

RNASTAR: An RNA STructural Alignment Repository that provides insight into the evolution of natural and artificial RNAs

JEREMY WIDMANN,¹ JESSE STOMBAUGH,¹ DANIEL MCDONALD,² JANA CHOCHOLOUSOVA,³ PAUL GARDNER,⁴ MATTHEW K. IYER,⁵ ZONGZHI LIU,⁶ CATHERINE A. LOZUPONE,¹ JOHN QUINN,⁷ SANDRA SMIT,⁸ SHANDY WIKMAN,⁹ JESSE R.R. ZANEVELD,¹⁰ and ROB KNIGHT^{1,11,12}

¹Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, Colorado 80309, USA

²Biofrontiers Institute, University of Colorado at Boulder, Boulder, Colorado 80309, USA

³Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague 6, Czech Republic

⁴School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand

⁵Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA

⁶Department of Pathology Informatics, School of Medicine, Yale University, New Haven, Connecticut 06510, USA

⁷Thermo Fisher Scientific, Lafayette, Colorado 80026, USA

⁸Laboratory of Bioinformatics, Wageningen University, 6700 AN Wageningen, The Netherlands

⁹Hägersten 12930, Sweden

¹⁰Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA

¹¹Howard Hughes Medical Institute, Boulder, Colorado 80309, USA

ABSTRACT

Automated RNA alignment algorithms often fail to recapture the essential conserved sites that are critical for function. To assist in the refinement of these algorithms, we manually curated a set of 148 alignments with a total of 9600 unique sequences, in which each alignment was backed by at least one crystal or NMR structure. These alignments included both naturally and artificially selected molecules. We used principles of isostericity to improve the alignments from an average of 83%–94% isosteric base pairs. We expect that this alignment collection will assist in a wide range of benchmarking efforts and provide new insight into evolutionary principles governing change in RNA structural motifs. The improved alignments have been contributed to the Rfam database.

Keywords: SPuNC; alignments; isostericity

INTRODUCTION

Multiple sequence alignments are critical for understanding evolutionary principles including phylogenetic relationships among sequences (Thompson et al. 2005; Brown et al. 2009) and functional principles such as critical active sites, or even elements of three-dimensional (3D) structures, revealed through patterns of conservation (Cruz and Westhof 2011). The alignment can even have more of an influence on the inferred phylogeny than does the phylogeny inference method (Loytynoja and Goldman 2008; Wong et al. 2008). In studies of RNA, large alignments such as those in

the CRW (Cannone et al. 2002) and in Rfam (Griffiths-Jones et al. 2003) have been useful for identifying new family members and inferring the secondary structures they contain. Tools such as INFERNAL (Nawrocki et al. 2009) have greatly assisted in this endeavor, especially as the databases continue to grow.

Improved alignments of natural and artificial RNAs will also increase our ability to test hypotheses about RNA evolution and architecture. Clear patterns of nucleotide composition have been noted in both natural and artificial RNA families (Wang and Hickey 2002; Gan et al. 2003; Gevertz et al. 2005; Knight et al. 2005; Smit et al. 2006, 2007, 2009), and one fascinating question is thus whether RNAs shaped by natural selection share similar features with those artificially selected in the laboratory. Comparing natural and artificial RNAs is important because such comparisons tell us whether we are seeing contingent features of organisms

¹²Corresponding author

E-mail rob.knight@colorado.edu

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.032052.111>.

as they have evolved on Earth, or universal principles of RNA architecture (Yarus and Welch 2000). Artificial RNAs also provide ideal test cases for homology comparison methods because they provide a test set of sequences that are known to be nonhomologous with each other, or with any natural RNA. A key question is whether tertiary motifs (Batey *et al.* 1999; Leontis and Westhof 2002) reliably recur among different classes of RNAs and can be used as universal building blocks for synthetic biology of functional RNAs (Jaeger and Chworos 2006; Jossinet *et al.* 2010).

There has been substantial progress toward automated alignment methods, although they are still relatively in-

accurate, especially for distantly related RNAs (Gardner *et al.* 2005). Most alignment programs do not incorporate features such as isostericity (Leontis and Westhof 1998) and compositional preference (Smit *et al.* 2009) that are known to be important in RNA evolution. BoulderALE (Stombaugh *et al.* 2011) incorporates both of these features, allowing construction of manually curated, high-quality alignments that can be used to improve algorithms for automated methods.

When evaluating an alignment of RNA molecules, nucleotides are aligned based on conservation, which can be at the level of the nucleotide or at the level of structure

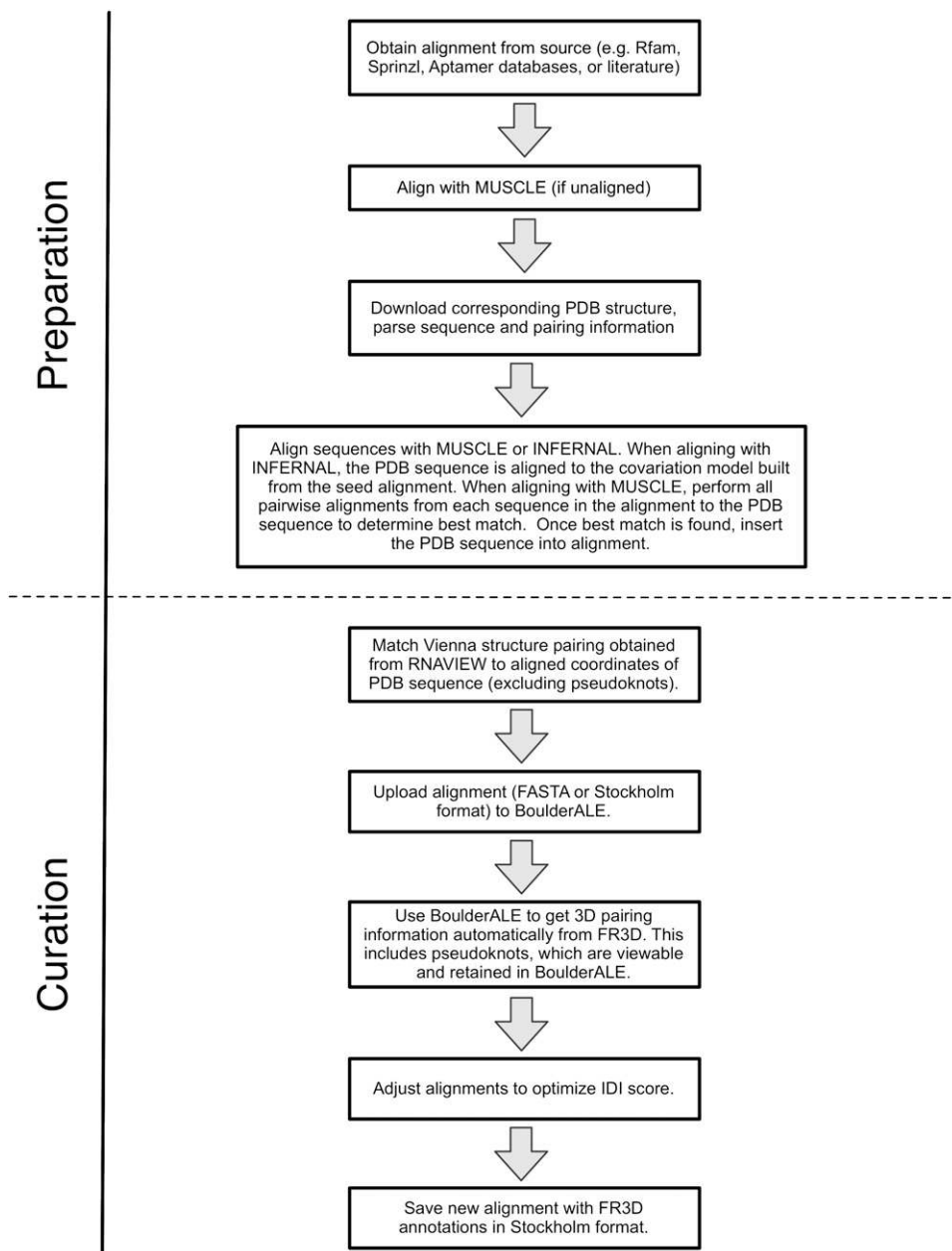


FIGURE 1. Overview of workflow for alignment.

(secondary or 3D). The evaluation of an alignment at the level of structure consists of understanding the nucleotide interactions and the effects of nucleotide mutations on those interactions. For instance, if NT1 and NT2 form a specific base pairing, a mutation of NT1 could affect the base-pairing interaction to NT2. Leontis et al. (2002) classified all RNA base-pairing interactions into 12 geometric families. Then, using qualitative methods, they identified the base pairs within a family that could be easily substituted for one another without disrupting the structure, otherwise known as isostericity (Leontis et al. 2002). In a more recent publication, Stombaugh et al. (2009) extended this notion by developing the IsoDiscrepancy Index (IDI), a quantitative method for classifying each base pair into an isosteric group. When determining the IDI between two base pairs, the method determines three attributes: (1) if the C1'–C1' distances between the interacting nucleotides are nearly identical; (2) if the corresponding nucleotides form hydrogen bonds between equivalent atoms; and (3) if the rotational matrices between corresponding nucleotides are nearly identical (Stombaugh et al. 2009). Using these three attributes, the IDI between two base pairs can be calculated, where lower IDIs (<2) refer to isosteric base pairs.

Therefore, we constructed a large collection of crystal and NMR structures that were related to multiple sequence alignments using the procedure shown in Figure 1. This collection of manually curated alignments backed by experimentally

determined atomic-resolution structures provides us both with an ideal training set for further algorithm development and with seeds for more sensitive database searches.

RESULTS AND DISCUSSION

We chose the 3D structures by manually examining all NMR and atomic-resolution crystal structures in the PDB containing RNA with a resolution <4.1 Å up to October 2011 (except for the 5S rRNA PDB 1YL3, which was 5.5 Å). Base-pair information was derived from each structure using FR3D (Sarver et al. 2008). Redundant sequences, defined as structures with identical base composition and base pairing, were dropped from the data set, typically by choosing the most recent and/or highest-resolution structure. Structures for which no homologous sequences could be found in Rfam (Griffiths-Jones et al. 2003), the tRNA database (Juhling et al. 2009), the Aptamer Database (Lee et al. 2004), or as readable figures in the literature (Guo et al. 1993; Burgstaller and Famulok 1994; Jenison et al. 1994; Pan et al. 1994; Wallis et al. 1995; Wang and Rando 1995; Famulok and Huttenhofer 1996; Jiang et al. 1996, 1997, 1999; Wang et al. 1996; Yang et al. 1996; Zimmermann et al. 1997; Wilson et al. 1998; Kim et al. 1999, 2000; Seelig and Jaschke 1999; Giedroc et al. 2000; Collins 2002; Lafontaine et al. 2002; Wang and Hickey 2002; Licis and van Duin 2006) were excluded from the analysis.

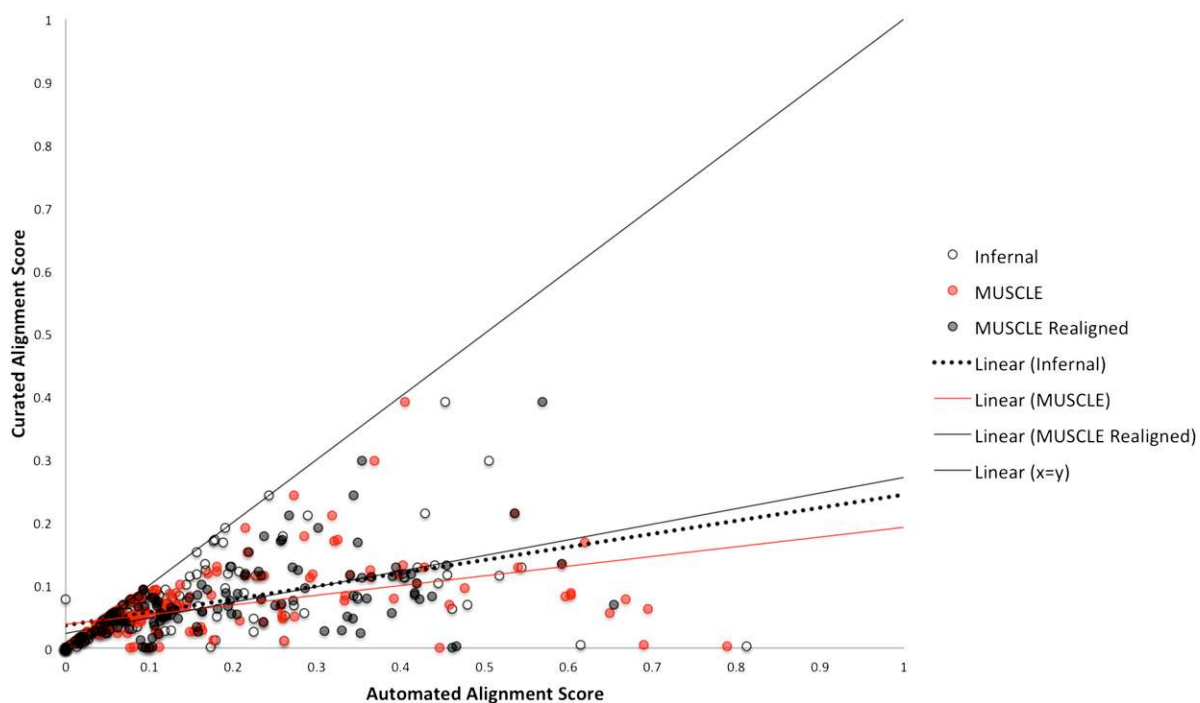


FIGURE 2. Comparison of original and improved alignments. The manually curated alignment scores (*y*-axis) are compared with each of three kinds of automated alignment (*x*-axis): inserting the PDB sequence with INFERNAL (white), inserting the PDB sequence with MUSCLE (red), and building the alignment de novo with MUSCLE (black). Scores are based on fraction of non-isosteric base pairs. Data are supplied in Supplemental Table S2.

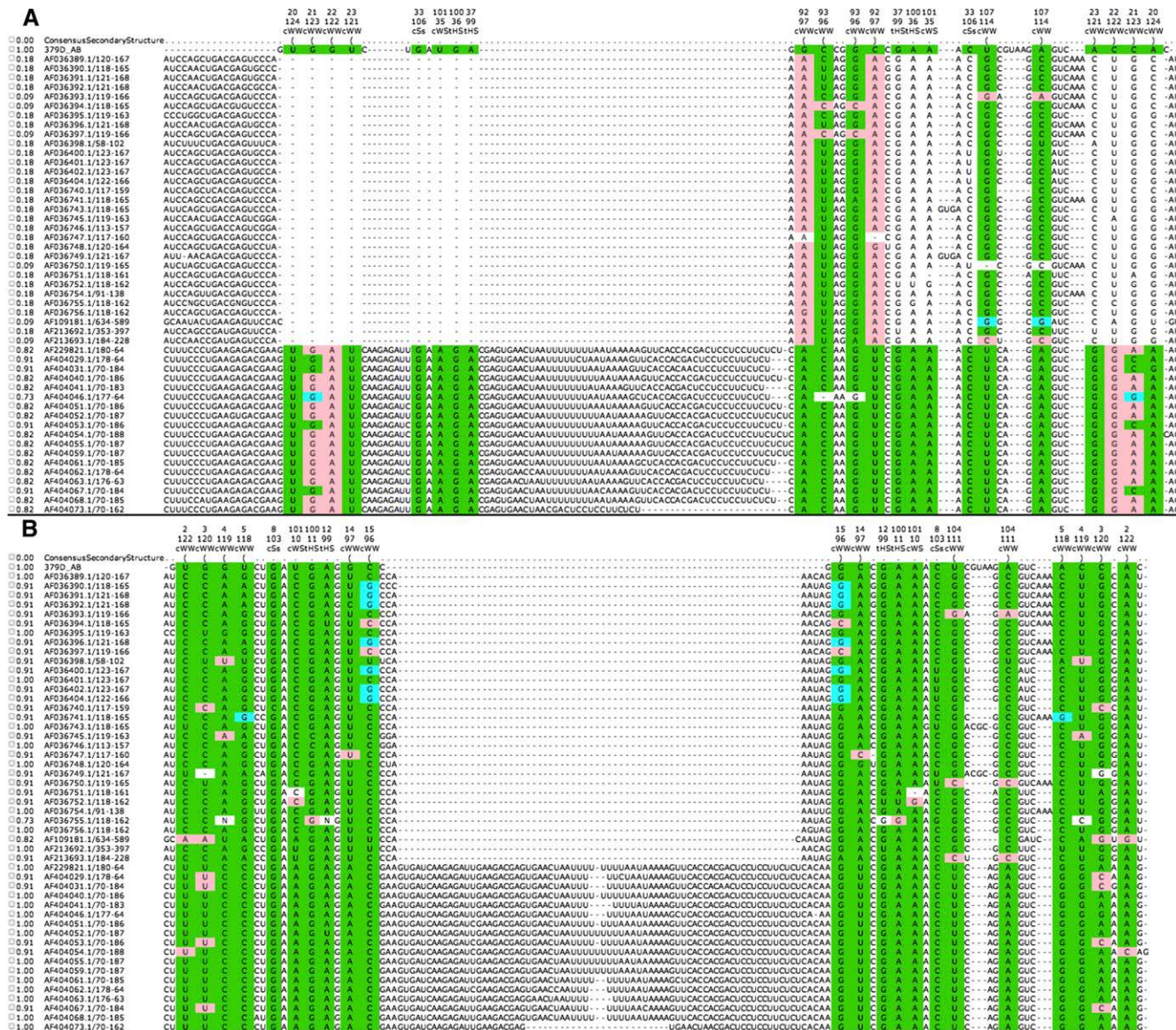


FIGURE 3. BoulderALE screenshots showing Hammerhead ribozyme alignment (Rfam: RF00163), where MUSCLE was used to align the corresponding crystal structure (PDB: 379D) sequence (A) versus the manually curated alignment (B). The colors from BoulderALE highlight isosteric (green), non-isosteric (pink), and not allowed (blue) covariations with respect to the 3D structure. For this alignment, there is an element expansion, and as you can see in A, MUSCLE aligned the X-ray crystal structure to a portion of the insertion. For the manual alignment (B), we shifted the X-ray crystal structure to align with the appropriate corresponding region.

The manually curated alignments were substantially improved over automated alignments produced using MUSCLE (Edgar 2004) or INFERNAL (Nawrocki et al. 2009), with essentially all showing an improvement in the fraction of non-isosteric base pairs (Fig. 2). An example of the Hammerhead ribozyme MUSCLE alignment versus the manually curated alignment is shown in Figure 3. For these alignments, the crystal structure sequence (PDB: 379D) was aligned to a homologous set of sequences from Rfam (RF00163): Note the substantially lower number of gaps and increased number of aligned positions in the manually curated alignment, which improve the IDI

scores for a given alignment. On average, alignments in which the manual curation affected a greater number of positions also improved more substantially (Fig. 4), as measured by IDI score (Stombaugh et al. 2009): The curated alignments had an average IDI score of 0.94, compared with an average score of 0.87 for INFERNAL alignments and 0.83 for the two MUSCLE alignment methods (see Materials and Methods). Relative to the automatically generated alignments, the IDI scores of the curated alignment improved 32% of the sequences relative to INFERNAL, 39% relative to MUSCLE, and 49% relative to the MUSCLE realigned method. The IDI scores of

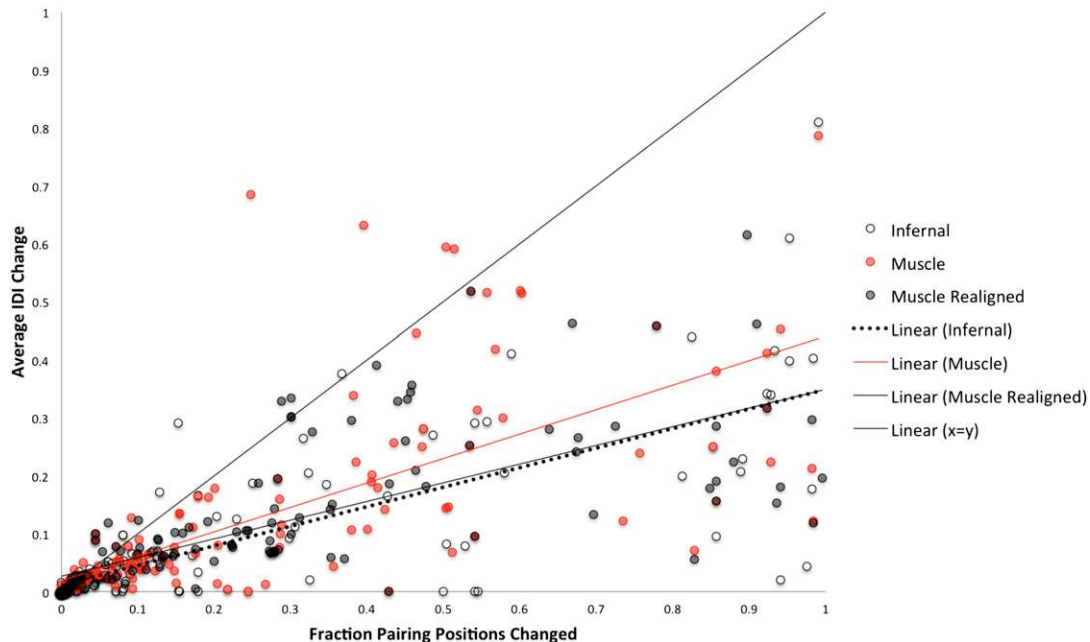


FIGURE 4. Alignments that underwent greater change during the manual curation process also improve more as shown by the change in IDI between the curated alignment and the three kinds of automated alignments: INFERNAL (white), MUSCLE (red), and de novo MUSCLE (black). The y -axis shows average IDI score change, and the x -axis shows the average fraction of changed base-pairing positions within an alignment. Data are supplied in Supplemental Table S3.

the curated alignment decreased in 1.6% of the sequences relative to INFERNAL, 1.0% relative to MUSCLE, and 1.6% relative to the MUSCLE realigned method. Finally, the SPuNC scores, which calculate how well the RNA secondary structure predicted by an alignment matches the known compositional preferences for that secondary structure type (Smit et al. 2009), were substantially improved in the curated alignments over the automated alignments (Table 1). Consequently, manual curation substantially improved the overall alignment quality, as shown by two distinct measures.

Overall, 146 of the 148 alignments showed equal or improved IDI scores. The two exceptions were special cases. The valine tRNA alignment (RST00143.sto), which applies the base-pairing information from PDB ID: 1J2B, only aligns optimally when base-pairing information from all available crystal structures is taken into consideration, possibly suggesting structural variation. For this alignment there were three corresponding X-ray crystal structures, thus we inserted all three sequences from those structures and applied the FR3D base-pairing information for each structure independently to determine the quality of the manually curated alignment. The VS ribozyme alignment (RST00145.sto) using the base-pairing information from PDB ID: 1HWQ has a different issue: The automated alignment has the first base pair at the start of extremely long sequences, then inserts ~ 100 bases until the next base pair on both sides, thereby getting a perfect IDI score. In the curated alignment, the closing base pair is next to all of the other base

pairs, producing a non-isosteric substitution. However, this substitution is more likely as the true alignment and is sterically acceptable at the end of the helix.

As an example of the utility of a structure-backed alignment database incorporating both natural and artificial RNAs and using consistent methodology, we compared natural RNA families with artificial RNA families in terms of their rates of change of GC content across specific structural categories. On average, the total GC content did not differ substantially between natural (Fig. 5A) and artificial (Fig. 5B) RNAs ($t = 1.29$, $p = 0.198$). When we look at the responses to altered GC in the multiple sequence alignment within each category, we see a remarkable degree of universality in the response. Figure 6 shows the scatterplots of total GC content of natural sequences (Fig. 6A) and artificial sequences (Fig. 6B) against GC content of each structural category (stems, loops, bulges). For each structural category, the slopes of regression were determined and are represented as histograms in Figure 7 separated by structural category—stem (Fig. 7A), loop (Fig. 7B), and bulge (Fig. 7C). The t -test comparing natural versus artificial distributions of slopes shows that the difference in responses in stems is significant between natural and artificial RNA families ($t = 2.63$, $p < 0.01$), but the difference in responses in bulges and loops is not significant ($p > 0.6$ in both cases). The apparent difference in stem responses is likely driven by the greater range of mutation pressures that genomes experience relative to artificial RNA pools. A more sophisticated ANCOVA anal-

TABLE 1. Averages from SPuNC output for manually curated alignments, INFERNAL automated alignments, MUSCLE automated alignments, and MUSCLE realigned alignments

Alignment	Unweighted extreme	Unweighted trained	Weighted extreme	Weighted trained
Curated	1.361414599	0.911091926	1.659870363	1.099245743
Infarnal	1.376352659	0.958821011	1.680768259	1.141649617
Muscle	1.414364163	0.988129238	1.737054868	1.159960456
Muscle realigned	1.597322008	1.07525067	1.962070865	1.308403591

The manually curated alignments as a whole comply far better with overall compositional preferences in different RNA structure regions as reported by Smit et al. (2008).

ysis, which separates out the effects of covariation in each category, suggests that interaction effects are at best weak (uncorrected interaction P -values are 0.02 for stems, 0.51 for loops, and 0.15 for bulges; none are statistically significant when corrected for multiple comparisons). Consequently, the results are consistent with the idea that universal patterns of compositional change under GC content variation hold for both natural and artificial RNA families.

CONCLUSIONS

Manual alignments, especially those backed by crystal structures, still substantially outperform automated techniques by a range of metrics, suggesting that substantial improvement in algorithms is still possible. Since scoring schemes such as IDI and SPuNC can detect the improvement in manually curated alignments, incorporation of these metrics of isostericity and sequence composition into automated alignment software will likely lead to improvements in automated techniques.

IDI scores could provide an important filter for motif searching in large sequence databases, such as those now generated by sequencing SELEX pools or by metagenomics. More broadly, improved manually curated alignments will assist with benchmarking different RNA alignment and structure prediction algorithms and provide a training set for ongoing development of these algorithms as well as providing us insight into how RNA molecules evolve.

MATERIALS AND METHODS

Our choice of alignments was based on a requirement that there was a corresponding crystal structure or NMR structure in the Protein Data Bank (PDB). A full list of alignments and their corresponding PDBs is found in Supplemental Table S1. We did not accept poor-resolution (>4.1 Å with the exception of 5S rRNA PDB 1YL3, which was 5.5 Å) or cryo-EM structures for our reference structures. Redundant structures and those superseded by newer structures were not included in the curated alignments.

Base-pair lists corresponding to each X-ray crystal structure were downloaded from the “Find RNA 3D” (FR3D) website (Sarver et al. 2008) (<http://rna.bgsu.edu/FR3D/AnalyzedStructures/>). FR3D

classifies all canonical and noncanonical base-pair interactions for a given RNA 3D structure using the Leontis and Westhof (2001) base-pair nomenclature, which has been adopted by the RNA Ontology Consortium as the standard annotation scheme for RNA base-pair interactions (Hoehndorf et al. 2011).

Structures that were identical or superseded by newer structures were eliminated from the analysis. Redundant sequences were eliminated from the analysis. Sequences introducing gaps in $>95\%$ of the positions in the alignment were also eliminated from the analysis. After these filter criteria, we ended

up with 9600 nonredundant sequences corresponding to 148 unique structures.

Sequences were aligned using INFERNAL 1.0.2 and MUSCLE 3.7. For the INFERNAL alignments containing a secondary structure, we aligned the PDB sequence to the alignment with

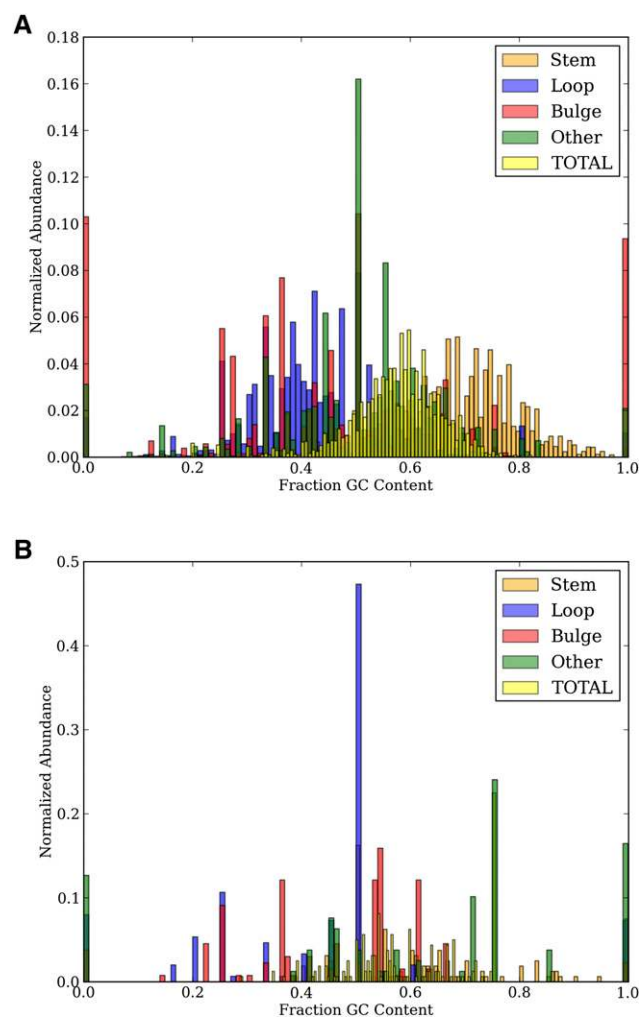


FIGURE 5. (A) Histogram of average GC content split up by structural category for naturally occurring sequences. (B) Histogram of average GC content split by structural category for artificially occurring sequences.

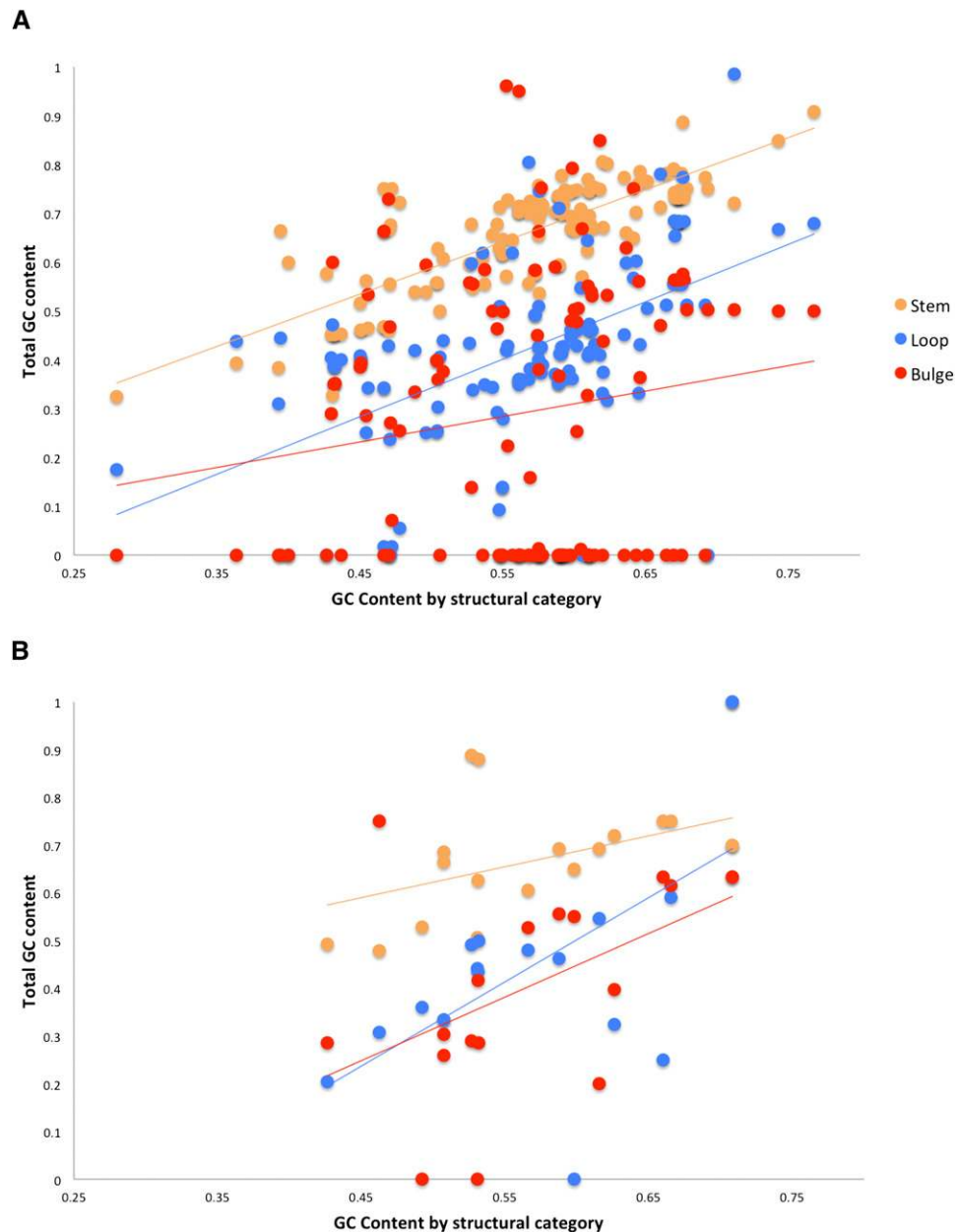


FIGURE 6. Scatterplot of total GC content (y -axis) of natural (A) and artificial sequences (B) against GC content of each structural category (stem, loop, bulge) of the same sequences on the x -axis.

default parameters. For the cases in which no secondary structure was present, we built a CM with `cmbuild` (using the `-ignorant` flag) and used an unpaired placeholder for the consensus secondary structure, then aligned the PDB sequence to this alignment with `cmalign`. MUSCLE alignments were produced using two methods: (1) find the best pairwise match in an existing alignment to the PDB sequence, then insert the PDB sequence into the full alignment with MUSCLE and align it to its best match; (2) using an existing alignment, remove all gaps in all sequences, then use MUSCLE to realign the entire alignment and insert the PDB sequence into this alignment using MUSCLE in the same way as the first method.

Curation of alignments was done using BoulderALE, where we were able to apply Watson-Crick and non-Watson-Crick base-

pair information onto the alignment. Using the base-pairing information, we were able to manually curate the alignment to optimize isostericity. For some cases, manual inspection of the X-ray structure was necessary to determine the reliability of specific base-pair interactions and for insight into the appropriate location for insertion/deletions.

We used several scoring schemes to assess the quality of the curated versus the automated alignments. The simplest way to score the alignments was to calculate the total entropy of the alignment. This is done by using the frequency of all nucleotides in each position (column) of the alignment to calculate the Shannon entropy for that position. The entropy values for each position can vary from 0 (absolutely conserved) to 2 (completely degener-

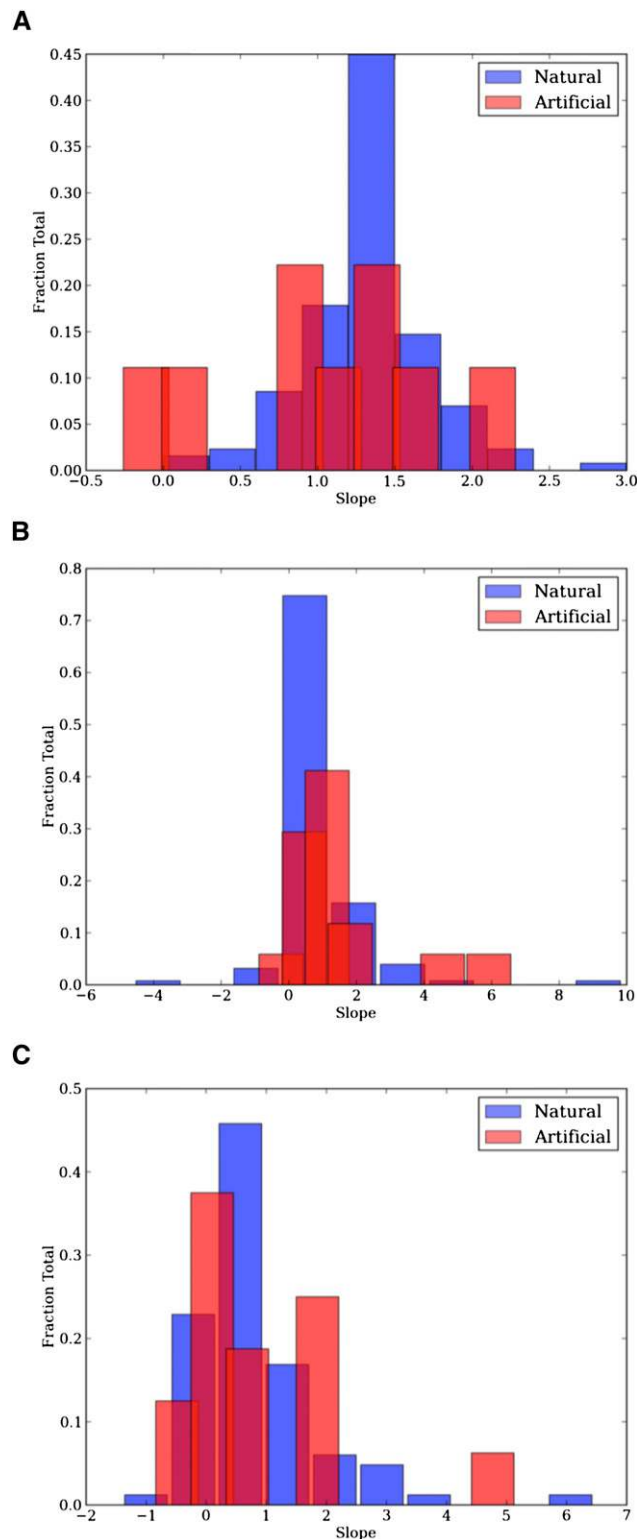


FIGURE 7. Histograms showing slopes of regression lines of GC content for each structural category (stem [A], loop [B], bulge [C]) versus total GC content. The responses to changes in GC content are extremely similar between natural and artificial RNA families.

ate). These values were then summed for the entire alignment. However, we found that this simple method lacked statistical power to discriminate even among visually very good and very bad alignments (data not shown).

We also scored the alignments based on isostericity of base pairs that are known to form in the crystal/NMR structures. Using the 3D base interaction annotations from FR3D (Sarver et al. 2008), we were able to assess the quality of the pairing regions of the alignments. Using the PDB sequence as a reference, for each sequence, each base pair was assigned a value of 1 for isosteric and near-isosteric or a value of 0 for non-isosteric or not allowed. The sequence was then given a score that represented the fraction isosteric/near-isosteric base pairs. The alignment score is the average of each sequence's score, ranging from 0.0 (completely non-isosteric/not allowed) to 1.0 (perfectly isosteric/near-isosteric).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Eric Nawrocki for his correspondence and Micah Hamady for manually entering alignments from the literature. We thank the National Institutes of Health (Grant HG4872 to R.K.); NASA Astrobiology (Grant NNX08AP60G to R.K.); the Howard Hughes Medical Institute, the RNA Ontology Consortium for travel (NSF 0443508), and the Ministry of Education, Youth and Sport of the Czech Republic (KONTAKT Grant ME09019, to J.C.).

Received December 30, 2011; accepted May 2, 2012.

REFERENCES

- Batey RT, Rambo RP, Doudna JA. 1999. Tertiary motifs in RNA structure and folding. *Angew Chem Int Ed Engl* **38**: 2326–2343.
- Brown JW, Birmingham A, Griffiths PE, Jossinet F, Kachouri-Lafond R, Knight R, Lang BF, Leontis N, Steger G, Stombaugh J, et al. 2009. The RNA structure alignment ontology. *RNA* **15**: 1623–1631.
- Burgstaller P, Famulok M. 1994. Isolation of RNA aptamers for biological cofactors by in vitro selection. *Angew Chem Int Ed Engl* **33**: 1084–1087.
- Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. 2002. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 2. doi: 10.1186/1471-2105-3-2.
- Collins RA. 2002. The *Neurospora* Varkud satellite ribozyme. *Biochem Soc Trans* **30**: 1122–1126.
- Cruz JA, Westhof E. 2011. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat Methods* **8**: 513–521.
- Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Famulok M, Huttenhofer A. 1996. In vitro selection analysis of neomycin binding RNAs with a mutagenized pool of variants of the 16S rRNA decoding region. *Biochemistry* **35**: 4265–4270.
- Gan HH, Pasquali S, Schlick T. 2003. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res* **31**: 2926–2943.

- Gardner PP, Wilm A, Washietl S. 2005. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res* **33**: 2433–2439.
- Gevertz J, Gan HH, Schlick T. 2005. In vitro RNA random pools are not structurally diverse: A computational analysis. *RNA* **11**: 853–863.
- Giedroc DP, Theimer CA, Nixon PL. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* **298**: 167–185.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: An RNA family database. *Nucleic Acids Res* **31**: 439–441.
- Guo HC, De Abreu DM, Tillier ER, Saville BJ, Olive JE, Collins RA. 1993. Nucleotide sequence requirements for self-cleavage of *Neurospora* VS RNA. *J Mol Biol* **232**: 351–361.
- Hoehndorf R, Batchelor C, Bittner T, Dumontier M, Eilbeck K, Knight R, Mungall CJ, Richardson JS, Stombaugh J, Westhof E, et al. 2011. The RNA Ontology (RNAO): An ontology for integrating RNA sequence and structure data. *Appl Ontol* **6**: 53–89.
- Jaeger L, Chworos A. 2006. The architectonics of programmable RNA and DNA nanostructures. *Curr Opin Struct Biol* **16**: 531–543.
- Jenison RD, Gill SC, Pardi A, Polisky B. 1994. High-resolution molecular discrimination by RNA. *Science* **263**: 1425–1429.
- Jiang F, Kumar RA, Jones RA, Patel DJ. 1996. Structural basis of RNA folding and recognition in an AMP–RNA aptamer complex. *Nature* **382**: 183–186.
- Jiang L, Suri AK, Fiala R, Patel DJ. 1997. Saccharide–RNA recognition in an aminoglycoside antibiotic–RNA aptamer complex. *Chem Biol* **4**: 35–50.
- Jiang L, Majumdar A, Hu W, Jaishree TJ, Xu W, Patel DJ. 1999. Saccharide–RNA recognition in a complex formed between neomycin B and an RNA aptamer. *Structure* **7**: 817–827.
- Jossinet F, Ludwig TE, Westhof E. 2010. Assemble: An interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**: 2057–2059.
- Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J. 2009. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**: D159–D162.
- Kim YG, Su L, Maas S, O'Neill A, Rich A. 1999. Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc Natl Acad Sci* **96**: 14234–14239.
- Kim YG, Maas S, Wang SC, Rich A. 2000. Mutational study reveals that tertiary interactions are conserved in ribosomal frameshifting pseudoknots of two luteoviruses. *RNA* **6**: 1157–1165.
- Knight R, De Sterck H, Markel R, Smit S, Oshmyansky A, Yarus M. 2005. Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. *Nucleic Acids Res* **33**: 5924–5935.
- Lafontaine DA, Norman DG, Lilley DM. 2002. Folding and catalysis by the VS ribozyme. *Biochimie* **84**: 889–896.
- Lee JF, Hesselberth JR, Meyers LA, Ellington AD. 2004. Aptamer database. *Nucleic Acids Res* **32**: D95–D100.
- Leontis NB, Westhof E. 1998. Conserved geometrical base-pairing patterns in RNA. *Q Rev Biophys* **31**: 399–455.
- Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512.
- Leontis NB, Westhof E. 2002. The annotation of RNA motifs. *Comp Funct Genomics* **3**: 518–524.
- Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30**: 3497–3531.
- Licis N, van Duin J. 2006. Structural constraints and mutational bias in the evolutionary restoration of a severe deletion in RNA phage MS2. *J Mol Evol* **63**: 314–329.
- Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**: 1632–1635.
- Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**: 1335–1337.
- Pan T, Dichtl B, Uhlenbeck OC. 1994. Properties of an in vitro selected Pb²⁺ cleavage motif. *Biochemistry* **33**: 9561–9565.
- Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. 2008. FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol* **56**: 215–252.
- Seelig B, Jäschke A. 1999. A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol* **6**: 167–176.
- Smit S, Yarus M, Knight R. 2006. Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA* **12**: 1–14.
- Smit S, Widmann J, Knight R. 2007. Evolutionary rates vary among rRNA structural elements. *Nucleic Acids Res* **35**: 3339–3354.
- Smit S, Knight R, Heringa J. 2009. RNA structure prediction from evolutionary patterns of nucleotide composition. *Nucleic Acids Res* **37**: 1378–1386.
- Stombaugh J, Zirbel CL, Westhof E, Leontis NB. 2009. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* **37**: 2294–2312.
- Stombaugh J, Widmann J, McDonald D, Knight R. 2011. Boulder Alignment Editor (ALE): A web-based RNA alignment tool. *Bioinformatics* **27**: 1706–1707.
- Thompson JD, Holbrook SR, Katoh K, Koehl P, Moras D, Westhof E, Poch O. 2005. MAO: A Multiple Alignment Ontology for nucleic acid and protein sequences. *Nucleic Acids Res* **33**: 4164–4171.
- Wallis MG, von Ahsen U, Schroeder R, Famulok M. 1995. A novel RNA motif for neomycin recognition. *Chem Biol* **2**: 543–552.
- Wang HC, Hickey DA. 2002. Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Res* **30**: 2501–2507.
- Wang Y, Rando RR. 1995. Specific binding of aminoglycoside antibiotics to RNA. *Chem Biol* **2**: 281–290.
- Wang Y, Killian J, Hamasaki K, Rando RR. 1996. RNA molecules that specifically and stoichiometrically bind aminoglycoside antibiotics with high affinities. *Biochemistry* **35**: 12338–12346.
- Wilson C, Nix J, Szostak J. 1998. Functional requirements for specific ligand recognition by a biotin-binding RNA pseudoknot. *Biochemistry* **37**: 14410–14419.
- Wong KM, Suchard MA, Huelsenbeck JP. 2008. Alignment uncertainty and genomic analysis. *Science* **319**: 473–476.
- Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M. 1996. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**: 1343–1347.
- Yarus M, Welch M. 2000. Peptidyl transferase: Ancient and exiguous. *Chem Biol* **7**: R187–R190.
- Zimmermann GR, Jenison RD, Wick CL, Simorre JP, Pardi A. 1997. Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat Struct Biol* **4**: 644–649.



RNA

A PUBLICATION OF THE RNA SOCIETY

RNASTAR: An RNA STructural Alignment Repository that provides insight into the evolution of natural and artificial RNAs

Jeremy Widmann, Jesse Stombaugh, Daniel McDonald, et al.

RNA 2012 18: 1319-1327 originally published online May 29, 2012
Access the most recent version at doi:[10.1261/rna.032052.111](https://doi.org/10.1261/rna.032052.111)

-
- Supplemental Material** <http://rnajournal.cshlp.org/content/suppl/2012/05/07/rna.032052.111.DC1>
- References** This article cites 52 articles, 10 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/18/7/1319.full.html#ref-list-1>
- Open Access** Freely available online through the *RNA* Open Access option.
- License** Freely available online through the RNA Open Access option.
- Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).
-



To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
