMBO J* 10, 3105–3111. [PubMed: 1915284]
- (40). Matzov D, Aibara S, Basu A, Zimmerman E, Bashan A, Yap M-NF, Amunts A, and Yonath AE (2017) The cryo-EM structure of hibernating 100S ribosome dimer from pathogenic *Staphylococcus aureus*, *Nat Commun* 8, 723. [PubMed: 28959035]
- (41). Leontis NB, and Westhof E (2001) Geometric nomenclature and classification of RNA base pairs, *RNA* 7, 499–512. [PubMed: 11345429]
- (42). Lemieux S, and Major F (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire, *Nucleic Acids Res* 30, 4250–4263. [PubMed: 12364604]
- (43). Nagaswamy U, Larios-Sanz M, Hury J, Collins S, Zhang Z, Zhao Q, and Fox GE (2002) NCIR: a database of non-canonical interactions in known RNA structures, *Nucleic Acids Res* 30, 395–397. [PubMed: 11752347]
- (44). Cheng AC, Chen WW, Fuhrmann CN, and Frankel AD (2003) Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains, *J Mol Biol* 327, 781–796. [PubMed: 12654263]
- (45). Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, and Westhof E (2003) Tools for the automatic identification and classification of RNA base pairs, *Nucleic Acids Res* 31, 3450–3460. [PubMed: 12824344]
- (46). Sarver M, Zirbel CL, Stombaugh J, Mokdad A, and Leontis NB (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures, *J Math Biol* 56, 215–252. [PubMed: 17694311]
- (47). Mukherjee S, Bansal M, and Bhattacharyya D (2006) Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis, *J Comput Aided Mol Des* 20, 629–645. [PubMed: 17124630]
- (48). Roy A, Panigrahi S, Bhattacharyya M, and Bhattacharyya D (2008) Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies, *J Phys Chem B* 112, 3786–3796. [PubMed: 18318519]
- (49). Zacharias M, and Sklenar H (2000) Conformational deformability of RNA: a harmonic mode analysis, *Biophys J* 78, 2528–2542. [PubMed: 10777750]
- (50). Aytenfisu AH, Spasic A, Seetin MG, Serafini J, and Mathews DH (2014) Modified Amber force field correctly models the conformational preference for tandem GA pairs in RNA, *J Chem Theory Comput* 10, 1292–1301. [PubMed: 24803859]
- (51). Bottaro S, Palma FD, and Bussi G (2014) The role of nucleobase interactions in RNA structure and dynamics, *Nucleic Acids Res* 42, 13306–13314. [PubMed: 25355509]
- (52). Das R, Karanicolas J, and Baker D (2010) Atomic accuracy in predicting and designing noncanonical RNA structure, *Nat Methods* 7, 291–294. [PubMed: 20190761]

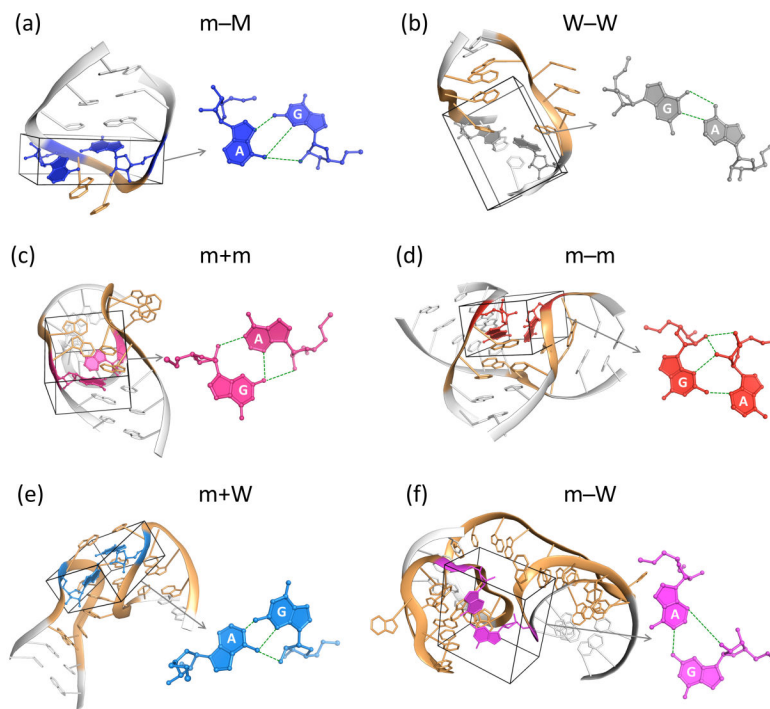


- (53). Bernauer J, Huang X, Sim AYL, and Levitt M (2011) Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation, *RNA* 17, 1066–1075. [PubMed: 21521828]
- (54). Morgado CA, Svozil D, Turner DH, and Šponer J (2012) Understanding the role of base stacking in nucleic acids. MD and QM analysis of tandem GA base pairs in RNA duplexes, *Phys Chem Chem Phys* 14, 12580–12591. [PubMed: 22722325]
- (55). SantaLucia J Jr, and Turner DH (1993) Structure of (rGGCGAGCC)<sub>2</sub> in solution from NMR and restrained molecular dynamics, *Biochemistry* 32, 12612–12623. [PubMed: 8251479]
- (56). Wu M, and Turner DH (1996) Solution structure of (rGCGGACGC)<sub>2</sub> by two-dimensional NMR and the iterative relaxation matrix approach, *Biochemistry* 35, 9677–9689. [PubMed: 8703939]
- (57). Tolbert BS, Kennedy SD, Schroeder SJ, Krugh TR, and Turner DH (2007) NMR structures of (rGCUGAGGCU)<sub>2</sub> and (rGCGGAUGCU)<sub>2</sub>: probing the structural features that shape the thermodynamic stability of GA pairs, *Biochemistry* 46, 1511–1522. [PubMed: 17279616]
- (58). Berger KD, Kennedy SD, and Turner DH (2019) Nuclear magnetic resonance reveals that GU base pairs flanking internal loops can adopt diverse structures, *Biochemistry* 58, 1094–1108. [PubMed: 30702283]
- (59). Leontis NB, and Zirbel CL (2012) Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking, In *RNA 3D Structure Analysis and Prediction* (Leontis N, and Westhof E, Eds.), pp 281–298, Springer, Berlin Heidelberg.
- (60). Kerpedjiev P, Hammer S, and Hofacker IL (2015) Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams, *Bioinformatics* 31, 3377–3379. [PubMed: 26099263]
- (61). Olson WK, Gorin AA, Lu X-J, Hock LM, and Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc Natl Acad Sci, USA* 95, 11163–11168. [PubMed: 9736707]
- (62). Sweeney BA, Roy P, and Leontis NB (2015) An introduction to recurrent nucleotide interactions in RNA, *Wiley Interdiscip Rev RNA* 6, 17–45. [PubMed: 25664365]
- (63). Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, and Doudna JA (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing, *Science* 273, 1678–1685. [PubMed: 8781224]
- (64). Huang L, Wang J, and Lilley DMJ (2016) A critical base pair in k-turns determines the conformational class adopted, and correlates with biological function, *Nucleic Acids Res* 44, 5390–5398. [PubMed: 27016741]
- (65). Lu M, and Steitz TA (2000) Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Å resolution, *Proc Natl Acad Sci, USA* 97, 2023–2028. [PubMed: 10696113]
- (66). Shalev-Benami M, Zhang Y, Matzov D, Halfon Y, Zackay A, Rozenberg H, Zimmerman E, Bashan A, Jaffe CL, Yonath A, and Skiniotis G (2016) 2.8-Å cryo-EM structure of the large ribosomal subunit from the Eukaryotic parasite *Leishmania*, *Cell Rep* 16, 288–294. [PubMed: 27373148]
- (67). Ben-Shem A, Loubresse N. G. d., Melnikov S, Jenner L, Yusupova G, and Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution, *Science* 334, 1524–1529. [PubMed: 22096102]
- (68). Agalarov SC, Prasad GS, Funke PM, Stout CD, and Williamson JR (2000) Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain, *Science* 288, 107–113. [PubMed: 10753109]
- (69). Cocozaki AI, Altman RB, Huang J, Burman ET, Kazmirski SL, Doig P, Prince DB, Blanchard SC, Cate JHD, and Ferguson AD (2016) Resistance mutations generate divergent antibiotic susceptibility profiles against translation inhibitors, *Proc Natl Acad Sci, USA* 113, 8188–8193. [PubMed: 27382179]
- (70). Serganov A, Huang L, and Patel DJ (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch, *Nature* 458, 233–237. [PubMed: 19169240]

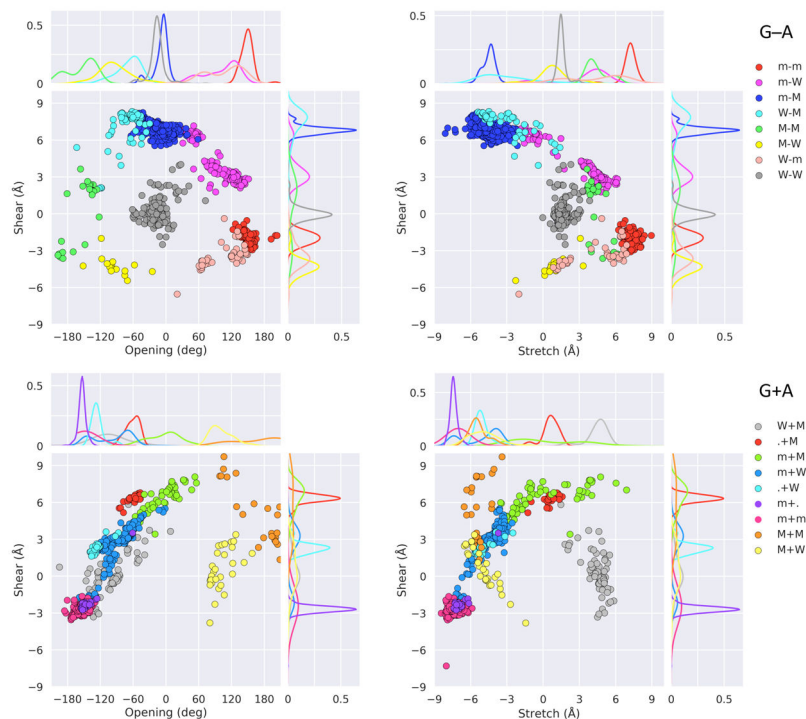
- (71). Huang L, Ishibe-Murakami S, Patel DJ, and Serganov A (2011) Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch, *Proc Natl Acad Sci, USA* 108, 14801–14806. [PubMed: 21873197]
- (72). Doherty EA, Batey RT, Masquida B, and Doudna JA (2001) A universal mode of helix packing in RNA, *Nat Struct Biol* 8, 339–343. [PubMed: 11276255]
- (73). Cash DD, and Feigon J (2017) Structure and folding of the *Tetrahymena* telomerase RNA pseudoknot, *Nucleic Acids Res* 45, 482–495. [PubMed: 27899638]
- (74). Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, Boniecki MJ, Bujnicki JM, Chen S-J, Cheng CY, Chou F-C, Ferré-D'Amaré AR, Das R, Dawson WK, Ding F, Dokholyan NV, Dunin-Horkawicz S, Geniesse C, Kappel K, Kladwang W, Krokhotin A, Łach GE, Major F, Mann TH, Magnus M, Pachulska-Wieczorek K, Patel DJ, Piccirilli JA, Popena M, Purzycka KJ, Ren A, Rice GM, Santalucia J Jr, Sarzynska J, Szachniuk M, Tandon A, Trausch JJ, Tian S, Wang J, Weeks KM, Williams II B, Xiao Y, Xu X, Zhang D, Zok T, and Westhof E (2017) RNA-puzzles round III: 3D RNA structure prediction of five riboswitches and one ribozyme, *RNA* 23, 655–672. [PubMed: 28138060]
- (75). Watkins AM, Geniesse C, Kladwang W, Zakrevsky Paul, Jaeger L, and Das R (2018) Blind prediction of noncanonical RNA structure at atomic accuracy, *Sci Adv* 4, eaar5316. [PubMed: 29806027]
- (76). Abu Almakarem AS, Petrov AI, Stombaugh J, Zirbel CL, and Leontis NB (2012) Comprehensive survey and geometric classification of base triples in RNA structures, *Nucleic Acids Res* 40, 1407–1423. [PubMed: 22053086]
- (77). Serra MJ, Lyttle MH, Axenson TJ, Schadt CA, and Turner DH (1993) RNA hairpin loop stability depends on closing base pair, *Nucleic Acids Res* 21, 3845–3849. [PubMed: 7690127]
- (78). Sykes MT, and Williamson JR (2009) A complex assembly landscape for the 30S ribosomal subunit, *Annu Rev Biophys* 38, 197–215. [PubMed: 19416066]



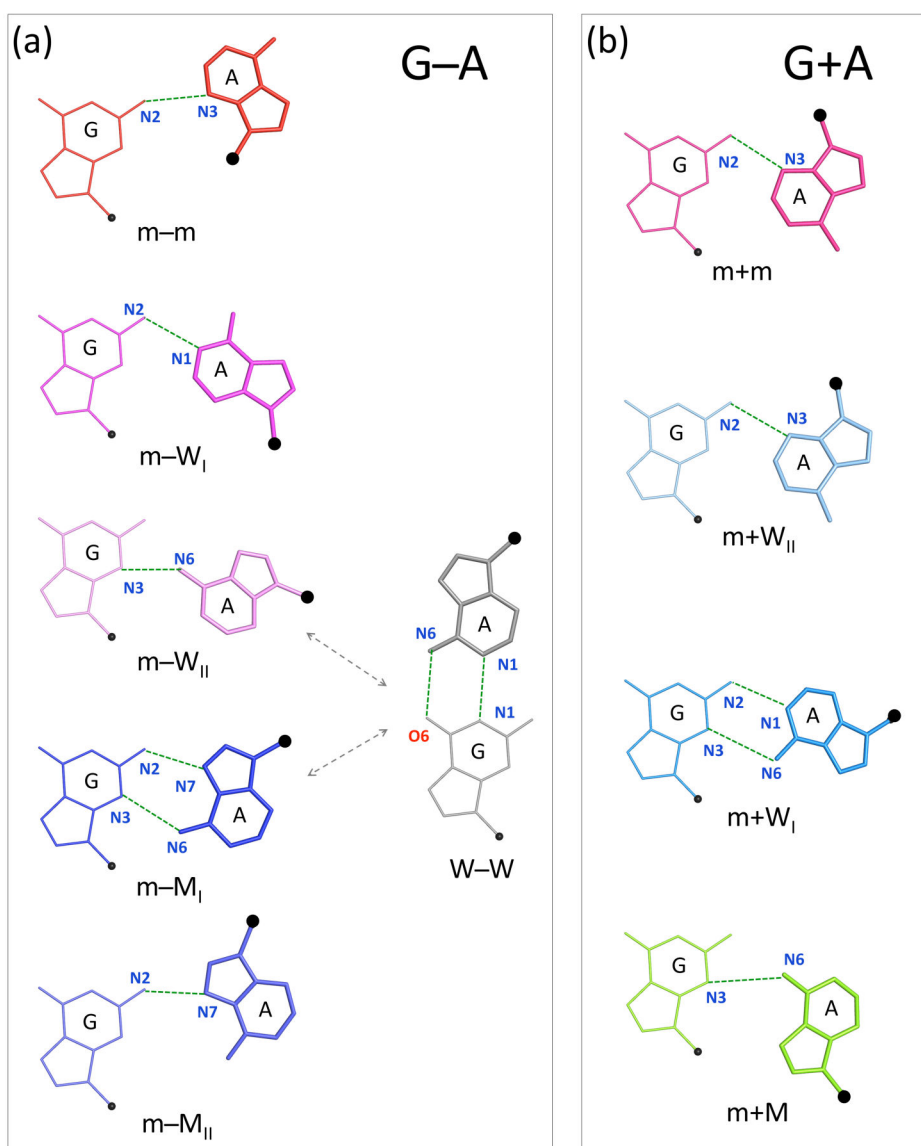
**Figure 1.** Comparison of the hydrogen-bonding interactions, chemical structures, and relative spatial arrangements of nucleotides in the six dominant modes of G-A pairing and in a canonical Watson-Crick G-C pair. Hydrogen bonds shown by thin dashed lines, with arrows directed toward base/backbone atoms capable of accepting protons. Structures generated with 3DNA<sup>24</sup> and rendered in PyMOL ([www.pymol.org](http://www.pymol.org)) using the average rigid-body parameters in Table 1 and a canonical A-RNA backbone.<sup>34</sup> Structures depicted in the standard reference frame of G.<sup>30</sup> Color-coding denotes the mode of base association: sheared m-M (dark blue); imino W-W (gray); m+m (pink); m-m (red); m+W (light blue); m-W (magenta); canonical (white), where the combinations of signs and letters denote the orientation (parallel +/ antiparallel -) and the approximate sites (minor m/major M/Watson-Crick W edges) of base association.<sup>27</sup> Interestingly, less than 10% of the + states reflect an anti-to-syn sugar-base rearrangement, and these examples all involve adenine.



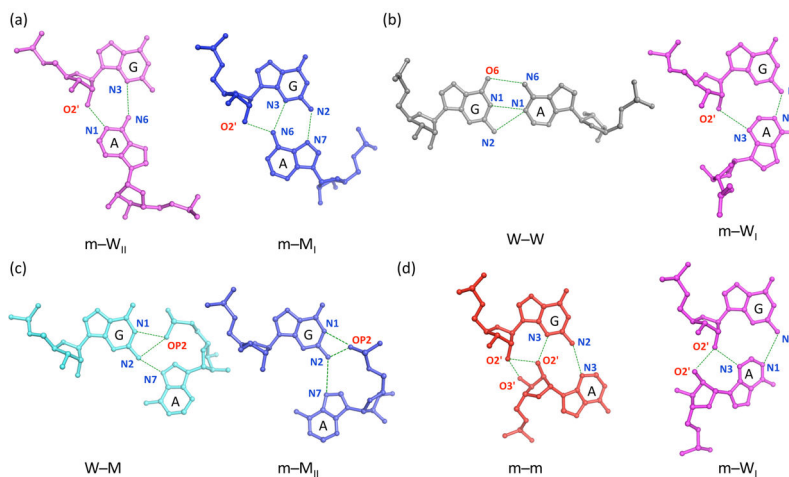
**Figure 2.** Molecular images of RNA secondary structural motifs incorporating each of the dominant modes of G-A base pairing. Motifs correspond to one of the common settings of the designated pairs: (a) sheared m-M pair closing the GNRA tetraloop at the end of the P4 helix of the *glmS* ribozyme bound to glucosamine 6-phosphate;<sup>35</sup> (b) imino W-W pair at the end of an asymmetric internal loop in the *Thermus thermophilus* 70S ribosome in complex with the hibernation factor pY;<sup>36</sup> (c) m+m pair linking the G in a double-helical stem and the A in an internal loop of the complex of *Thermus thermophilus* ribosomal protein L1 with a fragment of the L1 RNA from *Methanococcus vannielii*;<sup>37</sup> (d) m-m pair joining a stem and single-stranded fragment within a three-way junction of the *Thermoanaerobacter tengcongensis* ydaO riboswitch bound to cyclic di-AMP;<sup>38</sup> (e) m+W pair connecting the D and T hairpin loops of yeast initiator transfer RNA;<sup>39</sup> (f) m-W pair linking two hairpin loops of the 5S rRNA within the structure of the hibernating 100S ribosome dimer from pathogenic *Staphylococcus aureus*.<sup>40</sup> G-A pairs color-coded as in Figure 1 and shown at the nucleotide level within each structural element and in a separate local depiction to the right of each example. Loops depicted in gold and canonical base pairs at the ends of loops or within double-helical stems in white. Ribbons connect phosphorus atoms in successive nucleotides. See Table S3 for respective Protein Data Bank identifiers, chain names, and residue numbers of the depicted pairs and Figure S2 for simple secondary structural diagrams of the associated motifs.



**Figure 3.** Scatter plots of the rigid-body components — Shear, Stretch, Opening — that distinguish the modes of G-A association in RNA-containing structures. Smooth curves on the edges of the scatter plots are the normalized densities of individual parameters. Points with the magnitude of Opening in excess of  $180^\circ$  include requisite changes in the signs of Shear, Stretch, Buckle, and Propeller. Color-coding of dominant pairs matches that in Figures 1, 2. Secondary states with 16 or more structural examples are noted by related hues. Images depict the spread of values in the Shear-Opening (left) and Shear-Stretch (right) planes for antiparallel G-A (top) and parallel G+A (bottom) arrangements.

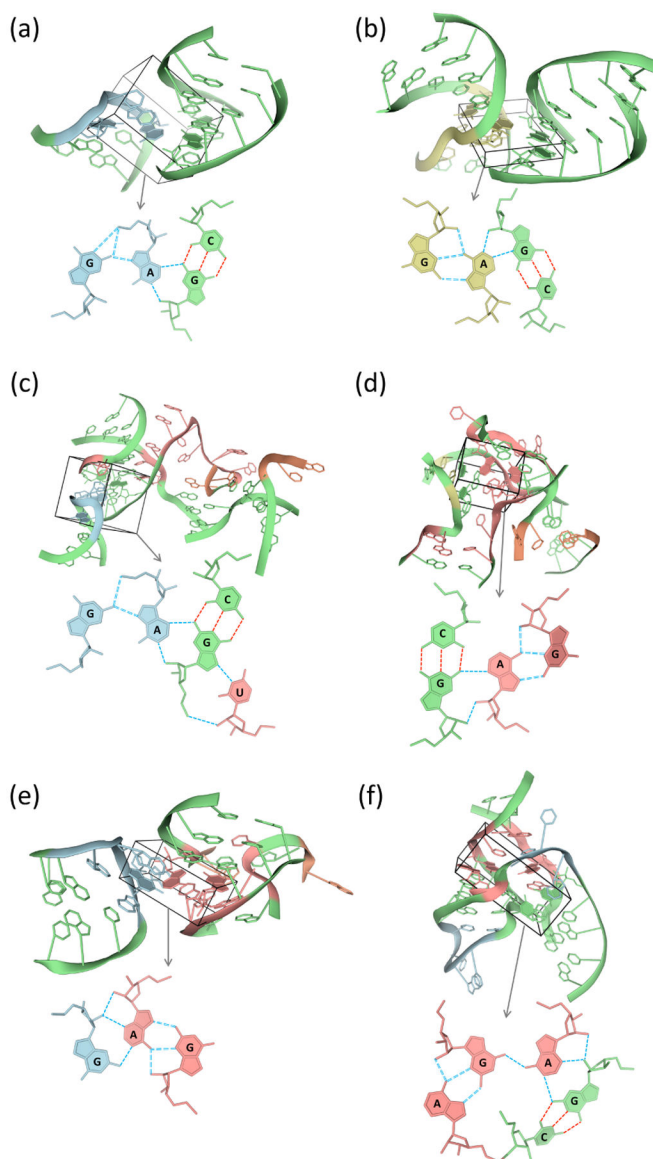


**Figure 4.** Succession of configurations illustrating the rigid-body motions that transform the associations of G and A between different pairing modes. Images of adenine oriented with respect to a common coordinate frame on guanine. Structures generated with 3DNA<sup>24</sup> using the average rigid-body parameters reported in Table 1. Base pairs color-coded by interaction mode (Figures 1-3), with the minor (II) substates of  $m\pm W$  and  $m-M$  pairs noted by lighter hues. Pathways connecting (a) antiparallel  $m-m$ ,  $m-W_I$ ,  $m-W_{II}$ ,  $W-W$ ,  $m-M_I$ ,  $m-M_{II}$  states; (b) parallel  $m+m$ ,  $m+W_{II}$ ,  $m+W_I$ ,  $m+M$  states. Note the counterclockwise rotation of ribose C1' atoms (darkened spheres) along the top-to-bottom transformation of antiparallel pairs and the clockwise rotation along the corresponding progression of parallel pairs. Hydrogen bonds between base atoms depicted by thin dashed lines.



**Figure 5.**

Molecular images illustrating hydrogen bonds (dashed lines) shared between different forms of G-A pairing: (a) the N3...N6 interaction common to m-W<sub>II</sub> and m-M<sub>I</sub> base-paired arrangements found respectively in a variant of the SAM-I riboswitch<sup>64</sup> and the complex of *Escherichia coli* ribosomal protein L25 with a 5S rRNA fragment;<sup>65</sup> (b) the respective N2...N1 associations in m-W<sub>I</sub> and W-W pairs in the *Leishmania donovani* large ribosomal subunit<sup>66</sup> and the *Saccharomyces cerevisiae* 80S ribosome;<sup>67</sup> (c) the N2...N7, N1...OP2, and N2...OP2 hydrogen bonds respectively stabilizing W-M and m-M<sub>II</sub> pairs in the central domain of the *Thermus thermophilus* 30S ribosomal subunit<sup>68</sup> and in the complex of the *Thermus thermophilus* 70S ribosome with hibernation factor pY;<sup>36</sup> (d) the respective paired association of 2'-hydroxyl groups, one from G and the other from A, adopted in m-m and m-W<sub>I</sub> pairs within the complex of tetracycline with the U1052G-mutated 70S *Escherichia coli* ribosome.<sup>69</sup>



**Figure 6.** Molecular images of multipler complexes of tetracycline with the U1052G-mutated 70S *Escherichia coli* ribosome.<sup>69</sup> Loops incorporating m-M sheared pairs are linked by different associations of G and A (highlighted within boxes) to other secondary structural units. A local representation of the linked bases is shown below each global depiction of associated secondary structures. Examples include: (a) a tetraplex with m+m pairing between a hairpin loop and a double-helical stem; (b) a tetraplex with m+m pairing between an internal loop and a double-helical stem; (c) a pentaplex with m+W pairing between a hairpin loop and a double-helical stem that is linked in turn to a junction; (d) a tetraplex with m-m link pairing between an internal loop and a 5-way junction; (e) a triplex with m-W pairing between a hairpin loop and a 3-way junction; (f) a pentaplex with .-M and m-W pairing within a 7-way junction. G-A pairs and secondary structural motifs color-coded as in Figure 2. See Table S3 for respective



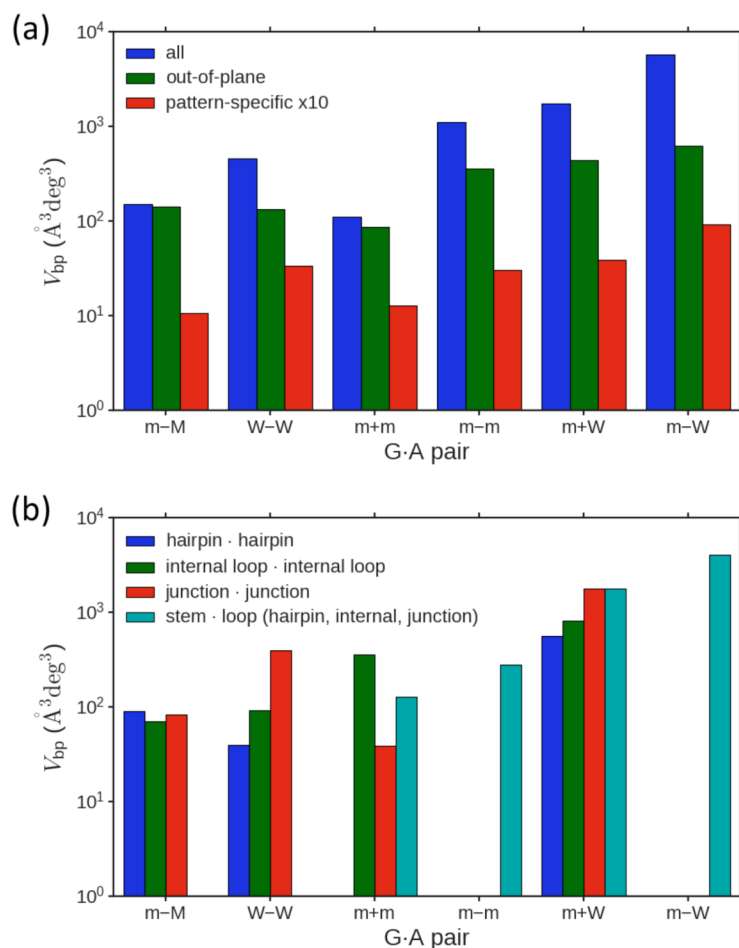
Protein Data Bank identifiers, chain names, and residue numbers of depicted G-A-linked multiplets. Color-coding of bases and hydrogen bonds matches that in the corresponding secondary structural diagrams in Figure S8.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 7.** Histograms of G·A deformability expressed in terms of the volume of conformation space occupied by the two bases in the major pairing schemes: (a) comparative contributions of pattern-specific (Shear, Stretch, Opening – red bars), out-of-plane (Stagger, Buckle, Propeller – green bars), and all six rigid-body parameters (blue bars – values listed in Table 1) to base-pair mobility; (b) relative deformability of G·A pairs in different tertiary structural settings, each with 44 or more structural examples. Base pairs grouped in terms of the observed setting of G followed by that of A. Note the logarithmic scale of volume and the tenfold ( $\times 10$ ) enhancement of pattern-specific values.

Table 1.

Rigid-body parameters of G·A pairings in RNA structures<sup>\*</sup>

G·A pair	# <sup>†</sup>	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (deg)	Propeller (deg)	Opening (deg)	$V_{bp}$ (Å <sup>3</sup> deg <sup>3</sup> )
m-m	141	-1.84 $\pm$ 0.86	7.19 $\pm$ 0.62	0.01 $\pm$ 1.54	-2.0 $\pm$ 25.3	-8.9 $\pm$ 22.1	150.0 $\pm$ 15.3	1116
m-W	153	3.74 $\pm$ 1.37	3.44 $\pm$ 2.17	0.33 $\pm$ 1.28	-8.0 $\pm$ 30.2	-16.0 $\pm$ 26.6	101.5 $\pm$ 31.1	5851
m-W <sub>I</sub>	125	3.14 $\pm$ 0.57	4.39 $\pm$ 0.79	0.17 $\pm$ 1.31	-3.4 $\pm$ 29.9	-16.6 $\pm$ 27.5	113.1 $\pm$ 20.6	1700
m-W <sub>II</sub>	28	6.40 $\pm$ 0.34	-0.83 $\pm$ 0.74	1.04 $\pm$ 0.84	-28.2 $\pm$ 23.1	-13.2 $\pm$ 22.0	49.6 $\pm$ 8.8	274
m-M sheared	1338	6.83 $\pm$ 0.38	-4.48 $\pm$ 0.58	0.37 $\pm$ 0.57	5.8 $\pm$ 16.0	-3.0 $\pm$ 14.9	-8.5 $\pm$ 14.1	153
m-M <sub>I</sub>	1216	6.75 $\pm$ 0.28	-4.39 $\pm$ 0.50	0.35 $\pm$ 0.54	5.4 $\pm$ 15.8	-2.2 $\pm$ 14.9	-5.0 $\pm$ 8.7	73
m-M <sub>II</sub>	122	7.63 $\pm$ 0.35	-5.33 $\pm$ 0.59	0.62 $\pm$ 0.77	10.1 $\pm$ 17.5	-10.4 $\pm$ 12.0	-43.9 $\pm$ 6.3	132
W-M	59	7.10 $\pm$ 1.52	-2.84 $\pm$ 2.36	-0.08 $\pm$ 1.13	-1.5 $\pm$ 25.4	-4.0 $\pm$ 23.7	-69.6 $\pm$ 24.5	14440
M-M	25	0.37 $\pm$ 2.52	4.49 $\pm$ 0.86	0.11 $\pm$ 1.66	7.0 $\pm$ 25.0	5.5 $\pm$ 19.5	-154.1 $\pm$ 25.5	1870
M-W	16	-4.24 $\pm$ 0.74	0.89 $\pm$ 1.23	-0.2 $\pm$ 0.98	6.5 $\pm$ 20.7	0.3 $\pm$ 15.3	-93.4 $\pm$ 28.6	386
W-m	38	-3.39 $\pm$ 1.05	4.27 $\pm$ 2.43	0.35 $\pm$ 1.69	7.0 $\pm$ 31.3	-0.2 $\pm$ 36.9	106.2 $\pm$ 33.7	6687
W-W imino	344	0.07 $\pm$ 0.72	1.55 $\pm$ 0.42	-0.33 $\pm$ 0.55	6.7 $\pm$ 15.2	-9.6 $\pm$ 16.8	-16.4 $\pm$ 14.0	465
m+M	53	6.42 $\pm$ 1.09	0.29 $\pm$ 2.79	0.38 $\pm$ 1.39	-15.5 $\pm$ 24.3	11.4 $\pm$ 23.9	-3.3 $\pm$ 29.2	15556
.+W	27	2.35 $\pm$ 0.48	-4.94 $\pm$ 0.73	-0.34 $\pm$ 1.47	5.1 $\pm$ 34.3	8.2 $\pm$ 18.8	-125.2 $\pm$ 12.0	206
m+W	388	1.59 $\pm$ 2.54	-5.11 $\pm$ 1.58	-0.04 $\pm$ 0.90	-4.4 $\pm$ 22.7	11.8 $\pm$ 21.5	-97.6 $\pm$ 35.0	1762
m+W <sub>I</sub>	267	3.23 $\pm$ 0.72	-4.12 $\pm$ 0.64	-0.12 $\pm$ 0.82	0.1 $\pm$ 22.1	14.0 $\pm$ 24.5	-77.2 $\pm$ 19.9	553
m+W <sub>II</sub>	121	-2.02 $\pm$ 0.64	-7.30 $\pm$ 0.35	0.15 $\pm$ 1.04	-13.8 $\pm$ 21.2	6.7 $\pm$ 11.5	-142.6 $\pm$ 10.0	54
m+.	55	-2.31 $\pm$ 0.82	-6.95 $\pm$ 0.51	1.21 $\pm$ 1.36	-35.7 $\pm$ 31.7	1.6 $\pm$ 9.9	-146.6 $\pm$ 13.1	20
m+m	395	-2.70 $\pm$ 0.35	-7.38 $\pm$ 0.33	-0.08 $\pm$ 1.42	-11.1 $\pm$ 29.5	8.7 $\pm$ 18.3	-153.8 $\pm$ 9.8	111
W+M	69	0.29 $\pm$ 1.68	4.47 $\pm$ 0.90	0.22 $\pm$ 1.05	-7.2 $\pm$ 19.7	8.3 $\pm$ 25.7	-104.8 $\pm$ 39.2	14442
.+M	24	6.25 $\pm$ 0.38	0.61 $\pm$ 0.67	-0.01 $\pm$ 1.10	-8.2 $\pm$ 15.5	1.7 $\pm$ 21.7	-61.2 $\pm$ 11.8	58
M+M	26	0.19 $\pm$ 6.06	1.08 $\pm$ 5.39	0.17 $\pm$ 0.94	-12.2 $\pm$ 26.4	5.4 $\pm$ 18.0	-30.2 $\pm$ 142.0	56463
M+W	29	0.58 $\pm$ 1.70	-4.8 $\pm$ 1.20	-0.29 $\pm$ 0.92	5.5 $\pm$ 22.1	4.3 $\pm$ 16.4	102.2 $\pm$ 20.2	1960

<sup>\*</sup> Structures taken from the dataset of non-redundant, high-resolution RNA crystal structures curated by Leontis and Zirbel.<sup>59</sup> Similar results are found with other datasets, e.g., the RNA09 coordinate files curated by the Richardson group at URL:<http://kinemage.biochem.duke.edu/databases/madb.php>.

<sup>†</sup> Counts do not distinguish *cis* (*c*) and *trans* (*t*) arrangements of C1'-N9...N9-C1' angles reported by the software (Table S2) and used by others<sup>9,41</sup> to distinguish pairing modes. This 'chemical' descriptor does not always match the  $\pm$  directional criteria used here, e.g., the m-W states include a small number of *t* arrangements and the m+W states some *c* arrangements.

Table 2.

Secondary structural composition of the six dominant forms of G-A pairing<sup>\*†</sup>

1292 1338 <sup>‡</sup> m-M	blg <sub>A</sub>	hpn <sub>A</sub>	int <sub>A</sub>	jct <sub>A</sub>	stm <sub>A</sub>	323 344 <sup>‡</sup> W-W	blg <sub>A</sub>	hpn <sub>A</sub>	int <sub>A</sub>	jct <sub>A</sub>	stm <sub>A</sub>
blg <sub>G</sub>	12 <sub>3</sub>		3	1		blg <sub>G</sub>	2				1
hpn <sub>G</sub>	5	351 <sub>6</sub>		8		hpn <sub>G</sub>		20 <sub>2</sub>		3	
int <sub>G</sub>		2	627 <sub>5</sub>	1		int <sub>G</sub>			169		
jct <sub>G</sub>				261		jct <sub>G</sub>	1	1		118	2
stm <sub>G</sub>	12	3	4	2		stm <sub>G</sub>	1	1		1	3 <sub>1</sub>
496 395 <sup>‡</sup> m+m	blg <sub>A</sub>	hpn <sub>A</sub>	int <sub>A</sub>	jct <sub>A</sub>	stm <sub>A</sub>	197 141 <sup>‡</sup> m-m	blg <sub>A</sub>	hpn <sub>A</sub>	int <sub>A</sub>	jct <sub>A</sub>	stm <sub>A</sub>
blg <sub>G</sub>	2	2	3	12		blg <sub>G</sub>		5	4	5	
hpn <sub>G</sub>		16 <sub>14</sub>	1	12		hpn <sub>G</sub>	3	9 <sub>5</sub>			2
int <sub>G</sub>		16	44 <sub>17</sub>	8		int <sub>G</sub>	2	15	10	9	1
jct <sub>G</sub>		17	7	68 <sub>19</sub>	1	jct <sub>G</sub>		11		22 <sub>5</sub>	
stm <sub>G</sub>	15	77	97	98		stm <sub>G</sub>	5	31	28	35	
500 388 <sup>‡</sup> m+W	blg <sub>A</sub>	hpn <sub>A</sub>	int <sub>A</sub>	jct <sub>A</sub>	stm <sub>A</sub>	187 153 <sup>‡</sup> m-W	blg <sub>A</sub>	hpn <sub>A</sub>	int <sub>A</sub>	jct <sub>A</sub>	stm <sub>A</sub>
blg <sub>G</sub>	6	3	12	8		blg <sub>G</sub>	5 <sub>2</sub>	6	1	1	
hpn <sub>G</sub>	1	81 <sub>58</sub>	4	17		hpn <sub>G</sub>		13 <sub>7</sub>	1	7	
int <sub>G</sub>	3	9	55 <sub>13</sub>	26		int <sub>G</sub>	1	7	17 <sub>6</sub>		
jct <sub>G</sub>	1	13	7	71 <sub>1</sub>		jct <sub>G</sub>		7	2	30	1
stm <sub>G</sub>	10	78	36	59		stm <sub>G</sub>	12	35	18	23	

\* Number of observations of the guanine and adenine in identified pairs in five common secondary structural forms: blg – bulge loop; hpn – hairpin loop; int – internal loop; jct – junction loop; stm – canonical double-helical stem. The subscript following each acronym refers to the base in the designated secondary structure, e.g., stm<sub>G</sub> refers to a guanine in a canonical double-helical stem. The dataset includes a total of 149 G–A and 112 G+A pairs outside the 25 listed categories, i.e., 90% of the 1108 G+A and 93% of the G–A pairs occur within the listed structural categories.

<sup>†</sup>The numerical entries in the set of grids correspond to the number of examples with a given combination of secondary structural states, e.g., 351 examples of sheared m–M pairs with both bases (hpn<sub>G</sub>-hpn<sub>A</sub>) found in a hairpin loop. The subscripted value for this and other diagonal entries denotes the number of examples where the paired bases occur in different ( ) hairpins, here the “6” referring to six cases of m–M pairs with G in one hairpin and A in another. Blanks and missing subscripts refer to null values.

<sup>‡</sup>Number of G-A pairs of the specified type in the survey of high-resolution structures followed by the total number of pairs of the same type in one of the 25 categories. See text for discussion of numerical differences.