High-throughput dissection of the thermodynamic and conformational properties of a ubiquitous class of RNA tertiary contact motifs

Steve L. Bonilla[a,b], Sarah K. Denny[c,d], John H. Shin[a,b], Aurora Alvarez-Buylla[c], William J. Greenleaf[c,d], and Daniel Herschlag[a,b,d,1]

[a]Department of Biochemistry, School of Medicine, Stanford University, Stanford, CA 94305; [b]Department of Chemical Engineering, Stanford University, Stanford, CA 94305; [c]Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305; and [d]Biophysics Program, Stanford University, Stanford, CA 94305

Edited by Michael F. Summers, University of Maryland, Baltimore County, Baltimore, MD, and approved June 29, 2021 (received for review May 16, 2021)

Despite RNA's diverse secondary and tertiary structures and its complex conformational changes, nature utilizes a limited set of structural "motifs"—helices, junctions, and tertiary contact modules—to build diverse functional RNAs. Thus, in-depth descriptions of a relatively small universe of RNA motifs may lead to predictive models of RNA tertiary conformational landscapes. Motifs may have different properties depending on sequence and secondary structure, giving rise to subclasses that expand the universe of RNA building blocks. Yet we know very little about motif subclasses, given the challenges in mapping conformational properties in high throughput. Previously, we used "RNA on a massively parallel array" (RNA-MaP), a quantitative, high-throughput technique, to study thousands of helices and two-way junctions. Here, we adapt RNA-MaP to study the thermodynamic and conformational properties of tetraloop/tetraloop receptor (TL/TLR) tertiary contact motifs, analyzing 1,493 TLR sequences from different classes. Clustering analyses revealed variability in TL specificity, stability, and conformational behavior. Nevertheless, natural GAAA/11ntR TL/TLRs, while varying in tertiary stability by ∼2.5 kcal/mol, exhibited conserved TL specificity and conformational properties. Thus, RNAs may tune stability without altering the overall structure of these TL/TLRs. Furthermore, their stability correlated with natural frequency, suggesting thermodynamics as the dominant selection pressure. In contrast, other TL/TLRs displayed heterogenous conformational behavior and appear to not be under strong thermodynamic selection. Our results build toward a generalizable model of RNA-folding thermodynamics based on the properties of isolated motifs, and our characterized TL/TLR library can be used to engineer RNAs with predictable thermodynamic and conformational behavior.

RNA structure | RNA folding | high-throughput biochemistry | tertiary motifs | RNA nanotechnology

RNAs fold into compact tertiary structures and undergo dynamic conformational changes to function in essential processes, including translation, premessenger RNA splicing, and genome maintenance (1–4). To accomplish these functions, RNAs typically fold by a hierarchical process that starts with the formation of stable helices (i.e., secondary structure) followed by compaction and formation of long-range tertiary contacts between prefolded helices (5–7). Because of the importance of structure for function, decades of study have been dedicated to dissecting the forces that dictate folding, determining steps and intermediates in the folding process, and characterizing the final folded functional structures (5).

We now have a reasonably predictive understanding of secondary structure formation from sequence covariation and thermodynamic "nearest neighbors" parameters for base pairing (8, 9). However, our understanding of RNA tertiary structure and folding is far less developed. In part, this is because, in comparison to secondary structure, tertiary structure formation is a less local problem, with more degrees of freedom and more interdependent interacting groups and thus requires more terms to describe its energetics (10). Further complicating predictive models, RNAs exist as conformational ensembles dictated by thermodynamic Boltzmann-weighted distributions, rather than as single static structures, and do so in both their unfolded and folded states so that static structural descriptions are inadequate (11, 12). Despite these complexities, simplifications arising from RNA's modular architecture facilitate the analysis of RNA structure and may ultimately lead to the development of a predictive understanding of RNA-folding energetics (5, 13).

Diverse RNAs are built using a common set of tertiary structural building blocks or motifs (14, 15). These include A-form helices formed by base pairing, junctions that connect and orient helices in a three-dimensional (3D) space, and tertiary contacts that form long-range interactions between distal regions. Some of these motifs appear highly modular, largely conserving their 3D conformational ensemble across different RNA contexts (14, 16–19). For example, crystal structures of the canonical tetraloop/tetraloop receptor (TL/TLR) tertiary contact motif (also referred to as the GAAA/11ntR motif) are superimposable (Fig. 1*A*) (5, 20, 21). Similarly, certain junction motifs hold their emanating helices in their native orientations, even when isolated outside of their natural context (19, 22–24). These observations suggest that RNA structure is largely dictated by the intrinsic conformational properties of its constituent motifs and that a generalizable model based

Significance

We would like to have a predictive understanding of how RNAs form and interconvert between intricate three-dimensional shapes that are essential to function. Despite great structural diversity, RNAs are built from a limited set of structural motifs so that a predictive model may arise from the in-depth understanding of these building blocks. We used a quantitative, high-throughput approach to investigate the thermodynamic and conformational properties of the tetraloop/tetraloop receptor (TL/TLR), a ubiquitous RNA tertiary contact motif. Our investigation of 1,493 TLRs and two TLs revealed classes of TL/TLR thermodynamic behavior. One class displayed conserved conformation across sequences and appears to be thermodynamically selected in biology. Other classes are conformationally diverse and may be selected for flexibility instead of stability.

Author contributions: S.L.B., S.K.D., W.J.G., and D.H. designed research; S.L.B., S.K.D., and A.A.B. performed research; W.J.G. contributed new reagents/analytic tools; S.L.B., S.K.D., and J.H.S. analyzed data; and S.L.B. and D.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: herschla@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2109085118/-/DCSupplemental.

Published August 9, 2021.

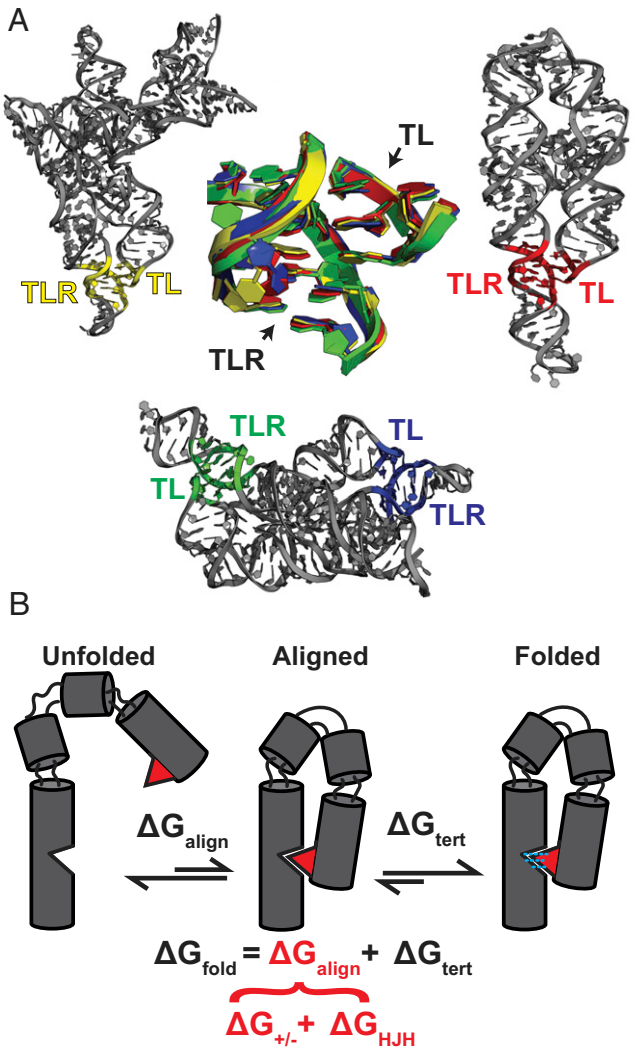

Fig. 1. Modular RNA structure and energetics. (*A*) Tertiary structure of the canonical GAAA/11ntR TL/TLR. Crystal structures of GAAA/11ntR from RNase P RNA (*Top Left*; Protein Data Bank [PDB] 1NBS), P4-P6 domain (*Top Right*; PDB 1GID), and *Azoarcus* group I intron (*Bottom*; PDB 1ZZN), superimposed using the PyMOL molecular graphics system (version 2.0 Schrödinger, LLC). (*B*) Reconstitution Model of RNA folding (13). Folding free energy (ΔG_{fold}) is decomposed into an internal conformational search (ΔG_{align}) and the formation of stabilizing tertiary interactions (ΔG_{tert}). ΔG_{align} can be further decomposed into a contribution from electrostatics and ion interactions ($\Delta G_{+/-}$) and a contribution from the conformational properties of helices, junctions, and tertiary contact motifs comprising the RNA (ΔG_{HJH}; HJH stands for helix-junction-helix). Thus, in this model, the conformational ensemble of the motifs is dictated by properties intrinsic to the motif such as sequence and topology, while environmental factors such as ions affect the relative stabilities of the conformations within the ensemble.

on RNA's architectural modularity may emerge from an understanding of these motifs.

In particular, RNA's modular architecture and prior energetic analyses support a reconstitution model of RNA folding (5, 13, 25). In this model, an RNA is considered as a collection of motifs connected in space, with the conformational ensemble of each motif specified by properties intrinsic to the motif (Fig. 1*B*). Two major components dictate the free energy for the formation of a folded RNA (ΔG_{fold}): 1) the probability that the collection of structural motifs will align the tertiary contact interfaces (ΔG_{align}) and 2) the free energy associated with the formation of stabilizing tertiary interactions (ΔG_{tert}; Fig. 1*B*). ΔG_{align} is dictated by the intrinsic conformational ensembles of each of the helices, junctions, and tertiary contacts comprising the RNA and the order in which these elements are connected to each other; ΔG_{tert} is determined by the hydrogen bonds, base stacking, and other interactions that form and break upon the formation of the tertiary contacts. Evidence supporting the reconstitution model includes the ability to quantitatively predict the effect of specific mutations on the folding of the *Tetrahymena* group I intron P4-P6 domain from the effects of the same mutations in the context of a simplified construct that isolates the tertiary contact elements (13).

As the reconstitution model and other models based on RNA structural modularity rely on an in-depth quantitative understanding of the intrinsic properties of helices, junctions, and tertiary contact motifs that recur in folded RNA structures, a major goal toward understanding RNA conformational ensembles must be to characterize the conformational and thermodynamic properties of the many recurring motifs and sequences.

We recently developed a high-throughput methodology to study the conformational and thermodynamic properties of RNA structural motifs across sequence space (26). Our method relies on the tertiary assembly of an RNA heterodimer engineered by Jaeger and coworkers (27, 28), referred to as "tectoRNA," that is stabilized by the formation of two orthogonal TL/TLR modules: GAAA/11ntR$_{wt}$ and GGAA/R1 (Fig. 2*A*). The stability of the dimer (ΔG_{bind}) depends on the probability that the structural motifs comprising the tectoRNA correctly and simultaneously align both TLs to their cognate TLRs (Fig. 2*B*) (26). The alignment process and probability can be conceptualized in terms of an internal conformational search that is analogous to that undergone during intramolecular RNA folding (ΔG_{align}; Fig. 1*B*), and it depends on the intrinsic conformational properties of the structural motifs comprising the tectoRNA construct (Fig. 2*B*). The value of ΔG_{bind} is also determined by the strength of the tertiary interactions within the TL/TLR modules (ΔG_{tert}), in the same way the formation of tertiary contacts in RNA intramolecular folding determines folding stability (Fig. 1*B*).

In our previous studies, we used "RNA on a massively parallel array" (RNA-MaP) to insert thousands of helix and junction motif sequences into the tectoRNA construct and measure the effect of these insertions on ΔG_{bind} (Fig. 2*C*) (10, 26). Each motif insertion was made across multiple, structural versions of the tectoRNA, varying in length, sequence, and/or secondary structure to produce a series of ΔG_{bind} values for each motif sequence; each of these ΔG_{bind} series was referred to as a "thermodynamic fingerprint." Thermodynamic fingerprints reported on the conformational properties of each helix and junction sequence in the library as follows: Helices and junctions with similar conformational properties produced similar thermodynamic fingerprints, as their insertion similarly affected the TL/TLR alignment probabilities, whereas those with different conformational properties can yield distinct fingerprints. In our present study, we employed this approach to dissect the thermodynamic and conformational properties of TL/TLR tertiary contacts, a ubiquitous and important class of RNA structural motifs (Fig. 2*D*).

TL/TLRs stabilize the compact structure of diverse RNAs, including ribosomal RNAs, group I and II introns, RNase P RNAs, and riboswitches, and have been used in RNA nanotechnology (27, 29–31). TL/TLRs vary in sequence and architecture, and three major types are commonly found in nature (Fig. 2*D*) (21, 31). The simplest class of TL/TLR consists of GNRA TLs (where N is any nucleotide [nt] and R is a purine) making hydrogen bond networks with the minor groove of base-paired helices. These TL/TLRs are often referred to as GNRA/minor groove interactions. The second class consist of GAAA TLs making hydrogen bonds and stacking interactions with an 11-nt semiconserved TLR (11ntR) (GAAA/11ntR; Fig. 2*D*). The third class consists of GAAA TLs interacting with a 12-nt TLR that shares some secondary structural features

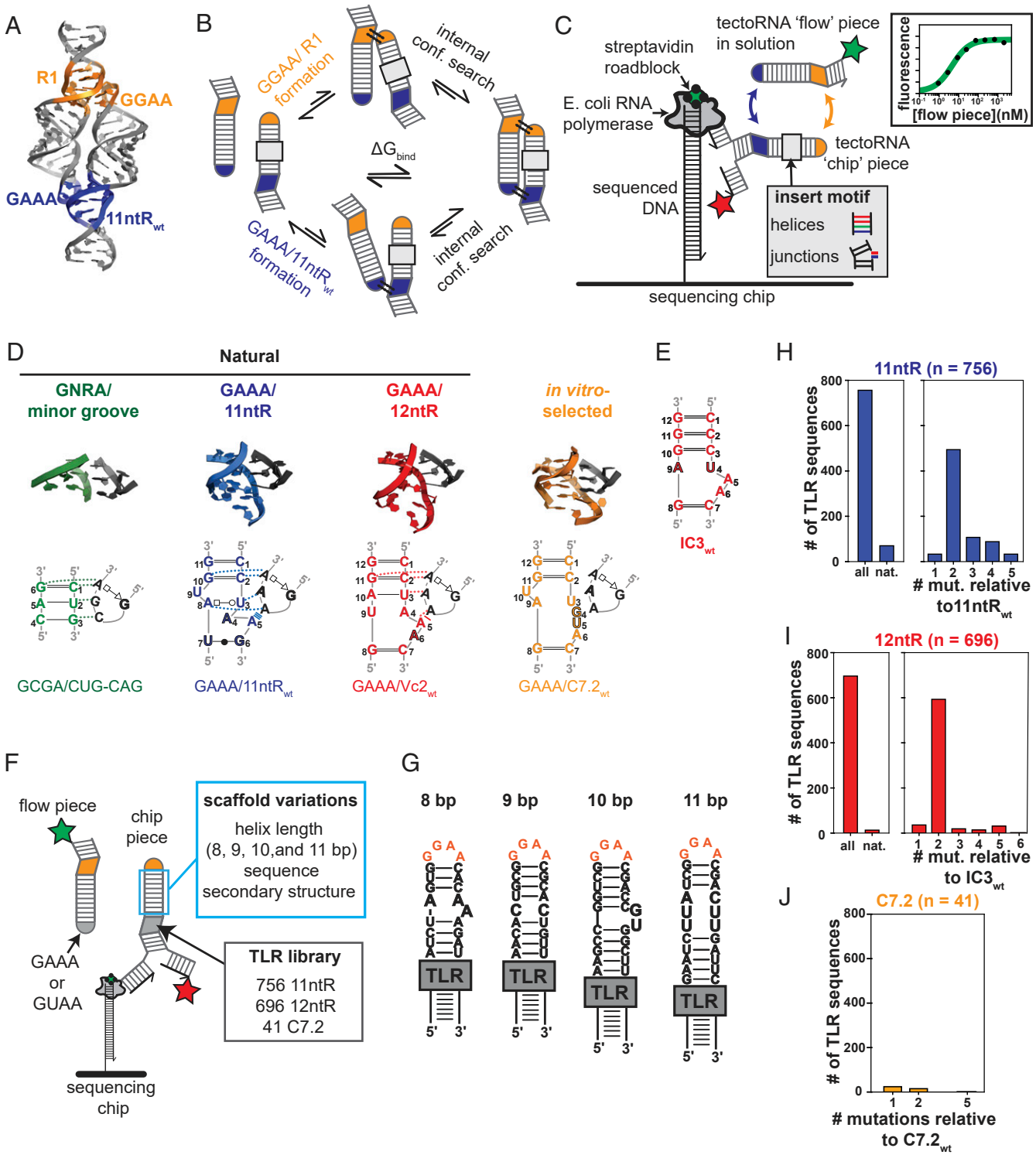

Fig. 2. High-throughput studies of RNA structural motifs. (*A*) Tertiary structure of tectoRNA dimer. Structure shown is that of homodimer with two identical GAAA/11ntR$_{wt}$ TL/TLRs (49). One of the GAAA/11ntR$_{wt}$ is replaced by an orthogonal GGAA/R1 TL/TLR in our studies to ensure heterodimer formation (28). (*B*) Pathways for the formation of bound tectoRNA dimer. After the formation of one of the tertiary contacts, an internal conformational search aligns the second TL to its cognate TLR. (*C*) High-throughput RNA-MaP platform used to characterize RNA structural motifs. Tens of thousands of "chip" piece sequences are transcribed from sequenced DNA on an Illumina chip (37). Stars represent fluorescent probes. Junctions and/or short helical elements are inserted into the helical domain of the chip piece (26). (*D*) Sequence, secondary, and tertiary structure of representative TL/TLRs belonging to different types. Tertiary structures are from GCGA TL docking into CUG-CAG (Protein Data Bank [PDB] 3IGI), GGAA/11ntR$_{wt}$ (PDB 1GID), and GAAA/Vc2$_{wt}$ (PDB 3IRW). The GAAA/C7.2$_{wt}$ structure was modeled computationally (50). (*E*) Sequence of IC3$_{wt}$ TLR motif. (*F*) Schematic for characterization of the TL/TLR library via RNA-MaP. Varying TLR sequences were inserted in place of the 11ntR$_{wt}$ sequence in the chip piece. Scaffold variants for the insertion of each TLR sequence were generated by altering the length, sequence, and/or secondary structure of the helical domain of the chip piece. (*G*) Examples of scaffolds of different lengths. The length of the chip piece is defined as the number of bp between the GGAA TL and the TLR, including canonical and noncanonical (i.e., mismatches) bp. Residues without opposing residues in complementary strand (e.g., bulges) do not contribute to the reported length. (*H–J*) Distribution of TLR sequences in the library. Natural 11ntR and 12ntR TLR variants were obtained from databases and published sequence alignments of functional RNAs (39–42).

with 11ntRs (GAAA/12ntR; Fig. 2*D*). In addition to natural TL/TLRs, in vitro selection studies have generated stable TL/TLR variants not observed in nature (28, 30). Although diverse in sequence, secondary structure, and complexity, there are several commonalities between the three different types of natural TL/TLRs described above. They all involve the interactions of GNRA TLs that dock into the minor groove of the TLR helix by formation of so-called A-minor interactions (30, 31). In the present study, we investigate the properties of a wide range of natural and mutant 11ntR (n = 756) and 12ntR (n = 696) TLRs, a smaller subset of variants of an in vitro selected TLR (n = 41), and a representative minor groove TLR.

The structures, relative stabilities, and/or specificities of several representative TL/TLR sequences have been characterized previously (20, 28, 30–33). These include the canonical GAAA/11ntR$_{wt}$ of the P4-P6 domain of the *Tetrahymena* group I intron (Fig. 2*D*, blue), a GAAA/12ntR variant found in *Vibrio cholera* cyclic-di-GMP riboswitches, referred to as GAAA/Vc2$_{wt}$ (Fig. 2*D*, red), a GAAA/12ntR commonly found in subtype IC3 group I introns, referred to as GAAA/IC3$_{wt}$ (Fig. 2*E*) (34), and the in vitro–selected GAAA/C7.2$_{wt}$ (Fig. 2*D*, orange). The thermodynamic properties of additional, artificial TL/TLRs generated by in vitro selection or by rational design have also been characterized (28, 30, 35). However, compared to the great number of potential TL/TLR sequences, only a small fraction of the thermodynamic and conformational space of TL/TLRs has been explored. To determine the diversity of the thermodynamic and conformational landscape of TL/TLRs and obtain insights into how nature selects among this diversity to build functional RNAs with different, functional, dynamic, and conformational properties, we used the RNA-MaP high-throughput platform to measure the relative stability, specificity, and conformational properties of a library of natural and designed TL/TLR sequences (Fig. 2*F*). From these results, we were able to classify and compare the conformational and thermodynamic behavior of TLRs, results that help build toward a general ability to describe RNA-folding thermodynamics, from the properties of its constituents, that provide biological insights and that can be of immediate use in RNA engineering.

Results and Discussion

RNA-MaP Enables High-Throughput Studies of TL/TLRs in tectoRNA Scaffolds. In our previous RNA-MaP studies of helix and junction motifs, we kept the sequence of the TL/TLRs constant to probe helices and junctions (Fig. 2*C*) (10, 26). In the present study, we systematically altered the sequence of the TLR in the tectoRNA chip piece and the sequence of the TL in the flow piece, with the GGAA/R1 TL/TLR kept constant to focus on a single TL/TLR (Fig. 2*F*). We inserted a library of 1,493 TLR sequences into chip pieces, and we synthesized two flow pieces, one with a GAAA TL and another with a GUAA TL. The flow pieces with the different TL sequences allowed us to probe the specificity of each TLR in the library. In addition, each TLR was inserted into multiple chip pieces differing in length, sequence, and/or secondary structure. These structural variations of the chip piece are referred to as "scaffolds" (Fig. 2*F*). The length of the scaffolds spanned 8 to 11 bp, with examples shown in Fig. 2*G*. The scaffold sequences were chosen to span a wide range of tectoRNA stabilities and thus conformational preferences, based on our previous studies (10, 26); those studies showed large effects on stability from changes in scaffold length and more moderate effects from changes in sequence and/or secondary structure. The library was designed with each of the 1,493 TLR sequences inserted into five scaffolds and a subset (n = 277) inserted into 50 scaffolds for deeper characterization. As explained in detail in *Thermodynamic Fingerprints Report on TL/TLR Conformational Properties* and following our prior studies (26), ΔG_{bind} measurements across multiple scaffolds produced thermodynamic fingerprints that report on the 3D conformational properties of the TL/TLRs in the library.

Clusters of DNA encoding the chip pieces were expressed and sequenced on an Illumina chip (36, 37). The chip was then installed into a custom-built, fluorescence microscope, and the locations and sequences of the DNA clusters were mapped using a previously described method (36, 38). For statistical rigor, variants that were expressed in fewer than five clusters were not analyzed (26). RNA clusters were generated in situ by transcribing the sequenced DNA clusters and stalling the RNA polymerase using a streptavidin "roadblock" and were visualized by annealing a fluorescent oligo to a common flanking sequence. A fluorescently labeled flow piece was introduced at a series of concentrations (0.91 nM to 2.0 μM), and dissociation constants were generated by measuring the change in the fluorescence of each cluster as a function of flow piece concentration after equilibration. Based on prior control experiments, we set an upper limit of −7.1 kcal/mol for the reliable measurement of ΔG_{bind} (26). Values greater than this threshold nevertheless showed expected trends and are shown in figures for completeness but were not used in quantitative analyses.

A TL/TLR Library to Probe Diversity of Natural and Mutant TL/TLRs. To investigate and compare the thermodynamic and conformational properties of TL/TLRs belonging to different structural types and gain insights into how the sequence of the motif may determine these properties, we designed a library of 1,493 TLR sequences and systematically introduced them into the tectoRNA chip piece. The 1,493 TLRs included natural and mutant 11ntR (n = 756) and 12ntR (n = 696) TLRs and variants of in vitro–selected C7.2$_{wt}$ (n = 41) (Fig. 2 *H–J*). Natural TLR sequences were found in databases and publications containing sequence alignments of group I and II introns, RNase P RNA, and cyclic di-GMP riboswitches (*SI Appendix*, Tables S1 and S2) (39–42). We refer to 11ntR and 12ntR sequences using 11ntR$_{wt}$ and IC3$_{wt}$ as references, respectively (Fig. 2 *D* and *E*). Mutations are noted using subscripts with the position and identity of the mutation. For example, a mutant of 11ntR$_{wt}$ in which position 4 is mutated to G is referred to as 11ntR$_{4G}$, and a mutant of IC3$_{wt}$ in which position 1 is mutated to C is referred to as IC3$_{1C}$. Multiple mutations are noted sequentially (e.g., 11ntR$_{4G/7C}$). The IC3$_{U3/A4/U9/A10}$ receptor is referred to as Vc2$_{wt}$ to be consistent with previous literature (31). Mutants of C7.2$_{wt}$ are described using the same nomenclature. The library contained TLRs with up to six mutations, relative to the reference sequences, with the large majority being double mutants of 11ntR$_{wt}$ and IC3$_{wt}$ (Fig. 2 *H–J*).

Thermodynamic Fingerprints Report on TL/TLR Conformational Properties. As noted in *RNA-MaP Enables High-Throughput Studies of TL/TLRs in tectoRNA Scaffolds*, each TLR was inserted into multiple chip piece scaffolds varying in length, sequence, and/or secondary structure (Fig. 2*F*). The value of ΔG_{bind} depends on the ability of the structural elements comprising tectoRNA to correctly align the two TL/TLRs (Fig. 2*B*). Therefore, scaffolds with the same TL/TLRs but different conformational properties produce different ΔG_{bind} values, and these differences, in turn, depend on the alignment preferences of the TL/TLRs. As these concepts are central to our studies, we elaborate them below.

ΔG_{bind} depends on the structural properties of the scaffold. Consider the folding pathway in which GGAA/R1 forms first and is followed by the formation of GAAA/11ntR$_{wt}$ (top pathway; Fig. 2*B*). As the final bound state always has both TL/TLRs formed and ΔG_{bind} is a state function dependent only on the initial and final states, the choice of pathway does not affect our analysis (26, 28). Also, intermediates with a single TL/TLR formed are too unstable to be detectable, so we observe the initial and final states experimentally (26, 28).

After the docking of GGAA/R1, the tectoRNA undergoes an internal conformational search to align the second TL/TLR (Fig. 3*A*), analogous to that undergone in intramolecular RNA folding (Fig. 1*B*). The structural properties of the scaffold (e.g.,

its length and any internal junctions) determine the space sampled during this internal conformational search and, therefore, the probability of aligning the TL/TLRs to form the bound tectoRNA complex (Fig. 3*A*). As the thermodynamic stability of a complex is given by the probability of forming the complex over the probability of breaking it, tectoRNA scaffolds that differ conformationally produce different ΔG_{bind} values. Binding measurements across a series of scaffolds produce a series of ΔG_{bind} values that we refer to as a thermodynamic fingerprint (Fig. 3*B*). The shape of a thermodynamic fingerprint depends on the structural properties of the scaffolds and on the particular TL/TLR (i.e., on the alignment preferences of the TL/TLRs, as explained next).

Thermodynamic fingerprints report on the 3D conformational properties of the TL/TLRs. Changing the TL/TLR in tectoRNA can have two effects on ΔG_{bind}. It can 1) change the strength of the tertiary interactions formed and 2) change the alignment required for TL/TLR formation. To illustrate these points, consider the three hypothetical TL/TLRs embedded in identical scaffolds shown in Fig. 3*C*. The behaviors of these hypothetical TL/TLRs are used to illustrate expected trends and interpret the experimental thermodynamic fingerprints presented in the following sections.

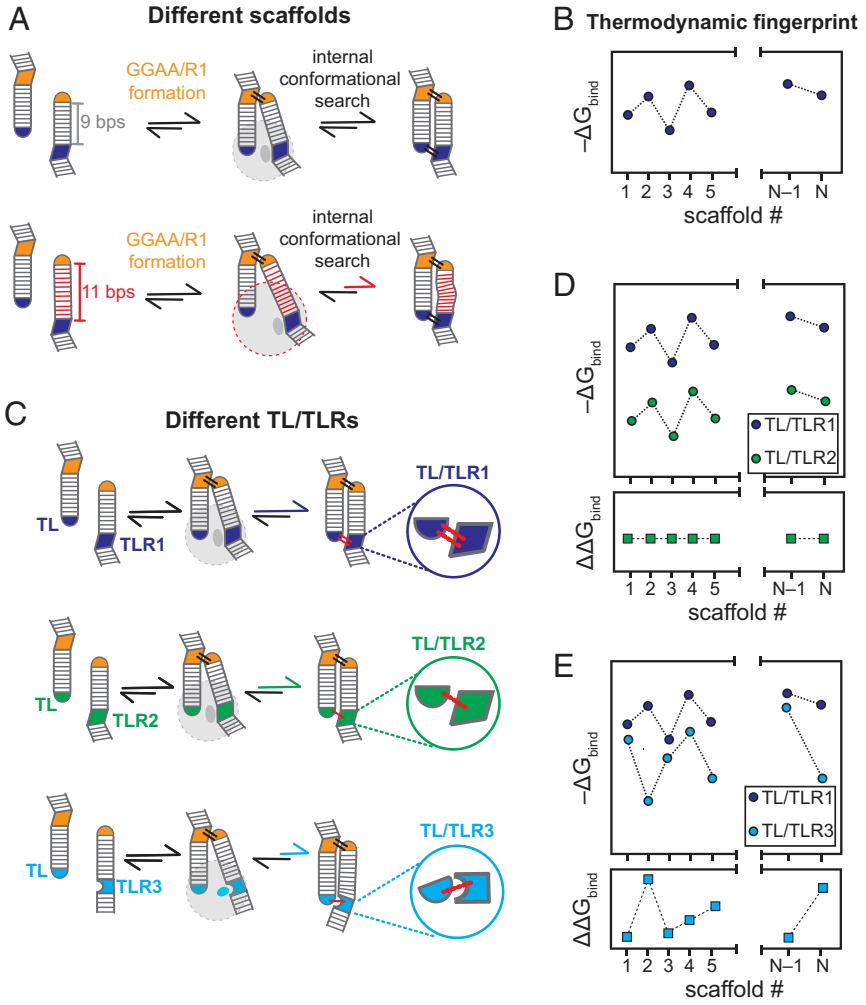

Fig. 3. Thermodynamic fingerprints report on TL/TLR 3D conformational properties. (*A*) Structural changes to the scaffold affect ΔG_{bind} by altering the internal conformational search to align the TL/TLRs. The figure shows the schematic comparing the internal conformational search undergone by a short scaffold (9 bp, *Top*) and a long scaffold (11 bp, *Bottom*). The large gray circle represents the conformational space explored by a particular scaffold during the internal conformational search, and the smaller darker gray area represents the part of this space where the TL and TLR are "correctly" aligned for the formation of the TL/TLR. In this simplified schematic (with equal probabilities throughout the 2D gray space), the fraction of conformational space, leading to productive formation of the TL/TLR, is greater for the 9 bp than for the 11-bp scaffold. Thus, the 9-bp scaffold has a greater probability of aligning the TL/TLR, resulting in higher affinity compared to the 11-bp scaffold. In actual cases, the relative occupancy across the available conformations also affects the probability of binding with each scaffold. (*B*) Measurements of ΔG_{bind} across distinct scaffolds produce thermodynamic fingerprints that report on the alignment properties of the TL/TLRs. Each datapoint corresponds to a distinct scaffold varying in length, secondary structure, and/or sequence. Depending on its conformational properties, each scaffold has a distinct probability of aligning the TL/TLRs, resulting in the distinct ΔG_{bind} values. (*C*) Differences between TL/TLR variants can affect ΔG_{bind} by affecting the strength of the tertiary interactions and/or the alignment preferences. Cartoons represent three distinct TL/TLR sequences embedded within the same scaffold. TL/TLR1 and TL/TLR2 have the same alignment preferences but differ in the strength of their tertiary interactions. In this case, the difference in stability between the TL/TLRs is independent of the structural context in which the comparison is made. TL/TLR3 has different alignment preferences compared to the other two TL/TLRs. In this case, the difference in stability between TL/TL3 and the other TL/TLRs depends on the structural context in which the comparison is made. (*D*) TL/TLRs with the same alignment preferences produce thermodynamic fingerprints with the same shape. The strength differences between TL/TLR1 and TL/TLR2 are expressed by a constant offset in their stability ($\Delta\Delta G_{bind}$). (*E*) TL/TLRs that differ in alignment preferences produce thermodynamic fingerprints with different shapes, such that $\Delta\Delta G_{bind}$ depends on the scaffold identity.

Hypothetical motifs TL/TLR1 and TL/TLR2 differ in the strength of their tertiary interactions but have the same alignment preferences (Fig. 3*C*, blue versus green). TL/TLR3 has different alignment preferences (Fig. 3*C*, cyan). The differences in strength between TL/TLR1 and TL/TLR2 will produce different ΔG_{bind} values but, because the final alignments are the same, this difference, $\Delta\Delta G_{bind}$, is independent of the structural context in which TL/TLR1 and TL/TLR2 are embedded. Thus, $\Delta\Delta G_{bind}$ is constant across scaffolds and the thermodynamic fingerprints of these TL/TLRs have the same shape, with a constant offset given by $\Delta\Delta G_{bind}$ (Fig. 3*D*).

In contrast, because TL/TLR3 has a different preferred alignment, $\Delta\Delta G_{bind}$ between TL/TLR1 and TL/TLR3 is dependent on the structural context (Fig. 3*C*) and, because of this, the shape of their thermodynamic fingerprints differs (Fig. 3*E*). Thus, comparing the shapes of the thermodynamic fingerprints for two different TL/TLR sequences provides information on the similarity of their 3D conformational properties.

To clarify, the fingerprints provide information on the "global" conformation of the TL/TLRs (i.e., the alignment and orientation of the TL with respect to the TLR) and not on the internal configuration of the TL/TLRs. Thus, in principle, two TL/TLRs can have a different set of interactions (e.g., base pairs [bp] and base stacking within the TLR tertiary hydrogen bonds) but the same alignment preferences and, therefore, similar thermodynamic fingerprints. Additionally, for simplicity, we have been referring to the final aligned TL/TLR as a single state, but TL/TLRs also exist as conformational ensembles in their folded state. Thus, the shapes of the thermodynamic fingerprints report on differences and similarities in the global conformational ensembles of the bound TL/TLRs. Each TL/TLR conformational ensemble can consist of multiple states (some more probable than others) that connect the TL to its TLR. In the simplest case, if more states are available for the bound TL/TLR (i.e., a broad conformational ensemble), many scaffolds would have similar probabilities to align the TL/TLRs so that changing one scaffold for another would not yield a major energetic penalty and a flatter thermodynamic fingerprint would be expected. In contrast, a TL/TLR with a highly specific (or restrictive) conformation (i.e., a narrow TL/TLR conformational ensemble) is expected to be more sensitive to changes to the geometry of the scaffold.

Different Thermodynamic and Conformational Behavior of TL/TLRs that Represent Distinct Structural Classes. We began our analysis by comparing the thermodynamic fingerprints of a representative subset of TL/TLR variants across 50 possible common scaffolds and across three conditions: 30 mM $MgCl_2$ (Fig. 4*A*), 5 mM $MgCl_2$/150 mM KCl, and 5 mM $MgCl_2$ (*SI Appendix*, Fig. S1). This set of TLRs included wild-type (wt) TLR sequences $11ntR_{wt}$, $IC3_{wt}$, $Vc2_{wt}$, and $C7.2_{wt}$. We also included a second representative 11ntR, the triple mutant $11ntR_{1G/9G/11C}$, which is found in several group I intron and RNase P RNA variants (*SI Appendix*, Table S1). In addition, to compare the TL/TLRs studied here to the simpler GNRA/minor groove TL/TLR interactions, we performed thermodynamic measurements with chip constructs in which the TLR is replaced by tandem bps 5′-CU/AG-3′ ("bp helix").

Thermodynamic fingerprints for each representative TLR binding to GAAA were generated by plotting ΔG_{bind} across scaffolds, with the scaffolds ordered according to length (Fig. 4*A* and *SI Appendix*, Fig. S1). To compare the TL/TLR fingerprints, we calculated Pearson's correlation coefficients (r) and rmsd (Fig. 4*A* and *SI Appendix*, Fig. S1). In addition to reporting on similarities and differences in the shapes of the fingerprints, the value of r also depends on the stability of the fingerprints being compared, because low-stability fingerprints have greater measurement error (*SI Appendix*, Fig. S2) and contain fewer values within the threshold (–7.1 kcal/mol). As a result, TLRs with the same intrinsic correlation display lower correlations in practice when they are less stable. To account for this effect, we simulated theoretically identical fingerprints across a range of average stabilities and used the correlation values obtained from these simulations to guide the interpretation of our data (*SI Appendix*, Fig. S3). As an additional way to visualize and compare the TL/TLRs, we calculated $\Delta\Delta G_{bind}$ with respect to $GAAA/11ntR_{wt}$ (Fig. 4*B*). As explained in *Thermodynamic Fingerprints Report on TL/TLR Conformational Properties*, $\Delta\Delta G_{bind}$ is constant across scaffolds for TL/TLRs with the same conformational properties and varies across scaffolds for TL/TLRs with different conformational properties (Fig. 3 *D* and *E*).

Remarkably, the shapes of the thermodynamic fingerprints for $GAAA/11ntR_{wt}$ and $GAAA/11ntR_{1G/9G/11C}$ are nearly identical (rmsd = 0.28 kcal/mol; r = 0.95; Fig. 4*A*), despite a stability difference of over 2 kcal/mol ($\Delta\Delta G_{bind,avg}$ = 2.35 ± 0.28 kcal/mol; Fig. 4*B*). The similarity of these fingerprints was maintained under lower salt conditions (*SI Appendix*, Fig. S1), although with a weaker correlation coefficient due to lower stability of the $GAAA/11ntR_{1G/9G/11C}$ TL/TLR (*SI Appendix*, Fig. S3). The parallel fingerprints of these TL/TLRs indicate that $GAAA/11ntR_{1G/9G/11C}$ and $GAAA/11ntR_{wt}$ have indistinguishable alignment preferences and 3D conformational properties, even though they differ in sequence at three positions and in tertiary interaction strength by 2.4 kcal/mol.

$GAAA/C7.2_{wt}$ also displayed a thermodynamic fingerprint shape nearly identical to $GAAA/11ntR_{wt}$ (rmsd = 0.23 kcal/mol; r = 0.97; Fig. 4*A*) and, therefore, a constant $\Delta\Delta G_{bind}$ across scaffolds (Fig. 4*B*) with near-identical stabilities, despite the differences in length and sequence of this TLR (Fig. 2*D*).

The value of $\Delta\Delta G_{bind}$ for $GAAA/11ntR_{1G/9G/11C}$ (i.e., the difference in stability between $GAAA/11ntR_{1G/9G/11C}$ and $GAAA/11ntR_{wt}$) was independent of ionic conditions (*SI Appendix*, Fig. S4). In contrast, the magnitude of $\Delta\Delta G_{bind}$ for $GAAA/C7.2_{wt}$ depended on whether K^+ was present. Specifically, $\Delta\Delta G_{bind}$ was nearly the same with 30 mM $MgCl_2$ (–0.28 ± 0.24 kcal/mol) and 5 mM $MgCl_2$ (–0.44 ± 0.23 kcal/mol) but differed with 5 mM $MgCl_2$/150 mM KCl ($\Delta\Delta G_{bind}$ = 1.42 ± 0.31 kcal/mol; *SI Appendix*, Fig. S4). This effect arises from a relative stabilization of $GAAA/11ntR_{wt}$ upon the addition of KCl. Specifically, in the presence of 5 mM $MgCl_2$, the added KCl stabilized $GAAA/11ntR_{wt}$ by 0.5 kcal/mol but destabilized $GAAA/C7.2_{wt}$ by 1.3 kcal/mol (*SI Appendix*, Fig. S5). The destabilization of $GAAA/C7.2_{wt}$ by KCl presumably arises from partial replacement of Mg^{2+} by K^+ in the RNA ion atmosphere, diminishing the net preferential Mg^{2+} stabilization of the compact folded state (43); the preferential K^+ stabilization for the $11ntR_{wt}$ presumably arises from additional energy provided by the specific K^+ binding site seen in $GAAA/11ntR_{wt}$ crystal structures (44, 45).

Despite the effect of K^+ on the magnitude of $\Delta\Delta G_{bind}$ for $GAAA/C7.2_{wt}$, the shape of its thermodynamic fingerprint is independent of ionic conditions (Fig. 4*A* and *SI Appendix*, Fig. S1). Indeed, the shapes of the thermodynamic fingerprints are largely independent of ionic conditions across the TLR library (*SI Appendix*, Fig. S6), even though the $GAAA/11ntR_{wt}$ TL/TLR and likely other TL/TLRs bind K^+. Thus, the presence or absence of a K^+ binding site does not alter the global conformation of a TL/TLR and its interactions with connecting RNA elements, even though there are likely local changes around the K^+ binding site. As noted in *Thermodynamic Fingerprints Report on TL/TLR Conformational Properties*, our experiments across scaffolds report on global conformational ensembles determined by the sets of orientations of helices emanating from the TL/TLR; this property defines how the TL/TLR structurally and energetically interacts with the RNA within which is it embedded and is thus of particular interest in predicting and engineering RNA tertiary structure. Furthermore, the independence of the shape of the thermodynamic fingerprints from ionic conditions supports the reconstitution model of RNA folding (Fig. 1*B*). In this model,

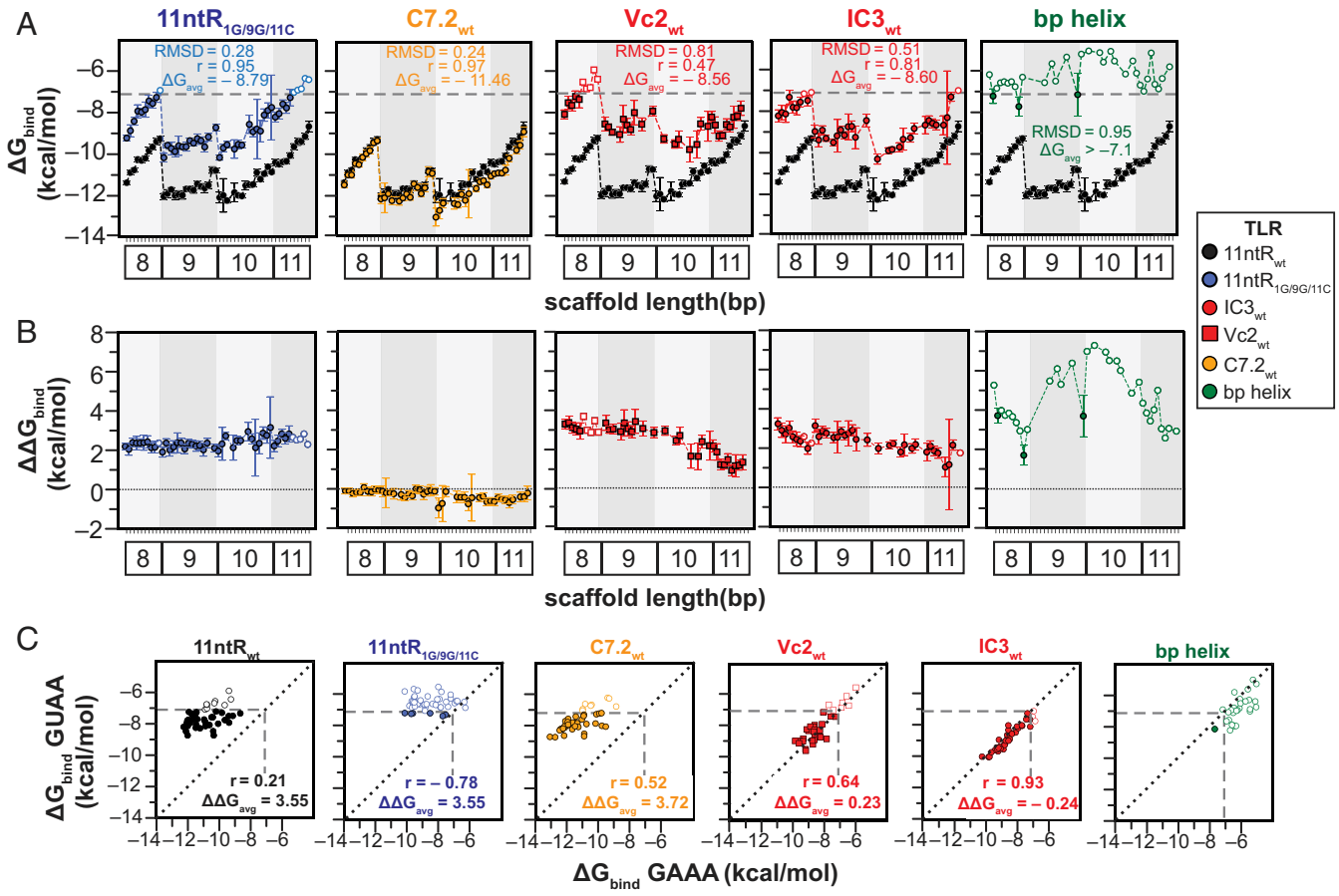

Fig. 4. Thermodynamic and conformational properties of representative TL/TLRs across scaffolds. (*A*) ΔG_{bind} for TLRs binding to GAAA TL across scaffolds. Dashed gray line indicates the threshold of –7.1 kcal/mol. Open symbols are ΔG_{bind} above this threshold. Dashed colored lines serve as guides for the thermodynamic fingerprints. To compare the fingerprints, they were superimposed to that of GAAA/11ntR$_{wt}$ by shifting them by a constant ΔG_{bind} while minimizing rmsd. This minimal rmsd is shown as is the correlation constant, *r*. Values above threshold were not used to calculate the correlation coefficient (*r*). ΔG_{avg} is the median. Values of *r* for TLRs with ΔG_{avg} > –7.1 kcal/mol are not reliable and therefore were not considered. Solution conditions: 89 mM Tris-Borate, pH 8.0, 30 mM $MgCl_2$, 0.01 mg/mL yeast tRNA, and 0.01% Tween 20. Measurements at different ionic conditions are provided in *SI Appendix*, Fig. S3. (*B*) $\Delta\Delta G_{bind}$ relative to GAAA/11ntR$_{wt}$, calculated as the differences in ΔG_{bind} showed in *A*. Open symbols correspond values below stability threshold and are lower limits. (*C*) ΔG_{bind} for GAAA versus GUAA TLs for each representative TLR across its scaffolds. Dashed gray lines indicate a threshold of –7.1 kcal/mol, as above. $\Delta\Delta G_{avg}$ is the average difference in stability between binding to GUAA and GAAA in units of kcal/mol (e.g., on average, constructs with GAAA/11ntR$_{wt}$ are 3.55 kcal/mol more stable than constructs with GUAA/11ntR$_{wt}$).

the energetics associated with electrostatics and ion binding ($\Delta G_{+/-}$; Fig. 1*B*) are separable from the conformational properties of the RNA elements (ΔG_{HJH}; Fig. 1*B*) and the tertiary interactions (ΔG_{tert}; Fig. 1*B*), greatly simplifying the challenge of predicting folding energies (13).

Returning to our representative TL/TLRs, in contrast to the very similar fingerprints for GAAA/11ntR$_{wt}$, GAAA/11ntR$_{1G/9G/11C}$, and GAAA/C7.2$_{wt}$, the GAAA/Vc2$_{wt}$ fingerprint differed considerably (Fig. 4*A*; rmsd = 0.81 kcal/mol; $r = 0.47$; see also *SI Appendix*, Fig. S2). Accordingly, $\Delta\Delta G_{bind}$ for GAAA/Vc2$_{wt}$ was not constant, varying from 0.9 to 3.7 kcal/mol across scaffolds (Fig. 4*B*). In particular, scaffold length differentially affected the probability of aligning GAAA/Vc2$_{wt}$ versus GAAA/11ntR$_{wt}$ (Fig. 4*B*), strongly suggesting differences in the alignment preferences of these TL/TLRs with respect to the bend, rotational angles, and/or number of binding modes of the two TL/TLRs. The fingerprint of GAAA/IC3$_{wt}$ also differed from that of GAAA/11ntR$_{wt}$ (Fig. 4*A*; rmsd = 0.51 kcal/mol; $r = 0.81$; see also *SI Appendix*, Fig. S2); although the differences were more subtle than for GAAA/Vc2$_{wt}$, a clear trend in $\Delta\Delta G_{bind}$ with scaffold length was observed (Fig. 4*B*).

In summary, our comparison of the fingerprints of representative TL/TLRs suggest that GAAA/11ntR$_{wt}$, GAAA/11ntR$_{1G/9G/11C}$, and GAAA/C7.2$_{wt}$ have very similar alignment properties, whereas those of GAAA/Vc2$_{wt}$ and GAAA/IC3$_{wt}$ differ markedly (Fig. 4 *A* and *B*). Nevertheless, close examination of the GAAA/11ntR$_{wt}$ and GAAA/C7.2$_{wt}$ fingerprints revealed, despite the fingerprints overlay with 8- and 9-bp scaffolds, a small but consistent difference with 10- and 11-bp scaffolds (Fig. 4*B*). These subtle distinctions highlight the power of multiple, quantitative thermodynamic comparisons via high-throughput RNA-MaP experiments to reveal even subtle differences in conformational properties. This ability is further highlighted by analyses of ΔG_{bind} averaged across scaffold lengths (*SI Appendix*, Fig. S7).

TL Specificity across TLR Variants. Natural and artificial TL/TLRs have been selected to bind various GNRA TLs (where N is any nt and R is a purine) (21, 28, 30, 35). In particular, 11ntR$_{wt}$ has been shown to have high specificity toward the GAAA TL, while IC3$_{wt}$ binds with similar stability to various GNRA TLs (34, 46). Because our library is mostly comprised of 11ntR$_{wt}$ and IC3$_{wt}$ mutants, we expected to see variation in TL specificity. The high-throughput capabilities of RNA-MaP allowed us to further investigate TL

sequence specificity across different structural contexts, potentially revealing effects on the conformational preferences of the TL/TLRs caused by differences in the sequence of the TL. To do this, we compared ΔG_{bind} values obtained with GAAA and GUAA for the different TLRs across the 50 scaffolds (Fig. 4*C*).

$11ntR_{wt}$, $11ntR_{1G/9G/11C}$, and $C7.2_{wt}$ TLRs displayed much weaker binding for the GUAA TL than for the GAAA TL (Fig. 4*C*). For example, tectoRNA constructs are on average 3.55 kcal/mol more stable with GAAA/$11ntR_{wt}$ than with GUAA/$11ntR_{wt}$ (Fig. 4*C*). In the crystal structures of GAAA/$11ntR_{wt}$, the second A of the GAAA TL makes two sequence-specific hydrogen bonds with A_8, which is part of the core of the TLR, as well as stacking interactions with the A_5 of the AA platform (*SI Appendix*, Fig. S8); thus, the large thermodynamic effect of mutating the TL can be rationalized in terms of disruption to these interactions.

In agreement with the literature, $IC3_{wt}$ and $Vc2_{wt}$ had no significant preference for GUAA or GAAA TL (Fig. 4*C*). The simplest model to account for these results is an absence of sequence-specific interactions between the second residue of the TL and the TLR. Examination of the crystal structure of GAAA/$Vc2_{wt}$ suggests the formation of a single hydrogen bond between the second A of the GAAA TL and A_5 of $Vc2_{wt}$ TLR, as well as base stacking interactions with A_5 (*SI Appendix*, Fig. S8). However, in comparison with GAAA/$11ntR_{wt}$, these interactions appear much less sequence specific, and any nt may be able to be rearranged to form similar interactions. Alternatively, the single hydrogen bond formed between the TL and the TLR may not have a significant net contribution to the stability of the folded TL/TLR and therefore not provide sequence specificity.

Interestingly, the TLRs also differed in the shapes of their thermodynamic fingerprints when the TL sequence was changed. For $IC3_{wt}$, ΔG_{bind} for GAAA and GUAA are highly correlated across scaffolds ($r = 0.93$; Fig. 4*C*). This observation and the similar stabilities for GAAA and GUAA suggest that the conformational ensemble of TL/$IC3_{wt}$ is the same with GAAA and GUAA and that GAAA/ $IC3_{wt}$ and GUAA/ $IC3_{wt}$ can be considered essentially identical motifs with respect to their thermodynamic and conformational properties. For $Vc2_{wt}$, the correlation is weaker ($r = 0.64$; Fig. 4*C*), suggesting that the conformational ensemble of the TL/TLR is moderately affected by the mutation to the TL, although without causing major disruption as the stabilities are similar for GAAA and GUAA.

For $11ntR_{wt}$, $11ntR_{1G/9G/11C}$, and $C7.2_{wt}$, mutation of the TL appears to have a greater effect on the TL/TLR conformational ensembles. In particular, the range of affinities with GUAA was much narrower than that with GAAA (Fig. 4*C*). These results suggest that the A to U substitution broadens the TL/TLR conformational ensemble so that scaffolds with different conformational properties now have similar probabilities of aligning the GUAA to the TLR. The complex network of interactions in the GAAA/$11ntR_{wt}$ motif likely narrows the conformational ensemble, such that the orientation of the TL with respect to the TLR is highly specified to meet the geometrical constraints required to make these interactions. In contrast, the GUAA TL makes fewer interactions with the TLR, and the TL can assume a broader range of orientations with respect to the TLR and still be able to form interactions with similar energies. In other words, having fewer tertiary interactions lowers the sequence specificity and appears to make it more likely to have similar energies across a broader range of conformational states.

This comparative, thermodynamic information and additional information about mutational effects may be valuable in structural modeling and in relating structural models to energetics. Most practically, these data can be used to design RNAs with different tertiary stability and specificity. With these future applications in mind, we provide a list of the 1,493 TLR sequences in our library with binding affinities for GAAA and GUAA TLs and correlation coefficients relative to GAAA/$11ntR_{wt}$ (Supplementary Dataset).

Base-Paired Helices Acting as TLRs. GNRA TLs docking into the minor groove of base-paired helices—so called GNRA/minor groove interactions—are common in nature (21), yet tectoRNA constructs containing a base-paired helix as a TLR (Fig. 4*A*, bp helix, green) bound weakly across all scaffolds, with most ΔG_{bind} values above the –7.1 kcal/mol threshold regardless of whether the TL was GAAA or GUAA (Fig. 4*D*, green). This type of TL/TLRs is commonly observed in large RNA–protein complexes such as the ribosome, while more stable TL/TLRs such as the GAAA/$11ntR_{wt}$ are not observed there (21). Perhaps stable TL/TLRs such as the GAAA/$11ntR_{wt}$ are selected for smaller structured RNAs that have only sparsely distributed tertiary interactions in which more stability is needed from a more limited set of interactions, whereas GNRA/minor groove interactions are sufficiently stable in structural contexts with additional stability provided by nearby tertiary contacts and binding proteins. In addition, GNRA/minor groove interactions are more probable to evolve compared to more complex TL/TLRs, as any helical minor groove can serve as potential TLR (47). Another possible advantage of GNRA/minor groove interactions may arise from their ability to readily form and break and/or conformationally rearrange to allow the formation of a series of functional states.

Classification of TL/TLR Thermodynamic and Conformational Diversity. To investigate the thermodynamic and conformational properties of all 1,493 TLR variants in the library more broadly and in an unbiased fashion, we clustered their thermodynamic fingerprints across the five common scaffolds and two TL sequences. These five scaffolds had a common length (10 bp) but varied in their secondary structure and sequence (Fig. 5*A*). Three were fully base paired ("bp;" Fig. 5*A*), while the other two had unpaired residues ("bulge;" Fig. 5*A*). Based on our previous studies of helices and junctions (26), these scaffolds were expected to probe different conformational spaces and therefore reveal differences and similarities in the conformational ensembles of the TL/TLRs, as discussed in *Thermodynamic Fingerprints Report on TL/TLR Conformational Properties*.

We first calculated average stability in order to separate TLRs that were too weak for this analysis (Fig. 5*B*). TLR sequences that on average destabilized tectoRNA binding below the threshold (Fig. 5*B*; n_{weak} = 387 of 1,493 total) were not included in the clustering analysis, as their thermodynamic fingerprints do not contain enough information to be reliably compared. To cluster the rest of the TLRs based on the shape of their fingerprints rather than on their average stability, each fingerprint was shifted by subtracting its mean value to produce $\Delta\Delta G_{rel}$ values. This is approximately equivalent to putting fingerprints on the same scale so that differences in their shape guide the clustering algorithm. To reduce redundancy in the information content of the dataset, the $\Delta\Delta G_{rel}$ values were then used in a principal component analysis (PCA). The first two principal components (PCs), representing >80% of the variance, were grouped by hierarchical clustering using Euclidean distances and Ward's method (Fig. 5 *C* and *D*). The clustering revealed two major TLR classes (A and B; Fig. 5*D*) that we further classified into five subgroups (1 to 5; Fig. 5 *D*–*F*) based on the variance ratio criterion, which takes into account the between-clusters dispersion and intercluster dispersion (*SI Appendix*, Fig. S9). TLR sequences grouped according to the clustering results are provided in *SI Appendix*, Table S3.

Fig. 5*E* shows the thermodynamic fingerprints grouped according to the clustering analysis. As expected, TLRs with similar fingerprints clustered near one another. TLRs derived from common parents tended to also cluster nearby (Fig. 5*F*). The observation that the naturally observed 11ntR variants clustered closely (Fig. 5*F*) supports that biologically relevant differences and similarities in thermodynamic and conformational behavior have been captured through the clustering. One natural 11ntR with an unusual GC sequence replacing the canonical AA platform was an outlier and

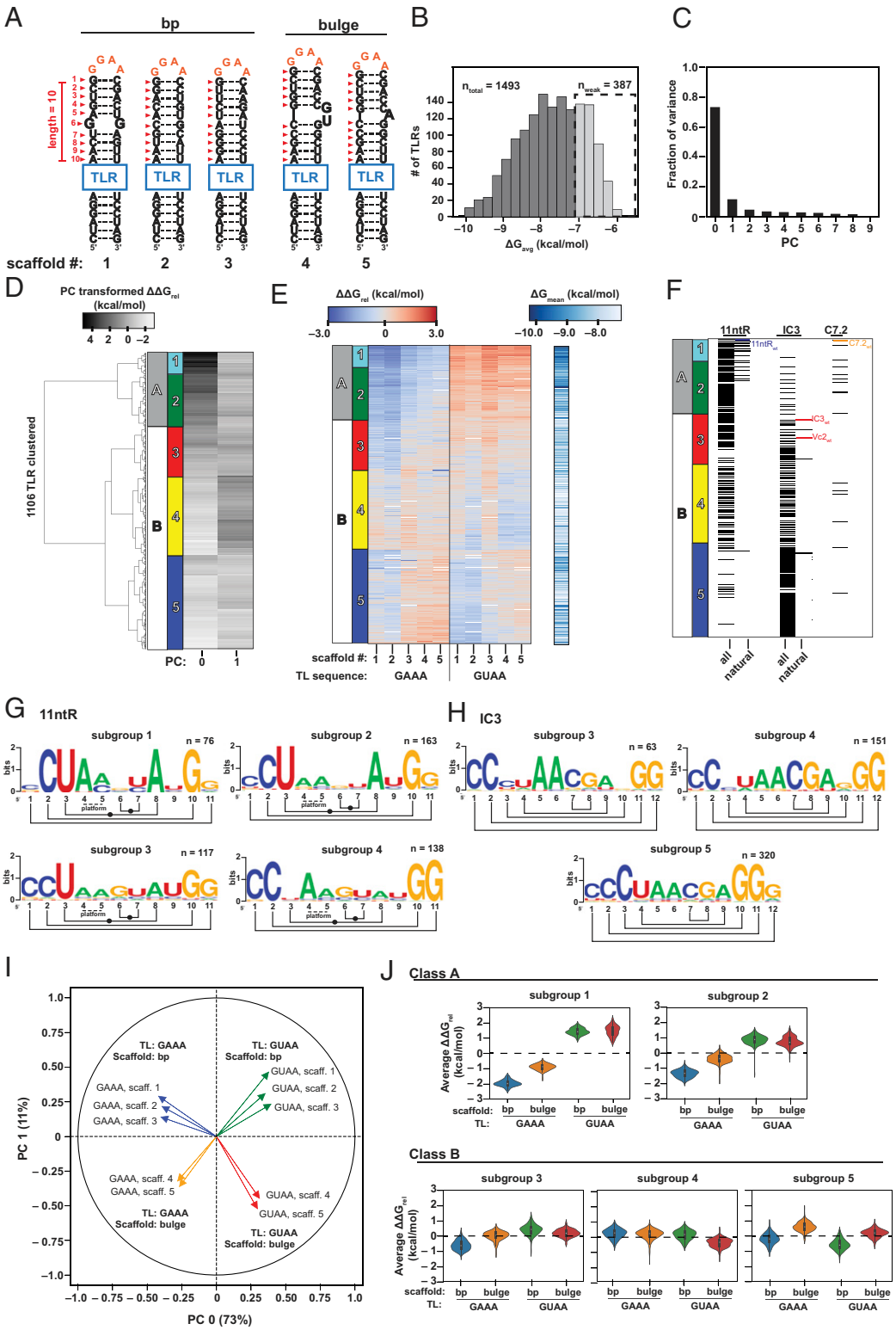

Fig. 5. Overview of TLR thermodynamic behavior. (*A*) Secondary structure and sequence of five scaffolds common to the entire TLR library. (*B*) Distribution of ΔG_{avg} for TLR sequences across five common scaffolds in *A* and two TL sequences. TLRs with $\Delta G_{avg} > -7.1$ kcal/mol ($n_{weak} = 387$) were not included in the clustering analysis. (*C*) Fraction of the variance represented by each of the PCs. (*D*) Hierarchical clustering of the first two PCs using Ward's method implemented in Python. (*E*) Hierarchical clustering of TLRs across five common scaffolds and TL sequences GAAA and GUAA. Prior to clustering, ΔG_{bind} values for each of the TLRs was scaled by ΔG_{mean} to produce $\Delta\Delta G_{rel}$ values. The $\Delta\Delta G_{rel}$ values were used in a PCA (*C* above), and the first two PCs were hierarchically clustered to generate the clustergrams shown. (*F*) Distribution of TLR mutants according to type (i.e., 11ntR, IC3, and C7.2) across clusters. (*G* and *H*) Sequence motifs representing variability among 11ntR (*G*) and IC3 (*H*) variants in the individual subgroups revealed by hierarchical clustering. Sequence motifs were generated using Weblogo (51). Subgroups 1 and 2 were mostly composed of 11ntR variants and subgroup 5 mostly of IC3 variants. (*I*) PCA loading plot showing the influence of the sequence of the binding TL and the secondary structure of the scaffold on the first two PCs. (*J*) Pseudofingerprints showing the average behavior of the TLR subgroups across TL sequence and scaffold secondary structure.

grouped in Class B; it is discussed in the following section. As expected, C7.2$_{wt}$ grouped with 11ntR$_{wt}$ (Fig. 5*F*), consistent with the nearly identical thermodynamic fingerprints of these TLRs across 50 scaffolds (Fig. 4*A*).

To describe commonalities and differences in the sequences of the TLRs in the classes and subgroups, we generated "sequence motifs" for 11ntR and IC3 variants in different subgroups (Fig. 5 *G* and *H*). Class A, containing subgroups 1 ($n_1 = 82$) and 2 ($n_2 = 195$), was composed mostly of 11ntR variants (86%) with only a small fraction of IC3 (5%) and C7.2 (9%) variants. The 11ntR variants in this class have the highly conserved core (composed of bp C_2-G_{10} and U_3-A_8) but varied at other locations (Fig. 5*G*). Whether an 11ntR variant clustered in subgroup 1 or 2 depended strongly on the identity of residue 4—which is part of the AA platform in the canonical 11ntR$_{wt}$—with subgroup 1 variants tending to maintain the conserved A at that location and variants in subgroup 2 having more variability (Fig. 5*G*).

Subgroups 3 ($n_3 = 186$) and 4 ($n_4 = 295$) were the most diverse and contained near-equal numbers of 11ntR (53%) and IC3 (44%) mutants (Fig. 5*F*). 11ntR variants in these subgroups had more variations in the core relative to those in Class A (Fig. 5*G*). For example, 11ntR variants in subgroup 3 have significantly more variability at core position 8 and subgroup 4 at positions 3 and 8 (Fig. 5*G*). Variations at these positions would be expected to greatly perturb the network of tertiary interactions with the GAAA TL observed in crystal structures (*SI Appendix*, Fig. S8). Interestingly, although IC3 and 11ntR variants in subgroups 3 and 4 have some commonalities (Fig. 5 *G* and *H*), such as two CG bp at the "top" of the TLR (C_1-G_{11} and C_2-G_{10} in 11ntR and C_1-G_{12} and C_2-G_{11} in IC3), there are no other obvious sequence similarities, suggesting that the behavior of these subgroups may be less specific and attainable with a larger variety of sequences or sequence types. Subgroup 5 ($n_5 = 348$) was composed mostly of IC3 variants (92%; Fig. 5*F*). In comparison to IC3 variants in subgroups 3 and 4, variants in this subgroup tended to retain the C_3-G_{10} bp (Fig. 5*H*).

Visual inspection of Fig. 5*E* strongly suggested that the sequence of the binding TL (GAAA versus GUAA) was a major determinant of the clustering of the TLRs—with class A showing strong preference for GAAA versus GUAA and class B showing weaker TL sequence specificity. Differences in scaffold preferences among the subgroups were more subtle. To further examine the physical features that determined the observed clustering, we analyzed how the identity of the scaffold and the sequence of the TL influenced the first two PCs (>80% of variance) using a loading plot of the PCA (Fig. 5*I*). Inspection of this plot reveals that PC 0 (describing 73% of the variance across the TLRs) is largely determined by the sequence of the TL (GAAA versus GUAA) and that PC 1 (describing 11% of the variance) is largely dictated by the overall secondary structure of the scaffold (bp scaffolds 1 to 3 versus bulge scaffolds 4 to 5; Fig. 5*A*) (Fig. 5*I*).

Given these observations, we averaged the $\Delta\Delta G_{rel}$ over these four categories (GAAA versus GUAA TLs and bp versus bulge scaffolds) to create simpler pseudofingerprints for our clusters (Fig. 5*J*). As expected from the clustering results (Fig. 5*D*), classes A (Fig. 5*J*, *Top*) and B (Fig. 5*J*, *Bottom*) differ substantially in TL sequence specificity, with TLRs in Class A tending to bind much stronger to GAAA than to GUAA, whereas those in Class B had no strong preference for the sequence of the TL. In the case of 11ntR variants, this change in specificity is likely related to the sequence of residues 3 and 8, as 11ntR variants in Class A (GAAA specific) have a conserved U_3-A_8, while those in Class B (non-TL sequence specific) have more sequence variability at those positions (Fig. 5*G*).

Class A splits into subgroups 1 and 2 based on the magnitude of their TL sequence specificity (Fig. 5*J*, *Top*). This difference is likely determined by the sequence of the dinucleotide platform, as the major difference in sequence between 11ntR in subgroups 1 and 2 is localized to residue 4 (Fig. 5*G*). In terms of scaffold preferences, subgroups 1 and 2 have similar preferences for bp versus bulge scaffolds, suggesting similar GAAA/TLR conformational ensembles throughout these subgroups.

Subgroups within Class B (Fig. 5*J*, *Bottom*) displayed small differences in their scaffold preferences, suggesting variability in the alignment properties of the TLRs. For example, subgroup 3 TLRs exhibited a slight preference for GAAA TLs and bp scaffolds; subgroup 4 TLRs, a preference for GUAA TLs and bulge scaffolds; and subgroup 5 TLRs, a preference for bp scaffolds over bulge regardless of the sequence of the TL. From a practical perspective, in the design of structured RNAs, one can use the sequence motifs (Fig. 5 *G* and *H*) and the groupings provided in *SI Appendix*, Table S2 to choose TLR sequences with similar or distinct thermodynamic and/or conformational properties (e.g., to dial stability up or down without changing conformational properties) and similar or different TL specificities.

GAAA/11ntR Variants in Nature. A central question at the intersection of biology and molecular physics is what molecular properties are selected via evolution. Extensive thermodynamic and conformational data for many related TL/TLRs can provide insights into this question.

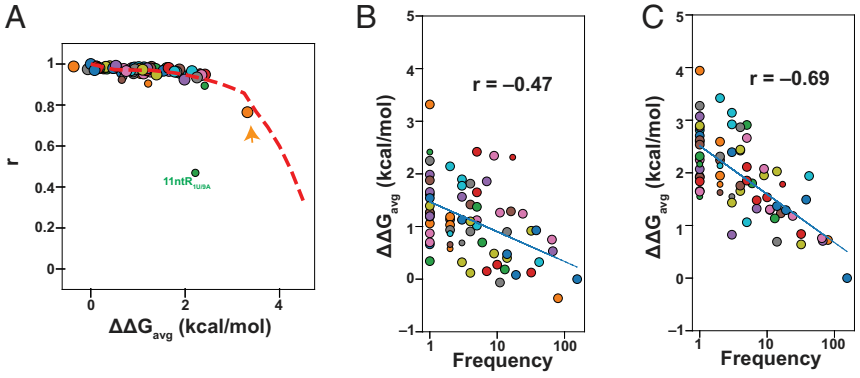

Fig. 6. Thermodynamic analysis of natural GAAA/11ntR variants. (*A*) Average relative stability and Pearson correlation coefficients of the fingerprints. $\Delta\Delta G_{avg}$ values are relative to GAAA/11ntR$_{wt}$. Larger symbols are for variants that were compared across 50 scaffolds and smaller symbols for those compared across five scaffolds. The red dashed line represents the decrease in correlation expected for theoretically identical fingerprints from error as stability decreases (*SI Appendix*, Fig. S2). (*B* and *C*) Correlation between relative stability in 30 mM $MgCl_2$ (*B*) or in 5 mM $MgCl_2$/150 mM KCl (*C*) and the frequency of TLR variants in natural RNA sequences.

The natural 11ntR variants ($n_{natural}$ = 70) grouped together in Class A, suggesting conserved thermodynamic and conformational properties. The exception was $11ntR_{4G/5C/9C}$, which has a GC substituting the AA platform and grouped in Class B (Fig. 5*E*). This substitution is likely to affect the conformation of the TLR and, consistent with this, the outlier is the least stable natural variant (Fig. 6*A*, orange arrow) and is found in only a single natural RNA sequence (*SI Appendix*, Table S1).

To look deeper into similarities and differences among natural variants of the $GAAA/11ntR_{wt}$, we compared their thermodynamic fingerprints (Fig. 6*A*), analyzing 61 of them across 50 scaffolds (Fig. 6*A*, larger symbols) and nine of them across five scaffolds (Fig. 6*A*, smaller symbols). The individual fingerprints are shown in *SI Appendix*, Fig. S10. The 11ntR variants differ in average stability over a range of >2.5 kcal/mol (Fig. 6*A*, $\Delta\Delta G_{avg}$ range), but the shapes of their thermodynamic fingerprints were highly similar (*SI Appendix*, Fig. S10), exhibiting high correlations, as predicted, for identical fingerprints for all variants except for one outlier (Fig. 6*A*). The outlier ($11ntR_{1U/9A}$; Fig. 6*A*) has an A replacing the canonical U bulge at position 9 that may interfere with the neighboring U_3-A_8 core bp, resulting in altered conformational properties. With the exception of this single outlier, the high correlation of the fingerprints and the clustering analysis suggests that the natural GAAA/11ntR variants have a conserved conformational ensemble, even though they differ substantially in their overall tertiary stability.

We observed that the GAAA/11ntR variants that were more stable also tended to be more prevalent. Indeed, the relative stability of the GAAA/11ntR variant strongly correlated with their frequency in natural RNA sequences (r = –0.5; $\Delta\Delta G_{avg}$ versus $\log_{10}$[Frequency]; Fig. 6*B* and *SI Appendix*, Table S1), and this correlation became even stronger in the presence of K^+, which binds to this TL/TLR and is the monovalent cation present intracellularly (r = –0.7; $\Delta\Delta G_{avg}$ versus $\log_{10}$[Frequency]; Fig. 6*C*). As the conformational properties of naturally occurring GAAA/11ntR TL/TLRs are conserved and their frequency corresponds to their stability, it appears that they are selected to maintain a particular conformational preference and to stabilize 3D structures with that conformational preference. It will be interesting to determine whether the outlier is selected or accommodated by other changes in its parent context.

In contrast, the $GAAA/IC3_{wt}$ motif is highly conserved in the IC3 subtype of group I introns, despite being substantially less stable than $GAAA/11ntR_{wt}$ (Fig. 4*B*). This different behavior may arise from a selective pressure other than thermodynamic stability or from high-evolutionary relatedness of the available sequences (*SI Appendix*, Table S2). Future bioinformatic studies are needed to provide a larger library of TL/TLR sequences and further insights into interconnections between conformation, stability, and natural occurrence. Finally, the frequent and widespread occurrence of GNRA/minor-groove interactions in ribosomal and other RNAs may arise, despite these interactions being weak and likely of lower specificity than other TL/TLRs, as these interactions will be more probable to evolve and/or may arise because of a selective pressure for more transient and more flexible interactions.

Conclusions and Implications

As diverse RNAs are built from recurring motifs, understanding motif behavior in terms of their energy landscapes or conformational ensembles can lead to predictive models of RNA-folding thermodynamics and kinetics, as laid out in the RNA reconstitution model (Fig. 1*B*) (13). RNA-MaP has proven to be a powerful tool to provide ensemble-level information about helices and junctions (10, 26), and here, we have extended this approach to address prevalent tertiary motifs, TL/TLRs (Figs. 1*A* and 2 *D* and *E*). Our dissection of these motifs entailed 69,528 thermodynamic measurements over 1,493 different TLR sequences and has allowed us to classify them by differences and similarities in their stabilities and their conformational properties (Figs. 4 and 5). In particular, the "thermodynamic fingerprints" of the TL/TLRs report on the ensemble of conformational states of the helices emanating from the TL/TLR tertiary motif that in turn defines the ability (or probability) of groups on those helices to make additional interactions.

Clustering of the thermodynamic fingerprints revealed similarities within and differences between the TL/TLR classes (Fig. 5*B*). The natural GAAA/11ntR variants clustered together, exhibiting highly similar conformational properties but varied substantially in their stabilities (Figs. 5*B* and 6*A*). The strong correlation between the stability and the natural frequency of this class of TL/TLRs implicates stability as the dominant selection pressure. Intriguingly, other TL/TLR sequences appear to be selected based on properties other than stability. More generally, RNA sequences with similar and dissimilar stability and conformation can be identified via RNA-MaP and inserted into complex RNAs in vitro and in vivo to carry out systematic structure–function studies and to extend these studies into the realm of conformational ensembles and their functional roles.

Our RNA-MaP data indicate similar conformational behavior of GAAA/11nt and GAAA/IC3 TL/TLRs but different behaviors with the GUAA TL present (Fig. 4). These and analogous, high-throughput quantitative data may provide constraints useful for modeling the GAAA/IC3 and other TL/TLRs for which we lack atomic structures.

In addition to building toward a generalizable model of RNA-folding thermodynamics based on the properties of isolated motifs and separable electrostatic contributions to folding (13), our library of characterized TL/TLR sequences can be used to engineer RNAs with predictable, thermodynamic, and conformational behavior and with predictable switching behavior, for example, upon the addition or removal of K^+ ions. For these purposes, we provide binding affinities and TL sequence specificities for the 1,493 TLR sequences in our library (Supplementary Dataset).

Materials and Methods

The preparation of the DNA library, in situ transcription of tectoRNA chip pieces, binding measurements, and data analysis were performed following published protocols (10, 26, 36, 38, 48) and are described in detail in *SI Appendix*. Briefly, a pool of DNA oligonucleotides coding for tectoRNA library was synthesized and PCR amplified. Then, the DNA was sequenced on an Illumina high-throughput sequencing chip. The sequencing chip was installed on a repurposed Ilumina GAIIx sequencer equipped with custom optics for fluorescence excitation and imaging and a fluidics system for the delivery of reagents to the chip surface. RNA was transcribed from the clusters of sequenced DNA on the chip by flowing a series of chemical and enzymatic reagents. The RNA remained associated to the DNA clusters by stalling the RNA polymerase using a streptavidin roadblock. DNA clusters and/or the RNA associated with each DNA cluster were visualized by hybridizing fluorescent oligos to sequences common to the entire library. The tectoRNA binding assays were performed by introducing the desired, fluorescently labeled flow piece at a series of concentrations and allowing it to equilibrate. Changes in the fluorescence of the clusters, caused by the binding of the fluorescent flow piece, were normalized and fit to binding curves (*SI Appendix*).

Data Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank members of the laboratory of D.H. and Hashim Al-Hashimi for discussions and review of the manuscript. This work was supported by the US NIH to D.H. and W.J.G. (Grant R01 GM132899). S.L.B. was supported by the NSF Graduate Research Fellowships Program under Grant DGE-114747.

1. M. B. Warf, J. A. Berglund, Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* **35**, 169–178 (2010).
2. J. L. Rinn, H. Y. Chang, Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
3. T. R. Mercer, J. S. Mattick, Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **20**, 300–307 (2013).
4. T. R. Cech, J. A. Steitz, The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).

5. D. Herschlag, S. Bonilla, N. Bisaria, The story of RNA folding, as told in epochs. *Cold Spring Harb. Perspect. Biol.* **10**, a032433 (2018).
6. I. Tinoco, Jr, C. Bustamante, How RNA folds. *J. Mol. Biol.* **293**, 271–281 (1999).
7. P. Brion, E. Westhof, Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137 (1997).
8. S. R. Eddy, R. Durbin, RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22**, 2079–2088 (1994).
9. D. H. Mathews, Revolutions in RNA secondary structure prediction. *J. Mol. Biol.* **359**, 526–532 (2006).
10. J. D. Yesselman *et al*., Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16847–16855 (2019).
11. L. R. Ganser, M. L. Kelly, D. Herschlag, H. M. Al-Hashimi, The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **20**, 474–489 (2019).
12. L. Salmon, S. Yang, H. M. Al-Hashimi, Advances in the determination of nucleic acid conformational ensembles. *Annu. Rev. Phys. Chem.* **65**, 293–316 (2014).
13. N. Bisaria, M. Greenfeld, C. Limouse, H. Mabuchi, D. Herschlag, Quantitative tests of a reconstitution model for RNA folding thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7688–E7696 (2017).
14. N. B. Leontis, A. Lescoute, E. Westhof, The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16**, 279–287 (2006).
15. R. T. Batey, R. P. Rambo, J. A. Doudna, Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed. Engl.* **38**, 2326–2343 (1999).
16. A. Lescoute, N. B. Leontis, C. Massire, E. Westhof, Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* **33**, 2395–2409 (2005).
17. C. Geary, A. Chworos, E. Verzemnieks, N. R. Voss, L. Jaeger, Composing RNA nanostructures from a syntax of RNA structural modules. *Nano Lett.* **17**, 7095–7101 (2017).
18. W. Grabow, L. Jaeger, RNA modularity for synthetic biology. *F1000Prime Rep.* **5**, 46 (2013).
19. J. D. Yesselman *et al*., Computational design of three-dimensional RNA structure and function. *Nat. Nanotechnol.* **14**, 866–873 (2019).
20. S. Bonilla *et al*., Single-molecule fluorescence reveals commonalities and distinctions among natural and in vitro-selected RNA tertiary motifs in a multistep folding pathway. *J. Am. Chem. Soc.* **139**, 18576–18589 (2017).
21. L. Wu, D. Chai, M. E. Fraser, S. Zimmerly, Structural variation and uniformity among tetraloop-receptor interactions and other loop-helix interactions in RNA crystal structures. *PLoS One* **7**, e49225 (2012).
22. D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz, The kink-turn: A new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
23. A. I. Murchie, J. B. Thomson, F. Walter, D. M. Lilley, Folding of the hairpin ribozyme in its natural conformation achieves close physical proximity of the loops. *Mol. Cell* **1**, 873–881 (1998).
24. J. B. Thomson, D. M. Lilley, The influence of junction conformation on RNA cleavage by the hairpin ribozyme in its natural junction form. *RNA* **5**, 180–187 (1999).
25. N. Bisaria *et al*., Kinetic and thermodynamic framework for P4-P6 RNA reveals tertiary motif modularity and modulation of the folding preferred pathway. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4956–E4965 (2016).
26. S. K. Denny *et al*., High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell* **174**, 377–390.e20 (2018).
27. L. Jaeger, E. Westhof, N. B. Leontis, TectoRNA: Modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.* **29**, 455–463 (2001).
28. C. Geary, S. Baudrey, L. Jaeger, Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res.* **36**, 1138–1152 (2008).
29. M. Costa, F. Michel, Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285 (1995).
30. M. Costa, F. Michel, Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: Comparison with in vivo evolution. *EMBO J.* **16**, 3289–3302 (1997).
31. J. L. Fiore, D. J. Nesbitt, An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* **46**, 223–264 (2013).
32. J. H. Hodak, C. D. Downey, J. L. Fiore, A. Pardi, D. J. Nesbitt, Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10505–10510 (2005). Correction in: *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13351 (2005).
33. E. R. Calkins *et al*., Deducing putative ancestral forms of GNRA/receptor interactions from the ribosome. *Nucleic Acids Res.* **47**, 480–494 (2019).
34. Y. Ikawa, D. Naito, N. Aono, H. Shiraishi, T. Inoue, A conserved motif in group IC3 introns is a new class of GNRA receptor. *Nucleic Acids Res.* **27**, 1859–1865 (1999).
35. P. Zakrevsky *et al*., In vitro selected GUAA tetraloop-binding receptors with structural plasticity and evolvability towards natural RNA structural modules. *Nucleic Acids Res.* **49**, 2289–2305 (2021).
36. J. D. Buenrostro *et al*., Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* **32**, 562–568 (2014).
37. S. K. Denny, W. J. Greenleaf, Linking RNA sequence, structure, and function on massively parallel high-throughput sequencers. *Cold Spring Harb. Perspect. Biol.* **11**, a032300 (2019).
38. R. She *et al*., Comprehensive and quantitative mapping of RNA-protein interactions across a transcribed eukaryotic genome. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3619–3624 (2017).
39. Y. Zhou *et al*., GISSD: Group I Intron Sequence and Structure Database. *Nucleic Acids Res.* **36**, D31–D37 (2008).
40. L. Dai, N. Toor, R. Olson, A. Keeping, S. Zimmerly, Database for mobile group II introns. *Nucleic Acids Res.* **31**, 424–426 (2003).
41. J. W. Brown, The Ribonuclease P Database. *Nucleic Acids Res.* **27**, 314 (1999).
42. N. Sudarsan *et al*., Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411–413 (2008).
43. R. Shiman, D. E. Draper, Stabilization of RNA tertiary structure by monovalent cations. *J. Mol. Biol.* **302**, 79–91 (2000).
44. S. Basu *et al*., A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **5**, 986–992 (1998).
45. N. Bisaria, D. Herschlag, Probing the kinetic and thermodynamic consequences of the tetraloop/tetraloop receptor monovalent ion-binding site in P4-P6 RNA by smFRET. *Biochem. Soc. Trans.* **43**, 172–178 (2015).
46. Y. Ikawa, K. Nohmi, S. Atsumi, H. Shiraishi, T. Inoue, A comparative study on two GNRA-tetraloop receptors: 11-nt and IC3 motifs. *J. Biochem.* **130**, 251–255 (2001).
47. K. Bokov, S. V. Steinberg, A hierarchical model for evolution of 23S ribosomal RNA. *Nature* **457**, 977–980 (2009).
48. I. Jarmoskaite *et al*., A quantitative and predictive model for RNA binding by human pumilio proteins. *Mol. Cell* **74**, 966–981.e18 (2019).
49. J. H. Davis *et al*., RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex. *J. Mol. Biol.* **351**, 371–382 (2005). Correction in: *J. Mol. Biol.* **360**, 742 (2006).
50. P. Sripakdeevong, W. Kladwang, R. Das, An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20573–20578 (2011).
51. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).