1138–1152 *Nucleic Acids Research, 2008, Vol. 36, No. 4* *Published online 23 December 2007*
doi:10.1093/nar/gkm1048

Comprehensive features of natural and *in vitro* selected GNRA tetraloop-binding receptors

Cody Geary, Stéphanie Baudrey and Luc Jaeger*

Department of Chemistry and Biochemistry, Biomolecular Science and Engineering Program, University of California at Santa Barbara, Santa Barbara, CA 93106-9510, USA

Received September 14, 2007; Revised October 24, 2007; Accepted November 5, 2007

ABSTRACT

Specific recognitions of GNRA tetraloops by small helical receptors are among the most widespread long-range packing interactions in large ribozymes. However, in contrast to GYRA and GAAA tetraloops, very few GNRA/receptor interactions have yet been identified to involve GGAA tetraloops in nature. A novel *in vitro* selection scheme based on a rigid self-assembling tectoRNA scaffold designed for isolation of intermolecular interactions with A-minor motifs has yielded new GGAA tetraloop-binding receptors with affinity in the nanomolar range. One of the selected receptors is a novel 12 nt RNA motif, (CCUGUG … AUCUGG), that recognizes GGAA tetraloop hairpin with a remarkable specificity and affinity. Its physical and chemical characteristics are comparable to those of the well-studied '11nt' GAAA tetraloop receptor motif. A second less specific motif (CCCAGCCC … GAUA GGG) binds GGRA tetraloops and appears to be related to group IC3 tetraloop receptors. Mutational, thermodynamic and comparative structural analysis suggests that natural and *in vitro* selected GNRA receptors can essentially be grouped in two major classes of GNRA binders. New insights about the evolution, recognition and structural modularity of GNRA and A-minor RNA–RNA interactions are proposed.

INTRODUCTION

Structured RNA molecules use intricate networks of tertiary interactions to promote the 3D packing of secondary structure elements defined by RNA helices (1,2). These interactions, corresponding to modular and recurrent tertiary motifs, are specified by conserved sets of nucleotides (nts) coding for non-canonical base pairings or specific backbone conformations (3,4). Long-range interactions occurring between sequence positions often separated by several hundreds of nts are among the most critical for the RNA conformational search. While the structural principles of some of them were initially identified by phylogenetic and experimental approaches (5–10), recent X-ray and NMR structures have provided detailed pictures allowing rationalization of their sequence conservation (11–14). Long-range interactions can be divided into four primary categories: classic Watson–Crick (WC) base pairs (bp) leading to pseudoknots and kissing loops (15,16), T-loop-mediated interactions (17,18), helix–helix interactions (19–21) and A-minor motif interactions (6,22–24). Of these, A-minor interactions are the most abundant of all identified long-range interactions in the ribosome, involving no less than two-thirds of the phylogenetically conserved (>95%) adenines (22,23).

The A-minor interaction typically occurs between stacked adenines and the shallow minor groove of a helical receptor composed of two WC bps. The most common, type I, A-minor-binding pattern forms up to four hydrogen bonds between an adenine and the *N*1-C2-*N*3 edges of a WC bp, interacting in a *trans* Shallow-Groove:Shallow-Groove (SG:SG) bp (22,23). In many examples, A-minor interactions have been observed to be part of larger motifs involving GNRA and GNRA-like loop motifs (N stands for any base, R stands for purine) (10,24). GNRA/helix interactions are frequently found in Group I and Group II introns (6,11,25–27), RNase P (9,14,28) as well as ribosomal RNAs (2,12,29,30). For example, in the ribosome 50% of GNRA terminal loops are involved in A-minor interactions with helical receptors (12). While the helix is the smallest receptor for GNRA, more complex structured receptors may generate additional tertiary contacts leading to increased thermostability and higher selectivity towards one GNRA tetraloop.

Among the most recurrent GNRA/receptor interactions, the most widely studied is the GAAA/11nt receptor interaction (6,7,26,31–33). This interaction is remarkable

*To whom correspondence should be addressed. Tel: +1 8058933628; Fax: +1 8058934120; Email: Jaeger@chem.ucsb.edu
Present address:
Stéphanie Baudrey, Architecture et Réactivité de l'ARN, Université Louis Pasteur de Strasbourg, CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg, France.

because it is the strongest and most selective A-minor loop/receptor interaction yet characterized. Furthermore, the GAAA/11nt receptor interaction has proven to be very robust, as it folds properly outside of its natural context and has been characterized by NMR and EPR in isolation (34,35) and in the context of a rationally designed 30 kDa tectoRNA (36,37). The 11nt receptor motif is known to undergo a structural rearrangement upon the binding of its cognate GAAA target (34,38). It has been proposed that the GAAA/11nt receptor interactions act as a clamp for stabilizing large RNAs after they have folded (33). Studies performed on the P4/P6 domain of group I introns are in good agreement with this hypothesis, as the GAAA/11nt receptor motif contributes more to the thermodynamic stability of the native folded state than to the kinetics of folding (33). However, despite these advances, kinetic and thermodynamic data for RNA tertiary interactions are still challenging to obtain (33,38). Furthermore, knowledge about the specificity of GNRA recognition and the thermodynamics of GNRA/receptor interactions remains fragmentary.

In order to further understand the principles governing RNA tertiary interactions, Costa and Michel (7) devised an *in vitro* selection system based on the mutual recognition of a group I ribozyme and its substrate, to isolate receptors that recognize GUGA and GAAA tetraloops. While many of them were variants of previously identified receptor motifs, they also identified new receptor sequences with good binding affinity and new patterns of recognition for GNRA (7). One of the new receptors, called C7.34, preferentially recognizes GGAA and has sequence similarities with the natural group IC3 receptor, which is apparently much more permissive for binding different GNRAs (31,32).

This suggests that highly specific receptors for GNRA other than GYRA (Y stands for pyrimidine) and GAAA tetraloops may likely exist. Identification of new GNRA receptor sequences can thus facilitate the investigation of the relationship between structure and function in RNA. While these interactions may have diverse nt composition, they are likely to have similar structural organization following rules that can be characterized. For example, the bifurcated A:G *trans* Hoogsteen: Shallow-Groove (HG:SG) closing bp is conserved in all GNRA tetraloops (39). The accessible moiety of adenine from this conserved non-canonical pairing is involved in the formation of the A-minor motif. By comparing natural and *in vitro* selected receptors with different specificities for the second and third GNRA positions we might deduce additional structural rules for GNRA receptors.

Herein we report the selection of novel receptors that bind tightly and specifically to GGAA tetraloops, and provide insight about the way GNRA tetraloops might be recognized to form stable long-range interactions involving A-minor motifs. The design of our selection system takes advantage of the self-assembling tectoRNA scaffolding based on bimolecular GAAA/11nt receptor interactions that we have previously used as building blocks for nano-construction (40–42). Thus, in contrast to the Costa and Michel selection, our selection scheme is directly based on RNA self-assembly properties, irrespective of catalytic activity. After seven rounds of selection, we have identified and characterized three classes of receptors. Our molecular system has proved to be particularly suitable for comparing the thermodynamics of these new receptors with known GNRA/receptor interactions (40,43). Additionally, thermodynamic analysis of mutants has led to new insights about the structural modularity of GNRA receptors. Based on our findings, we discuss the implications of how RNA structures might be constrained during their evolution.

MATERIALS AND METHODS

Design, synthesis and preparation of tectoRNA molecules

TectoRNA constructs are based on scaffoldings that take advantage of bi-molecular receptor/GNRA interactions (40,41,43) and for which atomic model structures are presently available (36,40). The design of each tectoRNA was checked for proper folding with the program mFold (44) to maximize the stability of its secondary structure while minimizing the occurrence of alternative secondary structure folds.

RNA molecules were prepared by *in vitro* run-off transcription of PCR-generated DNA templates using home-made T7 RNA polymerase as described (40,43,45). Antisense and primer DNA sequences were ordered from IDT. After purification on denaturing polyacrylamide gel electrophoresis (PAGE), RNA molecules were dissolved in water as described (40,45). RNA constructs were labeled at their 3′ end with [5′-^{32}P]pCp as described (40,45).

In vitro selection experiment

The DNA library was generated by PCR amplification of a 73-mer DNA template Tmp (5′-AGAGAAAACC TACCAGC(N9)AAGAACCAAGTTTCCCCGGTTCTT (N8)GCCAGTAGATTTCCC-3′) with the forward primer Fwd (5′-TTCTAATACGACTCACTATAGGGA AATCTACTGGC-3′) which includes the T7 promoter, and the reverse primer Rev (5′-AGAGAAAACCTAC CAGC-3′). After run-off T7 transcription as described (40,43,45) (step 1, Figure S1B, Supplementary Data), the corresponding ^{32}P body-labeled RNA library (200 pmol) was denatured at 95°C for 1 min, cooled on ice for 2 min and annealed at 30°C for 2 min prior incubation at 10°C for 30 min in KCl/Mg^{2+} buffer [89 mM Tris–borate pH 8.3, 50 mM KCl, 15 mM $Mg(OAc)_2$, 10 μM (RNA), final concentration]. After addition of 1/10 vol of gel loading buffer (KCl/Mg^{2+} buffer with 50% glycerol, 0.01% bromophenol blue, 0.01% xylene cyanol), the RNA library was purified on native 7% (29:1) polyacrylamide gels in Mg^{2+} buffer to eliminate molecules forming RNA complexes in absence of target (step 2a, Figure S1B). For the two last rounds of selections, the RNA pool was submitted to an additional counter selection step by being incubated in presence of GYRA and GAGA probe targets (1 μM each) at 10°C for 30 min in KCl/Mg^{2+} buffer (step 2b, Figure S1B). Only RNA pool fractions that remained as monomers during the counter selection steps

were purified, eluted and ethanol precipitated. For the selection step (step 3, Figure S1B), the remaining molecules at 1 μM were mixed with GGAA target (1 μM), folded at 10°C for 30 min in presence of KCl/Mg^{2+} buffer as described above, and partitioned on native PAGE in Mg^{2+} buffer at 10°C. The RNA pool fractions bound to the GGAA target and that migrated at the same position as a heterodimer molecular control of similar size, were then purified, eluted and ethanol precipitated. These selected RNAs were then reverse transcribed at 48°C with superscript II (Gibco/brl) and PCR amplified with *Taq* DNA polymerase in presence of the rev and fwd primers to generate the corresponding DNA template (step 4, Figure S1B). After purification using the QiaQuick PCR purification kit (Qiagen), the resulting enriched DNA pool was transcribed into RNA for additional round of selection. After seven rounds of selection, the final DNA library was cloned using the Topo TA cloning kit (Invitrogen) and 43 individual clones were sequenced.

RNA self-assembly, native PAGE and (Pb^{2+})-induced cleavage

Dissociation constant (K_d) for homodimer constructs were determined as described (40). For heterodimer constructs, equimolar amounts of each RNA monomer at various concentrations (0.5–20 μM final) were mixed in water, denatured at 95°C for 1 min, cooled on ice for 2 min and annealed at 30°C for 5 min prior assembly at 30°C for 30 min after addition of Mg^{2+} buffer [89 mM Tris–borate pH 8.3, 15 mM $Mg(OAc)_2$ final concentration] and incubation at 10°C for 15 min prior native gel analysis. To monitor and analyze RNA migration on native 7% (29:1) polyacrylamide gels in Mg^{2+} buffer at 10°C, one of the monomers contained a fixed amount of 3′ end [^{32}P] pCp-labeled RNA (~0.25–0.5 nM final). Monomer and dimer bands were quantified with Imagequant and dimer formation was correlated with RNA concentration. K_ds were determined as the concentration at which half the RNA molecules are dimerized. Lead (Pb^{2+})-induced cleavage experiments were performed as described (40). For a detailed version of the protocols used in the article, see Supplementary Data.

RESULTS

In vitro selection of novel RNA–RNA interactions

The *in vitro* selection system is based on the tectoRNA scaffolding shown in Figure 1B (40,45). It consists of a partially randomized RNA library able to form heterodimers by assembling to a constant RNA target (Figure S1 in Supplementary Data). The selection library (2×10^{10} different sequences) was designed such that each molecule contains an internal loop of 17 randomized nts located 11 bp from a GAAA tetraloop (Figure S1A and B). The 11-nt receptor within the probe target, R(11nt):L(GGAA) (see Figure 2B for explanation of this notation), is expected to act as an anchor for GAAA tetraloop from the RNA pool to facilitate the selection of GGAA-binding motifs within a structural context similar to known receptor/GNRA-tetraloop interactions (Figure 1C).

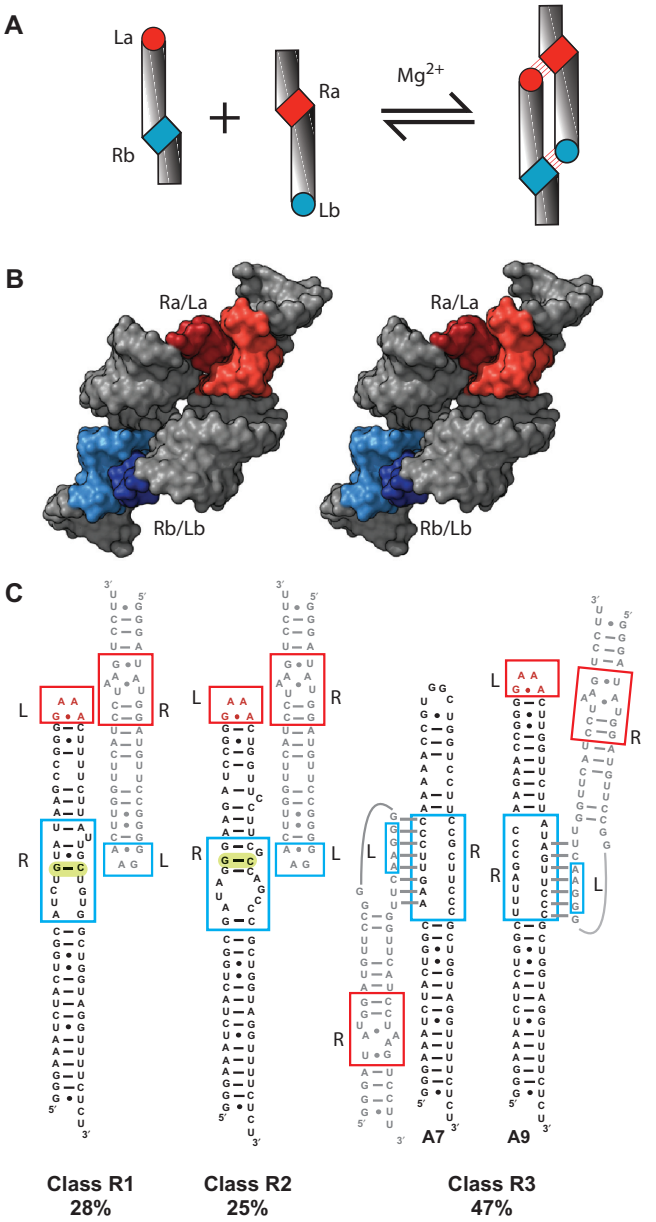

Figure 1. Principle of tectoRNA assembly and selected GGAA tectoRNA binders. (**A**) TectoRNA heterodimer assembly: Ra and Rb stand for the receptors of La and Lb tetraloops, respectively. Upon addition of Mg^{2+}, concomitant formation of the two loop/receptor interactions induces tectoRNA assembly. (**B**) 3D stereo image of a tectoRNA dimer according to the NMR structure obtained by Davis *et al.* (36). Molecular graphics were rendered using the PyMOL molecular graphics software package (67). (**C**) Secondary structure diagrams for class-R1, class-R2 and class-R3 binders isolated by SELEX. As class-R3 contained numerous different sequences, the two shown are representative of the pool. While class-R1 and class-R2 binders can fold and assemble to the GGAA probe according to the tectoRNA design principle shown in A, class-R3 is not likely to form loop/receptor interactions. Blue boxes are for the newly selected receptor/GGAA region. Red boxes are for the fixed 11nt/GAAA interaction. The G:C bp corresponding to 11 bp separation from the tetraloop is highlighted in green.

Selection was performed by native gel electrophoresis based on the differential electrophoretic mobility of dimer complexes versus monomers in presence of 15 mM Mg^{2+}, and 10°C (Figure S1B). For the first five rounds of

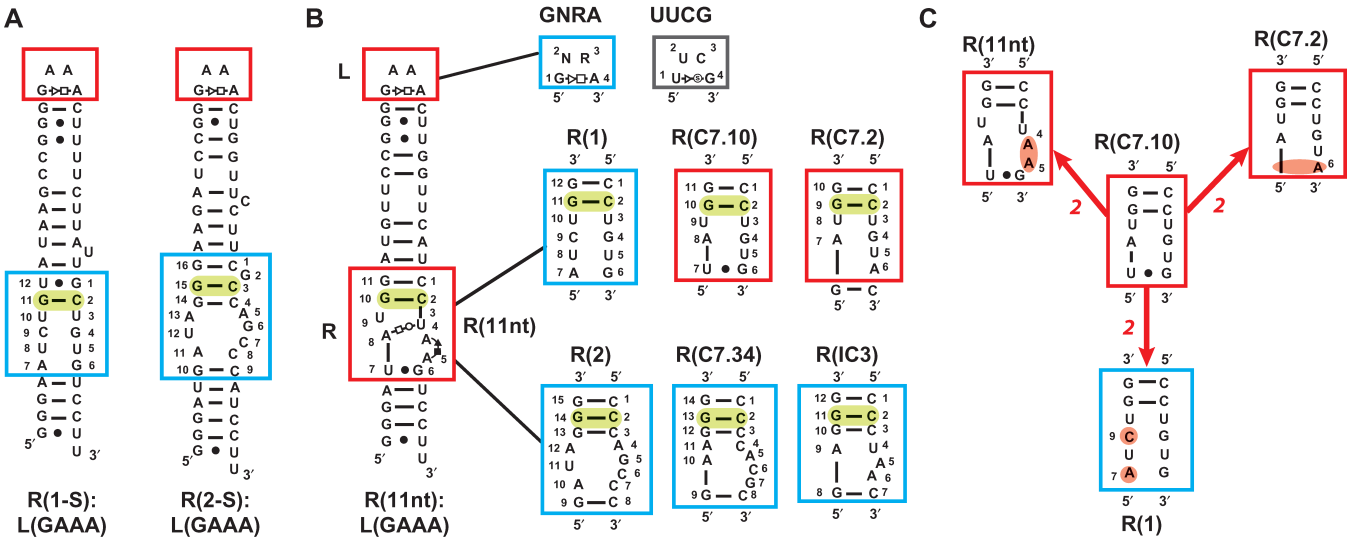

Figure 2. Secondary structure diagrams and nomenclature of tectoRNA molecules reported (See also Supplementary Data) (**A**) R(1-S) and R(2-S), the two shortened versions of class-R1 and class-R2 tectoRNAs, respectively. (**B**) Diagram of the anatomy of studied tectoRNA molecules. All constructs are built onto the same helical core, with only differences occurring in the loop region, designated as L(GNRA), and the receptor region, designated as R(receptor). TectoRNA constructs are named R(receptor):L(GNRA) with respect to the nature of their constitutive loop and receptors [e.g. the tectoRNA shown containing the 11nt receptor and the GAAA tetraloop is named R(11nt):L(GAAA)]. GAAA receptors are outlined in red. Other receptors are outlined in blue. The receptor G:C bp located at a distance of 11 bp form the GNRA tetraloop is highlighted in green. Nt positions within each receptor motif are numbered clockwise starting from the top right. (**C**) A possible mutational pathway links the sequence spaces associated to the selected R(1-S) and R(11nt) in six mutational steps. R(1) and R(C7.10) are two stable receptor/loop intermediates. Nts highlighted in red and orange are nt deletion and nt substitution, respectively.

selection-amplification, the RNA library was subjected to a counter-selection step to eliminate molecules that form RNA–RNA complexes in absence of GGAA target. The remaining molecules were then subjected to positive selection by being mixed with GGAA target and isolating RNA–RNA complexes that migrate more slowly than monomers on native polyacrylamide gels. Once separated, the population enriched in GGAA RNA binders was amplified by RT–PCR using primers complementary to the extended helical stem at the 5′–3′ end of the RNA library as described in the Materials and Methods section (Figure S1B). After initial pool enrichment, the pool was subjected to two additional selection-amplification rounds, each comprising an additional negative selection step to reduce the RNA pool affinity for other GNRA tetraloops. From the 43 individuals that were cloned and sequenced from the final selection pool, 21 different RNA sequences could be identified and grouped into three classes of different molecules (class-R1, 28%; class-R2, 25%; class-R3, 47%) (Figure S2 in Supplementary Data) (46,47).

Sequence comparison of selected receptors with known RNA–RNA interactions

The 19 different sequences comprising class-R3 (Figure S2) contain on the 5′ or/and 3′ side of the receptor region, five to eight contiguous nts that can form classic WC bps with five to eight contiguous nts within the sequence 5′-CGGGGGAACUUG-3′ of the GGAA probe (Figure 1). For all class-R3 binders, disruption of the structure of the GGAA tetraloop through partial unfolding of the stem is necessary to generate stable apical-loop/internal-loop complexes of up to 8 bp (Figure 1C). Interestingly, similar RNA–RNA interactions mediated by classic WC bps have been previously selected (16). In presence of the GGAA probe, class-R3 tectoRNAs assemble into RNA–RNA complexes that migrate slower on native gels than the control tectoRNA dimer (data not shown). This suggests that R3 binders form less-compact RNA–RNA complexes than those expected for tectoRNA dimers assembling through non-WC loop/receptor interactions. Apparently, point mutations and deletions introduced by PCR amplification during *in vitro* selection disrupt the structure of the GAAA terminal stem-loop of several class-R3 molecules without affecting their binding properties (data not shown). These observations suggest that most class R3 binders assemble with the GGAA probe without taking advantage of the GAAA/11nt loop–receptor anchor.

By contrast, class-R1 and class-R2 of GGAA binders are each represented by multiple copies of a unique RNA sequence signature without obvious WC base complementarity to the GGAA probe (Figure 1C, and Figure S2 in Supplementary Data). They assemble with the GGAA probe to form complexes that migrate as compact dimers like the control tectoRNA dimer (data not shown). Despite the presence of several mutations introduced in the constant stem-loop region during *in vitro* selection, class-R1 and class-R2 sequences can still fold into a GAAA stem-loop structure able to bind to the receptor probe. However, when nucleotide substitutions known to disrupt the formation of the GAAA tetraloop/11nt receptor interaction are introduced within class-R1 and

class-R2 molecules, these mutations prevent assembly to the GGAA probe. Additionally, both classes R1 and R2 present a conserved G:C bp located 11 bp from the tetraloop, the spacing determined to be ideal for favoring tectoRNA dimer assembly through loop-receptor interactions with A-minor motifs (Figures 1C and 2A) (40,41,45). All of these observations strongly suggest that both class R1 and R2 receptors assemble into compact GNRA/receptor dimers according to our original scaffolding design.

Class-R2 presents the sequence signature of previously selected class II GAAA binders, in which a (CCC... GGG) consensus sequence is followed by an asymmetric internal loop (7) (Figure 2B). This consensus sequence is also observed in the natural GAAA receptor R(IC3) from group IC3 introns (48) (Figure 2B). The receptor region of class-R2 shares 10 nt in common with the previously reported receptor R(C7.34), the 4 nt in variation being localized within the asymmetric internal loop.

Class-R1 presents some sequence similarities with the previously selected class I-B L(GAAA)-binding receptor R(C7.10) to which the 11nt motif belongs (7) (Figure 2B). In fact, based on sequence comparison receptors of class-R1 differ from R(C7.10) at only 4 nt positions. R(C7.10) is occasionally found in natural molecules (9) and differs from R(11nt) by only 2 nt positions (Figure 2C).

While each class of selected binding receptors display some degree of selectivity for L(GGAA) versus other L(GNRA) tetraloops, class-R3 receptors do not use GNRA/receptor interactions with A-minor motifs. Therefore we decided not to investigate class-R3 receptors further. Classes-R1 and R2 were subjected to additional analysis in order to better understand the structural constraints pertaining to these two classes of L(GGAA) receptors.

Optimization of the selected GGAA binders by rational design

First, the extended 15 bp stem at the 5′–3′ ends of the class-R1 and class-R2 molecules was reduced to 4 or 5 bp to generate the constructs R(1-S):L(GAAA) and R(2-S):L(GAAA) (Figure 2A). These shortened versions still assemble with the GGAA probe, R(11nt):L(GGAA), without any loss of binding affinity. At 15 mM Mg^{2+} and 10°C, apparent dissociation equilibrium constants (K_d) for the R(1-S) and R2(1-S) probe complexes are 196 and 100 nM, respectively.

During the seven rounds of SELEX (Figure S1), point mutations and deletions were acquired within the constant region of the tectoRNA scaffolding, potentially resulting in sub-optimal GGAA binder structures in the final pool. In an attempt to reverse any lost binding affinity, we reverted some of these mutations back to the sequence corresponding to the initial scaffolding context (Figure S3 in Supplementary Data).

The two improved class-R1 and R2 receptors were respectively named R(1) and R(2), and are now used as references in our structural study. R(1):L(GAAA) and R(2):L(GAAA) constructs, assemble with the R(11nt):L(GGAA) probe with Kds of 4.4 and 12.5 nM, respectively. After optimization, we note that the GGAA receptor R(1) is two mutational events away from the GAAA receptor R(C7.10), which is in turn only two mutational events away from R(11nt) (Figure 2C). This strongly suggests that R(1), R(11nt) and R(C7.10) likely belong to the same class of receptor. Furthermore, the 2 nt positions that differ between the sequence signature of R(1) and R(C7.10) may be responsible for the selectivity within this class of receptor.

	Tetraloop (Kd.app nM)		
Receptor	**L(GGAA)**	**L(GAAA)**	**L(GUAA)**
R(1)	40	> 100000	> 100000
R(2)	200	> 100000	n.d.
R(C7.34)	40	2000	n.d.
R(IC3)	5000	1500	800
R(C7.10)	> 100000	20	> 100000
R(C7.2)	> 100000	4	> 100000
R(11nt)	> 100000	3	> 100000

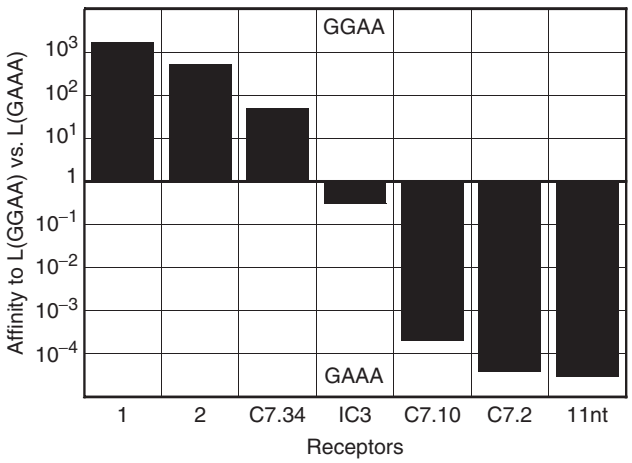

Figure 3. TectoRNA self-assembly into homodimers. **(Top)** Table with apparent equilibrium dissociation constants ($K_{d,app}$) for various homodimer constructs of the type: R(receptor):L(GAAA), R(receptor):L(GGAA) and R(receptor):L(GUAA). K_d values are in nM. **(Bottom)** Log-scale graph for the receptor affinity to L(GGAA) versus L(GAAA).

Using homodimers and heterodimers to measure the selectivity of GNRA/receptor interactions

We generated a series of homodimeric tectoRNAs for the purpose of comparing the stability and selectivity of R(1) and R(2) loop/receptor interactions with other natural or selected receptors [R(11nt), R(IC3), R(C7.34), R(C7.2) and R(C7.10); see Figure 2]. These receptors were tested as homodimers with the loops L(GAAA), L(GGAA) and L(GUAA). Both R(1) and R(2) are selective for L(GGAA) over L(GAAA) by at least two to three orders of magnitude, with apparent K_ds to L(GGAA) of 40 and 200 nM, respectively (Figure 3). The discriminating power of GGAA versus GAAA for R(1) and R(2) is almost of the same magnitude as the discriminating power of GAAA versus GGAA for R(11nt), R(C7.10) and R(C7.2) (Figure 3). By contrast, the receptors R(C7.34)

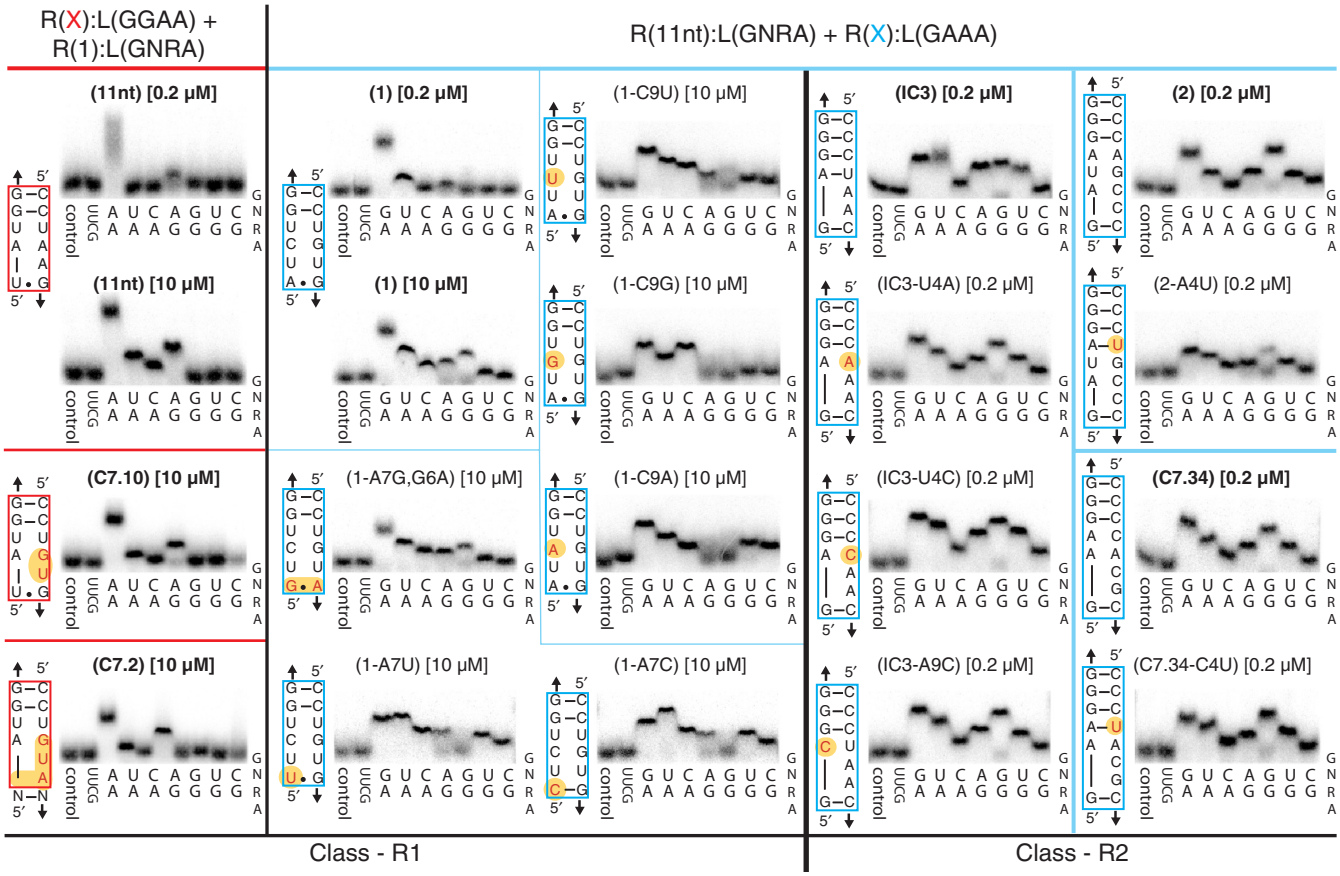

Figure 4. Patterns for the specificity of recognition of GNRA tetraloops visualized on native PAGE for various heterodimer tectoRNA constructs. R(X):L(GGAA) or R(X):L(GAAA) constructs are probed for the specificity of GNRA recognition of their receptor R(X), using a fixed and highly selective loop/receptor interaction involving either GGAA or GAAA tetraloops. With the exception of R(1):L(GGAA) and (11nt):L(GAAA), none of the other R(1):L(GNRA) or R(11nt):L(GNRA) constructs are able to self-dimerize, and can thus be used as probes to estimate the specificity of recognition of other receptor constructs. Consequently, while R(11nt):L(GGAA), R(C7.10):L(GGAA) and R(C7.2):L(GGAA) can be probed with R(1):L(GNRA) constructs, R(IC3):GAAA and other weaker receptors can be probed with R(11nt):L(GNRA) without self-dimerizing at the concentration of RNA tested. For each gel shown, the position UUCG corresponds to the R(11nt):L(UUCG) construct that migrates as a monomer. The control lane corresponds to R(X):LGGAA or R(X):LGAAA constructs that are tested in the absence of any probe and should not self-dimerize at the concentration of RNA tested. All other lanes are labeled according to the letters **NR** corresponding to the probes, R(1):L(G**NR**A) or R(11nt):L(G**NR**A), tested in the experiment. Positions that are mutated from the parent receptor are indicated in red with orange background. Numbering within each motif is according to Figure 2. Experiments were carried out in the presence of 15 mM $Mg(OAc)_2$ and 10°C, at tectoRNA constructs and probes of either 200 nM or 10 μM (indicated above each pattern), as described in the Materials and Methods section.

and R(IC3) are less specific. The natural receptor R(IC3) from group IC3 introns (31) has the lowest specificity, with a somewhat greater affinity to GUAA versus GAAA and GGAA (Figure 3). Receptor R(C7.34) is able to distinguish GGAA from GAAA by a factor of 50.

Using a larger set of self-dimers of the form R(11nt):L(GNRA) and R(1):L(GNRA), we found that both R(11nt) and R(1) are highly specific to their cognate tetraloop. All of the apparent K_ds of non-cognate loop/receptor interactions for these two receptors are above 100 μM (data not shown), making their precise measurement not possible. While probing tetraloop specificity with homodimers works well when the receptor binds the probe tetraloop strongly, the sensitivity falls off rapidly for testing weaker tetraloop/receptor interactions.

By using heterodimers locking one of the two loop/receptors, it is possible to greatly improve the measurement of non-cognate loop/receptor interactions. With both R(11nt) and R(1) receptors, we built multiple heterodimer tectoRNA systems holding either R(1)/GGAA or R(11nt)/GAAA constant as an anchor. These heterodimer-forming constructs were tested against a constant set of probes to measure by PAGE their pattern of specificity for GNRA tetraloops. Two systems were used, either R(GAAA receptor):L(GGAA) probed with R(11nt):L(GNRA) or R(GGAA receptor):L(GAAA) probed with R(1):L(GNRA) (Figure 4 and Table 1). Because one interaction is held constant in each case it is possible to test for only seven out of eight GNRA for each receptor.

The seven receptors tested can be divided in two categories based on their patterns of specificity for GNRA recognition (Figure 4). At 200 nM RNA concentration, the receptors R(1), R(11nt), R(C7.2) and R(C7.10) are all highly specific to a single cognate target, while R(2), R(IC3) and R(C7.34) bind multiple tetraloops (Table 1).

Table 1. Comparison of the specificity of recognition of GNRA tetraloops by various natural and selected receptors

TectoRNAs with fixed llnt/GAAA interaction	Receptor Class	R(llnt): L(GGAA)	R(llnt): L(GGGA)	R(llnt): L(GUAA)	R(llnt): L(GUGA)	R(llnt): L(GAGA)	R(llnt): L(GCRA)
R(1):L(GAAA)	R1	4.4	15 000	10 000	>20 000	>20 000	>20 000
R(2):L(GAAA)	R2	15	17.5	375	1000	500	>20 000
R(C7.34):L(GAAA)	R2	4	6	40	110	75	>20 000
R(IC3):L(GAAA)	R2	70	160	30	100	80	800
TectoRNAs with fixed R1/GGAA interaction	Receptor Class	R(l): L(GAAA)	R(l): L(GGGA)	R(l): L(GUAA)	R(l): L(GUGA)	R(l): L(GAGA)	R(l): L(GCRA)
R(llnt):L(GGAA)	R1	4.4	>20 000	>20 000	>20 000	10 000	>20 000
R(C7.2):L(GGAA)	R1	6.9	>20 000	>20 000	>20 000	8000	>20 000
R(C7.10):L(GGAA)	R1	61	>20 000	>20 000	>20 000	20 000	>20 000

Apparent equilibrium dissociation constants ($K_{d.app}$) were measured at 10°C in presence of 15 mM Mg^{2+} for various heterodimer constructs with either the 11nt/GAAA interaction or R(1)/GGAA interaction fixed. Reported values (in nM) are the average of at least two, typically three, independent measurements.

It is only at high concentrations (10 μM) of R(1), R(11nt), R(C7.2) and R(C7.10) that it is possible to see a small extent of non-cognate binding to other tetraloops. Despite their different specificity towards the second tetraloop position, the fact that R(1) and R(11nt)-related GAAA binders have similar behaviors and require an adenine at the third tetraloop position supports categorizing both as class-R1 receptors (Figure 4). Comparatively, R(2) and R(C7.34), which were shown to be selective for GGAA over GAAA in the homodimer study, are not able to distinguish GGGA from GGAA or GAGA from GUAA (Table 1). R(2), R(C7.34) and R(IC3) are all considered to be members of class-R2 because they have significant sequence similarities and can recognize either G or A at the third tetraloop position. R(IC3) receptor exhibits a less selective profile than R(2) and R(C7.34), however.

Measurements of the magnesium dependence of R(1):L(GGAA), R(11nt):L(GAAA), R(2):L(GGAA) and R(IC3):L(GAAA) homodimers corroborate a distinction between class-R1 and class-R2 receptors. For class-R1 receptors, we observed an apparent Hill constant of 2.1 for both R(1) and R(11nt) (data not shown), suggesting that at least one magnesium is bound by each receptor (37). By comparison, class-R2 receptors R(2) and R(IC3) have Hill constants of 1.2 and 1.4, respectively (data not shown). This suggests that the apparent binding affinities for Mg^{2+} are different within the two classes of receptors, and that the higher magnesium-dependent cooperativity for class-R1 receptors may be related to their better selectivity and affinity to target tetraloops.

Thermodynamic and structural analysis of class-R2 receptors

While class-R2 receptors all bind multiple GNRA targets, differences in their recognition patterns indicate that their conserved CCC:GGG sequence signature is not entirely responsible for GNRA specificity. Apparent K_d values measured for R(2), R(C7.34), and R(IC3) bound to various GNRA tetraloops indicates that the sequence of the internal loop can modulate the selectivity of the receptor towards particular GNRA (Table 1). R(C7.34) and R(IC3) display the same affinities towards GURA and GAGA, but differ in their ability to recognize GGAA, GGGA and GCRA. R(C7.34) is at least 15 times more selective for the recognition of GGRA than R(IC3). Its affinity for GCRA is decreased by a factor of at least 25, thus increasing the discrimination between GCRA and GGRA by a factor of 500 (Table 1). By contrast R(IC3) discriminates GCRA and GGRA by a factor of about 10 only. The best targets for R(2), like for R(C7.34), are GGRA tetraloops. The affinity of R(2) towards its cognate GGRA targets is decreased by a factor of three compared to R(C7.34). Its selectivity of recognition of GGRA versus GURA and GAGA (and possibly GCRA) is increased by a factor of about two with respect of R(C7.34), however.

It is remarkable that the single-point mutations U4A, A9C and to lesser extend U4C, changed the pattern of specificity of R(IC3) to closely resemble the one of R(C7.34). The selectivity for GGAA and GGGA of these R(IC3) variants is increased by a factor of over 10 (Figure 4). By comparison, the reverse of this point mutant, A4U in R(2) as well as C4U in R(C7.34), results in a decrease of selectivity of the receptor against non-cognate targets. However, these variants of R(2) and C(7.34) did not result in a loss of selectivity as dramatic as the one observed in R(IC3). While U4 and A9 in R(IC3) are clearly key positions for the specificity of recognition of GGAA and GGGA, it is also evident that the context of nucleotides surrounding these two positions is important. Further mutational studies are clearly required for identifying the pattern or rule of selectivity pertaining to class-R2 receptors (Geary and Jaeger, unpublished data).

To gain additional insight into the differences between class-R1 and class-R2 receptors, we tested two receptors from each class in a structural probing experiment. Pb(II) induced cleavages of receptors R(2) and R(IC3), both in presence and absence of probe, reveal that the internal loop regions of these two motifs remains accessible to solvent even when bound to a tetraloop target (Figure 5A and B). By comparison, R(1) and R(11nt) are highly protected in presence of their cognate target. Our data suggest that the nucleotides within class-R2 receptors are more flexible than those in class-R1 receptors, a probable explanation

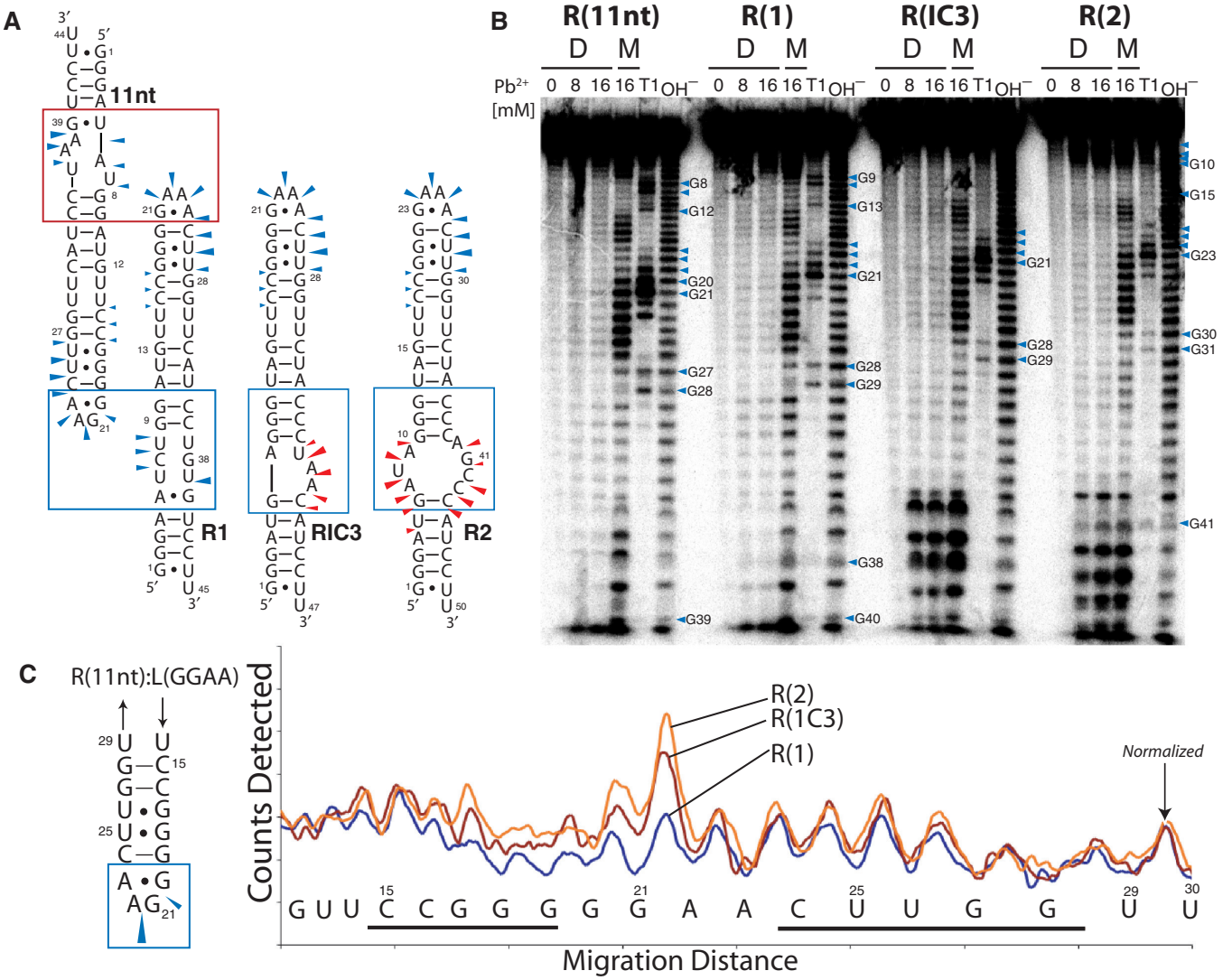

Figure 5. Lead(II)-induced cleavage patterns for R(11nt), R(1), R(2) and R(IC3) tectoRNAs in their monomeric or heterodimeric states. (**A**) 2D diagrams of tectoRNAs with reported phosphate positions cleaved in the monomer (M) and heterodimers (D). Phosphates that are cleaved in the monomers but are mostly protected in the hetero-dimers are indicated with blue arrows. Phosphate positions that are cleaved both in the monomer and heterodimer are indicated with red arrows. The 11nt/GAAA interaction motif is outlined in red. Other interacting motifs are outlined in blue and bind loop GGAA in a way similar to R(1). (**B**) Pb(II) cleavage patterns of 32P radiolabeled R(11nt):L(GGAA), R(1):L(GAAA), R(2):L(GAAA) or R(IC3):L(GGAA), either alone or bound to their non-radioactive cognate partner (as shown in A). M and D correspond to monomer and dimer lanes, respectively. Cleavage experiments were carried out at 0, 8 or 16 mM Pb(II) as described in the Materials and Methods section. T1 indicates RNaseT1 digestion (Gs are indicated with blue arrows). OH indicates alkaline hydrolysis ladder. (**C**) Superposition of Pb(II) cleavage profiles corresponding to the radiolabeled GGAA tetraloop when bound to R(1), R(IC3) or R(2) receptors. The phosphate cleavage site between position G21 and A22 of the tetraloop is protected to various extent depending on the receptor it is bound to. Nt numbering corresponds to nt position within the 5′ to 3′ RNA sequence according to (A). Plots were normalized to position U29.

for their looser specificity to GNRA tetraloops. It appears that class-R2 receptors generally seem to be more permissive to point mutation albeit less selective for specific GNRA targets (Geary and Jaeger, unpublished data). This could be a possible evolutionary advantage in organisms that evolve rapidly or are subjected to higher mutation rates.

Thermodynamic and structure analysis of class-R1 receptors

The patterns of specificity for the natural receptor R(11nt) and the two receptors previously selected by Costa and Michel (7), R(C7.10) and R(C7.2), are remarkably similar. While R(C7.2) differs from R(11nt) in four positions, it has identical apparent K_d and selectivity properties. By comparison, R(C7.10) is two mutations away from R(11nt), R(C7.2) and R(1), and retains the exact same selectivity as R(11 nt) but with 10-fold lower affinity to its target (Figure 2C). While the detailed atomic structures of R(11nt) R(C7.10) and R(C7.2) are unlikely to be the same, the availability of a detailed atomic structure of the 11nt/GAAA interaction (11) suggests that their overall structural organization is however likely to be very similar. This might also be the case for R(1).

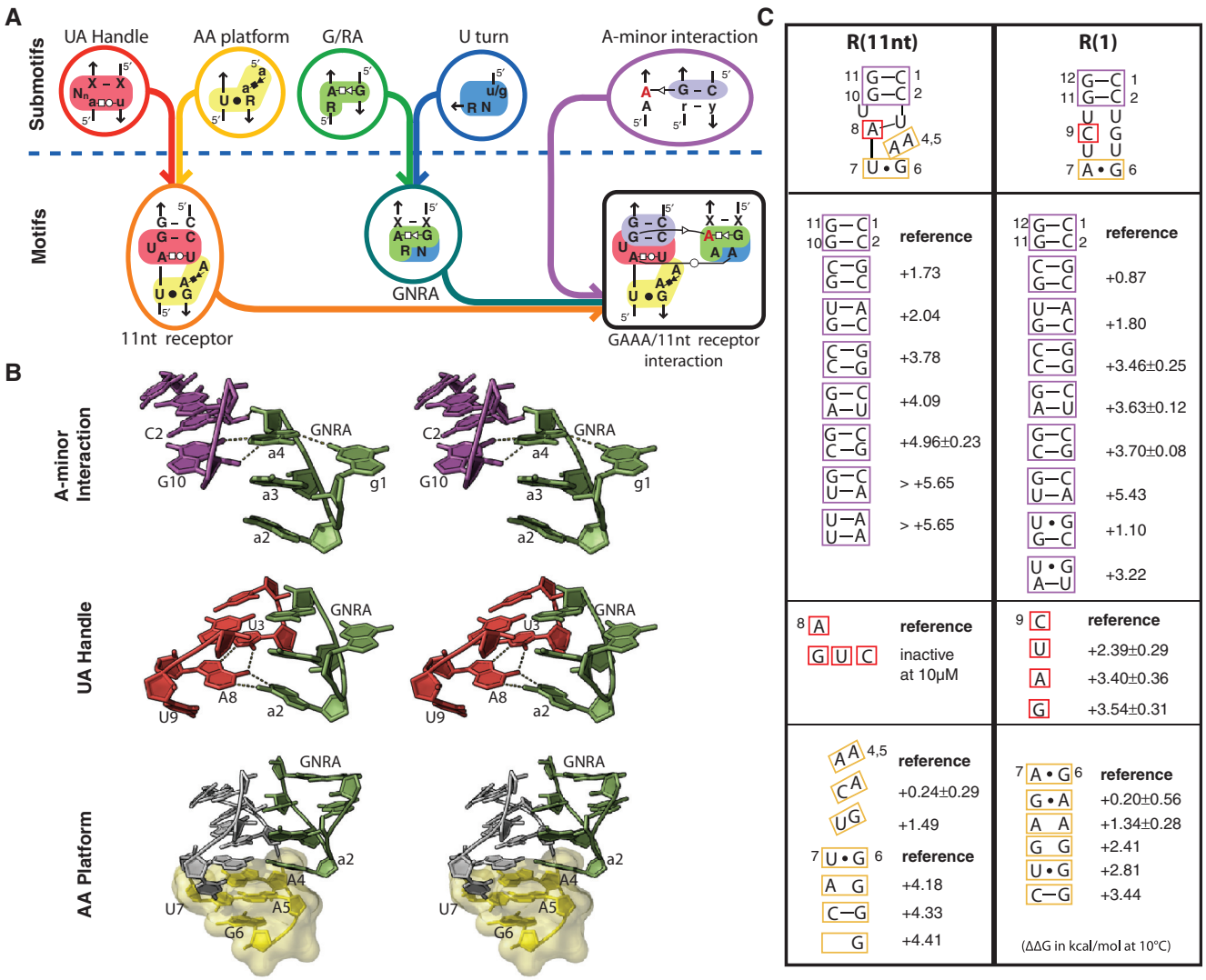

Figure 6. Structural modularity of the R(11nt)/GAAA and R(1)/GGAA interacting motifs. (**A**) At a structural level, the 11nt/GAAA interaction can be seen as a combination of five basic structural submotifs. The UA_handle and the AA platform motif combine to form the 11nt receptor. Combination of the U-turn and G/RA motifs leads to the GNRA tetraloop conformer that can interact with the 11nt motif through the A-minor motif. Base-pair interactions are indicated according to the Leontis and Westhof notation (68). (**B**) Cutaway view of the 11nt/tetraloop interaction (PDB ID: 1U6B) showing each submotif involved. Regions are color coded as shown in A. (**C**) Mutational and thermodynamic analysis of the R(11nt) and R(1) receptors within the R(11nt):L(GGAA) + R(1):L(GAAA) heterodimer context. The stability of various tectoRNA heterodimers presenting various mutations within each of the three submotif regions of the R(11nt) and R(1) receptors is compared to the wild-type reference and shown as $\Delta\Delta G$ in kcal/mol measured at 10°C and 15 mM $Mg(OAc)_2$.

These possible structural similarities prompted us to investigate the 11nt and R(1) receptors in more detail to better understand the structural elements responsible for their high selectivity and affinity towards GAAA or GGAA tetraloops.

Analysis of the 11nt receptor motif. According to the NMR and X-ray structures of the 11nt/GAAA interaction (11,34), the 11nt/GAAA interaction can be broken down into distinct structural submotifs that are recurrent in natural RNAs (Figure 6A). The structural makeup of the interaction is built around an A-minor interaction between the last adenine (a4) of the GAAA tetraloop and C2:G10 of R(11nt). Further stabilization occurs through specific H-bond contacts between the second adenine (a2) of the tetraloop and the UA-handle submotif (Verzemnieks and Jaeger, unpublished data) and a2 stacking on the AA platform submotif (49) (Figure 6B). Site-directed mutagenesis within the R(11 nt)/GAAA interaction was used to investigate the thermodynamic contributions of each of these elements to the overall stabilization of the interaction.

Seven mutants of the A-minor interaction region of the 11nt receptor were generated and compared to the original motif. All of these mutants destabilized the formation of heterodimers, corroborating the fact that the natural 11nt sequence signature is the most optimal (Figure 6C). Mutations within C2:G10 were the most detrimental to the stability, shifting the free energy by more than +3.78 kcal/mol. This is not surprising, as these bases

form four hydrogen bonds with a4 in a classic type-I A-minor interaction, in agreement with previous studies on this type of interaction (6,23). We additionally found that mutants to the top C1:G11 bp were destabilizing by at least +1.73 kcal/mol, despite the fact that there is no apparent tertiary contact observed in the crystal and NMR structures of this interaction (11,34,36) (Figure 6B and C). This suggests that C1:G11 is likely to be coupled to C2:G10, as both bps affect the thermodynamics of loop/receptor assembly. Further support of a potential coupling between these 2 bp is provided by the fact that a mutant reversing the entire A-minor-binding site by flipping both the top and bottom bps (mutant C1G:G11C, C2G:G10C) is one order of magnitude more stable than the reversal of C2:G10 alone. These findings are in agreement with a previous claim suggesting that the loop/receptor interaction forms in multiple steps with an initiation corresponding to the docking of the GAAA to the top C1:G11 bp before binding C2:G10 (34).

According to the crystallographic structure (11), the specificity of recognition of the GAAA tetraloop versus all the other tetraloops is provided by a *trans* WC:WC bp between A8 of R(11nt) and a2 at the second tetraloop position. None of the mutations to the A8 position in the receptor were able to promote tectoRNA assembly at concentrations as high as 10 μM (Figure 6C). Furthermore, none of the variants at position A8 were able to bind to any other GNRA tetraloops at 10 μM (data not shown). As A8 is also involved in a *trans* WC:HG interaction with U3, forming part of the UA handle submotif, this position is also critical for the proper folding of the 11nt receptor motif (Figure 6B). Due to the special geometry of this non-canonical bp, perfect isosteric replacements of this bp are not possible without unmodified bases (50).

The AA platform submotif consists of two adenines stacked on the top of a G:U wobble and contributes to tetraloop binding by forming a platform that favors purine base stacking (49). Surprisingly, the most detrimental mutations that we introduced within the AA platform submotif, were those to the G6:U7 wobble closing bp that is located at more than 7 Å from the GAAA (Figure 6C). The single-point mutation changing the G:U wobble into a G:C leads to +4.33 kcal/mol shift in free energy. While this G:U is not directly involved in any direct tertiary H-bond contact with the GAAA tetraloop, its mutation results in a destabilization of the same magnitude as mutations within the type-I A-minor interaction. Subsequent substitutions of the G:U with G:A or a single G nt had similar effect. Apparently, the G:U is involved in a metal-binding site that could not be easily mutated without significantly affecting the folding of the 11nt motif (37). Furthermore, in the structures of the 11nt motif, the G:U wobble also appears to compensate for extra helical twist imparted by the di-nucleotide platform. Substitutions of the AA nts within the platform by any geometrically similar bp should be possible without destabilizing the receptor. Drawing from natural variations and known nucleotide isostericity (50,51), AC as well as GU platforms have roughly the same predicted size and shape as the AA platform (Figure S4 in Supplementary Data). We found that replacing the platform with AC had no effect on the specificity or apparent K_d of the motif. Replacing it with GU [see R(C7.10)] decreased the binding energy by +1.5 kcal/mol while preserving the specificity (Figures 4 and 6C). This might explain why the AC platform variant of the 11nt motif is more frequently found in nature, while the GU platform variant has only been observed in RNA obtained by *in vitro* selection [R(C7.10)] (7). Additionally, our result suggests that the H-bond contact observed between U9 and A5 in the X-ray structure of the P4P6 domain does not seem to contribute much to the stability of the 11nt/GAAA interaction. Indeed, U9 is too far from A5C to make a similar contact (25). The loss of binding due to the GU platform in R(C7.10) is likely related to a potential alternative WC pairing between G4-U5 and A8-U9. It is interesting to note that R(C7.2), with a binding affinity comparable to the 11nt motif, has a GU platform and apparently avoids this alternative pairing by substituting G6:U7 bp by a single A6 (Figures 2 and 3).

Analysis of the R(1) receptor motif. The atomic structure and mode of binding of R(1) are unknown. Its resemblance to C7.10 strongly suggests that it might share a similar structural organization with the 11nt motif, however. To elucidate the structural buildup of R(1), site-specific mutations were introduced within the R(1)/GGAA interaction of the R(11nt):GGAA/R(1):GAAA heterodimer context to test the possible energetic contribution of structural elements expected to be equivalent to those found in the 11nt motif. As shown in Figure 6C, R(1) has two consecutive GCs able to act as binding site for A-minor interaction. Additionally, its residue C9, located at a position equivalent to A8 of R(11nt) and R(C7.10), could possibly interact with the GGAA tetraloop. Drawing from the sequence similarities observed between R(1) and C7.10 (Figure 2C), some nucleotides in R(1) might also be involved in the formation of a platform.

Mutations within bps C1:G12 and C2:G11 of R(1) had similar effects on tectoRNA assembly as those within the A-minor motif of R(11nt) (Figure 6C). Remarkably, the same trend in $\Delta\Delta G$s is observed, with mutations in C2:G11 being more detrimental than those in C1:G12. Our results strongly support the existence of an A-minor-binding site in R(1) that recognizes position a3 and a4 of the GGAA tetraloop, like in the structure of the 11nt/GAAA interaction. Interestingly, the R(1) double mutant C1G:G12C, C2G:G11C shows a partial compensation of the negative effect on binding that is observed with the single mutant C2G:G11C. This suggests that the coupling observed between the two CG bps of R(11nt) takes place in R(1) likewise (Figure 6C). However, the magnitude of the $\Delta\Delta G$ changes for R(1) variants is slightly less pronounced than the one for R(11nt) variants. This probably results from the different thermodynamics of the fixed loop/receptor interactions in each tectoRNA variant series.

To address specificity in R(1) we looked at R(1)C9, the nt in R(1) that is localized at the same physical position as R(11nt)A8, which is responsible for the remarkable

specificity of recognition of GAAA by the 11nt motif. Mutations of C9 dramatically alter the affinity of R(1) for its cognate GGAA tetraloop as well as the pattern of recognition for other GNRA tetraloops (Figures 4 and 6C). However, R(1)C9 variants are not as detrimental as R(11nt)A8 variants because they can still recognize the GGAA tetraloop with binding affinity reduced by 2.4–3.5 kcal/mol. This suggests that position C9 might not be as structurally constrained by its surrounding nts as position A8 within R(11nt). In a similar way as the contact involving A8 in the R(11nt):GAAA interaction, R(1)C9 could be involved in the formation of a C:G *trans* WC:WC with the second position of the GGAA tetraloop. With a $\Delta\Delta G$ greater than 5.65 kcal/mol, GCAA tetraloop is one of the GNRA least able to bind R(1). Interestingly, the R(1)C9G mutant improves the affinity of R(1) toward GCAA by at least 2 kcal/mol and binds to GCAA and GGAA with the same affinity (Figure 4). As the only *trans* WC:WC bp that is perfectly isosteric to a C:G is G:C (Figure S4 in Supplementary Data), this result is in good agreement with a direct contact between C9 and g2, the second position of GGAA. While C9 is our best candidate for specific recognition of the GGAA tetraloop versus the GNRA, one cannot rule out that other nucleotides within R(1) might contribute to the discrimination of GGAA versus other GNRA tetraloops, however.

To test whether some nts within R(1) are involved in the formation of a nucleotide platform that could potentially stabilize docking of the GGAA tetraloop, we investigated the effect of mutations on G6 and A7, hoping that these positions might be equivalent to the G:U closing bp of the AA platform of R(11nt). Interestingly, single-point mutations at either of these two positions have dramatic effects on the binding affinity and specificity of GNRA recognition (Figures 4 and 6C). By contrast, the double mutant that reversed G6:A7 into A6:G7 is neutral, suggesting the existence of a *cis* WC bp between G6 and A7. Widening of the backbone geometry seems important for the proper folding of the R(1) motif and binding of GGAA. When G6:A7 is changed into a G:U or a G:C bp, binding affinity to the GGAA tetraloop is reduced by 2.81–3.44 kcal/mol, respectively. However, G:G and A:A mismatches, which prevent WC pairing but can still widen the backbone helical geometry, are less destabilizing (Figure 6C). Additionally, the receptor specificity is also affected for the G6:A7U and G6:A7C variants. According to the structural organization of the 11nt motif and our data, R(1)G6:A7 is likely to be structurally equivalent to R(11nt)G6:U7 as both bps have similar phenotypes. Thus, while we cannot rule out that a direct physical contact might possibly occur between the GNRA and positions G6 and A7, our data are more consistent with the existence of a nucleotide platform in R(1). Drawing from the sequence similarities existing between R(1) and the closely related R(11nt) parent, R(C7.10), G4, U5 and U8 can potentially be involved in a GU platform that is stacked on the top of a widen G:A or A:G *cis* WC closing bp (Figure S4 in Supplementary Data).

Additional data favoring the existence of a platform in R(1) were provided by Pb(II) cleavage experiments (Figure 5). Pb(II) cleavage patterns of R(1) and R(11nt) in absence or presence of their respective target, indicate that both receptors are highly structured when bound to their cognate GNRA. While the profile of Pb(II) cleavage for unbound GGAA tetraloop shows that its backbone is cleaved between the second and third nucleotide position, the GGAA tetraloop bound to R(1) leads to a strong cleavage protection of the tetraloop at this site. By contrast, this protection is not observed to the same extent when the GGAA is bound to R(2) or R(IC3), suggesting that GGAA binds more tightly to R(1) than to R(2) and R(IC3) (Figure 5). The cleavage protection of the tetraloop in presence of R(1) is likely due to a decreased solvent accessibility of the tetraloop that might result from stacking of the tetraloop on a nucleotide platform in a structural context similar to the one of R(11nt). Further structural characterization of the R(1) receptor bound to it cognate GGAA tetraloop will be performed by NMR to verify this hypothesis.

DISCUSSION

Previous *in vitro* evolution systems for new RNA interactions were based on selections for structural and folding interactions required for catalytic activity (7,52,53), or selections based on structural compaction and RNA–RNA self-assembly (16,46,54,55). However, among the latter examples, selected tertiary interactions essentially took advantage of classic WC pairings for achieving high specificity of recognition of their targets. In the past, structural scaffoldings were successfully used to isolate new catalytic modules (47,56,57). Our selection, based on similar concepts, takes advantage of a rigid self-assembling scaffold design to allow for the direct selection of specific A-minor interactions promoting helical packing.

Although the selection was successful in isolating two A-minor dependent classes of receptors (R1 and R2), class-R3 forms classic WC pairings to bind the GNRA probe. Based on the highly stringent selection scheme used, this is not surprising as WC pairings are demonstrably the best way to achieve high target selectivity (15). In the case of class-R3, the binding of the GGAA probe through classic WC pairings requires the partial unfolding of the probe as well as the unfolding of its thermo-stable GGAA tetraloop. There are documented examples of these phenomena in nature. In the 23S rRNA the terminal GGAA tetraloop of helix H96 folds in an unusual conformation that is involved in the formation of two WC bps with another terminal loop in helix H57 (12). Moreover, class-R3 receptors are reminiscent of long-range tertiary pairings in Group I intron ribozymes that also require partial unfolding of stem-loops to form (8,58). Regardless, while loop–loop interactions often occur in nature, they are not as abundant as long-range interactions involving A-minor contacts (22,23).

The two A-minor dependent receptors that we have isolated shed light on the way GNRA receptors might be structurally organized to achieve high specificity and affinity. The use of selective heterodimer constructs has allowed us to investigate the thermodynamics of specific hydrogen bonds within class-R1 and R2 receptors as well

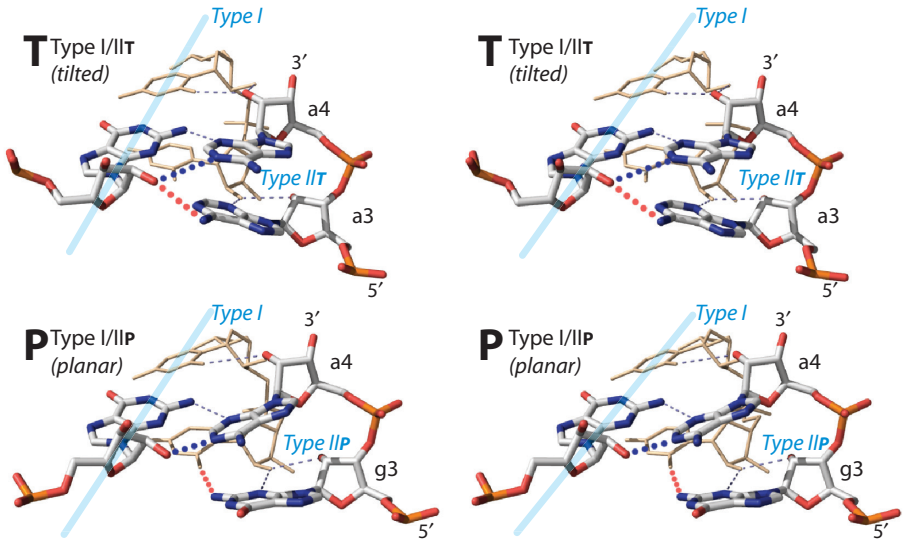

Figure 7. Stereographic images of the two classes of Type-I/II A-minor interactions. Top: **(T)** Type-I/IIT tilted configuration (GAAA/11nt interaction from PDB ID: 1HR2): both atoms N1 of a3 (blue H-bond) and N6 of a4 (red H-bond) of the GNRA tetraloop interact with the 2′OH of the canonical G from the A-minor Type-I triple. Bottom: **(P)** Type I/IIP planar configuration (GCGA/helix interaction from PDB ID: 2GCV). In this example, in addition to the classic Type II interaction, position O2 of g3 (red H-bond) can also interact with a O2 of the U of the Type IIP. In both cases, A4 can form a H-bond (blue) with the G of the Type-I A-minor triple bp.

as their modularity. Based on our data, the following general principles pertaining to the structural organization of GNRA receptors start to emerge: to achieve good binding affinity and selectivity, a platform module or/and a specific recognition module are necessary to be associated to the A-minor module (Figure 6, and Figure S5 in Supplementary Data). The A-minor module has similar energetic contributions within two different loop/receptor contexts [R(1) and R(11nt)]. Trends in the sequence preference of the type-I A-minor contacts suggest that the last nt position of the GNRA tetraloop recognizes the canonical C:G WC bp better than U:A, G:C and A:U. This result differs from the A-minor binding thermodynamic trend identified in the A-rich bulge of the P4-P6 Group I intron domain (59). Therefore, energetic contributions of A-minor motifs are likely to be context dependent.

In fact, the A-rich bulge and the 11nt motif have different hydrogen bonding patterns for their Type I/II A-minor interactions. Interestingly, both of these patterns are highly recurrent based on a survey of 42 GNRA/helix interactions from X-ray atomic structures (Table S1 in Supplementary Data). As shown in Figure 7, the overall shape of Type I and Type II interactions can be either planar (P) or tilted (T). In Type I/IIT, both GNRA positions R3 and a4 form H-bonds with the 2′OH of the canonical G belonging to the Type I A-minor bp. By contrast, in Type I/IIP, this 2′OH exclusively forms one H-bond with position a4. However, in both Type I/IIP and Type I/IIT, position R3 forms a classic ribose zipper (60) with the WC bp immediately below the type-I contact (Figure 7). Type-I/IIP and Type-I/IIT can lead to the formation of an equivalent number of H-bonds, but with different specificity of recognition (Figure 7). As predicted by the phylogeny (6,9), the presence of a G at the third GNRA position correlates with the presence of a U:A for the Type-II A-minor WC bp, allowing formation of one specific H-bond not present in Type-IIT (Figure 7). In summary, Type-I/IIP A-minor interactions are characteristic of GNRA/receptor interactions that can tolerate a G at the third GNRA position, as all GYGA/receptor interactions from X-ray structures are of Type-I/IIP (Table S1 in Supplementary Data). All Type-I/IIT interactions involve recognition of GNAA tetraloops (or two consecutive adenines) because they cannot tolerate a G in the third position due to steric clashes. Additionally, GYAA/receptors tend to favor Type-I/IIP versus Type-I/IIT, while the reverse trend is observed for GAAA/receptors (Table S1 in Supplementary Data).

By comparing the specificities of receptors R(1) and R(2) in light of the GNRA sequence trend observed for Type-I/II(P/T) interactions we can propose the following conclusions. The permissivity of R(2) for GGRA strongly suggests that R(2) forms Type-I/IIP A-minor interaction. By contrast, R(1), which is highly specific to GGAA, is more likely to take advantage of Type-I/IIT interaction to increase the specificity of recognition of the third GNRA nt position. In additional support of this hypothesis, R(11nt) and R(1) mutants that destabilize the receptor motif and decrease specificity still discriminate well against GNGA tetraloops (Figure 4). By contrast, mutants to class-R2 receptors R(IC3), R(2) and R(C7.34) remain permissive to GGRA. It remains to be determined if the thermodynamics of R(2) in the A-minor region will be similar to the A-rich bulge, as we would predict based on the sequence specificity of this class of receptor. However, both the thermodynamics measured for the R(1) A-minor interaction as well the lead cleavage data are comparable with the Type I/IIT data measured for R(11nt), while the lead cleavage data for the two class-R2 receptors R(2) and R(IC3) seem to indicate structural flexibility.

The P4-P6 domain A-rich bulge as well as the A-site/codon–anticodon interaction in the ribosome (61,62) involve Type-I/IIP A-minor contacts (Figure 7, and Figure S5 in Supplementary Data). They have evolved to discriminate against bp mismatches rather than against different classic WC bps. By contrast, the cooperative effect between different modules within loop/receptor interactions might amplify the constraints on the A-minor module. For example, in R(11nt) and R(1) we showed that the closing bp of the platform module, which is located at a distance over 10 Å from the A-minor site, has energetic contributions to binding of the same magnitude as the A-minor module itself. Platform motifs appear to stabilize Type-I/IIT A-minor interactions, as several examples from nature have this module besides the 11nt motif (e.g. P12-L13 from RNaseP and L39-H89 in the 23S rRNA in *Haloarcula marismortui*; Figure S5C and D in Supplementary Data). In all receptors with platforms, the nt recognition module is localized only one base pair below the Type-I A-minor interaction. However, in TypeI/IIP interactions the nt recognition module that interacts with the second GNRA position is always 2 bp below the Type-I A-minor contact (Figure S5B in Supplementary Data). Type-I/IIP configuration seems thus structurally less compatible with the formation of platforms. This suggests that the different receptor modules act synergistically for the proper orientation and recognition of the GNRA tetraloop, explaining the strict preference of the two adenines at the third and fourth positions of the GNRA as well as the canonical C:G bp forming a Type-I/IIT A-minor interaction in R(11nt) and R(1).

In summary, we propose a general rule that distinguishes two different classes of GNRA receptors based on the geometry of their A-minor Type-I/II interaction. Class-R1 corresponds to receptors forming Type-I/IIT and class-R2 corresponds to interactions of Type-I/IIP. Class-R1 (or Type-I/IIT) receptors are those specific for adenine at the third GNRA position. Class-R2 (or Type-I/IIP) receptors are those specific for purine at the third GNRA position. Therefore, while both classes can reach a good degree of specificity of recognition for the second nt GNRA position, class-R1 is able to reach a higher level of specificity because of its Type-I/IIT interaction.

While both class-R1 and class-R2 receptors are selected through natural and *in vitro* evolution processes, most selected artificial receptors still remain to be discovered in natural RNA sequences. Therefore, selected receptor sequence signatures might provide means to identify crucial tertiary contacts in RNA sequences for which the 3D structure is still unknown. The dominant selection pressure favoring the apparent emergence of GAAA/11nt versus GGAA/R(1) interactions in nature is unlikely related to the biophysical properties of R(11nt) and R(1) as these receptors have similar thermodynamics and binding selectivity for their respective cognate GNRA. Costa and Michel (7) suggested that the near absence of some of their selected receptors in nature might result from a lack of structural robustness, in the sense that the neutral network relating sequence to assembly function of selected receptors are highly reduced in plasticity (63). In other words, their selected receptors may lack the ability to be mutated without significantly decreasing their stability and function, explaining why they are apparently very rare in nature. However, some of our recent data indicate that this might not be the case, as point mutations within either the natural GAAA/11nt or selected GGAA/R(1) interactions have detrimental effects of similar magnitude. Conversely, both natural and selected class-R2 receptors are more tolerant to point mutations (Geary and Jaeger, unpublished data). Therefore, by being less constrained in their sequences, class-R2 receptors occupy a larger sequence space without losing much of their selectivity or affinity for their cognate GNRAs.

A more likely explanation for the preferred occurrence of known natural receptors may result from an evolutionary adaptation to avoid kinetic folding traps. Indeed, alternative folds may be more readily avoided through the usage of A-rich sequence signatures versus those involving Gs and Cs. For instance, phylogenetic data indicates a strong preference for AA versus AC and GU dinucleotide platforms in R(11nt) (7). AA and AC platforms were measured to be thermodynamically equivalent in the context of the R(11nt)/GAAA interaction, while GU led to a destabilization of 1 kcal/mol that is likely associated with a local alternative configuration of the internal loop (see Results section and Figure 6C). However, while local alternative folds can eventually be avoided by point mutations [see R(C7.2) compared to R(C7.10), Figure 2C], Gs and Cs that are not neutralized through stable base pairs within a local structural motif provide opportunity for long-range alternative pairings within much larger RNA molecules. This might be the case for the selected class-R1 and class-R2 receptors (Figure 2). Our lead cleavage data on R(1) in its bound and unbound state are compatible with the proposed induced fit mechanism of GAAA docking by R(11nt) (34,38). Therefore, the unbound state of R(1), which is likely un-structured, might be evolutionarily disfavored in large RNA structural contexts in comparison to the unbound state of R(11nt). Furthermore, in contrast to the natural A-rich internal loop of R(IC3), selected R(2) and R(C7.34) have several unpaired Gs and Cs that could increase the likelihood of misfolding RNA sequences *in vivo*. In the future, some of these hypotheses might be interesting to test by NMR or by studying RNA folding kinetics *in vivo*.

CONCLUSION

To solve the RNA-folding problem, the knowledge of the sequence space, kinetics and thermodynamics parameters of RNA tertiary interactions is invaluable for predicting the tertiary structure of an RNA from its sequence (1,4). By taking advantage of rationally designed tectoRNAs, we have demonstrated that meaningful data pertaining to GNRA/receptor interactions can be obtained. We believe that the same strategy can be applied to unravel the syntax of many other RNA tertiary motifs as well as their associated biophysical parameters.

Aside from being tools for investigating thermodynamic properties, the newly selected tectoRNAs expand our available library of nucleic acid building blocks for nanoconstruction (40–43, 64–66). RNA units with self-assembling interfaces based on selected class-R1 and class-R2 receptors were shown to form particles or/and filaments with controllable directionality and chirality (41,64). Additionally, heterodimer constructs taking advantage of R(1) were used to promote bottom up assembly of conductive gold-nanowires (42). In the future, RNA self-assembly principles will likely play a major role in the development of new responsive biomaterials and nanoparticles, with potential applications in biology and medicine (65,66).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

L.J. wishes to dedicate this work to Saint Patrick and Our Lady of Good Counsel. The authors thank Sam Butcher for critical reading of the manuscript and Eric Westhof and Neocles Leontis for their support in the early stage of this work. C.G., as the recipient of the distinguished Baker award 2007, wishes to present his gratitude to Ms Baker for the recognition of his present work at UCSB. This work was funded by National Science Foundation (DMR-05-20415 and CHE-0317154) and National Institutes of Health (R01 GM079604-01). Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Tinoco,I. Jr and Bustamante,C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
2. Noller,H.F. (2005) RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
3. Hendrix,D.K., Brenner,S.E. and Holbrook,S.R. (2005) RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.*, **38**, 221–243.
4. Leontis,N.B., Lescoute,A. and Westhof,E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.*, **16**, 279–287.
5. Michel,F., Jaeger,L., Westhof,E., Kuras,R., Tihy,F., Xu,M.Q. and Shub,D.A. (1992) Activation of the catalytic core of a group I intron by a remote 3′ splice junction. *Genes Dev.*, **6**, 1373–1385.
6. Jaeger,L., Michel,F. and Westhof,E. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.*, **236**, 1271–1276.
7. Costa,M. and Michel,F. (1997) Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.*, **16**, 3289–3302.
8. Lehnert,V., Jaeger,L., Michel,F. and Westhof,E. (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. *Chem. Biol.*, **3**, 993–1009.
9. Massire,C., Jaeger,L. and Westhof,E. (1997) Phylogenetic evidence for a new tertiary interaction in bacterial RNase P RNAs. *RNA*, **3**, 553–556.
10. Abramovitz,D.L. and Pyle,A.M. (1997) Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. *J. Mol. Biol.*, **266**, 493–506.
11. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
12. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, **289**, 905–920.
13. Adams,P.L., Stahley,M.R., Gill,M.L., Kosek,A.B., Wang,J. and Strobel,S.A. (2004) Crystal structure of a group I intron splicing intermediate. *RNA*, **10**, 1867–1887.
14. Torres-Larios,A., Swinger,K.K., Pan,T. and Mondragon,A. (2006) Structure of ribonuclease P–a universal ribozyme. *Curr. Opin. Struct. Biol.*, **16**, 327–335.
15. Brunel,C., Marquet,R., Romby,P. and Ehresmann,C. (2002) RNA loop–loop interactions as dynamic functional motifs. *Biochimie*, **84**, 925–944.
16. Aldaz-Carroll,L., Tallet,B., Dausse,E., Yurchenko,L. and Toulme,J.J. (2002) Apical loop–internal loop interactions: a new RNA–RNA recognition motif identified through in vitro selection against RNA hairpins of the hepatitis C virus mRNA. *Biochemistry*, **41**, 5883–5893.
17. Nagaswamy,U. and Fox,G.E. (2002) Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA*, **8**, 1112–1119.
18. Krasilnikov,A.S. and Mondragon,A. (2003) On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640–643.
19. Gagnon,M.G. and Steinberg,S.V. (2002) GU receptors of double helices mediate tRNA movement in the ribosome. *RNA*, **8**, 873–877.
20. Mokdad,A., Krasovska,M.V., Sponer,J. and Leontis,N.B. (2006) Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.*, **34**, 1326–1341.
21. Steinberg,S.V. and Boutorine,Y.I. (2007) G-ribo: a new structural motif in ribosomal RNA. *RNA*, **13**, 549–554.
22. Nissen,P., Ippolito,J.A., Ban,N., Moore,P.B. and Steitz,T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. US A*, **98**, 4899–4903.
23. Doherty,E.A., Batey,R.T., Masquida,B. and Doudna,J.A. (2001) A universal mode of helix packing in RNA. *Nat. Struct. Biol.*, **8**, 339–343.
24. Lee,J.C., Gutell,R.R. and Russell,R. (2006) The UAA/GAN internal loop motif: a new RNA structural element that forms a cross-strand AAA stack and long-range tertiary interactions. *J. Mol. Biol.*, **360**, 978–988.
25. Adams,P.L., Stahley,M.R., Kosek,A.B., Wang,J. and Strobel,S.A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature*, **430**, 45–50.
26. Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276–1285.
27. Costa,M., Deme,E., Jacquier,A. and Michel,F. (1997) Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.*, **267**, 520–536.
28. Brown,J.W., Nolan,J.M., Haas,E.S., Rubio,M.A., Major,F. and Pace,N.R. (1996) Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl Acad. Sci. USA*, **93**, 3001–3006.
29. Wimberly,B.T., Brodersen,D.E., Clemons,W.M.Jr., Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
30. Ramakrishnan,V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557–572.
31. Ikawa,Y., Naito,D., Aono,N., Shiraishi,H. and Inoue,T. (1999) A conserved motif in group IC3 introns is a new class of GNRA receptor. *Nucleic Acids Res.*, **27**, 1859–1865.
32. Ikawa,Y., Nohmi,K., Atsumi,S., Shiraishi,H. and Inoue,T. (2001) A comparative study on two GNRA-tetraloop receptors: 11-nt and IC3 motifs. *J. Biochem.*, **130**, 251–255.
33. Young,B.T. and Silverman,S.K. (2002) The GAAA tetraloop-receptor interaction contributes differentially to folding

thermodynamics and kinetics for the P4-P6 RNA domain. *Biochemistry*, **41**, 12271–12276.
34. Butcher,S.E., Dieckmann,T. and Feigon,J. (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.*, **16**, 7490–7499.
35. Qin,P.Z., Butcher,S.E., Feigon,J. and Hubbell,W.L. (2001) Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: a site-directed spin labeling study. *Biochemistry*, **40**, 6929–6936.
36. Davis,J.H., Tonelli,M., Scott,L.G., Jaeger,L., Williamson,J.R. and Butcher,S.E. (2005) RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex. *J. Mol. Biol.*, **351**, 371–382.
37. Davis,J.H., Foster,T.R., Tonelli,M. and Butcher,S.E. (2007) Role of metal ions in the tetraloop-receptor complex as analyzed by NMR. *RNA*, **13**, 76–86.
38. Hodak,J.H., Downey,C.D., Fiore,J.L., Pardi,A. and Nesbitt,D.J. (2005) Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl Acad. Sci. USA*, **102**, 10505–10510.
39. Correll,C.C. and Swinger,K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA*, **9**, 355–363.
40. Jaeger,L., Westhof,E. and Leontis,N.B. (2001) TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.*, **29**, 455–463.
41. Nasalean,L., Baudrey,S., Leontis,N.B. and Jaeger,L. (2006) Controlling RNA self-assembly to form filaments. *Nucleic Acids Res.*, **34**, 1381–1392.
42. Bates,A.D., Callen,B.P., Cooper,J.M., Cosstick,R., Geary,C., Glidle,A., Jaeger,L., Pearson,J.L., Proupin-Perez,M. *et al.* (2006) Construction and characterization of a gold nanoparticle wire assembled using Mg2+-dependent RNA–RNA interactions. *Nano Lett.*, **6**, 445–448.
43. Liu,B., Baudrey,S., Jaeger,L. and Bazan,G.C. (2004) Characterization of tectoRNA assembly with cationic conjugated polymers. *J. Am. Chem. Soc.*, **126**, 4076–4077.
44. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
45. Jaeger,L. and Leontis,N.B. (2000) Tecto-RNA: One-dimensional self-assembly through tertiary interactions. *Angew. Chemie. Int. Ed.*, **14**, 2521–2524.
46. Juneau,K. and Cech,T.R. (1999) In vitro selection of RNAs with increased tertiary structure stability. *RNA*, **5**, 1119–1129.
47. Jaeger,L., Wright,M.C. and Joyce,G.F. (1999) A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl Acad. Sci. USA*, **96**, 14712–14717.
48. Ikawa,Y., Naito,D., Shiraishi,H. and Inoue,T. (2000) Structure-function relationships of two closely related group IC3 intron ribozymes from Azoarcus and Synechococcus pre-tRNA. *Nucleic Acids Res.*, **28**, 3269–3277.
49. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.
50. Lescoute,A., Leontis,N.B., Massire,C. and Westhof,E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.*, **33**, 2395–2409.
51. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
52. Green,R., Ellington,A.D. and Szostak,J.W. (1990) In vitro genetic analysis of the Tetrahymena self-splicing intron. *Nature*, **347**, 406–408.
53. Treiber,D.K., Rook,M.S., Zarrinkar,P.P. and Williamson,J.R. (1998) Kinetic intermediates trapped by native interactions in RNA folding. *Science*, **279**, 1943–1946.
54. Duconge,F. and Toulme,J.J. (1999) In vitro selection identifies key determinants for loop–loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1. *RNA*, **5**, 1605–1614.
55. Lodmell,J.S., Ehresmann,C., Ehresmann,B. and Marquet,R. (2000) Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. *RNA*, **6**, 1267–1276.
56. Yoshioka,W., Ikawa,Y., Jaeger,L., Shiraishi,H. and Inoue,T. (2004) Generation of a catalytic module on a self-folding RNA. *RNA*, **10**, 1900–1906.
57. Ikawa,Y., Tsuda,K., Matsumura,S. and Inoue,T. (2004) De novo synthesis and development of an RNA enzyme. *Proc. Natl Acad. Sci. USA*, **101**, 13750–13755.
58. Li,Z. and Zhang,Y. (2005) Predicting the secondary structures and tertiary interactions of 211 group I introns in IE subgroup. *Nucleic Acids Res.*, **33**, 2118–2128.
59. Battle,D.J. and Doudna,J.A. (2002) Specificity of RNA–RNA helix recognition. *Proc. Natl Acad. Sci. USA*, **99**, 11676–11681.
60. Tamura,M. and Holbrook,S.R. (2002) Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.*, **320**, 455–474.
61. Ogle,J.M., Brodersen,D.E., Clemons,W.M.Jr., Tarry,M.J., Carter,A.P. and Ramakrishnan,V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, **292**, 897–902.
62. Lescoute,A. and Westhof,E. (2006) The A-minor motifs in the decoding recognition process. *Biochimie*, **88**, 993–999.
63. Ancel,L.W. and Fontana,W. (2000) Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.*, **288**, 242–283.
64. Afonin,K.A. and Leontis,N.B. (2006) Generating new specific RNA interaction interfaces using C-loops. *J. Am. Chem. Soc.*, **128**, 16131–16137.
65. Jaeger,L. and Chworos,A. (2006) The architectonics of programmable RNA and DNA nanostructures. *Curr. Opin. Struct. Biol.*, **16**, 531–543.
66. Chworos,A., Severcan,I., Koyfman,A.Y., Wienkam,P., Oroudjev,E., Hansma,H.G. and Jaeger,L. (2004) Building programmable jigsaw puzzles with RNA. *Science*, **306**, 2068–2072.
67. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, CA, USA.
68. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.