UNIVERSITY OF CALIFORNIA

Santa Barbara

Investigation of Self-Assembling Modules for RNA Architectonics

A dissertation submitted in partial

satisfaction of the requirements for the degree

Doctor of Philosophy in Chemistry

by

Erin Rebecca Calkins

Committee in charge:

Professor Luc Jaeger, Chair

Professor Kevin Plaxco

Professor Peter Ford

Professor Chris Hayes

December 2015

The dissertation of Erin Rebecca Calkins is approved.

_____

Kevin Plaxco

_____

Peter Ford

_____

Chris Hayes

_____

Luc Jaeger, Committee Chair

September 2015

Investigation of Self-Assembling Modules for RNA Architectonics

ACKNOWLEDGEMENTS

I would first and foremost like to thank my mentor, Luc Jaeger. Without his unconditional support and guidance, this body of work would not be possible. I am grateful for all his hard work and dedication to his research and his students. His knowledge and expertise in the field of RNA research and technology was an invaluable tool to me. I would also like to thank my lab mates, both past and present. Their constant support and collaboration was truly inspirational. Last and most certainly not least, I'd like to thank my family for which none of this would be possible. Thank you my love, for pushing me to be my best and to never give up. Special thanks to my children for all their love and understanding, I love you both to the moon and back.

VITA OF ERIN REBECCA CALKINS
September 2015

Education

9/94-9/00, B.S. Biology (Molecular, Cellular and Developmental Biology Emphasis), University of California at Santa Barbara

9/07-Present, Ph.D. Student in Biochemistry, University of California at Santa Barbara (expected graduation date, summer 2014)

Fall 2013, M.A. in Chemistry, Biochemistry emphasis from University of California, Santa Barbara

Professional Experience

1/14-Present Moorpark City College- Adjunct Instructor

- General Biology (3 semesters)- BioM01

8/10-Present Santa Barbara City College- Adjunct Instructor

- General Chemistry Lab (3 semesters)- CHEM 155
- Nanotechnology in Society Lab (1 semester)- PHSC 107
- Cellular Biology Lab (8 semesters)- BIO 103
- General Biology Lab (2 Semesters)-BIO 100

7/05-9/07 Marketing Associate/Copy Writer, The Walking Company (Santa Barbara, California)

- Marketing research, copy writer and editor for catalog and web based shoe company.

9/03-6/05 Middle School Science Teacher, Santa Barbara Christian School (Santa Barbara, California)

- Taught grades 6th through 8th, Earth Science, Biology, Chemistry, Astronomy and Pre-Algebra.

7/03-1/05 Research Associate, EpigenX Pharmaceuticals (Santa Barbara, California)

- Assay development and research designed around detecting compounds as potential drug targets for the inhibition of DAM (DNA methyltransferase) as novel antibiotics or cancer drugs.

9/00-1/03 Research Associate, Target Protein Technologies (San Diego, California)

- Development of drug delivery system by isolating tissue specific proteins for targeting.

Additional College Teaching Experience
University of California, Santa Barbara

- General Chemistry Lab (Chem 1AL), Teaching Assistant
- General Chemistry Lab (Chem 1BL)- 2 quarters, Teaching Assistant
- General Chemistry Lab (Chem 1CL), 2 quarters, Teaching Assistant
- RNA World (Chem 143), Teaching Assistant
- Organic Chemistry Lab (Chem 6BL), 2 quarters, Teaching Assistant
- Biochemistry (Chem 142A and 142C), Teaching Assistant
- Biochemistry Lab (Chem 125L), Teaching Assistant
- Biochemistry Lab (Chem 125L), Instructor of Record

Publications
Attenuation of loop-receptor interactions with psuedoknot formation. Afonin, K, Lin, YP, Calkins, ER and Jaeger, L (2012). Nucleic Acids Research. 40 (5): 2168-2180.

RNA-nanotechnology in nanomedicine. In "*Recent advances in nanomedicine and drug delivery"*. Grabow, W.W., Afonin, K.A., Zakrevsky, P., Walker, F.M., **Calkins, E.R**., Geary, C., Kasprzak, W., Bindewald, E., Shapiro, B.A. & Jaeger, L. (2012) (S. Thomas, A George and S. Mathew, eds), Apple Academic press, Inc.,Toronto, p 209-220, 274-278.

Downward causation by information control in micro-organisms. Jaeger, L and Calkins, ER (2011). Interface Focus (Royal Society Publishing). 2 (1): 26-41.

In vitro selected GUAA tetraloop receptor exhibiting structural similarities to naturally occurring RNA. Calkins, ER, Keleshian, V, Zakrevsky, P and Jaeger, L (2015). In Preparation.

Retro-Synthesis of GNRA Tetraloop Receptors Utilizing Structural Homology and Ration Design Principles. Calkins, ER, Keleshian, V, Zakrevsky, P and Jaeger, L (2015). In Preparation.

Presentations/Conferences
UCSB Advancement to Candidacy Exam (PASS)

- Title: I*n vitro* selection of novel classes of GNRA tetraloop receptors

Experimental Biology Poster Presentation (April 2010)-Anaheim

- Title: I*n vitro* selection of novel classes of GNRA tetraloop receptors

Grants/Fellowships/Scholarships/Other
STARS 100k Research Planning Grant

- Title: On the Reality of Top-Down Causation

Apprentice Researcher Mentor

- Awarded by CSEP (Center for Science and Engineering Partnerships)
- 8 week fellowship (Summer 2009)
- Designed and supervised in-lab project for high school student. Assisted student with understanding of project material, lab procedures and presentation of project details and results.

Graduate Teaching Scholar Fellowship

- Awarded by CNS (Center for Nanotechnology and Society)
- 3 quarter fellowship for the purpose of designing and teaching a course on Green Science and Sustainability at Santa Barbara City College

CAMP (California Alliance for Minority Participation) Fellowship

- Awarded by MRL (Materials Research Laboratory)
- 10 week fellowship (Summer 2011)
- Design, train and supervise research project for UCSB undergraduate student.

Jarrod Davidson Memorial Scholarship

- Awarded by the Chemistry Department to a graduate student who faces special challenges while raising a family.
- Awarded Spring 2013

ABSTRACT

Investigation of Self-Assembling Modules for RNA Architectonics

by

Erin Rebecca Calkins

Natural RNA found in the ribosome, RNaseP, riboswitches and ribozymes has inspired the design and characterization of many synthetic RNA structures. These architectures are composed of building blocks that are comprised of RNA units that serve as modules (or motifs) for the construction of novel, structural and functional RNA molecules. To better understand the types of interactions involved in generating motifs, characterization of both the secondary and tertiary structure is essential. Large, complex RNA require many cooperative units folding in a programmable way, stabilizing the overall structure. The research presented herein, demonstrates the ability of RNA units to preserve their function despite drastic sequence variation, while still maintaining their overall shape or topology. We have identified several classes of GNRA tetraloop receptors (both natural and synthetic) that can be characterized by their phenotypic behavior toward GNRA tetraloops. Phenotypic behavior, as well as comparative analysis of known crystal structures can elucidate structural detail of unknown receptors. We have also identified additional folding constraints that can prove beneficial in structure prediction and architectonic design of large structured RNA's providing a basis for characterization and implementation of novel design principles, enhancing the complexity of synthetic RNA.

Table of Contents

List of Figures

List of Tables

Chapter 1 RNA and Synthetic Biology

1.1 Introduction

1.1.1 *What is RNA?*

RNA (ribonucleic acid) is an extremely important biopolymer that has several functions in biological systems. These include the catalytic center in the ribosome used for protein synthesis [2], the genetic material in some viruses [3] and the regulation of genes (RNAi, attenuators and gene splicing) [4-6]. RNA is comprised of the four bases adenine, guanine, cytosine and uracil (A, G, C and U, respectively) connected via a glycosidic linkage to the 1' carbon of a ribose sugar. Polymers of RNA are linked together through a phosphodiester bond between the 3' hydroxyl of one ribose and the 5' phosphate of another ribose. The orientation of this covalent backbone offers a 5' to 3' directionality that allows for RNA to act as an informational molecule and encodes the primary sequence of proteins. It is also the primary sequence of RNA that directly determines the secondary and tertiary structure of the molecule, which gives the RNA an additional functional role.

The folding pathway from 1° to 2° to 3° structure is hierarchical in nature. The folding of RNA is initiated through π-stacking and Watson-Crick hydrogen bonding between bases to form stable helices (1° to 2°). The rapid step forming secondary structure is followed by slower step where tertiary structure is formed. This step requires metal ion condensation and begins with a conformational search of potential

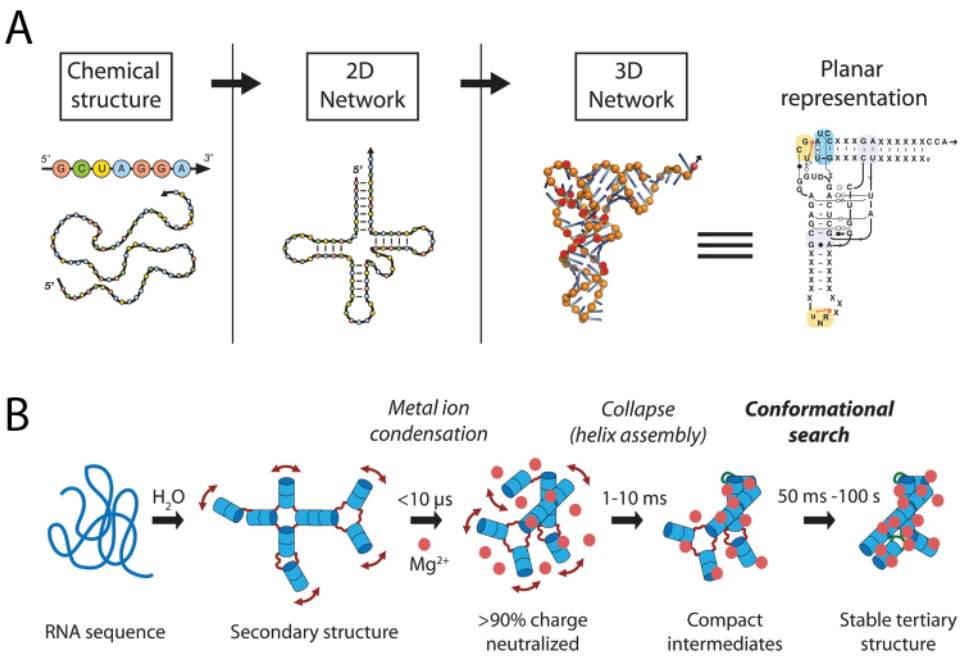


Figure 1.1 Hierarchical folding of RNA.
A) The primary nucleotide sequence dictates the secondary fold which is primarily direct by the formation of helices via classic Watson-Crick pairing. The secondary structure then further folds into the final 3D structure through non-canonical interactions. B) Metal ions are required for the proper folding of RNA by protecting the negative charges of the phosphate backbone. The addition of metal ions induces a collapse of the secondary structure and conformational search results in the final, stable native fold. (Grabow 2014)

folds until the final tertiary structure is formed (2° to 3°) [7]. Tertiary interactions occur through van der Waals contacts, π-stacking, coordination of metal ions and hydrogen bonding via unpaired regions of bases and/or the ribose sugar [8]. The conformational search for the final RNA fold can also involve long-range interactions that can occur both at the intermolecular and intramolecular level [9]. Typically, helices can be formed from canonical Watson-Crick base pairing (G:C and A:U) or through G:U wobble pairings, which also take advantage of the Watson-Crick edges. In addition to the classic pairings, RNA, is capable of interacting

through the nucleotides Hoogsteen (HG) edge as well as its Shallow Groove (SG) [9, 10]. While helices usually exist with base pairs in a cis orientation to one another, trans base pair interactions are also an integral part of RNA structure determination. With three possible edges capable of forming interactions through hydrogen bonds in two different orientations, there are many different interacting elements that can create a dynamic library of structures.

1.1.2 ***Synthetic Biology***

Recent advances in the fields of molecular and cellular biology have been pivotal in elucidating some of the complex interplay between cellular physiological processes. The field of synthetic biology takes advantage of these advances by employing non-natural molecules that can mimic natural functions in artificial systems [11, 12]. Molecular engineering concepts are employed as the means to build up pathways or even synthetic genomes [13]. Applications utilizing this technology include development of biosensors for generating synthetic networks, detection of new drugs (therapeutics) and biosynthesis of novel biofuels and biomaterials [14].

1.1.2.1 RNA as a Tool for Synthetic Biology

Proteins are typically thought of as having the role of structure, catalysis and regulation, whereas DNA is the carrier of genetic information. RNA, on the other hand bridges the gap between the two, functioning as the link between the genetic code and synthesis of proteins (mRNA, ribosome and tRNA). It has been well documented that RNA is so much more, acting as a catalyst (ribozyme), biosensors

(aptamers), performing the function of gene regulation (attenuators and siRNA) and as guide molecules to initiate chemical modifications (snoRNA). In addition, RNA's physiochemical properties, modular structure, limited number of building blocks and high number of possible interactions make it an excellent candidate for synthetic biology applications [15-19].

In large part, the emergence of this new field is due to high resolution crystallographic structures of naturally occurring RNA molecules, such as the ribosome, RNaseP, group I and group II introns [20-22]. Recurrent RNA structures, act as building blocks or scaffolds that can be built upon one another to create novel, functional RNA molecules. This type of nanoengineering has been implicated in applications for biosensors as well as in therapeutics for nanomedicine [23, 24]. Because RNA is capable of forming highly complex structures with diverse functionalities, it is an excellent model for synthetic biology applications. RNA Architectonics is a type of design strategy used in synthetic biology to systematically engineer, characterize and optimize engineered RNA nanoparticles.

1.2 RNA Motifs

The secondary structure of RNA is driven by the formation of helical elements that in turn generate a multitude of single stranded regions in the form of loops, bulges and junctions [25]. It was initially observed that some natural loop sequences as well as non-Watson Crick interactions were more frequent than expected if RNA sequence

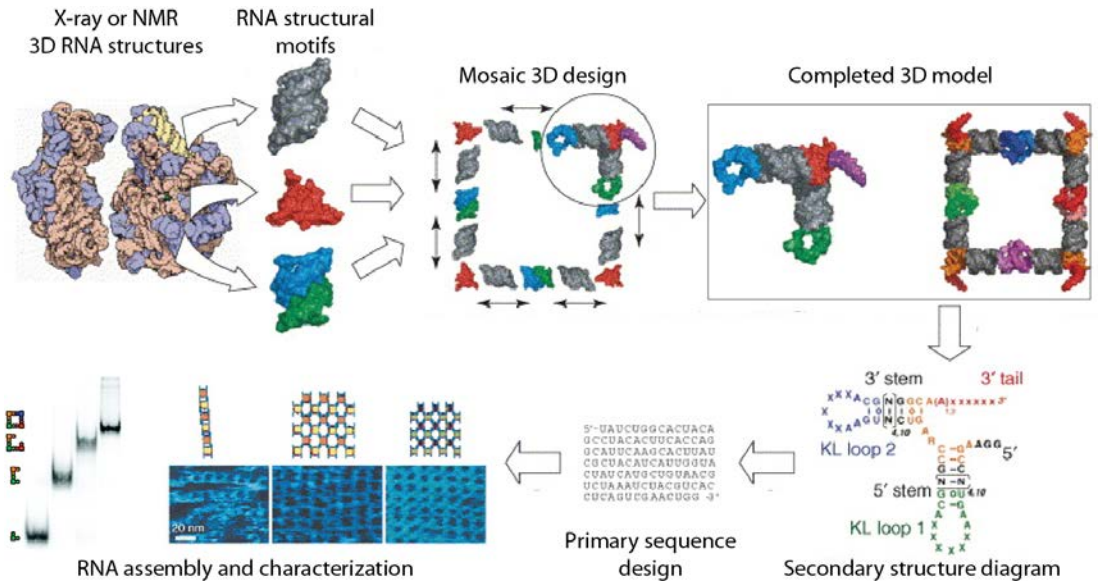


Figure 1.2 Rational design using architectonics.
Novel 3D RNA structures can be engineered through known crystallographic structures (motifs) These motifs can then be extracted as individual units to build up new structured RNA (in this case a square). From there, 2D structure can be determined and primary sequences can be optimized for RNA assembly. Methods for verification include native PAGE and atomic force microscopy (AFM). Figured adapted from (Jaeger 2006, Severcan 2009 and Chworos 2004)

was truly random [26]. A closer examination revealed common structural elements with distinctive conformations that are capable of adopting a recurrent folding pattern; RNA structural motifs. Examples of secondary structural motifs are terminal loops, which include triloops, tetraloops, T-loops and D-loops and internal loops, that consist of kink turns, hook turns and tetraloop receptors, just to name a few. These secondary structural motifs can then interact to form tertiary motifs that have well-defined base pairings and stacking interactions with a distinctive structural phosphodiester backbone topology [27]. These structural motifs include coaxial

helical stacks, A-minor motifs, ribose zippers, kissing loops, pseudoknots, loop-loop interactions and tetraloop/receptor interactions [28, 29].

Utilizing these structural motifs, it is possible to build larger structures in a predictable way due to the modularity of RNA components [30]. Each motif can act as a building block to generate a multitude of different interacting parts. In order to facilitate the engineering of increasingly complex structures, motifs can be characterized at both the local and global level [29, 31, 32]. Furthermore, understanding the parameters that influence the thermodynamic and kinetic folding of tertiary motifs will assist in the accurate design and prediction of novel, complex RNA structures.

1.3 Approaches for RNA Design

Several methods currently exist for the design of RNA nanostructures. Two approaches that are often utilized take advantage of well-characterized properties of RNA structure. First, the computer-aided method of single-stranded tiles that self-anneal to RNA sequences are used to generate stable scaffolds of predictable size and shape [33, 34]. The second strategy, RNA architectonics, involves the assembly of RNA subunits that can interact to produce macromolecular RNA architectures with novel structural and functional properties. These approaches have been well utilized to generate a multitude of functional RNAs, including RNA aptamers [35], siRNAs in the form of nanorings [36], as well as protein-RNA interactions [37]. Additionally, many computational programs have been designed in order to facilitate the design of larger, more complex RNA structures.

1.3.1 *RNA Architectonics*

RNA architectonics can be defined as the scientific investigation of the rules that govern RNA architectures in order to create novel molecules with predictable size and shape. There are several characteristics of RNA assembly that make the engineering of predictable nanostructures possible. A key component of RNA folding is that it is hierarchical in nature and provides the basis for RNA structure prediction (Figure 1.1). RNA is governed by two essential folding properties: it is able to spontaneously fold into a structure with a stable, well-defined 3D shape and it has the ability to form larger macromolecular structures by assembling multiple subunits [31]. In addition, RNA is capable of utilizing building blocks (or motifs) that can fold independently and then assemble into higher ordered structures. These types of modular subunits can be found in natural contexts, in addition to being engineered or selected for *in vitro* [38-40]. RNA motifs act as modular subunits that are capable of mediating tertiary interactions, which can then be utilized to engineer artificial self-assembling nanostructures [41-44].

1.3.2 *Self-annealing Tiles*

This tile strategy was first introduced as DNA origami by Nadrian Seeman in 1982 [45]. It is based on the self-assembly properties of nucleic acids by complimentary base pairing that can be utilized to construct complex, multi-dimensional structures. This paved the way for programmable self-assembly with RNA, which offers several advantages over that of DNA. It can be used to deliver siRNAs in therapeutic applications [46] or can be functionalized with the attachment

of aptamers [40]. However, the thermostability of the A-form helix as well as RNAs capability to form stable non-canonical base pairs, make it more apt to become trapped in alternative structures, making RNA tile scaffolds a bit more challenging to construct. These scaffolds almost exclusively use canonical Watson-Crick interactions and can be designed with the help of computer-based programs [33].

1.3.3 ***Computational Methods***

The manual manipulation of molecules can be done with programs like Swiss PDB viewer and Assemble, however this task is very time consuming and costly. Currently, there are many computational methods that take advantage of algorithms capable of either predicting self-assembly of RNA motifs or elucidating secondary structures from primary sequences [47-49]. There are also several databases that are continually expanding our repertoire of sequence space as it compares to structure.

Two general methods exist for RNA structure prediction; the nearest neighbor model and comparative sequence analysis. The nearest neighbor model is a method for secondary structure prediction that is largely based on thermodynamic models that calculate the free energy of all possible secondary structures, with the predicted structure having the lowest free energy [50]. The principle of free energy calculations are based on the energies of base pairing nucleotides in a sequence and the adjacent base pairs flanking each side [51, 52]. Examples of this type of structure prediction include Mfold [53], RNAstructure [54] and RNAfold [55], just to name a few.

Another type of computational RNA prediction deals with the way in which individual subunits come together in a 3D architecture [56]. RNA motif libraries, such as RNA 3D Motif Atlas, compile countless numbers of motifs from documented 3D structures. Sequence alignments from RNA families can be used to detect a motif from previously solved crystal structures based on primary sequence similarity [57]. There are two main underlying assumptions that this method considers; 1) different, homologous RNA sequences can adopt the same secondary and tertiary structure fold and 2) during evolution the secondary and tertiary structures are conserved, while the sequence may vary greatly. The covariation in sequence is a result of compensatory mutations that maintain the overall helical structure of the RNA [58]. This type of covariation acts as a detection method for structure conservation across species. With more and more crystal structures of RNAs being elucidated, the types of motifs and their interacting parts is broadening, allowing for the further refinement of structure prediction and design.

Each of these approaches to design as well as predict RNA structure, while extremely useful, do have their limitations. Mainly, secondary structure prediction is limited by the imprecision of the constraints imposed on the sequence/structure relationship and the incompleteness of the thermodynamic rules that govern RNA folding [59]. The drawback of comparative sequence analysis is that it relies on rules that ignore the phylogenetic relationship of the aligned sequences and any possible sequence divergence [60].

While computational modeling is certainly important to the field of RNA nanotechnology, experimental approaches can prove highly effective in validating

present computational models in addition to generating new rules to refine algorithmic constraints. Two well utilized strategies for the generation of novel RNAs that contain specific ligand binding properties, modular interactions, catalytic properties and/or other functional characteristics are rational design and SELEX (Systematic Evolution of Ligands by Exponential Enrichment). These two methods have the capability to generate novel RNAs that can be characterized as to their sequence, structure and function [61, 62].

1.4 Generation of Novel RNA Nanoparticles

Currently, two highly effective approaches are being utilized to generate functional RNAs with novel properties and unique sequences; rational design and SELEX. Often with the aid of computer modeling programs, rational design pieces together known motifs from crystallographic structures in unique ways. SELEX on the other hand, is a technique that takes advantage of a vast pool of randomly generated sequences that can be selected for *in vitro* through a variety of techniques.

1.4.1 ***Rational Design***

At the root of rational design is the inherent ability of RNA to fold in a hierarchical manner. As mentioned previously, the primary drive of RNA folding is through the formation of stable secondary structures that once folded can undergo a conformational search to find the desired tertiary structure. These self-folding subunits can then assemble to form larger, more complex quaternary structures. The general technique for this type of design is through inverse folding or reverse

engineering [63]. Figure 1.2 Rational design using architectonics. Figure 1.2 demonstrates the concept of this rational design method. By starting with 3D images of crystallographic structures, motifs can be extracted as separate entities that act as building blocks for the construction of new structured and/or functional architectures. Building up the new tectoRNA 3D model requires piecing together the motifs in a logical manner utilizing modeling programs like Swiss-Pdb Viewer or Assemble. This step involves local topologies (shapes) connecting in a way that the overall global structure is achieved irrespective of sequence. Next, the secondary structure diagram is generated such that it denotes the sequences that are required to conserve the desired 3D shape signature and those that can vary, i.e. stem sequences, kissing loop sequences, etc. To design the variable residues, sequences that stabilize the secondary structure as well as promote the formation of the desired 3D structure are chosen. Experimental testing will verify the effectiveness of the rational design and the primary sequence can be further optimized based on these results.

For example, the right angle motif from group I introns comprises a set of conserved sequences that form a right angle turn [64]. Severcan et al. extended the helical stems in either direction from the right angle structure and capped each stem with a kissing loop (another prominent motif found in many RNAs) [36, 65-67]. They were able to demonstrate with assembly assays that these individual right angle subunits were forming tectosquares, linked together via kissing loop interactions.

1.4.2 ***SELEX***

While nature provides us with numerous examples of RNA architectures that provide excellent examples of the sequence to structure connection, this is only a limited number of motifs. The SELEX (Systematic Evolution of Ligands by Exponential Enrichment) method allows for the screening of large pools of RNA sequences for a specific binding capability or function. This approach, although not originally designed to screen for oligonucleotides with novel functions [68], was quickly adapted for such applications as aptamer binding [69], catalytic activity [58, 70, 71] and RNA-RNA interactions [72, 73].

1.5 RNA and the Evolution of Biomolecules

In light of the ability for a "real-time" evolution of biomolecules, logically a connection to natural evolution became a topic of interest [74-76]. According to the RNA world hypothesis, RNA was the precursor biopolymer to that of DNA and proteins, where life relied solely upon the storing of genetic information and catalysis carried out by RNA molecules, with peptides, sugars, lipids and other small molecules being present. In order for this to be plausible, RNA would need to be able to polymerize nucleotides from an RNA template to produce the next generation. There is no known natural RNA that can perform this reaction, however *in vitro* evolution (SELEX) has produced several novel RNAs that can perform this task [77, 78]. Advances in this field of RNA nanotechnology have produced a wide variety of catalytic RNAs with new properties and functions. The use of pre-existing natural scaffolds as the starting material for new ribozymes [79] as well *in vitro*

selected ribozymes from randomized sequences, have enriched the repertoire of functional RNAs as well as given insight as to how these molecules have evolved. Much research is currently attempting to answer in a more concise way the details of how an RNA world could exist [80, 81] and the processes by which these molecules evolve.

One such approach to answering these questions is through the use of a fitness landscape [82, 83]. A fitness landscape associates a given genotype (sequence) with a phenotype (fitness). It is thought that evolution is a random "walk" through the sequence space and that selected sequences are chosen based on peaks of fitness [84]. This methodology is primary accomplished computationally, however the data, which is based on structural fitness, does not take into account function [55, 85]. Recently, new endeavors have shed light on the evolution of the structure/function relationship [82]. Presented herein is a unique look at fitness, structure and function as it pertains to the sequence space of RNA motifs and their interactions.

1.6 GNRA Tetraloop Receptors

1.6.1 *Natural context and structural evolution*

Most helical stems in structured RNAs are capped with a stabilizing tetraloop sequence: CUYG, UNCG and GNRA [86]. Each of these three classes has shown to contain specific geometries governed by networks of hydrogen bonds and base stacking [6, 87, 88]. The GNRA class has been shown to be extremely stable due to a U-turn motif between the G and the N of the tetraloop, however CUYG and UNCG

offer a higher degree of local stability over GNRA tetraloops, which initially suggested that their role is more functional [89]. This functionality is due the capability of GNRA tetraloops to form long-range tertiary interactions with distal helical elements [86, 89, 90]. It has been hypothesized that GNRA tetraloops were evolutionarily selected for their high stability at the level of the interaction, which enable them to facilitate the proper folding of the functional structure [91]. In particular, of the tetraloops, GNRA are the most abundant with the GAAA tetraloop comprising more than 50% of all the GNRA tetraloop sequences [92]. Because of the overwhelming distribution in favor of the GAAA tetraloop, much research has been focused on this loop and its interactions, in addition to answering the question, 'why is it so abundant compared to other tetraloops?'

GAAA tetraloops are often interacting with continuous helices (typically tandem C:G base pairs) and have been found in group I and group II introns [21, 73, 93, 94], RNaseP [95, 96] and the ribosome [97-100]. The A3 and A4 nucleotides of the GAAA tetraloop interact with the minor groove of the helix forming 7 out the 9 potential hydrogen bonds [101]. This type of interaction between adenine and the minor groove is known as the A-minor motif, which is by far the most common tertiary interaction in RNA structures [102, 103]. Because of the specific hydrogen bonding patterns, different GNRA receptors prefer different helical sequences. For example GNGA loops preferentially bind 5'-CU:AG-3'. This is because the C2 position of the G3 amino group would have a steric clash with a C:G base pair [93, 101]. This clash is alleviated by replacement of a U:A base pair in the helix [102].

In addition to helical receptors, there other receptor types that offer more stability and specificity. For example, the 11nt receptor is an asymmetric internal loop, that is one of the most stable and selective tertiary interactions found in group I and group II introns and RNaseP [21, 72, 94, 104], as well as in riboswitches [105, 106]. The interaction between the 11nt receptor and the GAAA tetraloop has been extensively studied and the specific interacting modules have been identified and characterized. The crystal structure of the P4-P6 domain of the group I intron of *Tetrahymena* showed that the 11nt-GAAA interaction was mediated via an A-minor motif as seen with helical receptors. The specificity of this interaction, however is facilitated by hydrogen bonds between the three adenines of the tetraloop and the internal loop of the receptor, where the adenines of the tetraloop are further stabilized by base stacking in the minor groove [21]. This receptor further demonstrates the modularity of RNA motifs by utilizing three submotifs through long-range interactions.

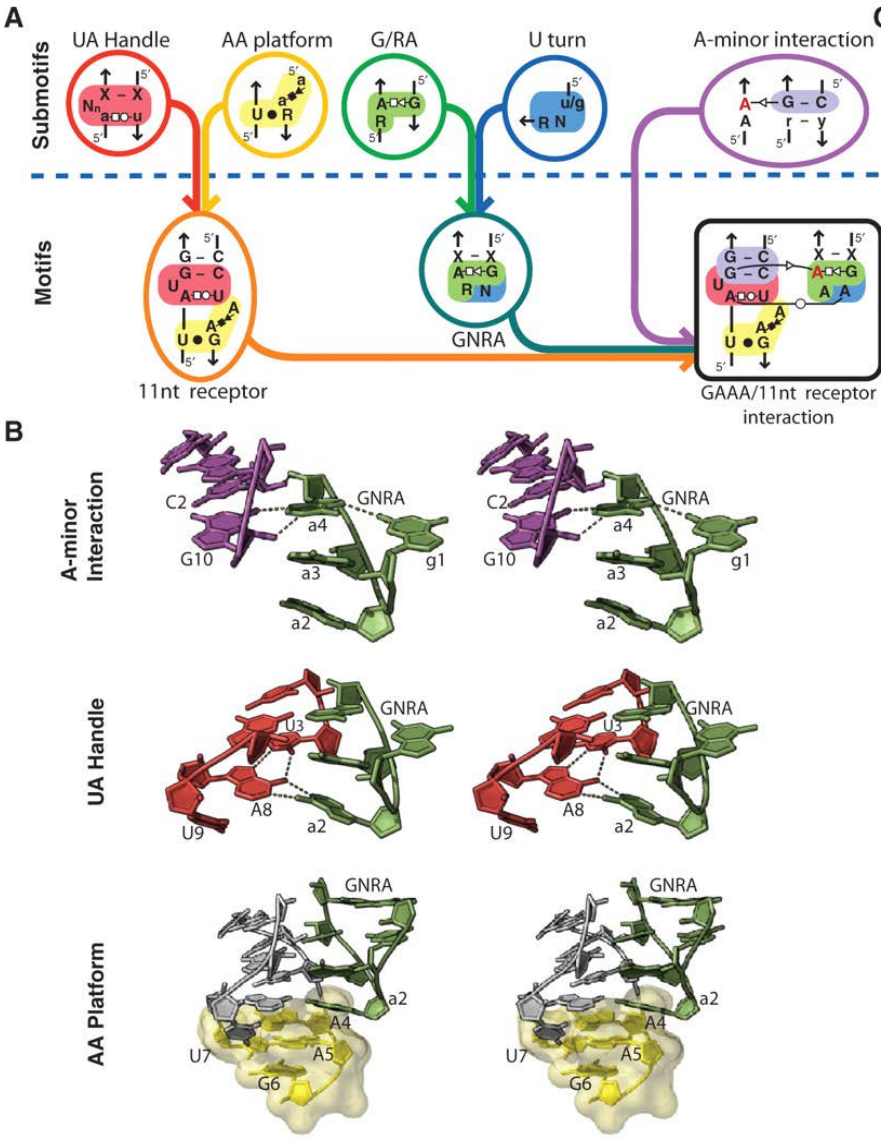


Figure 1.3 11nt Receptor Motif.
A) 2D rendering of the 11nt receptor is comprised of a group of submotifs that interact to generate the three fundamental motifs that have been found to interact in many natural and synthetic RNA receptors, A-minor interaction, UA Handle and dinucleotide platform. B) 3D images of the three motifs highlighting hydrogen bond interactions.

1.7 TectoRNA

The following research focuses on the long-range tertiary interactions of RNA receptor motifs, primarily GNRA receptors that self-assemble via A-minor interactions. The tectodimer approach designed by Jaeger et al. (2001) allowed us to not only probe the sequence space for novel functional sequences, it also acted as a means for structural and functional characterization. Chapter 2 introduces a novel way to retro-engineer a GNRA tetraloop receptor utilizing structural homology and rational design. We demonstrate the ability to rationally design GUAA tetraloop receptors by following topological guidelines for a naturally occurring 2x-bugle class of receptor found in 23S ribosomal RNA of bacteria and archaea [107].

Chapter 3 presents research that also utilizes the tectodimer system, but for selecting novel sequences of receptors capable of binding GNRA tetraloops via A-minor interactions. We characterized several novel classes of GUAA receptors and were able to elucidate structural data based off topological similarities of the naturally occurring protein binding RNA, S8 of the 16S ribosomal subunit [108].

Structural links between natural and artificial RNA motifs has prompted much research into the evolution of these long-range interactions. Clues as to the selection pressures that have attributed to the advent of novel design and functionality in nature is an invaluable tool in generating novel synthetic interactions for nanobiotechnology applications. One such observation shows a preference for A/U-rich sequences in long-range interactions. Chapter 4 provides data highlighting the importance of key features of these long-range interactions in the proper folding of a

structural RNA by utilizing attenuation sequences capable of turning off or on GNRA interactions. This study provides evidence for structural evolution as well as introduces novel design principles for long-range RNA interactions. Furthering the notion of structural evolution principles and providing a possible explanation for the mechanism of an increase in complexity of macromolecules, Chapter 5 introduces a biological framework for evidence in support of downward causation in microorganisms. The research herein additionally supports this notion, that functionally equivalent, higher order structures, rather than primary sequence, is more pivotal in establishing functionality. Moreover, this concept addresses the criteria that contribute to possible constraints in the development of molecular complexity.

Chapter 2 Retro-Synthesis of GNRA Tetraloop Receptors Utilizing Structural Homology and Ration Design Principles

2.1 Abstract

Structured RNA has been shown to adopt many different conformations and provide a vast repertoire of long-range interactions to assist in the folding of complex molecules such as the ribosome. Evolutionary evidence has shown that structural homology among functional RNAs is conserved thereby providing the framework for probing sequence space while maintain functionality. The 23S ribosome of *Thermus thermophilus* contains the H89 receptor that binds a GNRA-like GUAAG pentaloop and has been classified as a 2X-bulge class of receptor. Characterization of RNA receptors such as this has been made possible by utilizing a tectodimer assay developed by Jaeger et al (2001). The H89 interaction, however has been unsuccessful in its characterization due to post-transcriptional modifications, which *in vitro,* do not allow for the proper folding of the receptor. Utilizing structural homology and rational design principles we were able to "revive" the activity of this class of receptor by retro-synthesizing ancestral receptors that contained varying sequences (without post-transcriptional modification), yet maintained the overall structure and function of the 2x_bulge topology.

2.2 Introduction

Structured RNA molecules take advantage of multiple long-range tertiary interactions to stabilize their overall native conformation. The process of folding is facilitated by metal ions and is thought to initially involve a rapid

thermodynamically driven collapse of secondary structural elements, followed by a slower rearrangement, which ultimately leads to the native conformation [7, 31]. The final structural fold is a direct result of the interactions between recurrent RNA motifs [27, 109, 110], which are responsible for generating complex, three-dimensional RNA architectures from conserved sets of sequences involved in non-canonical base pairings and specific topological configurations [111, 112]. These motifs can be classified as either structural or interacting and can be built upon one another to form more complex motifs. Structural motifs include helices, hairpin loops, internal loops, and junctions [113-115]. Common RNA tertiary motifs responsible for long-range interactions include the ribose zipper, pseudoknot, D-loop, and GNRA tetraloop/receptor interactions [116-119].

The most abundant type of long-range RNA tertiary motifs is the tetraloop-receptor interaction [73, 103, 120, 121]. This motif has been identified in many RNA crystal structures, including the ribosome, RNaseP, group I and group II introns, as well as many ribozymes [21, 72, 122, 123]. It is often characterized by a terminal GNRA tetraloop interacting with the minor groove of tandem G-C pairs within a helix. It was first observed in group I introns [90, 93] and the hammerhead ribozyme [101], but since then various other types of receptors have been discovered [94, 124]. These interactions have been extensively studied and key components involving the structure and function of these motifs have been elucidated. Interestingly, two receptors found within the crystal structure of the ribosomal subunits of *Thermus thermophilus*, H89/L39 (23S) and the S8 binding domain (16S) of the S8 protein, show a strikingly similar fold (Figure 2.1A and B). Despite their similar folds, their

sequence is very different as well as their functional roles in the ribosome. The H89 receptor binds to an L39 terminal pentaloop and the S8 binding site interacts with the S8 protein on the opposite side, utilizing a different platform than the H89 receptor (Figure 2.1B). Figure 2.1C shows how the structures of the two receptors are superimposable with their interacting counterparts on the opposite sides of the receptor. Due to the strikingly similar topology between these two receptors and their seemingly unrelated sequences and functions, we proposed to characterize these structurally equivalent receptors.

In 2008, Geary et al characterized loop-receptor interactions using a tectodimer system developed by Jaeger and Leontis, which consisted of two monomer units each containing a loop and receptor that could bind to the other monomer in a synergistic manner (Figure 2.1D, E and Figure 8.1A and B). This system was used to carry out binding affinity experiments characterizing *in vitro* selected GNRA receptors utilizing the 11nt/GAAA interaction as an anchor. It was found that R1, a receptor selected from a pool of RNAs binding a GGAA tetraloop, had a comparable binding affinity for its cognate loop to that of the 11nt receptor with GAAA. This tectodimer scaffold proved very useful in characterizing the behaviors of various GNRA loop-receptor interactions.

Interestingly, the crystallographic structures of the L39 pentaloop and that of a GUAA tetraloop contain similar folds (Figure 2.1B-C), where the last three positions (UAA) of the loop are stacked and the guanine and the fourth adenine of the loop is in a trans Hoogsteen/shallow groove conformation. The last guanine in the L39 pentaloop is flipped out from the loop and appears to be obsolete in the interaction

between loop and receptor [107]. This would suggest that the H89 receptor could potentially form dimers in the tectoRNA context with its cognate loop, GUAAG. However, when the H89 receptor was tested *in vitro* it was found to be non-active in this context. This was most likely due to post-transcriptional modifications that prohibit the formation of the precise hydrogen bonding pattern that generates the proper three-dimensional shape which accommodates the docking of the loop. The advent of post-transcriptional modification in the evolutionary time-line has increased the repertoire of RNA interactions and has led to an increase in specificity and complexity [125]. Although sequence can be highly conserved among important RNA structures, it has actually been found that structure is more conserved in the long run [126, 127]. With this evidence, there is a high likelihood that the H89 receptor, may have had an ancestral structure that was very similar to what we see today, but without the post-transcriptional modifications.

Based on our present understanding of GNRA tetraloop/receptor interactions and their defining characteristics, such as the minor groove interaction, docking platform

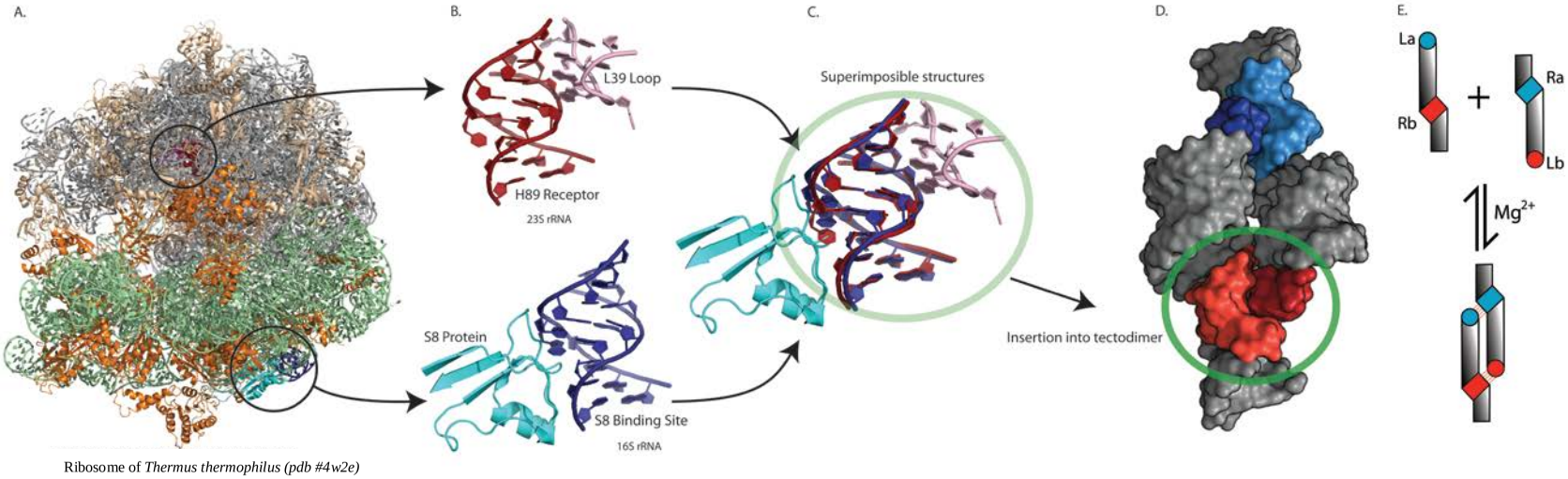
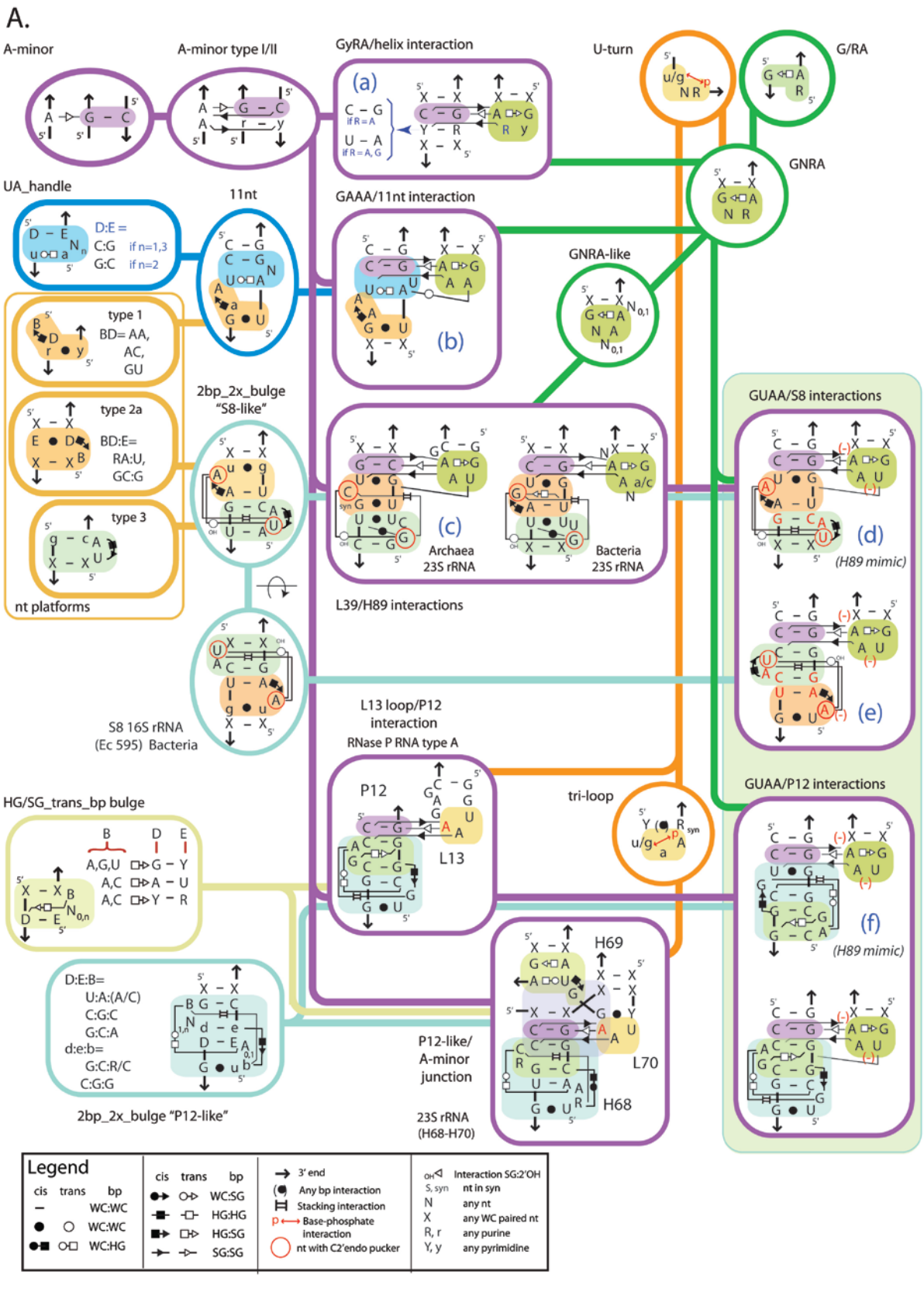


Figure 2.1 H89 and S8 receptors from the 16S ribosomal RNA of *Thermus thermophiles*
(A) Crystallographic structure of the ribsome from *Thermus thermophiles* highlighting the L39/H89 (pink/red) and S8 protein/S8 binding site (turquoise/blue) interactions. (B) Close up view of the L39/H89 and S8 protein interactions. (C) Superimposable images of the H89 and S8 receptors showing their similar topology. The L39 loop and the S8 protein, however are utilizing different platforms (opposite sides). (D) PDB_ID: 2JYJ, interactions such as the H89 or S8 receptor can be inserted into the tectoRNA dimer system as a way to characterize their interactions. (E)E) Monomer units each containing two separate loop receptor interactions can form dimers in the presence of Mg^{2+}.

and specific recognition module, we proposed to “revive” the activity of the H89 receptor in a modular fashion by retro-synthesizing a potential ancestral H89 interaction. It is hypothesized that H89 can systematically be reengineered to fold properly and assemble via long-range interactions without post-transcriptional modifications, thus creating structurally and functionally equivalent receptors to the natural ones we currently find in the ribosome. To further validate our characterization, several other natural receptors that contain similar functional motifs with slightly different structural architectures could potentially offer additional modular subunits for fine tuning functionality and creating receptors with adaptable characteristics. For example, the L13/P12 loop/receptor interaction from RNaseP recognizes a GNRA-like loop (Figure 2.2A), but is also taking advantage of a platform module [72]. The L70/H68 loop receptor interaction is another example of a variation of the platform module that could be useful in this study for comparison (Figure 2.2A). These modular platform subunits, while topologically different share a common functionality that indicates a certain propensity for interchangeability that could generate novel receptor interactions capable of altering selectivity and binding affinity.

Our work suggests that by using structural modularity principles, it is possible to carefully re-engineer parts of the ribosomal RNA capable of assembling and folding independently from proteins and posttranscriptional modifications. The retro-synthesized H89 receptor could potentially mimic an ancestral form of the ribosomal RNA [128].


A.
A-minor
A-minor type I/II
GyRA/helix interaction
U-turn
G/RA
GNRA
UA_handle
11nt
GAAA/11nt interaction
GNRA-like
type 1
BD= AA, AC, GU
type 2a
BD:E= RA:U, GC:G
type 3
nt platforms
2bp_2x_bulge "S8-like"
L39/H89 interactions
Archaea 23S rRNA
Bacteria 23S rRNA
GUAA/S8 interactions
(H89 mimic)
S8 16S rRNA (Ec 595) Bacteria
L13 loop/P12 interaction
RNase P RNA type A
P12
L13
tri-loop
GUAA/P12 interactions
HG/SG_trans_bp bulge
H69
L70
H68
P12-like/ A-minor junction
23S rRNA (H68-H70)
2bp_2x_bulge "P12-like"
Legend
cis trans bp
WC:WC
WC:WC
WC:HG
cis trans bp
WC:SG
HG:HG
HG:SG
SG:SG
3' end
Any bp interaction
Stacking interaction
Base-phosphate interaction
nt with C2'endo pucker
Interaction SG:2'OH
nt in syn
any nt
any WC paired nt
any purine
any pyrimidine

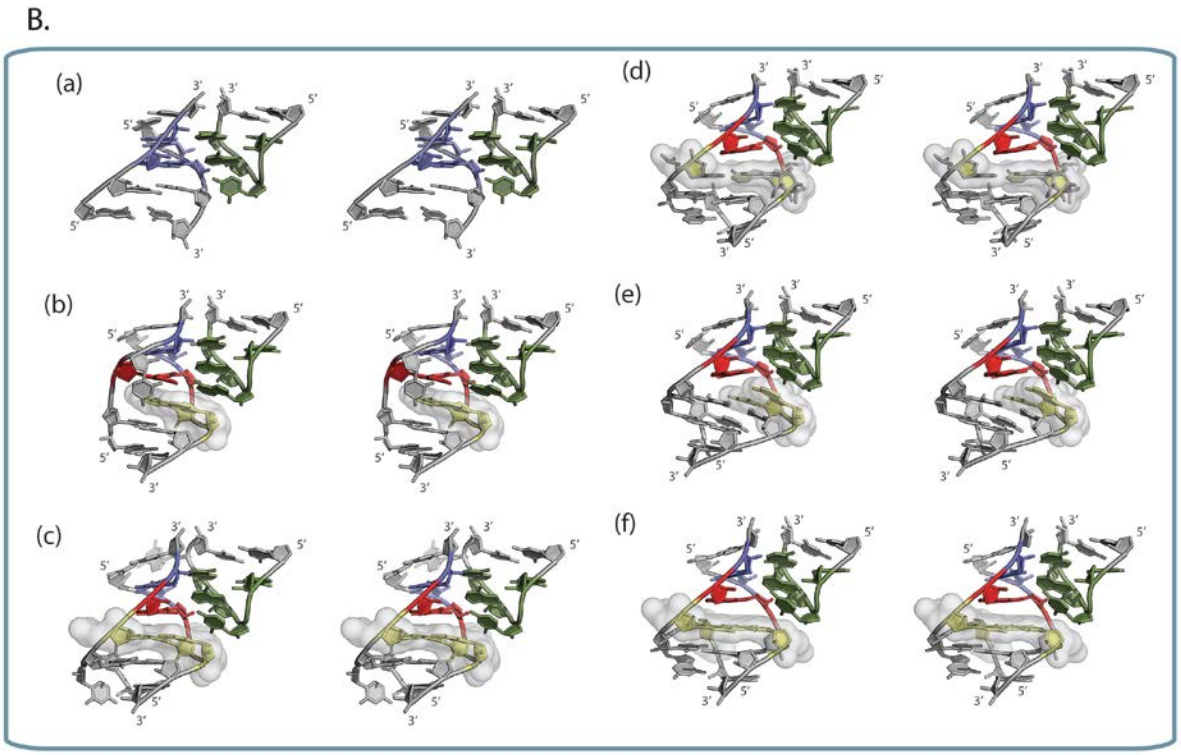


Figure 2.2 Motif Network for GNRA Receptors
(A) GNRA receptor network that demonstrates how motifs are built up to generate various interactions. Submotifs, such the A-minor, UA_Handle and nucleotide platforms are common structural elements for different classes of GNRA receptors. (B) 3D structures for selected receptors binding to a GNRA tetraloop. (a) Helical receptor (b) 11nt receptor (c) H89 receptor (d) S8/GUAA interaction (e) inverted S8/GUAA interaction (f) inverted P12 receptor

2.3 Results/Discussion

2.3.1 *Isosteric receptors: H89 and S8.a1*

The modules that make up the components of these loop/receptor interactions are comprised of units from a variety of different, highly recurrent submotifs (Figure 2.2A). The motifs themselves are built up in a hierarchical manner such that submotifs or functional modules can be combined to create larger, more complex

tertiary interactions. The network that comprises the GAAA/11nt and the L39/H89 interactions both contain functional modules that, while not identical have similar building blocks that perform similar functions (binding GNRA or GNRA-like loops). The network demonstrates the importance of structural homology over sequence homology. In other words, it's the nature of the shape of the module and its interacting parts rather than its sequence that dictate the function [129]. This is further demonstrated by the fact that the S8.a1 receptor (natural sequence found in ribosome), which differs from the H89 receptor at 8 out of the 11 nucleotide positions, can fold with nearly identical topology (Figure 2.1C).

The X-ray structure of the L39/H89 interaction shows a GNRA-like loop binding a receptor containing a 2x_bulge motif [107]. This motif consists of two base triple interactions formed by a two base pair helix flanked on either side by a 1-2 nucleotide bulge (Figure 2.2A (c)). The archaea L39 loop (GUAAG), folds as a GNRA tetraloop and interacts with H89 in a fashion comparable to the GAAA/11nt interaction. Both receptors utilize a type I/IIT A-minor interaction, a platform module and a recognition module, which would suggest the archaea L39/H89 loop/receptor should be able to demonstrate binding in the tectoRNA context. Additionally, the L39/H89 loop/receptor from *E. coli*, while utilizing an interaction with a hexaloop (GAUAAG) and containing a slightly different receptor sequence also folds much like the archaea S8 interaction (Figure 2.2A). Despite these structural similarities, the H89 interaction has yet to show any function *in vitro*, indicating that perhaps post-transcriptional modifications provide the means for proper folding in the natural context. To further validate the modularity of these

interactions, we modeled the S8.a1 rRNA protein receptor interacting with a GUAA tetraloop to demonstrate its ability to potentially act as a GNRA receptor Figure 2.2A and 2B (d)). S8.a1 was oriented in such a way to reflect a similar orientation to that of H89. H89 and S8.a1 show a high degree of structural homology with nearly superimposable images despite their differences in nucleotide sequence (Figure 2.1C). The models of H89 and S8.a1 both place the GNRA tetraloop in perfect position for stacking in the minor groove of the receptor. In addition, the second position of the GNRA tetraloop is able to stack on a platform and make additional hydrogen contacts, thus generating a recognition module. Due to the nearly identical topology of the H89 and S8.a1 receptors, it is plausible that S8.a1 may be a suitable model for our re-engineered H89 receptor.

2.3.2 ***Other GNRA-like receptors: P12.a1 and H68.a1***

Other natural receptors bind non-GNRA loops and do so with different structural platforms. Two such interactions involve binding a triloop (H68 from the 23S rRNA) and a pseudo-triloop involving a U-turn motif (P12 from RNaseP). Both of these interactions take place between the UAA nucleotides in the terminal loop, indicating the potential for generating an interaction with a GUAA tetraloop. In order to further demonstrate the modularity of these types of interactions, we looked at P12.a1 and H68.a1 (P12 and H68 in an inverse orientation to that of the natural receptors) (Figure 2.2A and B (f)). The P12.a1 positions the receptor in an orientation similar to that of S8.a1 and H89. The G/U wobble is in the same position, as well as the number of nucleotides contained within each bulge (1 on top and 2 on bottom).

H68.a1, on the other hand has a nearly identical sequence to that of the stacking platform of S8.a1 with the addition of an extra adenine in the bulge (Figure 2.3A). In the crystal structure of H68, this adenine is flipped away from the loop/receptor interaction and thus is hypothesized to not interfere with binding in the H68.a1 model [107]. While the natural P12 and H68 orientations recognize very different loop sequences, the overall topology of their loop/receptor interactions are remarkably similar to one another and are likely to behave in a similar fashion [107]. It was predicted that while these receptors are most likely capable of assembling with GNRA tetraloops, they will not interact as well as the S8.a1 receptor due slight differences in the topology of the P12.a1 and H68.a1. The S8 receptor (as well as the H89 receptor) utilizes a cis shallow groove:Hoogsteen interaction between the single nucleotide bulge (A) and the WC base pair directly adjacent to it, A:U (Figure 2.2A (d)). The P12 and the H68 receptors on the other hand contain interactions between the bulge sequences and the WC base pair that is not directly adjacent to the bulge, where both bulges contain contacts with the WC base pairs between the bulges. This slight difference in interacting nucleotides shifts the platform sequence of P12 upward (toward the A-minor interaction), which would create a steric clash with the 2^{nd} position of the GNRA tetraloop (Figure 2.4B (e)). Due to the inherent flexibility of these receptors, we would expect the P12.a1 and H68.a1 sequences to interact with the GNRA tetraloop, but with lower affinity.

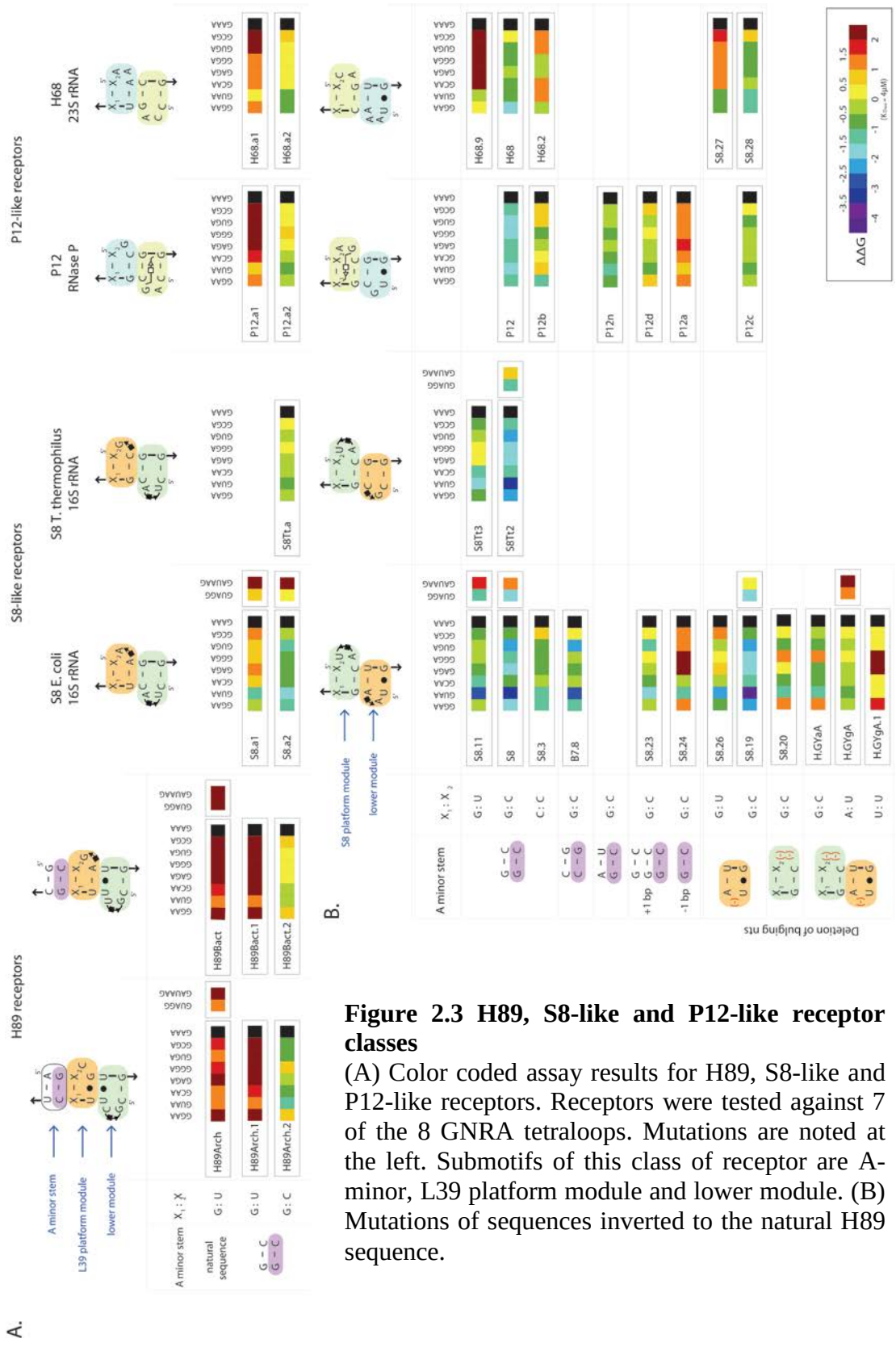


Figure 2.3 H89, S8-like and P12-like receptor classes

(A) Color coded assay results for H89, S8-like and P12-like receptors. Receptors were tested against 7 of the 8 GNRA tetraloops. Mutations are noted at the left. Submotifs of this class of receptor are A-minor, L39 platform module and lower module. (B) Mutations of sequences inverted to the natural H89 sequence.

2.3.3 *Additional modeling considerations*

The initial modeling of S8.a1 was done in such a way that the symmetry of the sequences and the position of the G/U wobble were identical to that of H89 (Figure 2.2A (c and d)). However, the sequence of S8 that truly interacts as a platform with the S8 protein is contained within the lower module of the S8.a1 model, the dinucleotide sequence UA (Figure 2.1B). To model S8 in an alternative orientation, which places the interacting platform of the S8 protein in the same position as a docking platform of a GNRA tetraloop, we inverted the sequence to yield an additional receptor essentially swapping the platform modules and replacing the G/U wobble with a Watson Crick G-C (S8) (Figure 2.2A (e)). The S8 receptor contains a 2-fold symmetry and it is for this reason that both orientations of this receptor are expected to bind well to the GUAA tetraloop. This receptor (S8) mimics the previous one, such that the GNRA tetraloop is able to properly dock in the minor groove and the orientation of the platform is in an optimal position for stacking (Figure 2.2B (e)). Both orientations of the S8 model offer potential candidates for optimal GNRA receptors. However, the S8 orientation perhaps may have a slightly better stacking with the GUAA loop due to the fact that in the S8.a1 model (L39 platform), the stacking occurs over the ribose of the adenine in the two base pair stem, while in the S8 model the second position of the loop is able to stack directly over the nucleotide base of the adenine in the two nucleotide bulge, which is more energetically favorable (Figure 2.4A-C (a and b)). This type of stacking is very similar to an interaction that takes place between the 11nt receptor and the GAAA loop.

Additionally, the P12.a1 and H68.a1 receptors can also be oriented in the inverse direction, which would place their sequences in a manner consistent with the natural receptors and their cognate loops, P12 and H68 (Figure 2.2). From the crystallographic structure of P12, it would appear that an interaction with a GNRA tetraloop, would not properly accommodate the second position of the loop because of the spacing between the platform module, thus creating a steric clash (Figure 2.4B and C (e)). In this instance the precise spacing of the loop/receptor interaction is slightly different from S8 due to an alternative hydrogen bonding pattern that shifts the platform up by approximately 1 base pair. The platform module in this instance demonstrates hydrogen bonding patterns that overlap the upper and lower modules, essentially linking the two. This results in an alternative backbone topology and indicates that the upper and lower modules may not be as interchangeable as the S8 modules, which do not interact with one another and can act as separate entities. The fact that P12 interacts with a tri-loop rather than a tetraloop would further support this notion of steric hindrance on the second position of a GNRA tetraloop, since a triloop does not interfere with the platform nucleotide (Figure 2.4B and C (d)). At second sight, due to the flexible nature of the helical stems, it is quite possible for the A-minor interaction to shift up one base pair (with the proper G-C), thus allowing for the stacking of the loop on the platform rather than sterically impeding the interaction (Figure 2.4B and C (f)). This orientation, however inhibits the second position of loop from stacking on the platform due to the shifting of the A-minor interaction.

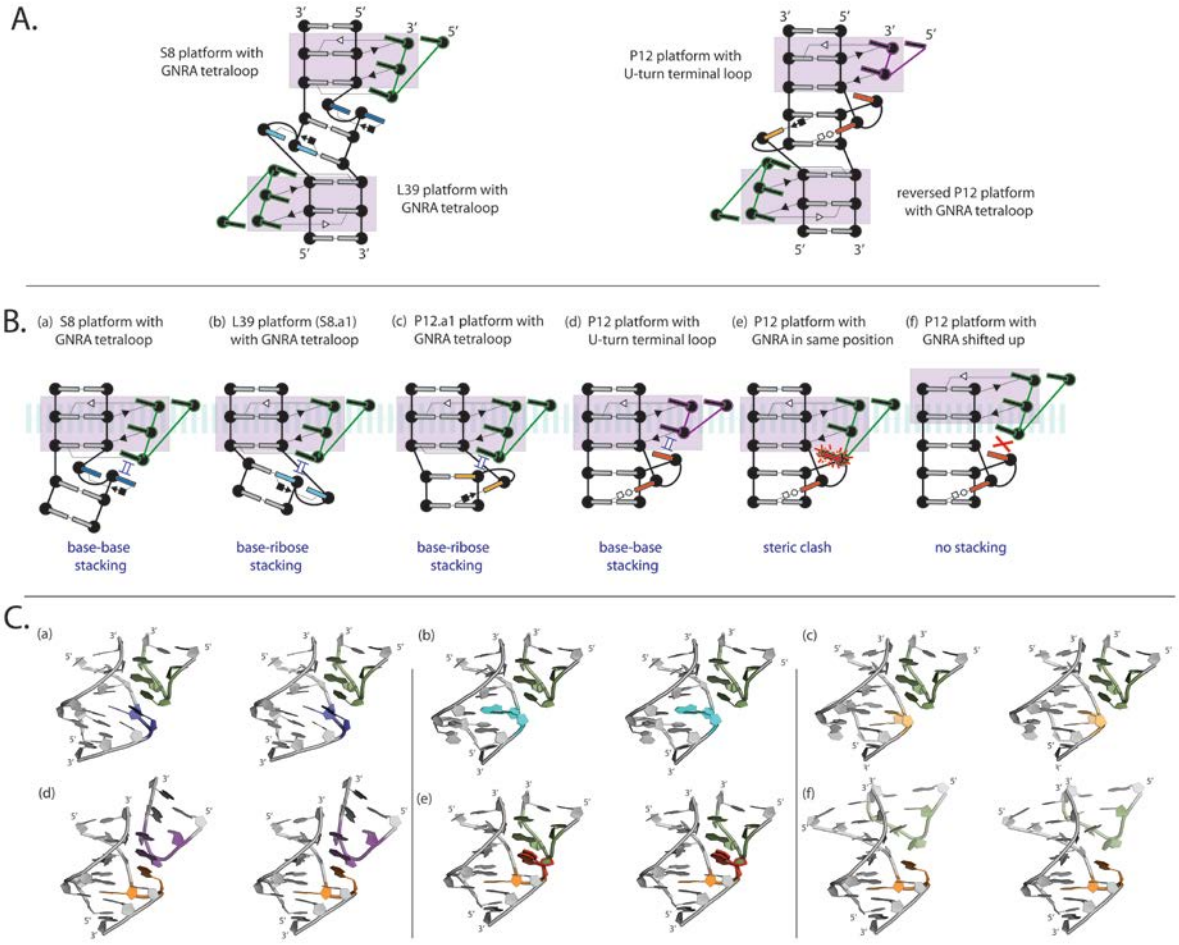


Figure 2.4 Comparison of S8 and P12 platform Structure

(A) 2D image of each platform (upper and lower) for S8 and P12. Orientation of the platform is slightly altered depending on the interaction of the nucleotides with the base pairs within the motif. The S8 platform is a dinucleotide sequence that interacts side-by-side via a cis shallow groove:hoogsteen interaction. The S8 reverse platform (also known as the L39 platform) is a single nucleotide platform that interacts with the base pair directly adjacent to it, also with a cis shallow groove:hoogsteen interaction. The P12 platform is a dinucleotide sequence where the platform is stabilized by the 5' nucleotide interacting with the second base pair in a trans Watson Crick:hoogsteen orientation. The reverse P12 is a single nucleotide interacting the second base pair via cis shallow groove:hoogsteen interaction. (B) Interactions as they would occur in the tectodimer context. (a) base-base stacking occurs with the S8 platform as opposed to base-ribose stacking in the L39 platform (b). The reverse p12 platform (c) stacks over a ribose with a GNRA tetraloop, the P12 platform can accommodate a base-base stack but only with a triloop (d). In this context, a GNRA tetraloop interaction would create a steric clash between the platform and the second position of the GNRA tetraloop (e). The flexibility of the tectodimer could accommodate a GNRA tetraloop, however, the twist that would occur as a result would eliminate the stacking interaction. (C) Molecular models for each other interactions described in (B).

2.3.4 ***Repairing the A-minor of H89 sequences***

Initially looking at the sequences of the H89 receptors, it is evident that the secondary structure is formed most likely by post-transcriptional modifications due to the lack of cognate interactions. Comparatively, the S8 receptor has traditional pairings that can be predicted by simply observing the primary sequence. Utilizing the components of the other natural receptors with structural homology to H89 (S8, P12 and H68), we set out to rationally design a GUAA receptor with comparable binding affinities to that of the 11nt/GAAA interaction. To improve upon the binding affinity of the H89 receptors to GNRA tetraloops in the tectoRNA context, we started by repairing the three base pairs in the stem involving the A-minor interaction by replacing them with sequences of natural helical receptors. The A-minor stem sequence of the archaea and bacterial H89 receptors contains the sequence 5'-AGU…GCU-3' and 5'-GCU…GGC-3', respectfully. H89arch.1 and H89bact.1 were replaced with the sequence 5'-CCU…GGG-3' and H89arch.2 and H89bact.2 were replaced with the sequence 5'-CCC…GGG-3' (Figure 2.3A). Previous characterization of A-minor interactions also demonstrated that the XGG:CCX confers the best A-minor sequence with the receptors cognate loop [72]. The results of this simple change were remarkable in that we saw almost a 50 fold decrease in K_d for GUAA. Additionally, the profiles for H89arch.2 and H89bact.2 have a nearly identical profile to a helical receptor containing the 5'-CCC…GGG-3' (H.GYaA) sequence in the A-minor (Figure 2.3A and B). Being that the H89 is unable to adopt a 2x_bulge motif without post-transcriptional modifications, it is

likely that the H89 receptor with a repaired A-minor stem would act much like a helical receptor.

2.3.5 *K_d and specificity profiles for receptors with orientation comparable to H89*

Both H89 receptors were tested in the tectodimer context with the GNRA loops as well as the L39 loops from both archaea and bacteria (Table 8.1-Table 8.3 and Figure 2.3A). As previously demonstrated, the binding affinity toward any loop was either very weak or undetectable, even with the cognate loop sequence (not GNRA). Within this context we tested both the S8.a1 receptor (from *Escherichi coli*) and the S8Tt.a sequence from *Thermus thermophilus*. S8.a1 was tested with the 7 GNRA tetraloops as well as the L39 loops from *H. marismortui* and *E. coli.* The profile for this receptor is clearly selective for the GUAA tetraloop, as expected, with a K_d of approximately 350nM, but the affinity is lower than the 11nt/GAAA interaction, which has a K_d of approximately 1nM in this context (Table 8.3) [72]. With the exception of GGAA, the rest of the loops had K_d's in the uM range (20-40uM). Additionally, this receptor did not bind well to the L39 loops, with the archaea loop having a K_d around 12uM and the bacterial loop showing no detectable binding. S8Tt.a, which differs from S8.a1 in the stacking module having a G-C closing base pair rather than a G/U wobble and a guanine bulge rather than an adenine (Figure 2.3A), showed very little selectivity toward GUAA, with K_d's ranging from 750nM for GUAA to 4400nM for GCGA.

We then tested the P12.a1 construct, which as predicted did not bind well with any of the GNRA tetraloops. Although, it showed a slight preference for GUAA, all

K_d's were in the uM range. Likewise, the H68.a1 receptor showed selectivity for GUAA, but the K_d was well above that of S8.a1 at 7900nM. This data indicates that the three classes of 2x_bulge receptors (S8, P12 and H68) are unique receptors despite seemingly similar secondary structure.

H89 contains a G/U wobble base pair directly below the A-minor interaction, as does S8.a1, P12.a1 and H68.a1. Other studies of artificially selected receptors have shown that a G-C in this position is optimal for loop binding [72] and Calkins (2015-in preparation). In order to test the effects of the base pair below the A-minor, we changed the G/U wobble to a G-C base pair in all three constructs, S8.a2, P12.a2 and H68.a2. The profile for S8.a2 drastically changes by slightly decreasing the K_d for GUAA, 220nM, but greatly decreasing the K_d for every other loop making the profile seemingly non-selective for GUAA (Figure 2.3). Contrastingly, this mutation for both P12.a2 and H68.a2, while greatly improving the binding affinity for all loops, did not affect the loop specificity profiles. This is further indication that the P12 and H68 receptors can be categorized as a separate class from the S8-like sequences.

The S8.a1, P12.a1 and H68.a1 receptors were initially designed so that their modular components were in a similar orientation to that of H89 (i.e. position of the G/U wobble and the number of nucleotides in the bulges). However, the orientation initially tested is not the natural one for the receptor in context with its binding partner. Having already demonstrated that the S8.a1 receptor shows affinity for the GUAA tetraloop in the tectodimer context, we further explored the receptors potential by inverting the sequences to restore them to their natural orientations (S8,

P12 and H68) (Figure 2.3B). This is further justified by the fact that there is symmetry within the topology of the receptor, where either orientation could potentially selectively bind GUAA. The results for S8, P12 and H68 were remarkable in that S8 had a K_d for GUAA of 19nM, which is more than 10 times better than S8.a1 and S8.a2. However, all loops showed a dramatic increase in binding affinity. P12, showed a 100 fold decrease in its K_d for GUAA to 275nM, but showed absolutely no selectivity, binding all loops with approximately the same K_d. H68 showed a 10 fold decrease in K_d for GUAA (688nM), however its selectivity shifted toward GGAA, with a K_d of 487nM. This new insight to the behavior of the three distinct modular receptors offered us an avenue for exploration and optimization of these GNRA receptors. In addition, with these three receptors interacting with all loops containing detectable binding affinities we were able to use these receptors as phenotypic baselines for categorizing receptor mutants, which is also a valuable tool for obtaining clues to receptor structure.

2.3.6 ***Flexibility of interactions (testing steric clash of p12)***

Knowing there is some inherent flexibility to helical stems, we wanted to explore the optimal stem length for these interactions. The 11nt/GAAA interaction acts to anchor the tectoRNA heterodimer such that the receptors being tested are in the proper orientation for binding the GNRA loop. The helical rotation of the monomer in relation to the GNRA tetraloop is an important design criterion for optimal binding. The S8 crystal structure shows the dinucleotide platform interacting with the S8 protein to be independent of the lower module of the receptor. Contrastingly,

the P12 receptor shows a network of hydrogen bonding between the upper and lower modules, which alters the receptors backbone slightly by shifting the platform up by approximately one base pair (Figure 2.4A). It is for this reason we hypothesize a smaller margin for variability in stem length for the P12 receptor when compared to S8 (Figure 2.4B and C). To test this, we created two mutants for both S8 and P12, each with an increase in stem length by one base pair (S8.23 and P12d) and a decrease in stem length by one base pair (S8.24 and P12a) (Figure 2.3B). For the S8 sequence, the results show a tolerance for stem length in either direction, with K_d's for GUAA of 19nM (S8), 133nM (S8.23) and 631nM (S8.24), where the optimal stem length appears to be 11bp from the GNRA tetraloop to the A-minor base pair, which is consistent with previous data obtained from this tectodimer system [39]. The P12 sequence on the other hand, was greatly destabilized by a decrease in stem length (P12a), with a K_d of 15uM for GUAA. Increasing the stem length (P12d) altered the affinity toward GUAA, but only slightly, increasing the K_d from 275nM (P12) to 734nM (P12d). These results indicate that perhaps the optimal stem length for the P12 receptor would between 11 and 12 base pairs. This would substantiate the model where shortening the stem would further create a steric clash between the second position of the GNRA tetraloop and the stacking platform module. This also confirms that the optimal stem length for S8-like and P12-like receptors is 11 base pairs from the loop to the A-minor interaction. Interestingly, P12d strongly resembles the profile for a helical receptor. In fact, the secondary structure of P12d can be rearranged to a helix. The sequence for P12 is nearly identical to P12d, with an insertion of a G-C base pair below the A-minor stem, however the loop profiles

for these two receptors do not show any similarity. Since P12 is also capable of rearranging into a helical structure, it is plausible that the docking of the loop into the receptor offers structural stability of the 2x_bulge conformation. This induced fit model has also been proposed for the docking of the GAAA loop and the 11nt receptor [122].

The P12n mutant was created to analyze the effects of the base pair above the canonical base pair involved in the A-minor interaction. Instead of the tandem GG:CC pairs, the A-minor stem was changed to GA:UC. The reasoning for this mutation is that the P12 interacting modules place the docking platform approximately 1 base pair higher, which in the tectodimer system would place the loop at or near the base pair above the typical A-minor interaction. By changing the G:C to a A:U, this mutation is thought to destabilize the GNRA/receptor interaction. In fact, the resulting K_d's for all the loops are indeed higher than for the P12 interaction. In addition, the loop specificity is minimally altered. This is a good indication that this base pair is involved in the A-minor/docking of the loop and not for specificity.

2.3.7 ***Modularity Characterization***

The S8 and H89 receptors each have two potential platforms that could be used to dock a GNRA tetraloop. We have demonstrated that each of these receptors contains a certain symmetry that allows for each to be inverted and still remain functional. By inverting the geometry of the receptor, the platform that is able to interact with the GNRA tetraloop has now been swapped (upper to lower). Figure

8.2 shows various S8 and H89 receptor modules and highlights the formation of the platform that is found. Each of these modules represents natural platforms where the atomic structure is available. In our study, we have utilized these platforms in our tectodimer assays and have characterized each with respect to S8-like and non-S8 like contexts.

2.3.7.1 Module mutations

Since the orientation of our optimized receptor utilizes the S8 platform module to interact with the GNRA loop and the L39 platform module to stabilize the fold (the inverted form of the natural H89 receptor), we will refer to the upper module as the S8 platform module and the lower module as the L39 platform module. In order to better understand the effect of the platform on the affinity and specificity of binding, we generated a series of mutants by altering the dinucleotide bulge sequence (Figure 2.5a). To establish a baseline for the significance of the bulge platform we tested a helical stem with no bulges (H.GYaA) (Figure 3B). This receptor had a typical GYRA binding pattern as seen in natural helical receptors, with a K_d for GUAA of 404nM. We also tested a receptor (S8.20) that deleted the platform, but kept the lower L39 module, which showed nearly an identical profile to that of the helix. Of the platforms tested, the AA platform generated a receptor that not only increased the K_d for the GUAA loop from the S8 sequence (7nM), but had high affinity for all the GNRA loops, as well as the Archaea loop, GUAAG (Figure 2.5a). Strikingly, the other platform sequences tested also had very low K_d's for all the GNRA loops, with the exception of S8.16 (UC platform). This platform sequence

drastically diminished the receptors ability to bind to GGRA loops, as well as GCGA. Figure 2.5 shows that, while the dinucleotide platform sequence does prefer binding a GUAA tetraloop, the profiles from one platform to the next are quite different with respect of the other GNRA tetraloops. This is further evidence in support of this platform module being utilized for selectivity [72].

Interestingly, CA was the only the platform that altered the specificity from GUAA to GGAA. This platform is the natural one found in the H68 receptor, which is also binding to GGAA better than GUAA. Figure 8.3 shows mutations from S8 to H68 via two different modular pathways that demonstrates several key points. First, the shift in specificity toward GGAA is a direct result of the change in platform sequence from UA to CA. In both pathways, the mutations occur in a different order and it is not until the mutation from UA to CA occurs that you see the shift in selectivity. Secondly, the change in base pair sequence in the stem of the stacking module from a G-C to C-G has very little effect on the receptor profile and binding affinity. Thirdly, because of the 6 fold decrease in affinity for both GUAA and GGAA in the mutation from S8.4 to H68, the bulge in the lower module prefers one A as opposed to two.

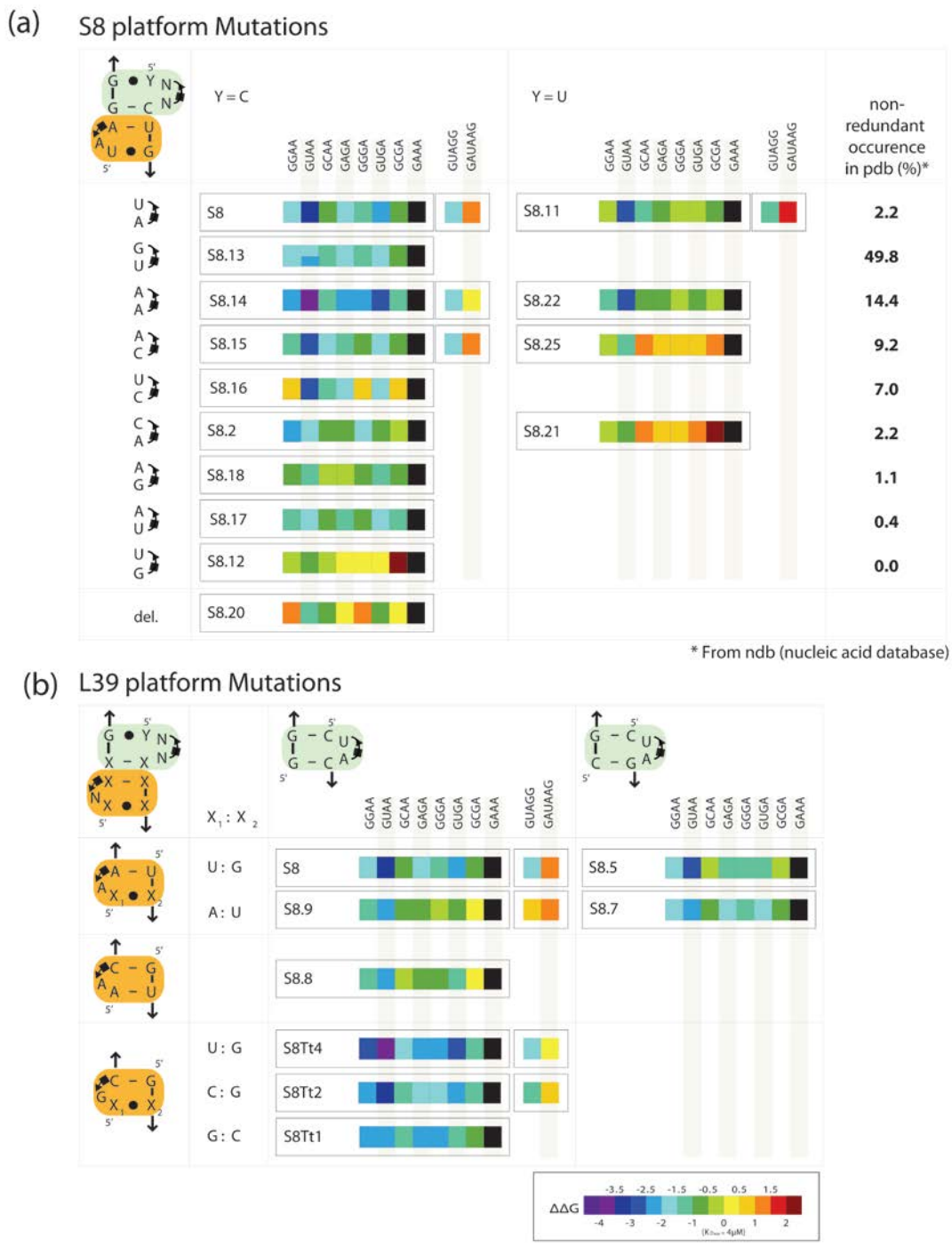


Figure 2.5 S8 and L39 Platform Mutations
(a) S8 dinucleotide platform mutations, including the non-redundant occurrence of naturally found platforms. Some mutants were also tested with both a G:C or a G:U base pair directly under the A-minor interaction. (b) L39 (lower module) mutations with different Watson Crick covariation.

We also compared the effect of the base pair just below the A-minor for platform sequences UA, AA, AC and CA (Figure 2.5a). We altered the last base pair by one point mutation changing the interaction from a G:C to a G:U. This mutation, while lowering slightly the K_d for GUAA, drastically increased the K_d for all other loops making a receptor with a G:U wobble more selective for the GUAA tetraloop. Additionally, according to the X-ray structure of the L39/H89 interaction of Archaea, it is noteworthy that a G:U base pair in this same position might be of some benefit for increasing the specificity of recognition of the GUAA tetraloop versus other GNRA loops. In fact, most H89 sequences have a G:U (or A:C) bp at this position.

Focusing on the S8 platform module, we wanted to see the effects of disrupting the base pair interaction just above the platform sequence. According to the crystallographic structures of P12 and H68, the A-minor interaction occurs lower than in S8, therefore a disruption of this base pair is expected to have a more detrimental effect due to a possible distortion of the A-minor interaction. To demonstrate this disruption we mutated the closing G-C base pair just below the A-minor to a C-C (S8.3, H68.2, P12b). The P12 and H68 receptors showed the greatest increase in K_d toward GUAA from 275nM to 10uM and 688nM to 25uM, respectively (Figure 2.3B). The S8 receptor (S8.3) had only a slight increase from 19nM to 487nM. These results would indicate the A-minor interactions of P12 and H68 are similar, whereas S8 interacts with the A-minor in a slightly alternative manor.

In altering the closing base pair in the lower module, the effect seen was minimal. Profiles for the most part were consistent with one another, where affinity was slightly altered. A U:G base pair seemed to confer the strongest affinity for GUAA when compared to A:U, C:G and G:C (Figure 2.5b) This would indicate that the upper and lower modules for the S8 class of receptor act independently of one another and only confer minor stabilization or the receptor /loop interaction.

2.3.7.2 Interchanging modular subunits

One of the main topological differences between the S8-like and P12-like receptors is dictated by the interaction between the upper and lower modules. The S8 structure shows the lower module acting independently from the stacking module (Figure 2.2A and Figure 8.4). The crystal structures show the stacking interactions between the upper and lower bulge sequences interact quite differently between the S8/H89-like fold and the P12/H68-like fold.

Figure 8.4 shows various natural bulge stack interactions of S8, P12 and H68. The S8 and H89 stacking interactions show that the platform for the second position of the GNRA tetraloop is formed by a side-by side cis Hoogsteen:Shallow Groove interaction. The P12 and H68, stacking platform is alternatively formed by a trans Hoogsteen:Watson Crick interaction with the base pair of the second module. This type of interaction, as previously noted pushes the platform up towards the A-minor interaction by approximately one base pair, thus creating a potential steric clash between the second position of the GNRA and the platform. There is further stabilization of the receptor by an additional interaction between the lower bulge and

the base pair of the upper module. It is for this reason, we hypothesized that mutations in the lower bulge sequence would have a greater effect on the phenotype of P12-like receptors. Receptors S8.19 and P12c are mutants of S8 and P12, respectively that remove the bulge in the lower module (Figure 2.3B). The results indicate that indeed, the secondary structures of S8 and P12 follow the predicted models, where the phenotype for P12 is greatly altered by the removal of the lower bulge, whereas S8 in fact maintains in phenotype. Interestingly, S8.19 slightly improved the K_d's for all the loops. This might suggest that the bulge in the lower module for S8-like receptors may have a destabilizing effect on the platform, perhaps because a structure with no lower bulge has no alternative pairings. This would suggest that the equilibrium between different secondary structures affects the ability of the tetraloop to bind to the receptor. P12c, on the other hand, not only altered the phenotype, but additionally destabilized the interactions with all the tetraloops. This further validates the 3D model and the importance of the interaction between the stacking bulge and the lower module of the P12 receptor.

Various mutants were created by interchanging the modular subunits from the S8, P12 and H68 receptors (Figure 2.66). There were very few P12-like receptors, as mutations in the upper or lower module resulted in S8-like receptors. S8-like behavior, on the other hand was much more prevalent indicating that this type of fold was much more robust as the individual modules acted in a more independent fashion. Interestingly the lower module of H68 when combined with the S8 upper module, had very different phenotypic behavior than S8, indicating that the additional adenine in the lower bulge affects the upper module.

2.3.7.3 Possible alternative folds

We originally had not tested the P12 receptor with a CA platform. Interestingly, one of the stem length mutants of P12.14 (P12h) showed a shift from GUAA to GGAA (Figure 2.7A). At first glance, this was unexpected because this receptor has an AC platform, which binds GUAA with greater affinity in both P12 and S8 contexts. However, upon closer examination, it is hypothesized that this receptor could contain an alternative fold (Figure 2.7A). The S8 receptor is also capable of adopting an alternative fold. In fact, Costa and Michel (1997) represented their selected B7.8 (which is nearly identical to S8) receptor as the lowest free-energy structure, which is not the S8 fold. This prompted us to look at alternative folds within these receptors. This alternative pairing in the stem results in a P12 receptor with a CA platform and has a profile that preferentially binds to GGAA. This further verifies that a CA platform (of H68) in this context creates a receptor that preferentially binds GGAA. We also looked at other receptor mutants that seemed to bind better to GGAA and took a closer look at the possible secondary structures that could form. P12b is another example where GGAA had the lowest K_d and in fact it could fold into a conformation with a CA platform (Figure 2.7B). S8.4 is an S8 mutant with a CA platform that also preferentially binds GGAA, however does not contain an alternative fold (Figure 2.7C).

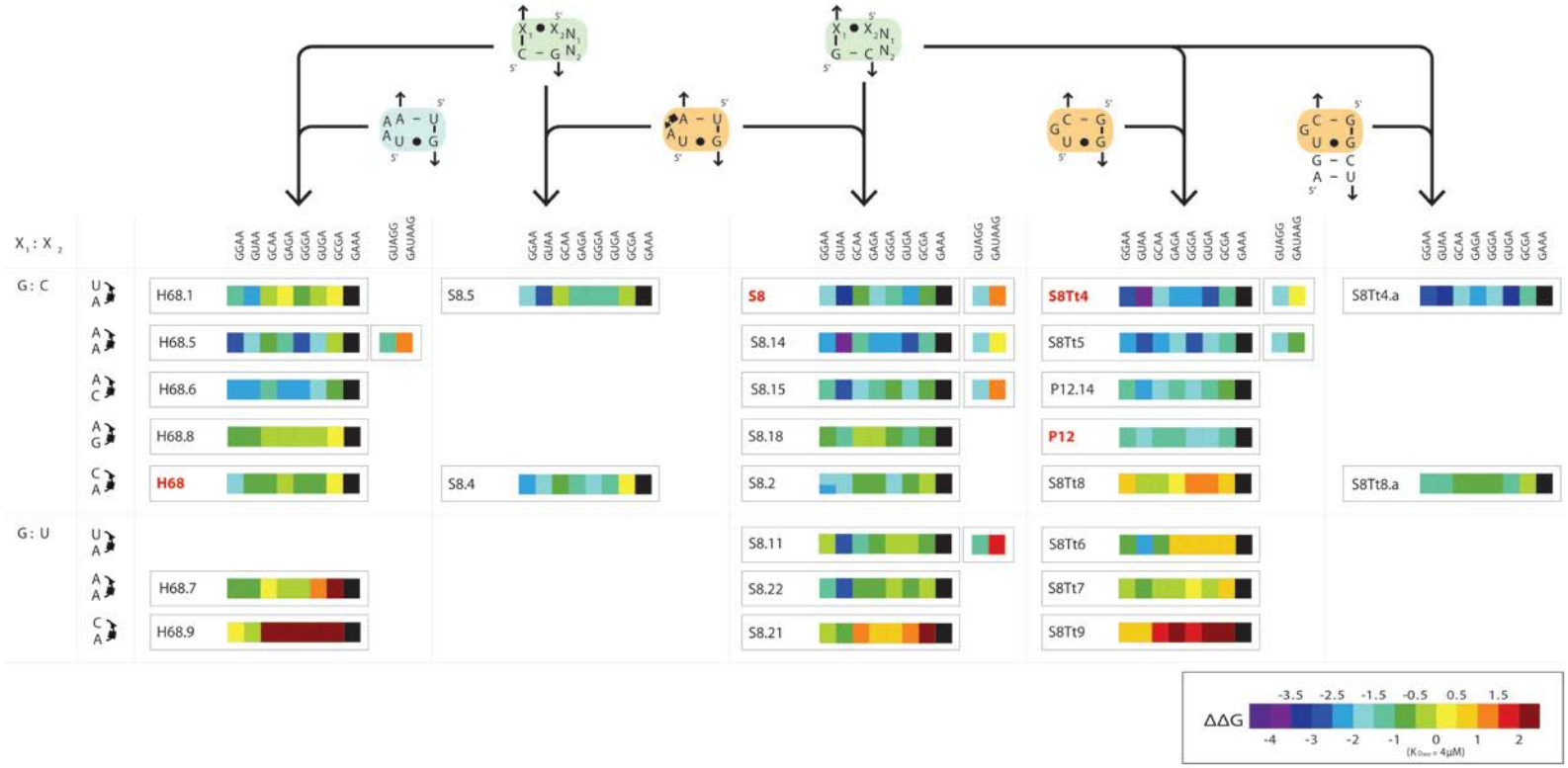


Figure 2.6 Modular component exchange
Mutants were generated by exchanging modular components from each of the three classes from this study (S8, P12 and H68). Modules in green (S8) are directly interacting with the GNRA tetraloop, whereas modules in either blue (H68) or orange (P12) are offering a distal effect (either stabilizing or destabilizing the interaction).

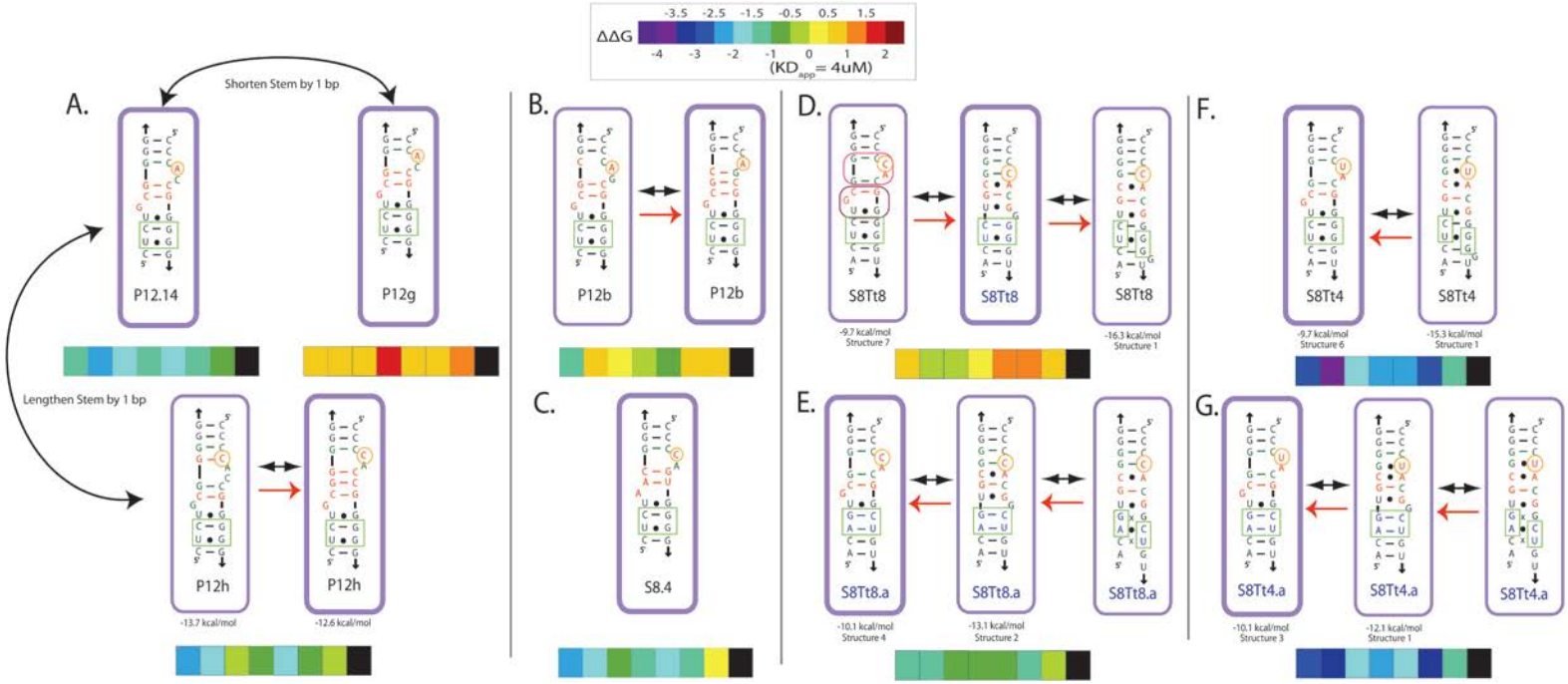


Figure 2.7 Alternative secondary structures

(A) Stem length mutants of P12.14 had unexpected results. Lengthening the stem (P12h) shifted selectivity toward GGAA, which was characteristic of CA platforms. Looking at the sequence, it is possible for there to be a rearrangement of the secondary structure to yield a receptor with a CA platform. (B) P12b was another instance where phenotypic behavior, which altered selectivity toward GGAA, prompted us to redraw the secondary structure to yield a CA platform. (C) S8 mutant with CA platform, also has selectivity for GGAA. (D) S8Tt4 was another S8 sequence from *Thermus thermophiles* that also had a CA platform, but did preferentially bind GGAA. It was noticed that this receptor could adopt an alternative fold, generating a helical stem. The profile for this receptor matched that of a helix (H.GYaA). (E) S8Tt4.a was a distal stem mutant that prevented alternative pairing, this mutation repaired the 2x_bulge motif. (F) S8Tt8 was a control for altering distal stem sequences. The same mutations in this receptor did not change the secondary structure. The predicted thermal stability values were calculated with quickfold (UNAfold) with RNA parameters 2.3 (http://mfold.rna.albany.edu/?q=DINAMelt?Quickfold

We then generated a P12 receptor (with normal stem length) with a CA platform (S8Tt8), to test whether its behavior mimicked that of P12h. Contrastingly, it demonstrated an unpredicted behavior that seemed to destabilize the loop/receptor interaction for all loops. S8Tt8 not only had a preference for the GUAA loop, but the binding affinity for GGAA was greatly destabilized, with a K_d of 12.3 uM. Upon closer examination of the primary sequence for S8Tt8, it was noticed that this receptor could also potentially adopt an alternative fold to yield a helical stem (Figure 2.7D). Likewise, the profile for this receptor is strikingly similar to that of Helix GYaA (Figure 2.3B). It was noticed that the stem sequence below the receptor of S8Tt8 promoted the formation of a helix over the predicted 2x_bulge secondary structure. To test whether the stem sequence had this predicted affect, we altered the primary sequence of the lower stem to drive the formation of the 2x_bulge structure over the helix (S8Tt8.a) (Figure 2.7E). This mutant not only restored the interaction between all GNRA tetraloops, but had a profile very similar to that of P12, with GUAA and GGAA tetraloops having comparable K_d's (370 nM and 529 nM, respectively).

Another receptor, S8Tt4, differs from S8Tt8 by one nucleotide in the platform module, however the profile behaves nothing like a helical receptor (Figure 2.7F). We altered the lower stem sequence of this receptor as well (S8Tt4.a) to see if the effects of our change were due to an alternative folding pathway (Figure 2.7F and G). This mutant had very little effect on the binding affinity and profile. Interestingly, UNAfold was used to calculate the energies of the predicted secondary structures for both S8Tt8 and S8Tt4. The program predicted, in both instances that

the helical structure would have an overall lower free energy when compared to the 2x_bulge structure, indicating that the sequence is most likely to adopt a helix in solution. The GNRA loop profiles however, do not confirm this prediction. A comparison of the loop specificity profiles, demonstrates that the S8Tt4 receptor is most likely adopting a non-helical structure since both S8Tt4 and S8Tt4.a share similar profiles. Note that calculated free energies for the helix are still lower than that of the 2x_bulge structure, however these programs do not always accurately predict the actual experimentally determined structure. This could be due to limitations in the parameters used to calculate the thermodynamically stable secondary structure or there is likely an induced fit mechanism at play.

We have already discussed the potential for the local secondary structures within these receptors sequences to transition between folds (S8Tt4, P12h, P12b, P12d and S8Tt8). To further demonstrate this phenomenon within these loop receptor classes, we mapped the phenotypic behavior in a sequential manner, with respect to its sequence. Figure 8.5 shows the mutational pathway from P12 to S8Tt2 demonstrating a single mutational connection between these natural sequences. Three point mutations separate the two types of receptors, with each demonstrating very different phenotypic behaviors. The specificity profiles as you walk through the sequence space suggest that perhaps some of the sequence variants may have the capability of folding in either conformation (P12-like or S8-like) depending on the context of the GNRA tetraloop. This is exemplified by the loop profiles for S8Tt5 and S8Tt4 although different, they contain the same sequence in the lower module and differ by only one point mutation in the stacking module (Figure 8.5). This

single point mutation shows a phenotype that seems to overlap the profiles for both S8 and P12. S8Tt4 has an identical profile to that of S8, where S8Tt5 is similar in profile to S8 and S8Tt4 with all loops but GAGA and GGGA, and instead mimic the profile for P12 with respect to these two loops. Being that neither profile is fully represented in the intermediate S8Tt5, but similar in some respects depending upon the context of the loop, this could signify a propensity for this receptor to fold in either orientation and adopt a fold upon docking of the loop. The thermodynamic stability of the local secondary structure most likely drives this behavior, in that lower stability at the local level will be induced by the docking of the loop and confer an overall global stability that dictates the phenotypic behavior.

2.4 Conclusion

Overall, this study further demonstrates one of the key fundamental principles that govern RNA functionality; the function of RNA is a result of structural elements that are not entirely dictated by their overall primary sequence. While motifs generally have sequence constraints, there is a certain flexibility at this level that allows for optimization, as long as the structural topology of the motif is maintained. By examining the 3D structures of the H89 and S8 receptors, we were able to isolate structural subcomponents and systematically alter the sequence of each module, while maintaining the overall organization of the receptor. The basis for our rational design began with the notion that structural homology could lead to similar phenotypic behaviors. While the sequences of H89 and S8 do not share any similarity, their structural topology does. We were able to demonstrate that the

phenotypic behavior (i.e. loop specificity and affinity), can be used as a tool to indicate different classes of receptors with similar secondary structures. It is remarkable the amount of sequence space we were able to explore and still generate a library of functional RNAs toward the GNRA tetraloops. Furthermore, this rational design strategy can be applied to other suboptimal structures to engineer novel higher ordered structures.

The purpose of our research was to rationally design a tetraloop receptor that is selective for GUAA, as well as shows a high affinity for this tetraloop. Looking at our results, the stacking platform sequence that generates the best binders for GUAA are UA and AA. S8.19, S8Tt4, S8M14, S8Tt2 and S8.11 have K_d's for GUAA of 1nM, 5nM, 7nM, 16nM and 21nM, respectively. From this it is clear that the sequence of the lower module influences the binding of the tetraloop, being that S8.19 removes the nucleotide bulge in the lower module. This result most likely occurs due to the ability of the alternative secondary structures to form as a result of the lower bulge. S8Tt4 with the G-U wobble closing base pair binds to GUAA with the greatest affinity, a K_d of 5nM. Based on our research the optimal rationally designed GUAA tetraloop receptor has proven to be an excellent candidate for a long-range tertiary interaction module. Several sequences with the 2x_bulge topology from our study are capable of binding GUAA tetraloops with high affinity and specificity. While the GAAA/11nt receptor interaction is the most abundant in nature, it is clear that many other tetraloop/receptor combinations are possible, with comparable binding affinities and specificities. Despite this finding, our results indicate that there is less specificity than found in the naturally occurring 11nt

receptor and the artificially selected R1 [72]. One possible explanation for this can be seen in Figure 8.6. While S8-like receptors are selective for GUAA, the L39/H89 interaction suggests that the second position of the GNRA tetraloop and the receptor could be mediated through the coordination of a Mg^{2+} ion, as opposed to direct base-base contact. A uracil in the second position of the GNRA could offer a more favorable coordination geometry, which explains the moderate selectivity profile of S9-like receptors.

Many such receptors have arisen from *in vitro* selection experiments, which have shown to bind GNRA tetraloops other than GAAA. Interestingly, Costa and Michel discovered the B7.8 receptor in an *in vitro* selection for GUGA tetraloop binders. After characterization of this receptor they realized B7.8 had a higher affinity toward the GUAA tetraloop. Upon closer inspection of the primary sequence of B7.8 it was noticed that this receptor was strikingly similar in sequence to that of S8, only differing in the A-minor sequence. Figure 8.7A show the Costa and Michel structure of the B7.8 with a three base pair bulge in the middle of the helical stem, as opposed to the 2x_bulge structure. This proposed structure is also the predicted secondary structure for this sequence as indicated by Mfold data. However, based on the crystal structure of S8 (Figure 8.7D) and our findings, we can redraw the secondary structure of B7.8 to contain the 2x_bulge motif. The thermodynamic energy of the 2x_bulge structure is 3.8kcal less stable than the most stable structure depicted by Costa and Michel. Additionally, a R5.58 receptor was selected for GUAA binding and was also found to adopt several alternative secondary structures, including the 2x_bulge (Figure 8.7B) (Calkins 2015, in preparation). Being that the specificity

profiles and K_d's for B7.8, R5.58 and S8 are nearly identical, this would indicate that the 2x-bulge is a robust structural motif preferred for binding GUAA tetraloops (Figure 8.7C) It is probable that in the absence of the cognate loop, its 2D structure is likely metastable. It is important to recognize that the interaction between the receptor and the tetraloop is directly affected by the equilibrium between the alternative folds. B7.8 and S8, while S8 is a slightly better binder for all of the loops, have identical specificity profiles, indicating that S8 too may be in equilibrium between alternative secondary structures.

We were able to demonstrate structural evidence between S8-like receptors and P12/H68-like receptors. There is a delineation between these two classes that can clearly be seen when analyzing loop specificity profiles and bar codes in comparison with sequence networks. Mutations within each class show that deviation from the parent receptor sequence can be emphasized through such analysis and receptors can be classified based on loop selectivity profiles and not sequence similarity. Based on our research, we have concluded several classes of receptors utilizing this framework. To highlight some of these classifications, one can look at the profiles for the pathways linking the natural receptors through the mutational network (P12 to S8Tt2-Figure 8.5and H68 to S8-Figure 8.3).

As more and more RNA motifs are characterized, our understanding of the fundamental nature and hierarchical organization of these dynamic molecules is rapidly broadening. The knowledge gained from studies such as these elucidates the principles that dictate these modular interactions and increases the repertoire of

rational design rules that could ultimately lead to an infinite number of engineering possibilities for these molecular entities.

Chapter 3 R5.58 *In vitro* selected GUAA tetraloop receptor exhibiting structural similarities to naturally occurring RNA

3.1 Abstract

The GNRA (N is any nucleotide, R is any purine) loop/receptor interaction is a key component in RNA macromolecular assembly for a variety of functional RNA molecules. Despite the possibility of having eight unique GNRA tetraloop/receptor interactions, we see in nature a predominant bias for the GAAA/11 nucleotide interaction. Successful *in vitro* selections have been done for GGAA and GUGA, however other tetraloop receptors have yet to be selected for. In this study, we performed an *in vitro* selection that revealed several new GUAA receptors that are specific and capable of binding with affinities comparable to the GAAA/11nt interaction. In addition, it was discovered these GUAA binding receptors exhibited a highly similar phenotypic behavior to previously rationally designed GUAA receptors and could adopt a similar secondary structure configuration that resembled the double-locked bulge motif found in other naturally occurring RNAs. These designed receptors were shown to have structural homology to the S8 ribosomal protein-binding region in the 16S rRNA and the H89_L39 interaction in the 23S rRNA. Using site-directed mutagenesis, native polyacrylamide gel electrophoresis, DMS chemical modification followed by primer extension, and 3D atomic modeling, we report a novel GUAA binding tetraloop receptor and demonstrate its structural isostericity to naturally occurring RNA analogues.

3.2 Introduction

In vitro selection is an invaluable tool that has produced novel solutions for many functional and structural biological constraints. Examples include the elucidation of a vast network of aptamers, ribozymes, DNAzymes, and novel enzymatic proteins ([130-132]. Previously, RNA receptor/tetraloop interactions have been successful in identifying new sequences that are capable of binding tetraloops with a high binding affinity, comparable to naturally occurring tetraloop receptors. The GNRA/receptor tertiary interaction in RNA (N is any nucleotide, R is any purine) is a widespread tertiary interaction unit utilized in structured RNA's, however nature has predominantly utilized only one out of the eight possible GNRA tetraloops. Each GNRA tetraloop shows only slight differences in thermodynamic stability [133], but naturally occurring structural RNA shows a strong bias toward the GAAA tetraloop. Along with this GAAA tetraloop a largely reoccurring 11 nucleotide (11nt) motif receptor has emerged as its natural cognate [21, 94, 122]. The GAAA/11nt interaction, as with all other loop-receptors, is stabilized through a long range A-minor interaction and is known to mediate the lowest energy conformational search of functional RNA molecules [72].

In spite of this natural bias toward the GAAA tetraloop, *in vitro* selections have revealed several receptors that are capable of binding to GGAA [72] and GUGA [73] with high affinity. In fact, the selection performed by Costa and Michel (1997) utilized an enzymatic construct based on a group I intron scaffold. The results of their selection revealed an excellent GUGA binding receptor (B7.8) that in fact was

shown to bind to GUAA with even greater affinity. It was also recently shown that a GUAA receptor could be rationally designed utilizing structural clues from the 16S ribosomal protein binding RNA, S8 (Calkins, 2015 in preparation). The sequence of S8 is nearly identical to that of B7.8 selected by Costa and Michel. In this study, we proposed an *in vitro* selection geared toward selecting GUAA receptors that are capable of binding GUAA with affinities comparable to that of the GAAA/11nt interaction, but utilizing a unique sequence signature, different from that of the B7.8 and S8 receptors.

We designed an *in vitro* selection based on the tectoRNA dimer approach to investigate alternative GUAA loop/receptor interactions [39, 134] (Figure 3.1). Geary et al previously utilized a similar scaffold to select the R1 and R2 receptors, both excellent GGAA binders. However, in their selection, almost half the receptors that survived did not bind via A-minor interactions, but instead through some alternative Watson-Crick pairing via partial unfolding of the helical receptor stem and disruption of the GGAA tetraloop probe. To avoid a high probability of "false positives" we modified the RNA scaffold for our GUAA selection (Figure 3.1). From this *in vitro* selection we reveal several new classes of GUAA tetraloop receptors containing unique primary sequences, as well as offer structural insights into the highly selected R5.58 receptor and it's strikingly similar folding properties to that of the naturally occurring S8 protein binding receptor in the 16S ribosomal RNA.

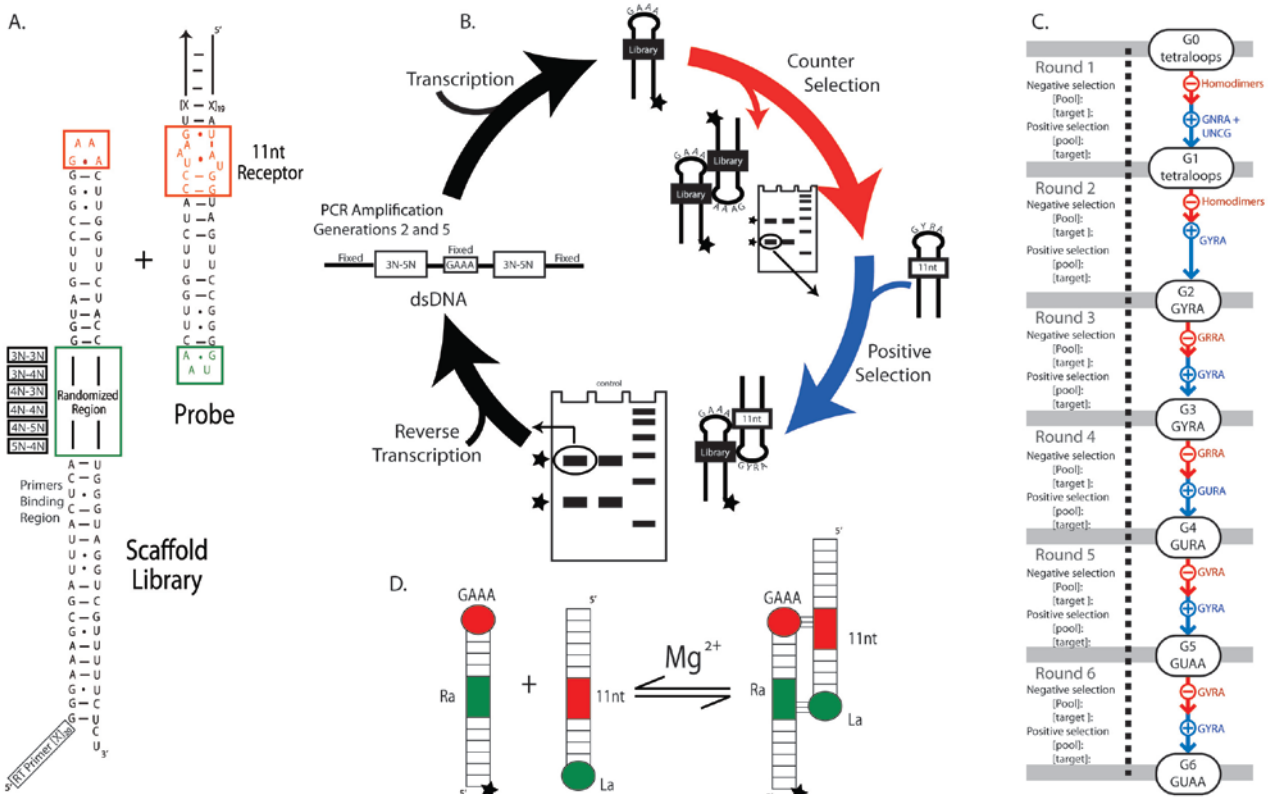


Figure 3.1 Selection Schematic for GUAA Receptors

(A) The scaffold library contained a randomized region that consisted of either 3N-3N, 3N-4N, 4N-3N, 4N-4N, 4N-5N, or 5N-4N mutations, a GAAA tetraloop and a reverse transcription binding region at the 5' end. The probe molecules contained various GNRA tetraloop and an 11nt receptor. The pool of randomized receptors was then subjected to 6 rounds of selection. The selection scheme is highlighted in (B) and the selection criteria are highlighted in (C). Each round had both a counter selection and a positive selection. The positive selection "winners" were the reverse transcribed and the pool was enriched via PCR at rounds 2 and 5. The selection criteria utilized a tectodimer system (D) developed by Jaeger et al (2001). The counter selection removed from the pool any hairpins that formed dimers in the absence of the probe (either via self-dimerization or non-A-minor interactions, such as kissing loop complexes.

3.3 Results

The *in vitro* selection was modeled after Geary et al (2008), which utilized a tectodimer A/B system consisting of two hairpins capped with a GNRA tetraloop (Figure 3.1D). In order to avoid a high number of non-A minor interactions as seen in the selection by Geary, we made a few minor modifications to the RNA scaffold. We did this by reducing the number of nucleotides being mutated (from 17 to 9) and maintained the A-minor interaction of G-C/G-C. Utilizing this robust self-assembling tectoRNA approach developed by Jaeger *et al*. [39], several new RNA tetraloop receptors were selected with remarkable affinity and selectivity for GUAA tetraloops. From the pool of selected receptors, we sequenced 35 in total, with 14 unique sequences. Of these 35 "winners", one receptor in particular was sequenced 15 times, R5.58 (Table 8.5). Of all the receptors sequenced, none had sequence similarities to the B7.8 or S8 receptors.

The K_d's for each of the selected receptors, as well as B7.8 and S8 (with GG:CC A-minor interactions) were calculated and selectivity bars and graphs were generated (using $\Delta\Delta G$'s) to assess the extent of binding toward 7 of the 8 GNRA tetraloops. In some instances, GAAA was tested if the receptor demonstrated homodimerization (Figure 3.2). Utilizing the specificity profiles we determined there were 4 classes of receptors that were obtained from the selection (a-d). Classes are defined based on the affinity of the receptor to each loop (Figure 3.2A) as well as the relational profile (shape of the graph) from one loop to another (Figure 3.2B). Of the selected receptors, R5.58 had the greatest affinity for the GUAA loop, demonstrating a K_d of

7 nM ± 2, which is comparable to the GAAA/11nt motif binding interaction previously tested in the tectoRNA-dimer system, with a K_d of 4nM [72]. Additionally, the R5.58 receptor comprised of 15 of the 35 selected sequences, which is consistent with its high affinity toward GUAA.

The R5a class, showed some of the most thermodynamically stable interactions with GUAA, however, all loops showed a high binding affinity, even GAAA. The R5b class also had high affinity for GUAA, albeit lower than the R5a class, but contrastingly showed greater selectivity, with the other loops having ΔΔG's all above the baseline ΔΔG value (0). Class R5c had a profile similar to that of class R5a, however, had a strikingly high affinity for GCAA by comparison. And finally, class R5d with one receptor (R5.16) showed extremely high affinity for GUAA, as well as the greatest selectivity.

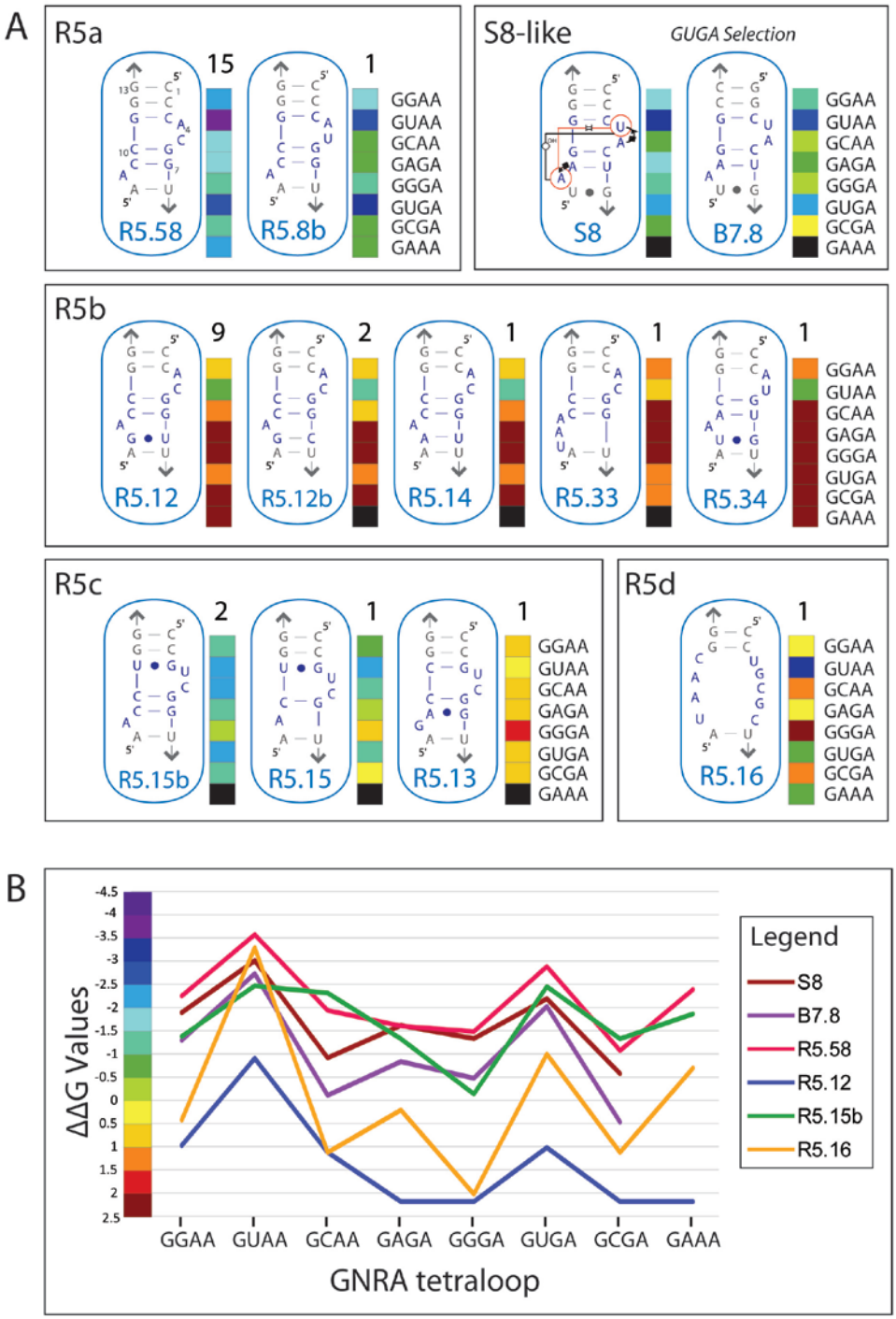


Figure 3.2 Selection "Winners" and Selectivity Profiles
(A) Sequences for selected GUAA receptors, along with sequences for S8 and B7.8 that provide a potential basis for secondary structure comparison. Color coded bar codes demonstrate the degree to which each receptor binds to GNRA tetraloops. The color patterns as well as the shape of the specificity profiles (B) allowed us to classify each receptor into one of 4 classes.

3.3.1 ***Secondary structure analysis***

Upon analyzing the sequences for potential secondary folds, it became apparent that both the selected B7.8 and R5.58 receptors could adopt a similar 2D fold resembling the naturally occurring 'double-locked bulge' (2bp_2x_bulge) motif described in Geary et al. 2011, which is found in RNase P, ribosomal RNA (H89), and S8 [107] (Figure 3.2A and B). The apparent similarity in secondary structure between *in vitro* selected receptors and a naturally occurring protein binding rRNA receptor (S8) capable of binding GUAA tetraloops prompted us to focus on the characterization of R5.58 being that it had the greatest affinity for GUAA and accounted for 15 out of the 35 selected sequences. In a previous study, it was shown that although the S8 binding domain interacts with a protein in its natural context, this same motif binds the tetraloop GUAA with great affinity (19nM) (Calkins, 2015-in preparation). In addition, the S8 binding affinity and specificity profile is nearly identical to that of R5.58 and B7.8 receptors, class R5a (Figure 3.2). Despite vast differences in sequence, the R5.58 selected receptor is capable of adopting a similar secondary fold to the natural S8 receptor, in addition to having a nearly identical phenotypic profile. Further analysis was performed in order to validate the isostericity between these GUAA receptors and prompted us to investigate the structural nature of this newly selected GUAA receptor (R5.58).

3.3.2 ***A-minor sequence comparison***

The sequence of B7.8 is identical to S8, except in the three base pairs involved in the A-minor interaction. S8 contains three base pairs in the A-minor with the

sequence 5'-UCA…UGA-3', whereas B7.8 contains the sequence 5'-GCC…GGC-3'. In our selection, we conserved the 2 base pairs in the A- minor to yield receptors with the sequence 5'-NGG…CCN-3'. In a previous study, both the S8 and B7.8 sequences were tested in the heterodimer context with a 5'-GGG…CCC-3' A-minor. This sequence is the same A-minor sequence seen in R5.58 and many other selected receptors. When comparing the selectivity profiles of S8 (with a 5'-GGG…CCC-3'), B7.8 and R5.58 (Figure 3.2B), you can see that although the K_d's are slightly altered between the three receptors, the overall selectivity profiles are nearly identical. Coincidentally, the 5'-GGG…CCC-3' A-minor sequence was also selected for in the Geary et al. (2008) selection (R2). R5.58 and S8 (both with the A-minor sequences of 5'-GGG…CCC-3') had K_d's for GUAA of approximately 7nM and 19nM, respectively, whereas the K_d for the selected B7.8 receptor, which has an A-minor sequence of 5'-GCC…GGC-3' was closer to 30nM. Since the sequences for S8 and B7.8 are identical in the bulge region and the only difference is the A-minor sequence, this would indicate that the A-minor region affects affinity only and not selectivity. In order to determine how the A-minor interactions affect the binding of this receptor class, we tested various combinations in the A-minor region (Figure 8.9). In our study mutations made to the base pairs X_1:X_{13} and X_2:X_{12} all maintained Watson Crick base pairs, since disruption of the A-minor helix would yield little to no interaction. The data indicates that mutations made to this region, while altering the affinity for tetraloops, does not drastically alter the relative specificity of the receptor to the 7 GNRA tetraloops (Figure 8.9A and B). The shape of the curves from Figure 8.9B maintain a fairly consistent pattern as opposed to mutations made

with in the X_3:X_{11} region (Figure 8.9C). Figure 8.9C shows receptors in red and orange with conserved Watson-Crick base pairs and have comparable specificity profiles to that of R5.58. Highly destabilizing mutations (involving adenine and guanine) are shown in shades of green and non-Watson-Crick interactions that maintained a neutral profile are highlighted in blue. Collectively, by maintaining the three A-minor Watson-Crick interactions specificity is conserved and affinity is only slightly altered. This supports previous data that demonstrated the A-minor modular component of this receptor class is not used for selectivity [72], therefore we chose to utilize the A-minor with the lowest K_d (greatest affinity), in this case 5'-GGG…CCC-3' to keep our data consistent with previously characterized receptors [72] and (Calkins, 2015-in preparation).

With a strong indication that R5.58 has the potential for a similar secondary structure as well as a nearly identical specificity profile to the S8-like sequence, further secondary structure analysis was needed. Additional verification that R5.58 is structurally equivalent to the S8 protein binding domain of the 16S rRNA could further be verified with probing experiments and mutational analysis.

3.3.3 ***Probing for modularity***

To further probe the structural relatedness between R5.58 and S8, a series of mutants were designed so that the modular components of each receptor could be swapped and tested. Figure 3.3 and Figure 3.4 show the secondary structure comparisons of R5.58 and S8 receptors (used in this study) indicating the modular components of this motif (Module I and Module II). The bar code data and

specificity profiles functioned as the phenotypic measurements for functional equivalence, which offered a robust platform for determining structural relatedness between the R5.58 and S8. The receptor was divided into two separate modules (I and II), each containing helical base pairs as well as a nucleotide bulge (Figure 3.3 and Figure 3.4). A series of tetraloop specificity gels were tested for various mutants of both R5.58 and S8, which allowed us to compare the effects of these mutations in each context.

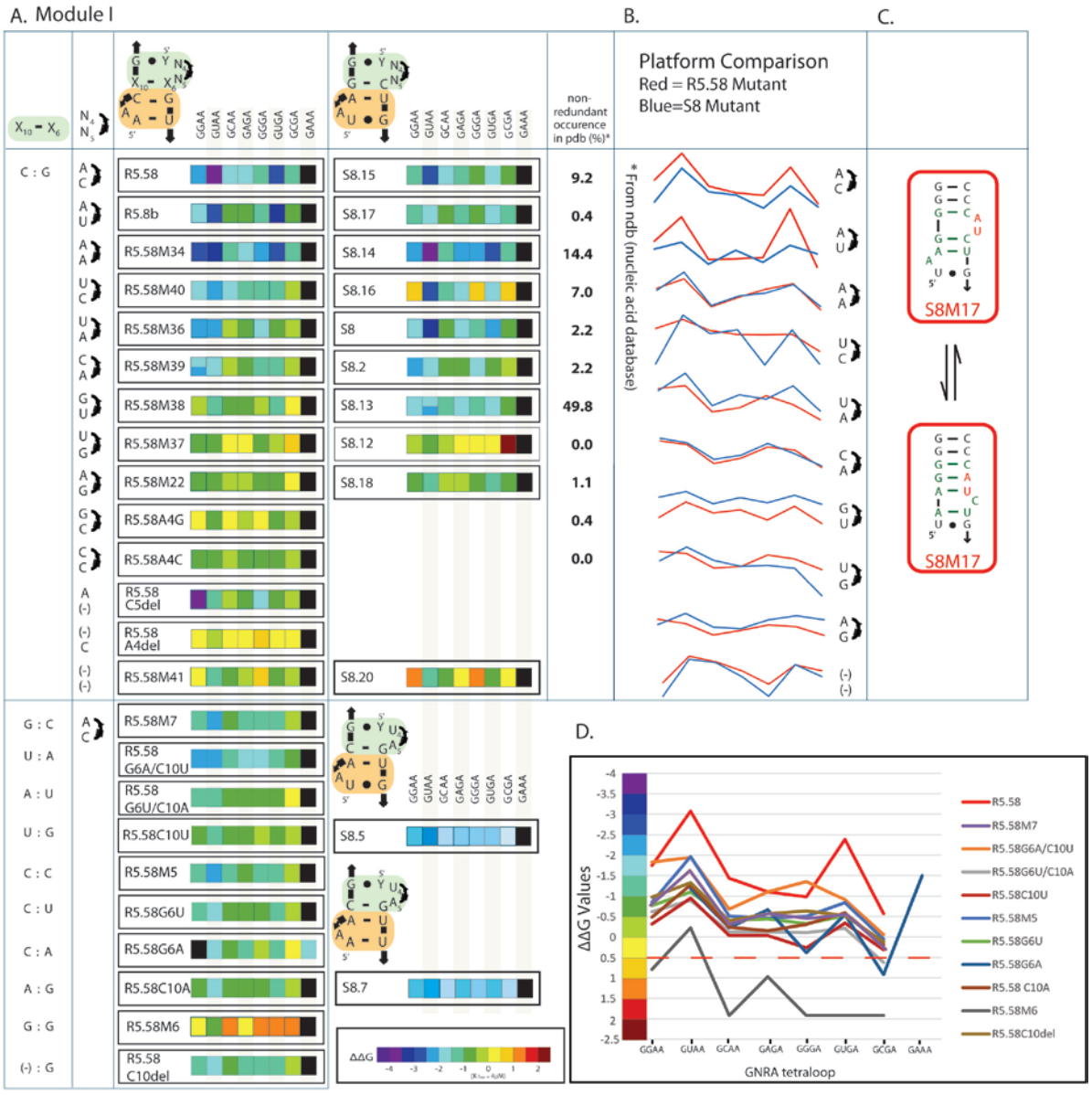


Figure 3.3 Module I Mutational Analysis

(A) Module I (in green) mutations. We compared mutations from both R5.58 and S8 within the dinucleotide platform, as well as mutations in base pair (X_{10}-X_6). (B) Specificity profiles that compare GNRA loop binding of platform mutants between S8 and R5.58 receptors. Non-redundant occurrence is shown. (C) S8.17 alternative secondary structure, which offers a possible explanation for why the AU platform mutation yielded a very different specificity profile for S8 when compared to R5.58. (D) Specificity profiles of X_{10}-X_6 mutants are compared which highlight the existence of Watson-Crick base pairing at this position.

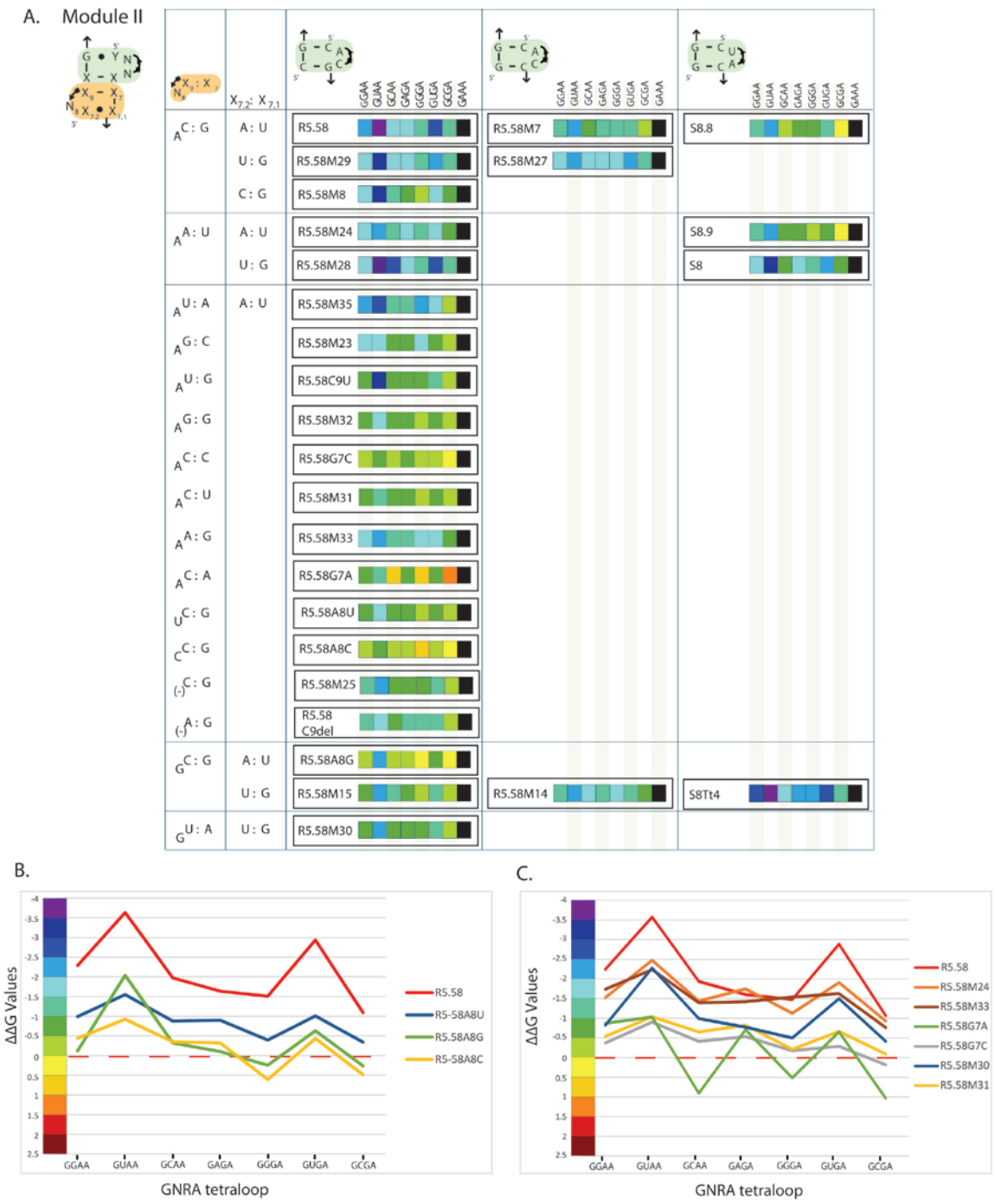


Figure 3.4 Module II Mutational Analysis

(A) Mutations in module II (in orange). Mutations in single nucleotide bulge showed bar codes with lower affinity than that of R5.58. We also analyzed mutations in the base pairs of module II to help demonstrate the importance of Watson-Crick parings at the positions X_9-X_7 and $X_{7.2}$-$X_{7.1}$. However, specificity profiles (B) are very similar, indicating that this position is for distal stability of the receptor and not for specificity. (C) The three most prevalent platforms are AC, AA and GU (Figure 3.3). The specificity profiles show a comparison of each of these platforms with and without a Watson-Crick interaction in base pair X_9-X_7.

3.3.4 ***Secondary structural probing: Dinucleotide platform analysis of Module I***

With data supporting the structural relatedness of R5.58 to S8, we proceeded to test the contribution of the dinucleotide bulge on tetraloop binding and whether it adopts a platform structure consisting of a *cis*-Hoogsteen/Sugar edge interaction, as found in the crystallographic structure of the S8 binding domain. Knowing that the AA platform is largely reoccurring in nature [135], we mutated both S8 and R5.58 to A4-A5, which strikingly gave nearly identical tetraloop specificity profiles (Figure 3.3A and B, mutants R5.58M34 and S8.14). The AA platform mutants also displayed strong binding to GUAA with K_d values that are comparable to the R5.58 and S8 wild types, 15 nM and 7 nM, respectively. Furthermore, mutations inserting a UG platform in both receptors (R5.58M37 and S8.12) greatly disrupted the binding affinity for GUAA, yielding K_d values of 1132 nM and 749 nM, respectively (Figure 3.3A and B). In the case of the R5.58M37, there is a potential for an alternative fold where G5 pairs with C10, resulting in a bulge containing only A4. Nevertheless, even in the instance of this alternative pairing, removing or destabilizing the dinucleotide platform has a negative impact on tetraloop binding. Furthermore, it appears that dinucleotide bulges without alternative folding pathways only promote tetraloop binding if a platform interaction exists. According to the isostericity matrix for *cis*-Hoogsteen/Sugar edge platform interactions [10], the UG platform is not a known interaction in naturally occurring RNAs, explaining the dramatic loss of binding in the R5.58M37 and S8.12 mutants. Moreover, all other naturally occurring platform mutants that were tested either fully or partially preserved binding affinity to GUAA. Of all the platforms tested, the AC, UA and AA platforms had excellent

binding toward GUAA in both receptor contexts. Interestingly, these platforms were selected for in B7.8 (UA platform) and R5.58 (AC platform), as well as the platform most widely used in nature found in the 11nt receptor (AA platform). To further validate importance of the dinucleotide bulge, the bar codes and specificity profiles were compared for S8 and R5.58 containing several different platform mutations (Figure 3.3A and B). In many instances the profiles for a particular platform confer very similar phenotypes in the mutated receptors. Another interesting finding is that in both receptor contexts, the CA platform showed a preferential selectivity toward GGAA (Figure 3.3A and B, mutants R5.58M39 and S8.2), as well as the profiles being nearly identical. This data is further evidence that the platform of this receptor class acts as a modular entity that can dictate the phenotypic behavior towards the GNRA tetraloops. This would indicate that the dinucleotide bulge is a significant module of these receptors and dictates specificity, but also that these two receptors are most likely adopting similar folds and are isosteric to one another. The AU platform is the only sequence where the patterns do not seem to be generally consistent between the two receptors (Figure 3.3B). This can most likely be explained by the fact that alternative pairings are possible in many of the receptors where there is interchangeability between the secondary structure of the nucleotides that are paired and in bulge. The S8 receptor with an AU bulge (S8.17) can pair in such a way that the receptor acts much like a helical receptor (Figure 3.3C), thereby altering the GNRA tetraloop profile. The R5.58 receptor with an AU bulge (R5.8b), which was selected for, is not capable of this alternative pairing structure, and it's GNRA tetraloop profile, is nearly identical to that of the wild-type. The phenotypic

differences seen in this instance are a direct result of the slightly differing thermostabilities of the loop-receptor interaction.

3.3.5 ***Secondary structural probing: Watson-Crick base pairing of module I***

With significant data towards the structural equivalence of R5.58 and S8, we chose to test the existence of Watson-Crick (WC) base pairs through a series of compensatory mutations within the R5.58 receptor. This would not only assist in verifying the secondary structure but offer additional clues as to any possible interactions taking place within and between module I and module II. Two sets of Watson-Crick mutations were tested at positions X_6:X_{10} and X_7:X_9 (Figures 3 and 4). According to Figure 3.3A and Figure 3.4A, mutants at each position support the existence of a WC interaction because all mismatch mutants are compensated by their respective WC partner (data for all loops can be seen in Figure 3.3A and Figure 3.4A). In module I, mismatches that promote destabilization of the 2-D structure at positions X_6:X_{10} were shown to decrease the affinity as well as alter the specificity profiles for the GNRA tetraloops (Figure 3.3D). For example, the mutant R5.58M6 (G:G mismatch) has a K_d of 1106 nM for GUAA possibly due to an alternative fold that creates a G10:C5 WC pair which would remove the dinucleotide bulge, thus destabilizing the platform. Interestingly, the binding was compensated by mutant R5.58M7 (K_d = 97 nM for GUAA), indicating the preference of a WC interaction at position X_6:X_{10}. To provide further evidence of the isostericity between R5.58 and S8, the lower base pairs of module I (X_6:X_{10}) were swapped (mutants R5.58M7 an S8.5). The pattern of tetraloop affinity in the receptor mutants indicates that these

receptors have a modular structural assembly consisting of separate parts that can be interchanged. In this instance, the profile for the mutants are nearly identical, however they alter from the wild type profiles. This would indicate that the mutation, in this case perhaps interacts with the dinucleotide bulge and it is the bulge along with this base pair that comprise the entire module I. While maintaining WC base pairs generally maintained GUAA binding, the profiles did alter somewhat and certain mismatches were tolerated quite well. This could be an indication that aside from the WC base pair, there were other interactions involved in this module.

3.3.6 ***Secondary structural probing: Module II***

Figure 3.4 shows that the closing base pair of module II in both R5.58 and S8 receptors affects the binding affinity (receptor mutants S8.9, R5.58M29 and R5.58M8) slightly destabilizes the tetraloop/receptor interactions for all 7 loops. Interestingly, R5.58 contains an A:U closing base pair, whereas S8 contains a U:G closing base pair, simply swapping these two base pairs slightly destabilizes the interaction for both receptors (Figure 3.4A). The 11nt receptor, a related natural receptor, similarly demonstrates a destabilizing trend when mutated at its closing base pair [72]. Because both mutants displayed a decrease in affinity toward all the GNRA tetraloops tested, it could be an indication that the closing base pair is in interaction with the lower base pair in the helix (positions 7 and 9), forming a modular component. To test this, the entire lower module was swapped (Figure 3.4A, mutants R5.58M28 and S8.8). Both receptors maintained nearly identical specificity profiles to that of the wild type, however by inserting the S8 lower

module into the R5.58 receptor, R5.58M28 shows a slight increase in the affinity for GUAA going from a K_d of 7 nM to 2.6 nM. This stabilizing effect suggests an interaction between the 5 nucleotides contained within module II. Mutational analysis of the WC positions provided promising results that prompted further secondary structure investigation utilizing DMS chemical modification.

Figure 3.4B shows four mutants that differ only in the single nucleotide contained within the module II bulge. Each of the four mutants demonstrates nearly identical profiles, however each offer varying degrees of thermostability of the dimer. To further examine module II, we tested various bulge components with and without a Watson-Crick interaction at positions X_7:X_9. In the crystallographic structure of S8 (Figure 3.2C), the module II nucleotide bulge A595 is shown to be part of a dinucleotide platform with position A596. The three most naturally prevalent dinucleotide platforms are AA, AC and GU [73, 135]. Several mutants were tested that preserved each of these platforms and compared each mutant with and without a Watson-Crick base pair at position X_7:X_9 (Figure 3.4C). Each receptor that preserved a Watson-Crick base pair (R5.58, R5.58M24 and R5.58M30) demonstrated nearly identical profiles and were all highly selective toward the GUAA tetraloop. Contrastingly, receptors containing these same platforms but with a disruption of the Watson-Crick base pair (R5.58G7C, R5.58M31, R5.58G7A and R5.58M30) showed a decreased in affinity for GUAA as well as a loss of selectivity.

3.3.7 ***DMS chemical modification***

DMS chemical probing is a powerful and versatile method that can demonstrate the existence of adenines and cytosines in single stranded bulged regions that are unprotected and solvent accessible [136]. Because R5.58 is primarily composed of adenine and cytosine, DMS modification followed by primer extension would provide an accurate prediction of secondary structure. According to Figure 3.5A, the DMS monomer lane of R5.58 shows a clear indication of modification at the platform region (A4:C5), the GAAA tetraloop (A36:A37:A38), and the bulged A8. Interestingly, C1-C3 and C9-C10 show complete protection in the monomer indicating the presence of WC interactions at those positions, supporting our secondary structure prediction. Additionally, the unprotected receptor nucleotides of the monomer lane (the two bulges) demonstrate a high degree of protection in the dimer, suggesting a tertiary structural interaction between monomers in this region. Upon dimer formation, the GAAA tetraloop is protected by the A-minor interaction with its cognate 11 nt receptor, while the protection pattern of the R5.58 receptor region is indicative of an induced fit. A8 and A4 are unprotected in the monomer, however upon dimer formation an induced fit mechanism offers N1 protection for both nucleotides. This is likely due to cross-strand hydrogen bonding offered by the 2'- hydroxyls of A8 and A4 to their respective N1 positions. This protection pattern is an indication of the structural signature that makes up the "S8-like" fold. Moreover, N3 protection of C5 appears to stem from the stacking interaction with the U residue in the GUAA tetraloop, indicating the influence of the dinucleotide platform on tetraloop binding and supporting the data in Figure 3.3. Furthermore, S8

also showed a comparable pattern of monomer and dimer DMS protection and provided further evidence of the structural isostericity with R5.58 (Figure 3.5B). Additional R5.58 mutants were tested to demonstrate the existence of important WC base pairs (Figure 3.5B). The DMS monomer and dimer profiles for WC mutants R5.58M12, R5.58M5, and R5.58M31 deviate from wild type R5.58 on both sides of the receptor, likely due to the conformational flexibility of the phosphate backbone offered by the disruption of the WC interactions. Additionally, R5.58M31 shows significant cleavage at position C_3, which in the monomer is protected in the other receptors because of Watson-Crick pairing with G11. R5.58M31 is capable of adopting an alternative fold as shown in Figure 3.5, which would account for this DMS profile. This data indicates the WC positions play a major role in stabilizing the proper conformational fold of the R5.58 receptor. Taken together with the thermodynamic analysis of receptor mutants, this data provides a strong indication of structural similarities between R5.58 and S8.

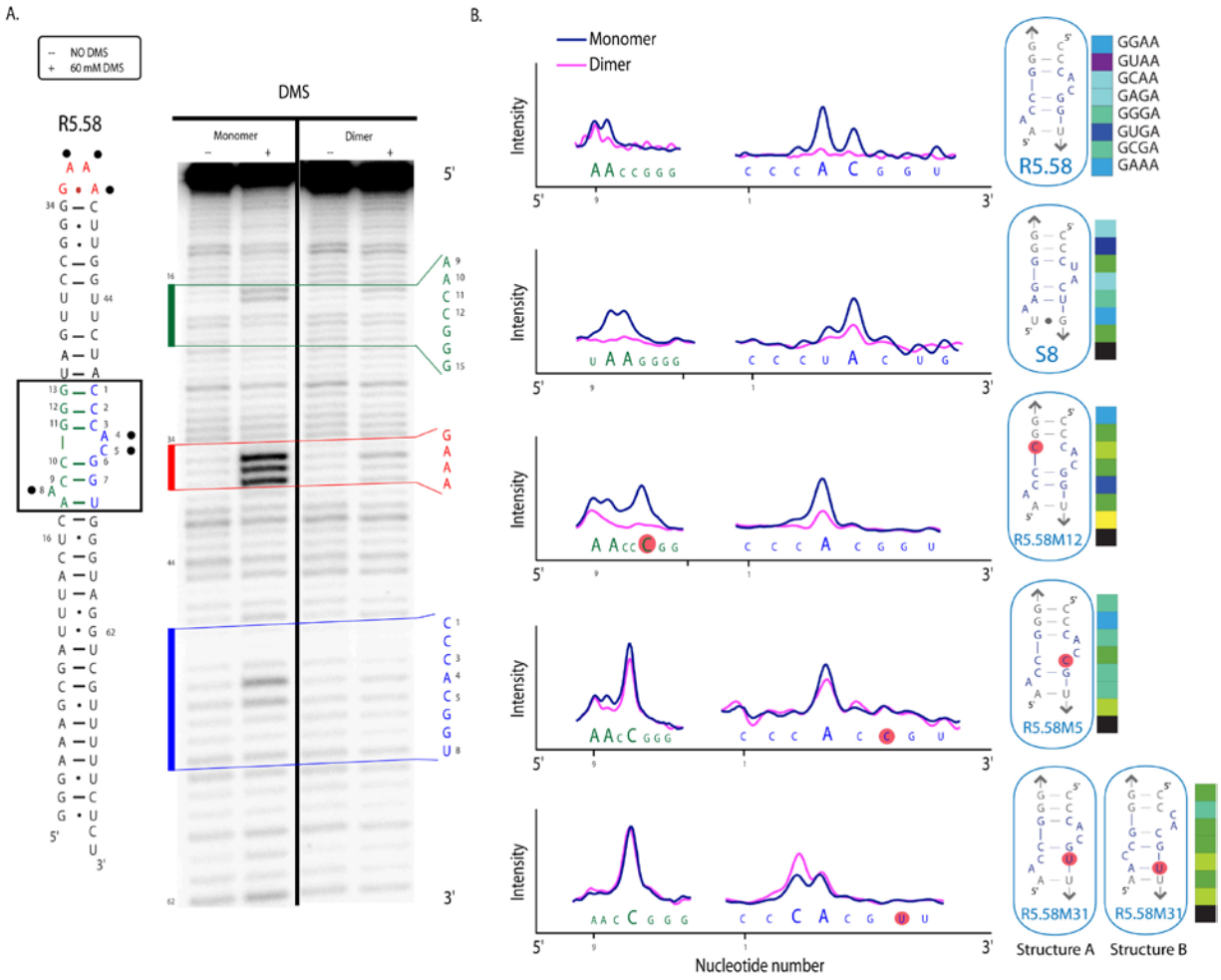


Figure 3.5 DMS chemical modification of wild type and mutant receptors.
(A) DMS modified primer extension profile of R5.58 in monomer and dimer R5.58/GUAA states. 'No DMS' samples were conducted in identical fashion as the DMS containing samples. Solid black circles represent predicted sites of modification based on the proposed secondary structure. DMS modification is accompanied by complete Sanger sequencing of R5.58 monomer. (B) Quanititated DMS profiles measured using ImageQuant software. Each monomer and dimer lane was subtracted from respective non-DMS lanes. The intensity corresponds to the amount of radiation. Receptor mutants with red circles indicate the location of the mutation. All dimer profiles correspond to the dimerization with L(GUAA) tetraloop tecto-RNA. Secondary structures for each receptor are shown at right. R5.58M31 shows two possible secondary structures. Structure B is confirmed by DMS modification due to significant cleavage at C_3 in the monomer.

3.4 Discussion

In vitro selection revealed several novel classes of GUAA tetraloop receptors with binding affinities rivaling that of the GAAA/11nt receptor interaction. Despite the expanding library of RNA receptors, there had previously been a lack of direct evidence to link structural homology to naturally occurring RNA analogues. Even without such direct evidence, previous work on class R1 receptors demonstrates chemical and physical characteristics comparable to the natural GAAA/11nt loop/receptor interaction, indicating that these artificial receptors maintain a degree of similarity to naturally occurring RNA. Moreover, the data presented herein outlines a new tetraloop receptor (R5.58) with remarkable affinity to the GUAA tetraloop and structural equivalence to the S8 ribosomal protein-binding site in the 16S rRNA. Interestingly, a previous *in vitro* selection, although designed to enrich populations of GUGA binders, elucidated a GUAA binding receptor (B7.8) [73]. However, the GUAA/B7.8 interaction was not completely understood and the secondary structure was originally reported in a way that would prevent optimal binding to GUAA, explaining why both Chauhan *et al*. and Ishikawa *et al*. achieved unfavorable results when the GUAA/B7.8 interaction was employed in their respective studies [137, 138]. Woodson and Ishikawa both mutated the B7.8 sequence from a U:G closing base pair to G:C and A:U closing base pairs, respectively, possibly explaining why these B7.8 mutants demonstrated impaired function in their respective contexts due to the apparent importance of this location and its interaction with other nucleotides. Because the sequence of the selected B7.8

is strikingly similar to the S8 protein binding site of the 16S rRNA of *E. coli*, we were able to infer a probable structure for the interaction of the selected GUAA/R5.58 receptor. Indeed, our data suggests that R5.58 and B7.8 are structurally isosteric to this naturally occurring S8 protein binding RNA. Although these receptors were subjected to different selection pressures, it appears that both the enzymatic and tectoRNA *in vitro* selection schemes yielded related structural solutions for the GURA tetraloop constraint (Figure 3.2 and Figure 3.6A). The S8 binding region in the 16S rRNA, however, is not known to interact with any kind of RNA hairpin loop. Rather, this region maintains important contacts with the S8 primary ribosomal protein and is responsible for the proper structural stability of the 16S rRNA and cooperative assembly of the 30S ribosomal subunit [139]. In addition to the natural S8 domain, the H89 domain of the 23S rRNA also contains a region that is structurally isosteric to the S8 protein-binding site of the 16S rRNA. Moreover, this region in H89 is known to interact with the L39 GNRA-like pentaloop through the type I/IIT A-minor interaction (*Haloarcula marismortui)*, similar to the GAAA/11nt interaction found in nature [107]. However, the H89 receptor was unable to show dimerization in this context and was therefore not subjected to mutational analysis (Calkins, 2015-in preparation). This is most likely due to post-transcriptional modifications that aid the proper folding of the functional receptor. This pattern of structural isostericity links both R5.58 and B7.8 to not one, but two naturally occurring loop receptors and is a striking example of topological equivalence between artificial and natural RNA.

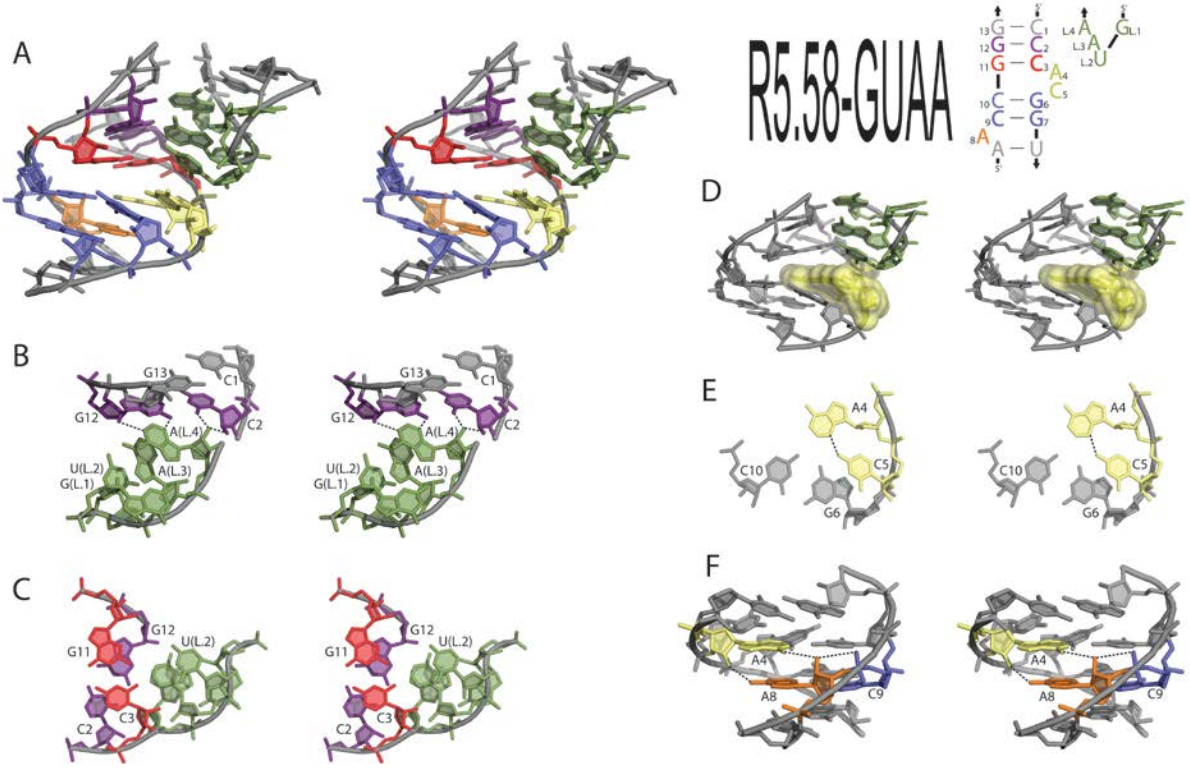


Figure 3.6 Stereoimages of R5.58/GUAA Model

(A) Stereoimage of R5.58 interacting a GUAA tetraloop (Secondary structure at right). GUAA loop shown in green, A-minor shown in purple, Watson-Crick interaction just below A-minor shown in red, dinucleotide bulge shown in yellow, internal Watson-Crick interactions shown in blue, lower nucleotide bulge shown in orange and closing base pair shown in grey. (B) Stereoimage zoomed in to the A-minor/GUAA tetraloop interaction. Hydrogen contacts are made with the G:C base pair of the A-minor and the 4th position of the GNRA tetraloop via shallow groove contacts. (C) Stereoimage showing the positioning of the second position of the GNRA tetraloop and the Watson-Crick base pair just below the A-minor creating binding pocket for the loop. (D) Stereoimage of GUAA loop stacking on AC platform. (E) Stereoimage of the top-down view of platform interactions with respect to Watson-Crick base pair C_{10}-G_6. (F) Stereoimage of hydrogen bond contacts being made between module I and module II.

3.4.1 ***Analysis of the S8 binding domain***

The S8 ribosomal protein-binding site is a small helical region centrally located within the 16S rRNA containing various non-canonical pairings and has been used as a platform to study RNA-protein interactions [140-142]. Furthermore, the mRNA transcript for S8 adopts a fold that is structurally identical to the S8 binding domain, which appears to be an S8 translational regulatory mechanism, suggesting that RNA outside the context of the ribosome has a potential to adopt an "S8 like" fold [143]. There are several interesting structural features that make up the S8 binding site fold (Figure 3.2C). According to the crystallographic structure, there is a U641:A642 platform module consisting of a *cis*-Hoogsteen/Sugar edge interaction. This is then followed by two base pairs, the second of which shows a base triple between the U641:A642 platform and A595. A595 is involved in an AA platform interaction with A596 and is stacked underneath U641 (Figure 3.2C). Interestingly, both A595 and U641 have C2' endo sugar puckers which deviate from the canonical C3' endo sugar conformation of an A-form RNA helix. Essentially, this modification creates hydrogen bonding between the C4 carbonyl of U641 and the 2'-hydroxyl of A595, and between the N1 of A595 and the 2'- hydroxyl of U641. This remarkable non-canonical interaction locks the adjacent UA and AA platforms and stabilizes the structural signature of the "S8-like" motif. Moreover, the H89 segment from the 23S rRNA is structurally isosteric to the S8 binding domain, indicating another instance of topological equivalence found in natural RNA structures [144]. H89 interacts with the L39 pentaloop in a type I/IIT A-minor fashion, and interestingly, this interaction takes place through the minor groove on the opposite side relative to the

S8 protein interaction in the 16S rRNA. The S8 binding site, both from a structural context as well as experimentally determined, can accommodate an A-minor like RNA interaction on its opposite end; however, this location is in close proximity to ribosomal protein S17 that perhaps acts as a protecting agent to prevent alternative A-minor RNA interactions [145].

3.4.2 *GUAA/R5.58 Interaction*

In our tectoRNA context, assembled heterodimers are maintained by two distinct interactions, one anchored by the GAAA/11nt interaction, and the other containing the GUAA/R5.58 interaction. Both interactions are necessary to maintain a dimeric species with distinct gel shifts in native PAGE experiments (Figure 8.8). With data indicating structural homology to natural RNA analogues, we were successfully able to create an atomic 3D model of the *in vitro* selected R5.58 receptor specific for the GUAA tetraloop (Figure 3.6A). This 3D model when compared to the crystallographic structure of S8 further demonstrates the structural isostericity of these two receptors. The GUAA docking into the R5.58 receptor most likely takes place through the classic type I/IIT A-minor interaction, composed of the third and fourth adenosine residues in the GUAA tetraloop hydrogen bonding in the minor groove of the G_{13}:C_1 and G_{12}:C_2 bps of the R5.58 receptor [103] (Figure 3.6B). As outlined by Geary *et al.*, GNRA tetraloops with an adenosine in the third position adopt a type I/II tilted A-minor configuration when in contact with a helical receptor [72]. Tetraloops with a guanine in the third position destabilize the interaction with a tilted configuration due to a steric clash between the carbonyl oxygen of G_3 of the

tetraloop and the 2'-OH of the guanine of the G_{12}:C_2 base pair of the A-minor. In the case of R5.58, the GUAA tetraloop binds with 5 and 10 times more selectivity than the GUGA and GGGA tetraloops, respectively, indicating that the GUAA A-minor interaction with R5.58 stabilizes a tilted configuration. Figure 3.6 C highlights the positioning of the tetraloop in the A-minor and the location of the second position of the GNRA tetraloop. Furthermore, the second U position in the GUAA tetraloop stacks on top of C_5 (Figure 3.6D and E). Although it is not clear why a uracil is favored over cytosine, it appears that the purines introduce steric clashes and are unable to adopt the proper configuration for optimal pi stacking. This stacking interaction also explains why C_5 shows N3 protection upon dimer formation in the DMS experiments (Figure 3.5). The stacking between A_4 and A_8 can be seen in Figure 3.6F, in addition to two additional ribose/N_6-H bonds, one from each adenine. This same type of interaction is seen in the crystallographic structure of S8, further validating that the lower module offers overall thermodynamic stability to the loop/receptor interaction. In addition, Figure 3.4C further demonstrates that the stacking interaction between A4, C_8 and A_9 as well as the Watson-Crick position at X_7:X_9 is important for offering stability in the loop receptor interaction, while being highly selective for GUAA. This is consistent with the crystallographic structure of S8, which reveals that the components of both module I and module II interact to further stabilize the loop/receptor tertiary interaction. Additionally, the C2' endo sugar puckers of A8 and A4 offer hydrogen bonding to their respective WC edges, explaining the A8 and A4 N1 protection observed upon dimer formation in the DMS experiment (Figure 3.5). Moreover,

upon binding to its cognate GUAA tetraloop, R5.58 induces a conformational change that is also known to occur in the GAAA/11nt interaction in the P4/P6 domain of group I introns [122]. Interestingly, the GUAA interaction occurs at the same site where the natural S8 protein resides and is on the opposite A-minor site from the H89_L39 interaction. This model suggests, and was verified in this study, that the S8 binding region can potentially accommodate an RNA A-minor interaction, however, it appears that this centrally located region of the 16S rRNA requires specific protein contacts to ensure proper assembly of the central domain [139, 146] and cooperative binding of S5 and S12 secondary ribosomal proteins [142]. Nevertheless, the interesting fold of both the S8 binding site and the H89_L39 region appear to be very flexible with respect to the type and location of binding partners. Recent investigation of the H89 receptor sequences from archaea and bacteria revealed designability within this receptor class that allowed for the rational engineering of novel GUAA tetraloop receptors (Calkins, 2015-in preparation). It was shown that the S8 receptor class contains modular units that can be honed to increase both selectivity and binding affinity toward the GUAA tetraloop. It was also shown, that the S8 receptor was capable of binding GUAA in both orientations with an affinity of 358nm (in the same orientation as H89), compared with 19nM as seen in this study (Calkins, 2015-in preparation).

Furthermore, the GUAA receptors in this study were discovered through *in vitro* selection schemes taking advantage of the GNRA A-minor interaction. However, in a previous study, an *in vitro* selection was conducted on the S8 protein binding site taking advantage of the S8 protein interaction [141]. This selection did not produce

any variants with sequence homology to the R5 class. Moreover, a comparative sequence alignment of Archaea and Bacteria 16S rRNA sequences in the S8 binding region demonstrate total conservation of A642 and G597:C643 [141, 142]. Our study employs mutations at these position that do not significantly affect tetraloop binding (Figure 3.3) suggesting that RNA/RNA interactions have a higher degree of plasticity than protein/RNA interactions with respect to sequence variation at this site.

3.4.3 ***Evolution of Structured RNA***

The double-locked bulge motif found in H89 in the 23S rRNA and S8 in the 16S rRNA, although structurally similar, differ in that the H89 receptor binds a GNRA-like pentaloop of RNA while the S8 receptor binds a ribosomal protein [107]. Additionally, the interacting nucleotides for each receptor take place on opposite sides of the receptor, essentially making the functional receptor orientation inverted to one another. Being structurally isosteric, it was highly probable that S8 would be able to bind to a GNRA tetraloop, specifically a GUAA loop, being that the H89 receptor binds a pentaloop with topological similarity to the GUAA tetraloop. However, it was striking how well S8 bound not only the GUAA tetraloop, but all the other tetraloops tested. In this study, an *in vitro* selection experiment has revealed yet another double-locked bulge receptor (R5.58) that binds to the GUAA tetraloop, with an affinity comparable to the 11nt/GAAA interaction. Interestingly, we have demonstrated that both artificially selected and naturally occurring receptors can share similar topological structures and differ at the level of the sequence quite

significantly, yet maintain a common function (GUAA binder). Figure 3.7 highlights several point mutation pathways that connect the artificial selected receptors and the natural S8 receptor. Each point mutation is represented by its fitness (or how well it binds each tetraloop). We determined the fitness parameter to be a K_d lower than 4000nM, where any interaction above that concentration is considered to be nonfunctional in a biological context. Not every pathway is considered successful, with certain mutations yielding little to no interaction between receptor and loop. However, it is demonstrated that similar classes of receptor (class R5a, R5b and R5c) allow for modification through point mutations, while still maintaining functionality. For example, nearly every selected receptor can withstand point mutations allowing a search for novel, more optimal sequences. Walking through the sequence space from R5.58 (turquoise) to S8 (yellow), it is clear that through different pathways, there can either exist a continuity in loop binding, or an abrupt shift or loss of binding for a particular loop. Mutant R5.58M7i1 yields a K_d of approximately 21 uM for the loop GUAA, which is likely a nonfunctional interaction. However, this mutant binds the loops GUGA, GAGA and GGAA with relatively high affinity (481 nM, 588 nM and 772 nM, respectively) offering an alternative loop/receptor pair. Our study also shows that in the instance of R5.34 (Figure 3.7, in pink), for the pathways we tested, we were unable to generate point mutants that connected the network in a continuous fashion. R5.34 seems to be an "island" with respect of functionality. Of course, not all pathways were tested, but it demonstrates that some sequences are more robust than others. For the pathways where there are successful links between distantly related receptors, this signifies a plausible neutral network

that connects artificially selected receptors to naturally occurring ones through point mutation leading to further evidence of the capability of RNA to maintain a very similar phenotype through many genotypic variations. This concept of neutral networks and functionally equivalent classes of receptors is vital to explain the adaptability and increase in complexity of functional RNA molecules [147]. Interestingly, we also saw neutral networks within the families of sequences selected from this study. Figure 8.10 highlights just a few the mutational pathways that link the selected receptors via genotype and phenotype. The robust nature of this RNA motif allows for structure/function to be maintained while the sequence space is simultaneously probed for new functionality. The receptors and pathways are color coded according to class, which can be traced and easily noted when a phenotypic shift occurs. This is evolutionarily significant in that it demonstrates the mutational robustness of this motif and further supports previous data that mutations at the level of a motif are what drive adaptation and increase the complexity of the RNA based structures [64, 148, 149].

3.5 Conclusion

Through an *in vitro* selection scheme taking advantage of the A-minor interaction, we developed a new GUAA selective tetraloop receptor (R5.58). We discovered an additional receptor (B7.8) from a previous *in vitro* selection with sequence homology to the S8 binding region in *E. coli* 16S rRNA, which offered a structural context for this class of receptor. Through a series of tetraloop specificity gel profiles, we demonstrated a high level of structural equivalence between the

artificial R5.58 receptor and the natural S8 receptor. Moreover, thermodynamic analysis of site directed mutants at important structural positions in R5.58 revealed a pattern consistent with our proposed secondary structure. Further DMS chemical modification followed by primer extension of wild type and receptor mutants revealed additional evidence for an "S8 like" fold and corroborated data obtained through native PAGE analysis. Finally, a 3D atomic model of R5.58 was generated to display the structural equivalence between related natural RNA sequences. Taken together, this data introduces a new artificially selected receptor for the GUAA tetraloop with a demonstrable structural link to naturally occurring RNA offering insight to potential evolutionary pathways that lead to an increase in complexity.

The loop/receptor interaction is crucial for the proper tertiary folding of functional catalytic RNA and these receptors can be employed to further explore the implications of this important tertiary interaction. Moreover, with the advent of RNA nanotechnology, the development of various selective tetraloop receptors enables greater regioselective control in nano-construction and can lead to programmable RNA architectures containing multiple tetraloops together with their cognate receptors [35, 40]. Further structural characterization of more tetraloop receptors is needed to investigate exactly how these various GNRA receptors are related in an effort to discover a bridge between naturally and artificially selected loop/receptor interactions.

FITNESS
ΔΔG (kcal/mol)
GUAA
GGAA
Legend
R5a
R5b
R5c
R5d
S8-Like
GGAA
GGGA
GUAA
GUGA
GCAA
GCGA
GAGA
GAAA

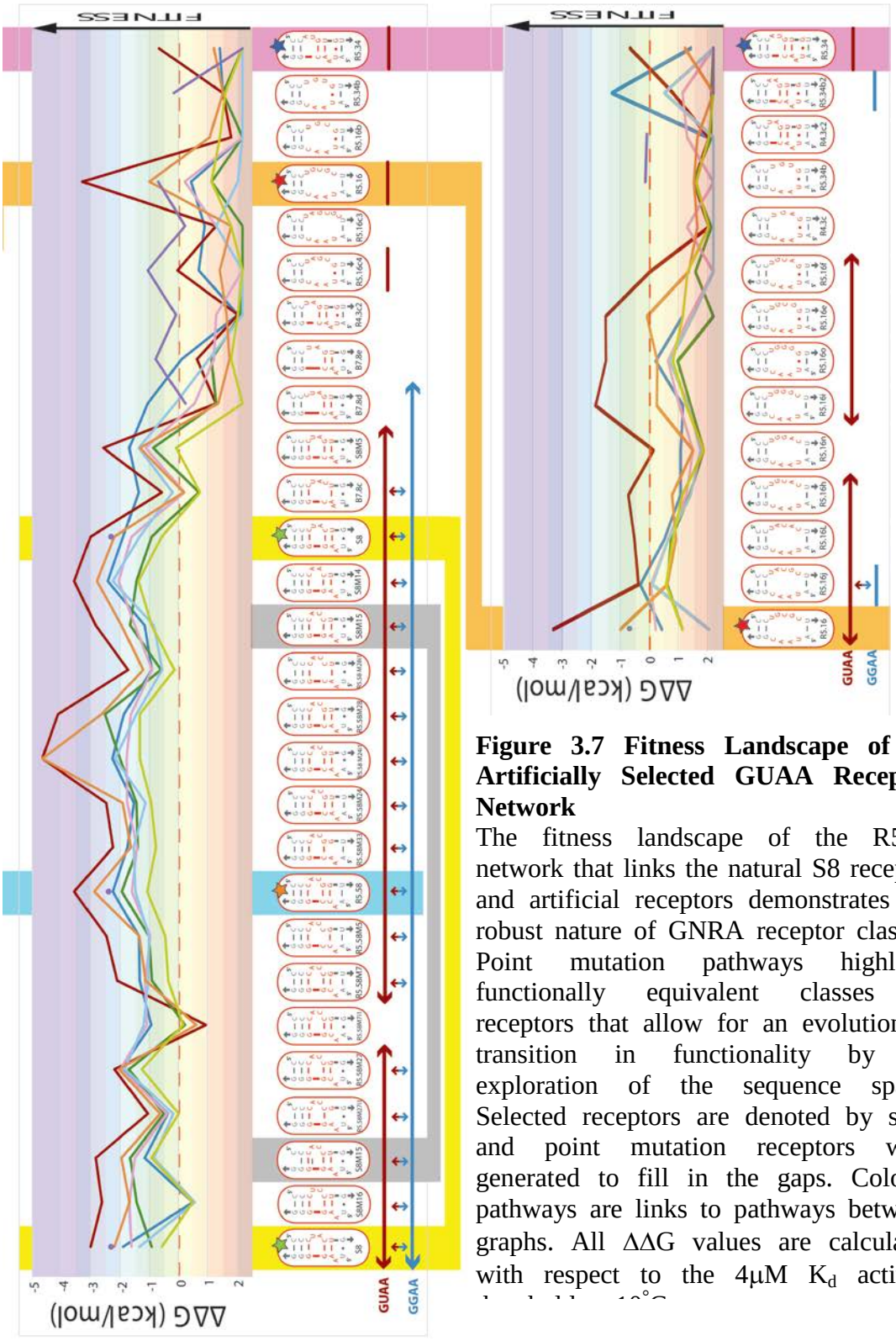


Figure 3.7 Fitness Landscape of an Artificially Selected GUAA Receptor Network

The fitness landscape of the R5.58 network that links the natural S8 receptor and artificial receptors demonstrates the robust nature of GNRA receptor classes. Point mutation pathways highlight functionally equivalent classes of receptors that allow for an evolutionary transition in functionality by an exploration of the sequence space. Selected receptors are denoted by stars and point mutation receptors were generated to fill in the gaps. Colored pathways are links to pathways between graphs. All ΔΔG values are calculated with respect to the 4μM K_d activity

Chapter 4 Attenuation of Loop-Receptor Interactions with Pseudoknot Formation

4.1 Abstract

RNA tetraloops can recognize receptors to mediate long-range interactions in stable natural RNAs. In vitro selected GNRA tetraloop/receptor interactions are usually more “G/C-rich” than their “A/U-rich” natural counterparts. They are not as widespread in nature despite comparable biophysical and chemical properties. Moreover, while AA, AC and GU dinucleotide platforms occur in natural GAAA/11nt receptors, the AA platform is somewhat preferred to the others. The apparent preference for “A/U-rich” GNRA/receptor interactions in nature might stem from an evolutionary adaptation to avoid folding traps at the level of the larger molecular context. To provide evidences in favor of this hypothesis, several riboswitches based on natural and artificial GNRA receptors were investigated in vitro for their ability to prevent intermolecular GNRA/receptor interactions by trapping the receptor sequence into an alternative intra-molecular pseudoknot (PK). Extent of attenuation determined by native gel shift assays and co-transcriptional assembly is correlated to the G/C content of the GNRA receptor. Our results shed light on the structural evolution of natural long-range interactions and provide design principles for RNA-based attenuator devices to be used in synthetic biology and RNA nanobiotechnology.

4.2 Introduction

In nature, long-range RNA interactions involving sequence positions often located hundreds of nucleotides away from each other, contribute to the folding of stable RNAs into functional three-dimensional (3D) structures [31, 97]. The most abundant of all identified long-range interactions are A-minor packing interactions, which occur between stacked adenines and the shallow groove of small helical receptors composed of at least two Watson-Crick (WC) base pairs (bp) [107]. In large ribozymes and riboswitches, A-minor interactions are often part of larger structural motifs involving GNRA tetraloops binding to helices or small receptors, with GYRA/helix and GAAA/11nt receptor interactions being the most widespread (N stands for any base, Y stands for pyrimidine and R stands for purine) [21, 73, 93-96, 98, 119, 150-152]. This observation has raised the question whether GNRA tetraloops other than GYRA and GAAA tetraloops can form equivalent specific long-range interactions. This was addressed, at least partially, when several new receptors for GUAA, GUGA, GAAA and GGAA tetraloops were identified by in vitro selection experiments [73, 153] and were shown to have thermodynamics and loop selectivity comparable to natural ones when tested in standard physiological conditions [153] (Calkins, Baudrey, Jaeger, unpublished data). Most of these in vitro selected loop/receptor interactions, including the GAAA/C7.2 [73], GAAA/C7.10 [73], GUAA/B7.8 [73], GGAA/R1 [153], and GGAA/R2 [153] loop/receptors, are not observed in known group I [154, 155] and group II introns [73, 94], RNase P RNAs [95, 156], molybdenum cofactor riboswitches [105], class I di-GMP

riboswitches [157] and ribosomal RNAs. In order to explain the evolutionary bias towards natural GNRA/receptor interactions versus those obtained by in vitro selection, other selection pressures than those for particular biochemical and biophysical properties should be at work during the structural evolution of natural RNA molecules.

In vitro selected loop/receptor interactions are typically more “G/C-rich” than their natural counterparts. Even naturally occurring GAAA/11nt receptors, which can accommodate AA, AC and GU dinucleotide platforms, strongly favor the AA platform versus all the others [73, 135]. This suggests that natural RNA motif sequences might be selected for their robustness toward intramolecular RNA misfolding rather than for their local thermodynamic stability or selectivity. Natural helical and “A/U-rich” GNRA receptors are more likely to avoid kinetic and thermodynamic folding traps at the level of larger sequence contexts than their artificial counterparts. On the other side, “G/C-rich” receptors might be more suited for designing artificial riboswitches able to attenuate formation of GNRA/receptor interactions.

Previously, we developed a self-assembling tectoRNA heterodimer system based on bimolecular GNRA/receptor interactions that was employed as building blocks for nano-constructions [39, 42, 43, 63, 107, 134, 153, 158] and in vitro selection of novel GGAA receptors [153] (Figure 4.1). Inspired by working principles from natural transcription attenuators [159-162], we have engineered several tectoRNA riboswitches able to adopt mutually exclusive structures that promote or inhibit formation of GNRA/receptor interactions (Figure 4.1). These

tectoRNA riboswitches are used to monitor the ability of several tetraloop/receptor motifs with different G/C content to be thermodynamically trapped by pseudoknot (PK) formation. The mechanism of attenuation of intermolecular GNRA/receptor interactions by intra-molecular PK is investigated by gel shift assays, lead cleavage probing, competition experiments and co-transcriptional assembly assays. While our data sheds new light on the structural evolution of GNRA/receptor interactions, it also provides new design principles for RNA-based switching devices suitable for synthetic biology and nanobiotechnology [163-165].

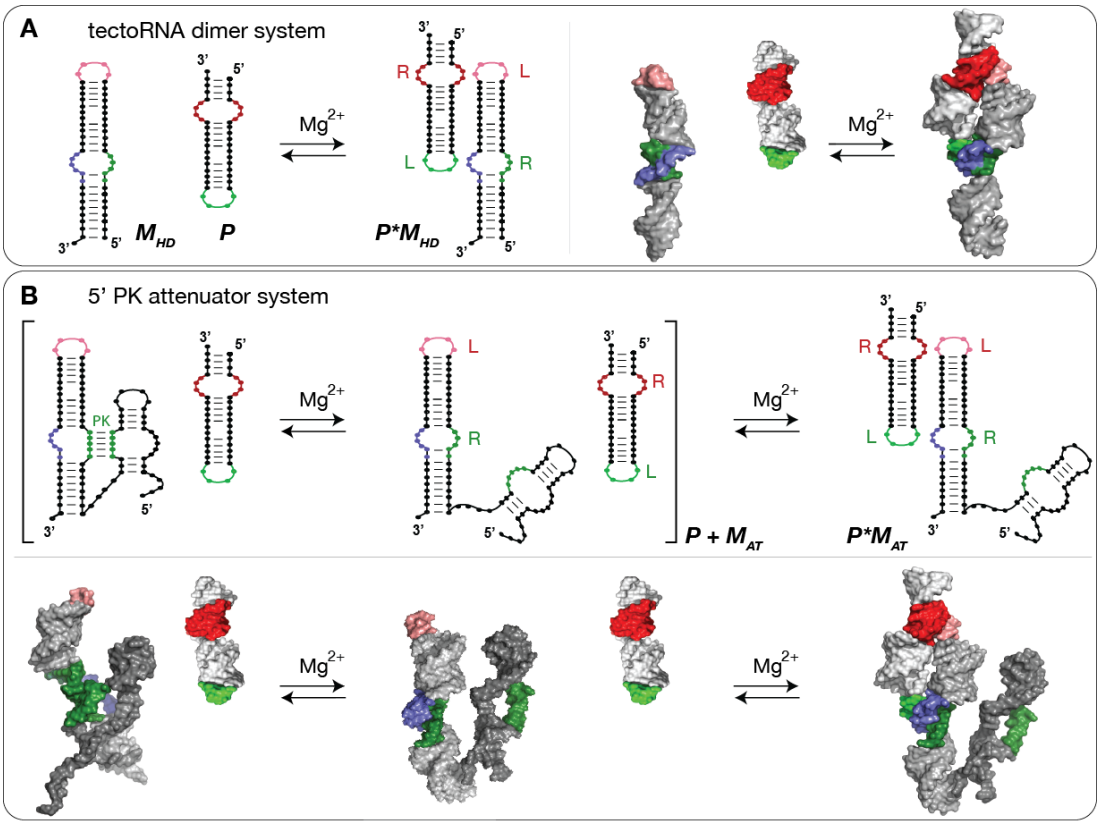


Figure 4.1 TectoRNA Dimer Schematic
Self-assembly equilibrium reactions for the tectoRNA systems reported. (A) TectoRNA heterodimer (HD) system: a HD-forming module (MHD) assembles with a probe (P) through GNRA/receptor interactions to form a heterodimer (P*MHD).

This system is used as control. (B) 5’ PK attenuator system: the tectoRNA attenuator (MAT), consisting of a HD-forming module linked to a PK-forming module, can assemble with a probe (P) through its HD-forming module (equilibrium on the right) to form the heterodimer (P*MAT). Attenuation of intermolecular self-assembly between the tectoRNA attenuator and the probe occurs when the PK-forming module interacts with the 5` side (in green) of the receptor of the HD-forming module to form a 5’ PK (equilibrium reaction between brackets). Interacting receptor (R) and loop (L) motifs as well as pseudoknot (PK) are indicated. Equilibrium reactions and 3D stereo view for the 3’ PK attenuator system are provided Figure 8.11.

4.3 Results

4.3.1 *Modular design of tectoRNA attenuators*

Each tectoRNA attenuator contains a “pseudoknot (PK)-forming” module linked in 5' to a "heterodimer (HD)-forming" module (Figure 4.1 and **Error! Reference source not found.**). The "HD-forming" module is based on a previous self-dimerizing tectoRNA construct [39, 43, 153] that consists of a GNRA receptor tectoRNA unit assembling through bimolecular GNRA/receptor interactions with a tectoRNA probe (Figure 4.1 and Figure 4.2).

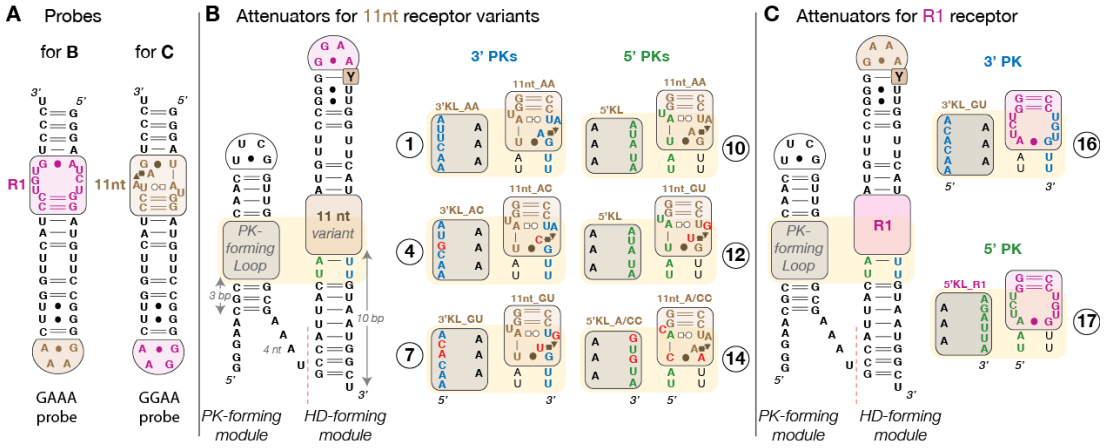


Figure 4.2 Secondary structure diagrams and nomenclature of tectoRNA attenuators reported.
(A) GAAA and GGAA molecular probes. (B) TectoRNA attenuators based on 11nt receptor variants: their HD-forming module can assemble with the GAAA probe. (C) TectoRNA attenuators based on the R1 receptor motif and assembling with the GGAA probe. Indicated RNA constructs (labeled 1 to 17) are those with combinations of PK-forming modules and HD-forming modules that can form 3'PK (between nucleotides in blue) or 5'PK (between nucleotides in green). Red nucleotides are positions that vary from molecules 1 and 10 (with the classic 11nt receptor and corresponding 3' and 5' PK-forming loops). Most constructs tested have a U at the pyrimidine position, which is labeled "Y" within the HD-forming module. Additional constructs with combinations of PK-forming modules and HD-forming modules leading to mismatched PKs or with Y=C have been tested (see Figure 4.3 and Figure 4.5 and Table 8.11). Base-pairings are indicated according to Leontis and Westhof annotation [9]. The 11nt_GU receptor has been reported in previous studies as receptor C7.10 [73, 153]. For a description of the modularity of 11nt receptor variants, see Figure 8.12.

Because of the high recognition specificity of GAAA and GGAA tetraloops by their cognate receptors, HD-forming modules cannot self-assemble in the absence of probe. The "PK-forming" module contains an internal loop called the PK-forming loop (or PKL) and promotes formation of an intramolecular PK with the 5' or 3' side of the receptor from the "HD-forming" module. This PK competitively inhibits assembly with the GRAA probe (Figure 4.1). In other words, pseudoknot formation attenuates tectoRNA dimer formation. While tectoRNA attenuators are based on the same structural scaffold, they essentially differ from one another at the level of their receptor and PKL sequences (Figure 4.2 and Figure 4.3C, Table 8.11). We designed a total of 17 different attenuators (numbered **1** to **17**) by combining five HD-forming modules based on the 11nt and R1 receptor sequences (11nt_AA, 11nt_AC, 11nt_GU and 11nt_A/CC)[73, 153], with various PK-forming modules containing different PKL loop sequences (3'KL_AA, 3'KL_AC, 3'KL_GU, 5'KL, 5'KL_A/CC

and 5'KL_R1), with up to 5 to 6 nucleotides complementary to the 3' or 5' sides of the receptor. The resulting intramolecular PKs are structurally and conceptually similar to the binding modality of stable intermolecular paranemic RNA molecules previously shown to require a minimum of 5 bps for self-assembly [41]. The effects of additional nucleotide variations were investigated within the context of some of

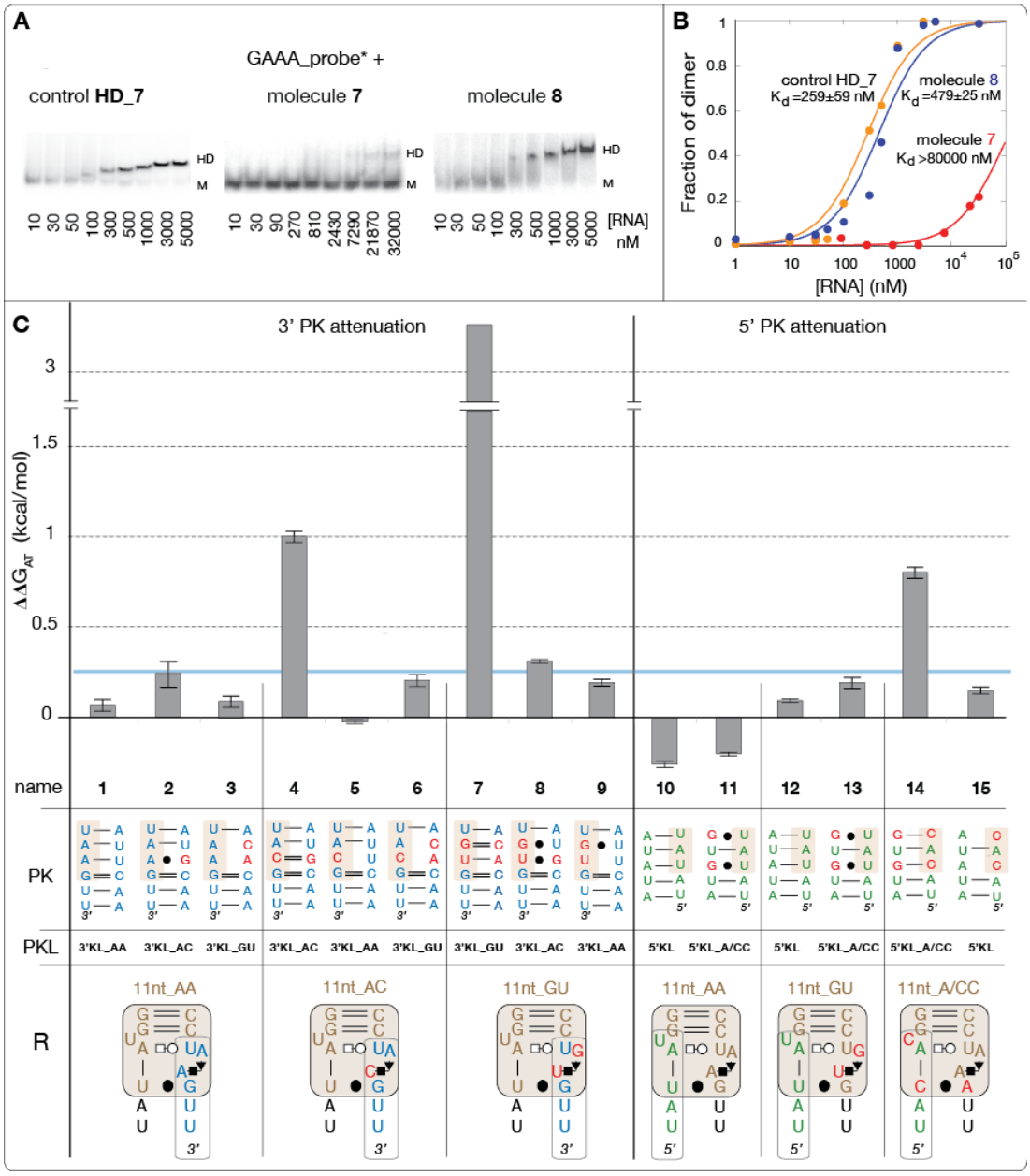


Figure 4.3 Thermodynamic analysis of tectoRNA attenuators based on 11nt receptor variants.
(**A**) Typical examples of native PAGE titration experiments at 15 mM $Mg(OAc)_2$ and 10°C, for the control DF_7 (11nt_GU HD-forming module alone), molecule 7 (11nt_GU HD-forming and 3'KL_GU PK-forming modules) and molecule 8 (11nt_GU HD-forming and 3'KL_AC PK-forming modules) in presence of equimolar concentrations of radiolabeled GAAA probe. M and HD indicate the position of monomers and heterodimers, respectively. (**B**) Titration curves with calculated equilibrium constants of dissociation (K_d's) corresponding to the tectoRNA assemblies in (A). (**C**) Free energies of attenuation of heterodimer formation for all attenuator constructs based on the 11nt receptor variants (see also Table 8.14). The free energies of attenuation ($\Delta\Delta G_{AT}$) were estimated at 10°C and 15 mM $Mg(OAc)_2$ as described in Materials and Methods. The sequence of the intra-molecular 3' or 5' PK, which competes with heterodimer formation by sequestering either the receptor 3' or 5' side, the name of the PK-forming module and the 2D structure and name of the receptor (R) from the HD-forming module are indicated for each tectoRNA attenuator tested (numbered 1 to 15). All constructs with $\Delta\Delta G$ values below the threshold of 0.25 kcal/mol (indicated by the blue line) are considered to have no significant attenuation. This threshold was estimated based on the range of standard error deviations observed through the study (Table 8.11).

these constructs. To check the influence of PKL size on PK formation, two additional adenines were introduced in the PKL, on the strand opposite to the PK forming strand (Molecules of the **1a** to **17a** series). To modulate the binding affinity of the HD-forming module for its cognate probe, the gGRAAu terminal loop of the HD-forming module was changed into a gGRAAc loop (Molecules of the **1"** to **17"** and **1a"** to **17a"** series).

4.3.2 *Characterization of tectoRNA attenuators based on GAAA/11nt receptor interactions*

Free energies of dimerization (ΔGHD) between HD-forming modules and their cognate probe can be derived from equilibrium constants of dissociation estimated

by native PAGE gel-shift assays as indicated in the Materials and Methods (Figure 4.3A and B,). By comparing ΔGHD of a particular HD-forming module in the tectoRNA attenuator context with the one in absence of linked PK-forming module, we can estimate for each tectoRNA attenuator the variation of free energy (ΔΔGAT) that corresponds to attenuation by PK formation (see Materials and Methods).

At 15mM $Mg(OAc)_2$ and 10°C, most HD-forming modules based on the 11nt receptor variants (Figure 8.12) assemble to the GAAA probe with very similar ΔGHD. ΔGHD's for the 11nt_AA, 11nt_AC and 11nt_GU modules are almost undistinguishable (see HD_1, HD_4 and HD_7 in Table 8.13) while the 11nt_A/CC receptor, which differs by three point mutations from the classic 11nt_AA receptor, leads to a minor decrease of 0.43 kcal/mol in binding affinity when compared to the 11nt_AA receptor (HD_14 in Figure 8.12). Overall, this result corroborates the isosteric nature of the AA, GU and AC dinucleotides platforms, which structurally contribute in a similar way to the stabilization of the GAAA tetraloop/11nt receptor interaction within the HD-forming module context.

By contrast, some of the tectoRNA attenuators based on these receptors display markedly different behaviors (Figure 4.3). Attenuators 1, 4, and 7 have PK-forming modules designed to form a PK of 6 Watson-Crick bps with the 3' sides of receptors 11nt_AA, 11nt_AC, and 11nt_GU, respectively. While no significant change in binding affinity is observed for molecule 1 with respect of HD_1, molecules 4 and 7 attenuate heterodimer formation by 1 and 3.3 kcal/mol, respectively. Other 3'PK attenuators with combinations of HD-forming and PK-forming modules that

introduce G:U bps or/and Watson-Crick mismatches in the PK do not present significant attenuation (molecules 2, 3, 5, 6, 8 and 9). This data suggests that attenuation is correlated to the stability of the PK that requires at least two G:C bps to efficiently compete with heterodimer formation. Similar results are provided by 5'PK attenuators 10, 12 and 14, designed to form pairings of 5 Watson-Crick bps with the 5' sides of receptors 11nt_AA, 11nt_AC, and 11nt_GU, respectively. Molecules 10 and 12 able to form a PK of 5 A:U bps are unable to trap the receptors but molecule 14, which can form a PK with 2 G:C bps, attenuates heterodimer formation by 0.80 kcal/mol (Figure 4.3). Not surprisingly, none of the attenuator combinations with PKs with G:U or A:C bps are able to compete with heterodimer formation.

Additional structural evidences for intra-molecular PK formation in attenuator 7 are provided by lead cleavage experiments (Figure 4.4 and Figure 8.13). Lead is widely used as a conformational probe for RNA because it preferentially cleaves the phosphodiester backbone in flexible regions or non-canonically paired motifs of RNA molecules [39, 41, 43, 134, 153]. Irrespective from absence or presence of GAAA probe, the 3' strand of the 11nt_GU receptor of attenuator 7 is strongly protected towards cleavage in comparison to the one of molecules 8 (or 9). Additionally, the 5' PK strand within the PK-forming module of 7 also shows enhanced protection towards lead cleavage relative to 8 (or 9). This strongly suggests that the PK is formed in attenuator 7 but not in 8 (or 9). By contrast, in presence of the GAAA probe, molecules 8 and 9 display partial protection of the receptor and

tetraloop regions from the HD-forming module, corroborating their assembly with the probe.

In summary, the 11nt receptor can easily accommodate sequence variations that are all able to efficiently promote self-assembly with the cognate GAAA tetraloop. However, the ability to trap its sequence in an alternative conformation like a PK, is highly dependent of the presence of Gs or Cs, G:C bps being much more effective than U:A bps for stabilizing alternative Watson-Crick pairings.

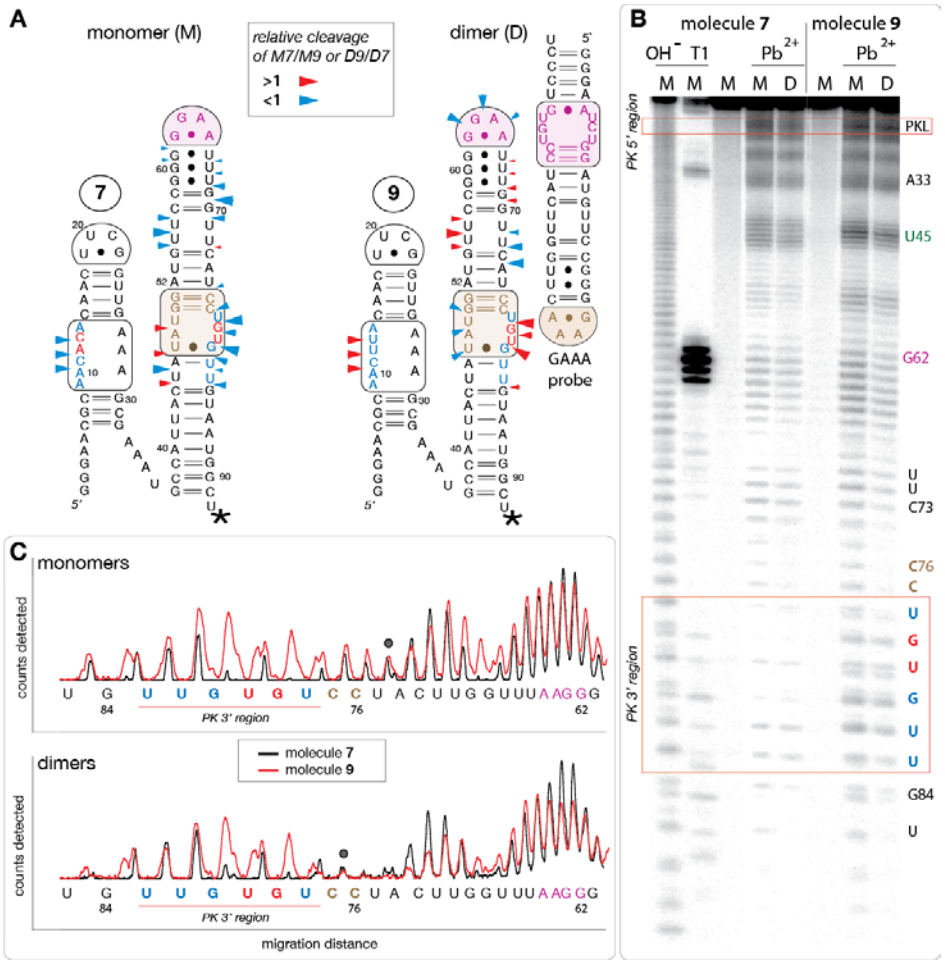


Figure 4.4 Lead(II)-induced cleavage patterns for tectoRNA attenuators 7 and 9 in their monomeric and heterodimeric states.

(**A**) 2D diagrams of tectoRNA attenuators with reported differential Pb(II) cleavage patterns in the monomeric (M) and heterodimeric (D) states. Phosphate positions in monomer 7 (M7) that show enhanced or reduced Pb(II) cleavage with respect to monomer 9 (M9) are indicated by red or blue arrows on the 2D diagram of **7**, respectively. Phosphate positions in heterodimer 9 (D9) that show enhanced or reduced Pb(II) cleavage with respect to heterodimer 7 (D7) are indicated by red or blue arrows on the 2D diagram of **9**, respectively. The size of the arrows is roughly proportionate to the difference in cleavage for M7 versus M9 or D9 versus D7. A star indicates the radiolabeled RNA 3' end. (**B**) Pb(II) cleavage patterns of ^{32}P radiolabeled molecules **7** and **9** either alone or bound to their non-radioactive cognate GAAA probe (as shown in (A)). M and D correspond to monomer and dimer lanes, respectively. Cleavage experiments (indicated by Pb^{2+}) were carried out as described in the Materials and Methods section; OH^- indicates alkaline hydrolysis ladder; T1 indicates RNaseT1 digestion. (**C**) Superposed lead cleavage profiles for monomers **7** and **9** (top) and for the corresponding heterodimers in presence of

GAAA probe (bottom). Black dots indicate positions used for normalization. Similar results are obtained by comparing attenuators **7** and **8** (Figure 8.13).

4.3.3 ***Modulating tectoRNA attenuation with an artificial receptor, point mutations and magnesium***

The GGAA/R1 receptor interaction was isolated by in vitro selection [153]. It is highly selective for the GGAA tetraloop and its affinity is comparable to GAAA/11nt receptor interactions [153]. Interestingly, R1 is four mutations away from the 11nt_AA receptor but only two mutations away from the 11nt_GU with which it shares an identical 3'-side (Figure 4.5A). R1 is, however, more "G/C-rich" than any of the 11nt variants. Consequently, the resulting attenuators **16** and **17** are expected to form stable intramolecular 3' and 5' PKs, respectively: according to Freier's table [166], the calculated stability of the 3'PK of **16** and 5'PK of **17** is -5.3 kcal/mol and -3.5 kcal/mol, respectively. As shown Figure 4.5B, molecules **16** and **17** attenuate heterodimer formation with their cognate GGAA probe by 2.16 and 1.91 kcal/mol, respectively. While both molecules **7** and **16** form the same 3'PK, attenuation with **16** is less dramatic than with **7** in presence of their respective probes; this could be explained by the fact that the **HD_16** heterodimer complex involving the GGAA/R1 interaction, is 0.5 kcal/mol more stable than the **HD_7** heterodimer complex involving the GAAA/11nt_GU interaction (Table 8.12). Therefore, for **16**, heterodimer formation is advantaged with respect of PK formation.

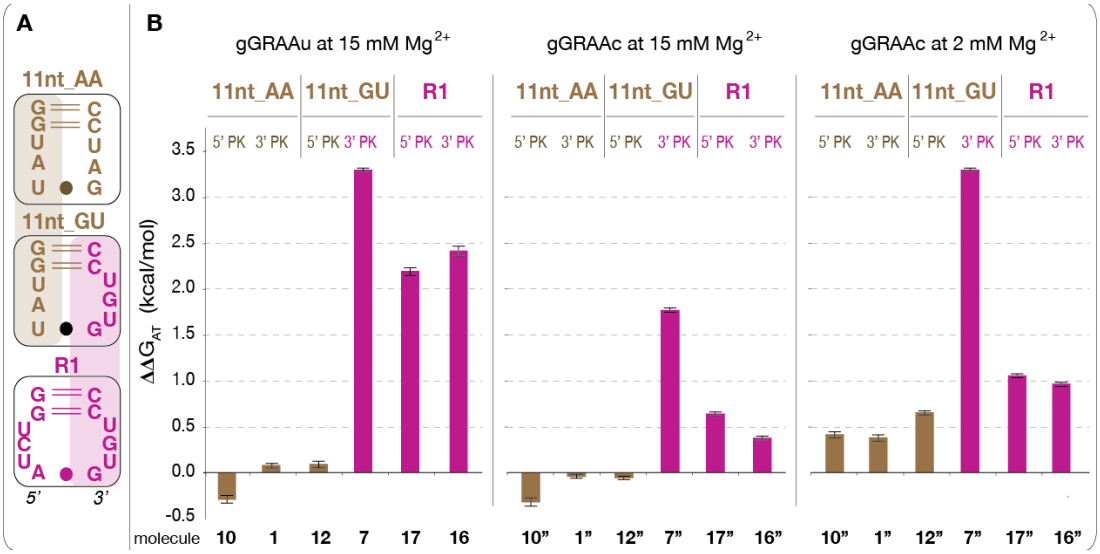


Figure 4.5 Thermodynamic analysis of tectoRNA attenuators based on the 11nt and R1 receptors.
(A) Sequence relationships between the R1 [153], 11nt_GU (or C7.10 [73, 153]) and 11nt receptors. (B) Free energies of attenuation of heterodimer formation for all attenuator constructs based on the 11nt and R1 receptors (see also Table S4). The free energies of attenuation ($\Delta\Delta G$) were estimated at 10°C and 2 or 15 mM $Mg(OAc)_2$ as described in Materials and Methods. Attenuator molecules 1", 7", 10", 12", 16" and 17" differ from molecules 1, 7, 10, 12, 16 and 17 by the presence of gGRAAc terminal loops (instead gGRAAu). This single nucleotide variation increases heterodimer stability. A similar series of attenuator constructs with PK-forming loops involving five As (instead of three) show similar attenuation results (Figure 8.14).

TectoRNA heterodimer assembly, which occurs through two GNRA/receptor interactions, is favored by a point mutation that changes the gGRAA**u** terminal loop of the HD-forming module into a gGRAA**c** loop. The thermodynamic stability of the resulting HD heterodimers is increased by 0.5 to 1.2 kcal/mol at 10°C and 15mM $Mg(OAc)_2$. This is likely due to small structural variations that favor the local stabilization incorporated within attenuators **7**, **16** and **17** (to give **7"**, **16"** and **17"**),

this mutation leads to a reduction of attenuation by 2 to 5 fold (Figure 4.5 Thermodynamic analysis of tectoRNA attenuators based on the 11nt and R1 receptors.B).

TectoRNA assembly is particularly sensitive to small variations in magnesium concentration [39, 43, 107]. By reducing magnesium concentration from 15 mM to 2 mM, the affinity between heterodimer modules and corresponding probes decreases by 1.5 to 2 kcal/mol (Figure 8.14). Intramolecular formation of 3' or 5' PK, which relies on the formation of canonical Watson-Crick bps, should not be as sensitive to magnesium as intermolecular formation of GNRA/receptor interactions. As expected, molecule **7''** and to a lesser extent, molecules **17"** and **16"**, attenuate HD formation more effectively at 2 mM than at 15 mM magnesium (Figure 4.5B).

TectoRNA attenuators of the 1a to 17a series, with two additional adenines in their PKL, were also tested in order to determine whether the size of the PKL could affect PK formation in conjunction with the nucleotide composition of the PK. The behavior of the 1a-17a attenuator series is overall comparable to the one of the 1-17 attenuator series (Figure 4.5 and Figure 8.14). At 15 mM magnesium, a small enhancement of attenuation is noticeable for molecules **1a**, **10a** and **12a** versus **1**, **10** and **12**. This indicates that "A/U-rich" PKs form better when the size of PKL is increased, probably because of less steric hindrance.

In summary, these results indicate that the most effective tectoRNA attenuators are those based on "G/C-rich" receptors such as the R1 and 11nt-GU receptors. The extent of tectoRNA attenuation can be modulated as a function of magnesium concentration as well as peripheral single point mutations in a predictable manner.

From a rational design point of view, our results suggest that artificial "G/C-rich" receptors are better suited than "A/U-rich" receptors for designing riboswitches that require folding into alternative RNA structures. However, from an evolutionary point of view, "G/C-rich" receptors might be disadvantageous because they are more prone to trap native RNA sequences into alternative undesirable structures.

4.3.4 *TectoRNA attenuation during in vitro transcription*

We have also investigated how tectoRNA attenuation could occur during in vitro RNA transcription in isothermal conditions (37°C) (see Materials and Methods). While all the experiments described above were performed in conditions usually favoring thermodynamic control versus kinetic control, co-transcriptional assembly experiments should be more representative of folding and assembly processes taking place within the cell [167, 168]. During the linear phase of RNA transcription, three different types of products are observed on native PAGE: the RNA probe, the tectoRNA attenuator, and the complex resulting from the intermolecular assembly between the probe and attenuator molecules (Figure 4.6A and Figure 8.15). Because of its smaller size, the probe product is transcribed in larger quantity than the attenuator product, explaining why a portion of it always remains unassembled. In presence of GAAA probe, the totality of attenuator **1 (**and **10**) products assembles to the probe (Figure 4.6B). By contrast, only ~60 percent of the attenuator **4** product forms a stable complex with the probe, suggesting that the remaining 40 percent is blocked into the PK conformation state (Figure 4.6B). In perfect agreement with previous data, attenuator **7** (and **7"**) demonstrates full attenuation of tectoRNA

assembly, while attenuator **9**, which differs from molecule **7** by only two point mutations within its intramolecular PK, assembles with the probe to its full extent (Figure 4.6B and Figure 8.15).

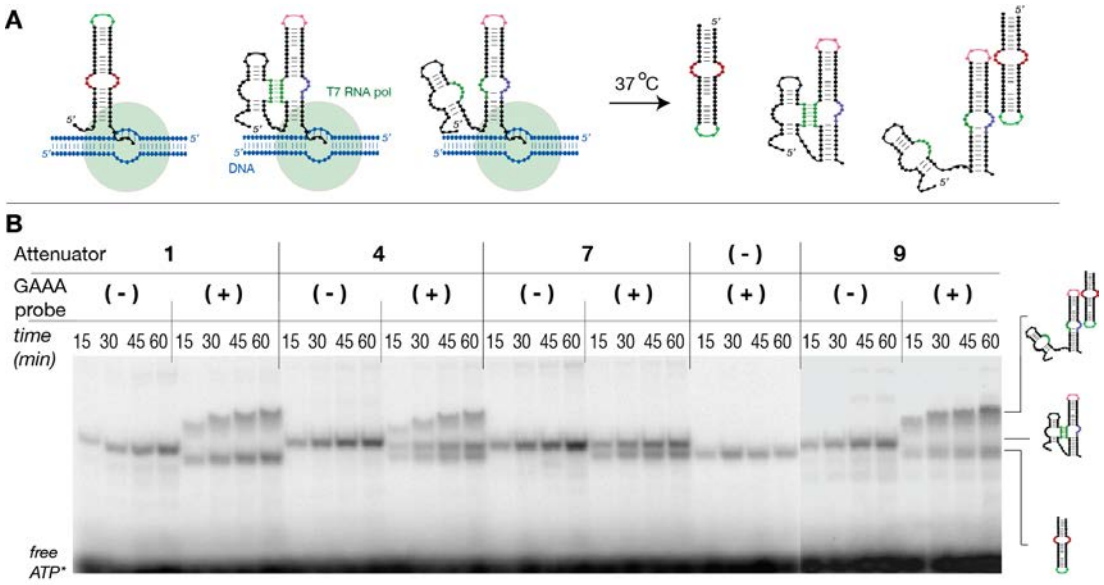


Figure 4.6 Co-transcriptional assemblies of tectoRNA attenuators.
1, **4**, **7**, and **9** in presence (or absence) of cognate GAAA probe. (**A**) Schematic illustrating the possible molecular states adopted by the tectoRNA attenuator system during its transcription from DNA templates (in blue) by T7 RNA polymerase (in green) at 37°C in presence of 10 mM Mg^{2+}. (**B**) Native PAGE analysis of different tectoRNA attenuator transcription mixtures at various times in presence (+) or absence (-) of GAAA probe: co-transcriptional assembly is monitored by RNA body-labeling with $\alpha[P^{32}]$[73]ATP and native PAGE is performed at 10°C and 10 mM $Mg(OAc)_2$ after quenching the transcription with DNase as described in Materials and Methods. See also Figure 8.15.

Interestingly, attenuators **14**, **16** and **17** assemble with their cognate probe to form complexes with faster gel mobility than those obtained with attenuator **1**, **9** and **10** (Figure 8.15). We have observed that tectoRNA complexes with higher K_ds (or lower affinities) typically migrate faster at lower RNA concentrations than those with lower K_d's (or higher affinities) [39, 153]. This behavior has been described as

resulting from monomers and heterodimers being in dynamic equilibrium [39, 153]. Our observation corroborates the fact that attenuators **14**, **16** and **17** bind less efficiently their cognate probe than their corresponding HD_forming modules. In these attenuators, formation of a transient intramolecular PK likely displaces the intermolecular assembly equilibrium towards the monomers.

Overall, co-transcriptional assembly data corroborate those obtained previously. To effectively attenuate tectoRNA assembly, the 3' and 5' PK base pairings need to have a calculated thermodynamic stability lower than -4 kcal/mol and -3 kcal/mol, respectively. Co-transcriptional data also suggest that attenuation can occur through two distinct mechanisms. In the mechanism shared by attenuators **4**, **7** and **7"**, the attenuator RNA product folds into stable PK_forming and HD_forming conformers, which are unable to interchange into one another. This is probably due to the PK_forming conformer acting as a folding trap, unable to switch into the HD_conformer. In the second mechanism shared by attenuator **14**, **16** and **17**, the PK_forming and HD_forming conformers are in dynamical equilibrium with one another, allowing the PK_forming conformer to switch into the HD_forming conformer. In the future, further work will be needed to unravel the dynamical and structural constraints favoring one mechanism versus the other.

4.3.5 ***Controlling tectoRNA attenuation with small RNA switches***

Moving towards more complex tectoRNA attenuator devices, we designed a tri-molecular system aiming at controlling tectoRNA attenuation with small RNA inhibitors acting as molecular switches (Figure 7A). TectoRNA attenuator **18** is

derived from attenuator **7**, from which it differs by four point mutations in the PKL, on the strand opposite to the PK forming sequence (Figure 7B). Two small RNAs, **SW(I)** and **SW(II)**, are designed to favor heterodimer formation between **18** and the **GAAA** probe by preventing intramolecular PK formation (Figure 7 A): they assemble to the PK-forming loop according to two equivalent structural modalities expected to perfectly mimic the assembly of the U65 ψ pocket (U65hp) of human U65 H/ACA snoRNA with its rRNA substrate [169, 170] (Figure 7 B). In presence of the GAAA probe, attenuator **18** has a more modest attenuation potential than attenuator **7** (-0.84 kcal/mol versus -3.24 kcal/mol): this possibly results from the formation of additional non-canonical bps, which might stabilize the internal structure of the PKL of **18** and disfavor PK formation. Nevertheless, when either **SW(I)** or **SW(II)** are added to the mix, the affinity of **18** for the **GAAA** probe significantly increases to be similar to the one observed for the **HD_7:GAAA** probe complex (~0.2 kcal/mol). Molecules **SW(I)** and **SW(II)** bind to the PK_forming module of **18** with K_d's of 7.3 nM and 16.8 nM, respectively. While these values are consistent with previously published results for similar binding interactions [41], they indicate that **SW(I)** and **SW(II)** can completely switch off the PK_forming module of **18** and prevent the formation of the intramolecular 3'PK with the 11nt_GU receptor of the HD_forming module. Therefore, these results provide further evidence in support of the mechanism of attenuation of heterodimer formation through intramolecular PK formation.

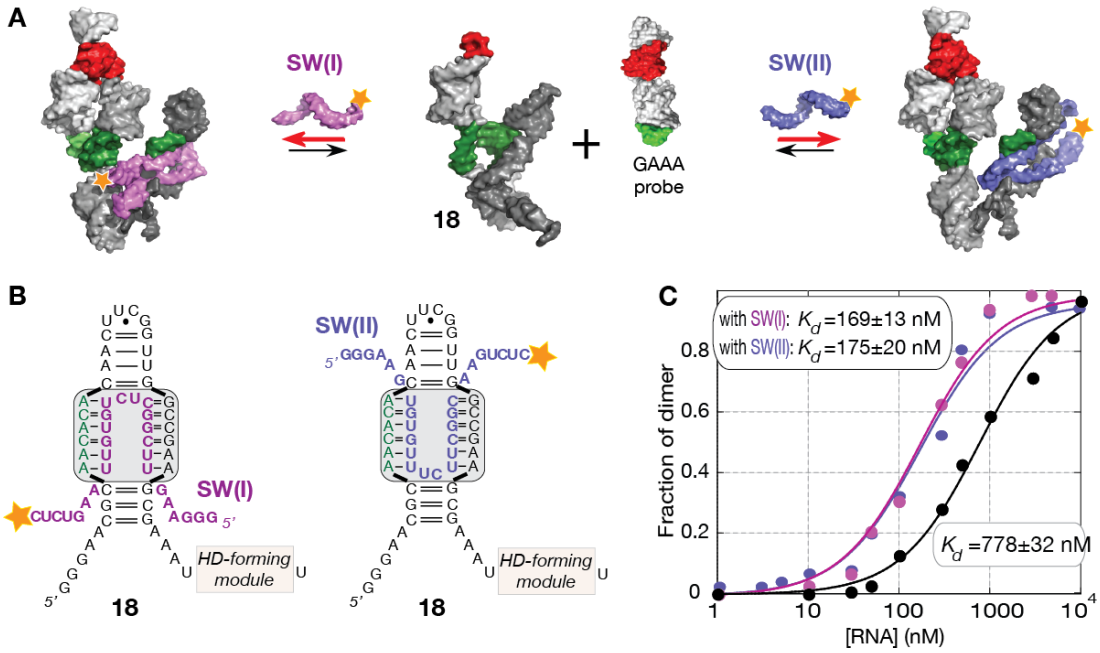


Figure 4.7 Switching off a tectoRNA attenuator with small RNA oligonucleotides.
(A) Schematic of the self-assembly equilibrium reaction of tectoRNA attenuator 18 in presence of two small RNAs (SW(I) and SW(II)) that favor heterodimer formation by preventing internal PK formation. The two small RNAs switch off the PK attenuator module by assembling to the PK-forming loop in a way expected to perfectly mimic the NMR structures of the pseudouridylation pocket of the Box H/ACA snoRNA bound to its rRNA substrate (PDB codes: 2p89 [169] and 2pcv [170]). (B) 2D structure diagrams of the PK-forming module of attenuator 18, bound to SW(I) and SW(II) small switching RNAs. Two possible equivalent binding modalities (boxed in grey) are shown. (C) Titration curves with calculated equilibrium constants of dissociation (K_d's) corresponding to the assembly of attenuator 18 with the GAAA probe in absence (black circles) or presence of switching RNAs (magenta circles for SW(I) and blue circles for SW(II)). Experiments were carried out at 15 mM Mg(OAc)2 and 10°C as described in Materials and Methods.

4.4 Discussion

4.4.1 *Implications for the rational design of RNA 3D structures*

Using a rationally designed molecular system based on tectoRNA self-assembly, we have demonstrated that the formation of GNRA/receptor tertiary interactions can be attenuated by formation of alternative PK structures: the higher the G/C content of the sequence signature of the GNRA receptor motif, the more easily this sequence can be trapped into an alternative pseudoknot structure that attenuates its ability to recognize a GNRA loop target. From a rational design point of view, these data make perfect sense because “G/C-rich” base pairings are thermodynamically more stable than “U/A-rich” base pairings. However, the data also highlight that it is a rather narrow thermodynamic threshold that determines whether the PK can effectively attenuate the formation of GNRA/receptor interactions in vitro. For instance, only one additional C or G in the nucleotide platform of the 11nt motif (11nt_GUand 11nt_AC versus 11nt_AA) is sufficient for stabilizing by 1.5-2 kcal/mol the resulting PKs and lead to attenuation. Nevertheless, if most PKs (with at least one G:C bp) were accurately predicted with the KineFold program [40], the overall quantitative extent of attenuation cannot yet be predicted from purely theoretical thermodynamic analysis. Indeed, our tectoRNA attenuation system depends on the thermodynamic stability of RNA tertiary interactions and structure motifs that essentially involve non-canonical base pair interactions and that are also very sensitive to divalent ion concentration. In vitro cotranscriptional self-assembly revealed that attenuation could proceed in isothermal

conditions according to two mechanisms that could be distinguished based on whether the PK_forming and the HD_forming conformers are in dynamical equilibrium or not. Previous studies [166, 168] have shown that kinetics and thermodynamics make different contributions to RNA folding in vitro and in vivo: it is therefore possible that the exchange between stable alternative tertiary structures might be more rapid in vivo than in vitro [166, 168]. Clearly, further work will be necessary to understand tectoRNA attenuation mechanisms in more detail, especially within the context of cells. In any case, our data already provide RNA modules and design principles that can be used for developing controllable artificial RNA nano-switches with tunable binding properties for nanobiotechnology and synthetic biology applications. For instance, we have demonstrated that RNA tertiary interactions can be specific target locations for designing RNA switches that allow precise modulation of the folding and assembly of these RNA molecules. Because of the large number of topologically equivalent GNRA/receptor interactions presently available [72, 73, 171], these interactions offer high structure designability for the rational design of RNA nanodevices [158]. The choice of a particular RNA self-assembling motif can vary depending on the intended design goal. A/U-rich RNA motifs can maximize a unique folding pathway by minimizing undesirable folding traps resulting from the formation of alternative base pairs. Alternatively, G/C-rich tertiary motifs with one or two Gs (or Cs) localized on the same strand, can be used to design artificial RNAs with distinct alternative conformational states. Pseudoknots involving 5 to 6 base pairs with two G:C bps are sufficient for competing with the formation of GNRA/receptor interactions.

However, as a difference of a few kcals can displace the equilibrium towards a unique molecular state, a good empirical understanding of the energetic balance between the thermodynamic strength of competing tertiary interactions is necessary for designing truly tunable devices.

4.4.2 *Implications for RNA structural evolution*

More importantly, our data provide possible clues for RNA structural evolution as they can explain the existence of particular sequence patterns coding for RNA tertiary interactions. GNRA mediated interactions in stable RNAs are essentially dominated by two families: GYRA/helix and GAAA/receptor interactions. For example, in the class I di-GMP riboswitches [157, 168, 172], ~35 % are GYRA/helix motifs and 65% are GAAA/11nt-like receptor motifs (Table S5A). Within the GAAA/11nt-like receptors, 64% have no more than one G or C, 24% have two G or C and only 12.3% have three G or C or more (Table 8.15B). Therefore, in the 11nt receptor family, “U/A-rich” rather than “G/C-rich” nucleotide compositions are favored at the level of the 11nt internal loop, with AA platforms (70.4%) being more abundant than AC (8.6%) and GU (11.3%) platforms. Interestingly, the thermodynamic stability and GNRA selectivity of the “G/C-rich” 11nt receptor variants are not significantly different from those of the more “A/U-rich” 11nt receptors (Table 8.12 and Figure 8.16B). Moreover, in addition to class I di-GMP riboswitches, several other natural RNA contexts like RNase P RNAs [95, 156], group I [173, 174] and group II introns [73, 94],contain GYRA/helix interactions, which are thermodynamically less stable than

most 11nt receptors, in place of GAAA/11nt motifs. Considering the range of observed riboswitch behaviors in response to the evolutionary need for precise genetic regulation, the sequence of the GNRA/receptor interaction from the di-GMP riboswitch may be tuned so that the riboswitch functions more as a dimmer or rheostat than a binary on/off switch [175]. In natural RNAs, the GAAA tetraloop is universally more abundant than any other GNRA loops [176, 177] (Table 8.15A). However, the predominance of the GUAA tetraloop over other GYRA and GRRA tetraloops is not consistent and might depend on the molecular and genomic context [176, 177]. As exemplified for the di-GMP riboswitch of class I, it is particularly striking that less than 1.5% of the GNRA/receptor interactions take advantage of other GRRA tetraloops (Table 8.15A) while it has been recently demonstrated that artificial GGRA receptors, such as the highly stable and selective R1 and R2receptors, could be isolated by SELEX [72]. In fact, using the RNAmotif software [133], we searched for the R1, R2 and 11nt_GU (C7.10) sequence signatures and did not identify any of them in known natural RNA sequences such as class I di-GMP riboswitch [157], molybdenum cofactor riboswitch [105], group I introns, group II introns and RNAse P RNA sequences from the rfam database [178]. While we cannot rule out the possibility that these receptors exist in some genomes, our search already suggests that they are much less common than the 11nt receptor. Based on the mere consideration of thermodynamics, kinetics and loop selectivity, it is not obvious to explain the strong bias towards “A/U-rich” GAAA/receptors or GYRA/helix receptors in natural stable RNAs. Clearly, higher order selection pressures imposed by the larger structural context of natural RNA

molecules, the kinetics constraints on the global folding of RNA inside the cell or the possible involvement of additional cellular components are likely at play [178]. Based on our present data, the most straightforward explanation is that the preferred occurrence of natural RNA motif sequences stems from an evolutionary adaptation that make them less prone to misfolding and therefore less likely to interfere with the folding of a large RNA sequence (through formation of alternative pairings or interactions with other regions of the RNA sequence). As we have clearly demonstrated that "G/C-rich" receptors are more likely to be trapped into alternative PK structures than "A/U-rich" receptors, we propose that, in cells, the natural GYRA/helix and "A/U-rich" GAAA/11nt receptor interactions result from two evolutionary strategies that minimize kinetic and thermodynamic folding traps in large RNA structural contexts. The first strategy, best exemplified by the "classic" GYRA/helix interaction, takes advantage of receptors that use Cs and Gs to maximize the formation of stable local Watson-Crick helical regions, preventing them to form long-range alternative pairings. The other strategy takes advantage of AU-rich internal loop motifs, like the 11nt receptor motif, that minimize the formation of stable alternative base pairings. Interestingly, the IC3 receptor motif, a natural GNRA receptor identified in IC3 group I introns [171], can be seen as a mix of both strategies, with Gs and Cs involved in local base pairs and U and As involved in a small asymmetrical internal loop [72, 171]. Recently, Mitra and colleagues [179] proposed that, in group I introns, the greater thermodynamic stability of a native conformation over nonnative structures, achieved through selection of strong tertiary interactions, comes at the expense of

slower folding to the catalytic conformation due to formation of long lived intermediates. Efficient folding is therefore achieved by balancing the gain in structural stability due to tertiary contact formation with the probability of misfolding due to loss of conformational freedom. As such, by contrast to other strong but G/C-rich GNRA/receptor interactions selected in vitro [72, 73], the GAAA/11nt motif offers a unique sequence pattern with great thermodynamic strength and lower probability to create stable alternative structures. In conclusion, when a structural motif is part of a large structural network, the avoidance of alternative folding traps could be a significant selective advantage during evolution. We have therefore an example of how the sequence information of the whole RNA molecule could affect from the top-down the sequence information of the small local structural part (58). Similar to the usage of codons in cells, many synonymous RNA structural motifs exist but they are not identical after all [180]. A good understanding of the structural designability of RNA motifs is therefore key for RNA architectonics [9, 35, 63, 107, 181]) and the future development of RNA synthetic biology and nanobiotechnology, especially when artificial RNA molecules need to operate in vivo.

Chapter 5 Downward Causation by Information Control in Microorganisms

5.1 Abstract

The concepts of functional equivalence classes and information control in living systems are useful to characterize downward (or top-down) causation by feedback information control in synthetic biology. Herein, we re-analyze published experiments of microbiology and synthetic biology that demonstrate the existence of several classes of functional equivalence in microbial organisms. Classes of functional equivalence from the bacterial operating system, which processes and controls the information encoded in the genome, can readily be interpreted as strong evidence, if not demonstration, of top-down causation by information control. The proposed biological framework reveals how this type of causality is put in action in the cellular operating system. Considerations on top-down causation by information control and adaptive selection can be useful for synthetic biology by delineating the irreducible set of properties that characterizes living systems. Through a "retro-synthetic" biology approach, these considerations could contribute to identifying the constraints behind the emergence of molecular complexity during the evolution of an ancient RNA/peptide world into a modern DNA/RNA/protein world. In conclusion, we propose top-down causations by information control and adaptive selection as the two types of downward causality absolutely necessary for life.

5.2 Introduction

Despite being at the core of natural sciences, the understanding of the nature of causation has attracted the attention of a rather small number of scientists. However, with the rise of systems biology and synthetic biology that aim at understanding biological systems in a more global, holistic way, causal concepts could be helpful to those interested by problems related to the emergence of complexity in life. Understanding the nature of causation in living systems can provide the necessary framework to characterize the relationships existing between different levels of biological complexity. As the perceptible world around us is based on matter, most scientists usually assume that causal effects stem from the bottom-up and that the complexity of living systems could be explained in terms of simple "bottom-up" and "same-level" causations. However, this present view likely reflects only parts of the truth and can be put to the test by the fields of systems biology [182, 183] and synthetic biology [14, 184-186]. For instance, the recent area of synthetic biology aims at designing, synthesizing and engineering new biological functions and systems not found in nature. Therefore, one can get a better understanding of how new functions and systems should be integrated to fulfill the properties of life.

For most scientists interested by questions related to the problem of emergence of complexity, it is generally accepted that the development of emergent properties, which is clearly a bottom-up (or upward) causality, is influenced by top-down (or downward) causality (e.g. references [187-190]). While bottom-up causation is the ability of lower levels of reality to have causal power over higher levels, top-down

causation (TDC) is typically defined as the ability of a higher hierarchic level to have causal power over lower levels. Therefore, while biological components (for example proteins and RNA molecules) have causal effects on the functioning of a whole biological system, it is also apparent that the whole, as a context, can have an effect on the lower components through boundaries and constraints that determine the outcomes of lower level causations ([189, 191] and references herein). Simple and concrete examples from biochemistry and molecular biology can demonstrate this phenomenon. For instance, a particular RNA sequence can fold into a functional three-dimensional conformation via an intricate network of hydrogen bonds between its constituent residues. In the context of the whole RNA sequence, the conformation of a set of residues is subject to the residues that surround it and can be energetically less favorable than the one that it might eventually adopt in the context of a smaller portion of the RNA sequence. Consequently, a complex RNA or protein illustrates how a network of structural interactions as a whole can affect the properties of its constitutive parts [192].

It was recently suggested by George Ellis [191, 193] that there are at least five different types of top-down causation taking place in natural sciences depending on the context: namely, algorithmic TDC; TDC via non-adaptive information control; TDC via adaptive selection; TDC via adaptive information control; and intelligent TDC. Recognizing these different forms of causation implies that other kinds of causes than physical and chemical interactions are effective in the real world [191, 193]. Because of its key implication in living systems, TDC by information control was investigated more extensively by Auletta, Ellis and Jaeger [192] in order to

propose an experimental and biological framework aimed at testing top-down causations at the bacterial level. TDC by information control occurs when a higher-level entity influences lower-level entities so as to attain a specific functional outcome (goal) through feedback information control loops [192]. The feedback control system depends essentially on the flow of information, coupled to an evaluation of that information relative to the particular functional outcome.

In Auletta et al. [192], we proposed that the establishment of classes of equivalence in living organisms could potentially be used as objective criteria for demonstrating the existence of TDC by information control. A class of functional equivalence is defined by a functional outcome (goal) that is operated by lower level components that can be different as long as they produce the same outcome. By exemplifying the conservation of functions rather than the conservation of modes of operations, the existence of classes of equivalence strongly suggests that it is the biological system as a whole that defines the boundaries and constraints within which a particular class of functional equivalence is established by natural selection. In other words, it is a functional need developed by the whole biological system that defines the constraints within which a particular class of equivalence is established. However, without understanding their respective roles within the context of a whole cell, functional equivalence classes are more symptoms of TDC rather than being direct proofs of TDC. It is therefore of prime importance to clearly establish the controlling instances behind the existence of an equivalence class and to understand how this class is established by natural selection. To address this question, it is beneficial to look at bacterial cells because of their simpler organization. Presently,

tool kits from synthetic biology allow the demonstration of the existence of functional equivalence classes in bacteria by experiments of complementation, which typically consist in the replacement of one functional gene by another one in a single organism [192] (see 8.4.1 in Supplemental Information). This can subsequently demonstrate TDC by information control if a good understanding of the controlling instances behind the existence of the class of equivalence studied can be provided.

In the present paper, we first define a bacteria cell based on the concepts of master functions, networks and flow of information. We then outline the concept of classes of functional equivalence and briefly describe the way they can be experimentally established in microbiology and synthetic biology. We then review several experiments published in the literature that present compelling experimental evidence for the existence of various functional equivalence classes in bacteria. By focusing on functional equivalence classes belonging to the cellular operating system of bacteria, we establish the information control instances behind their existence and propose that TDC by information control in conjunction with TDC by adaptive selection can to some extent explain how the functions from the cellular operating system came to be through evolution. As such, TDC by information control and adaptive selection is at the root of what characterizes living systems on Earth. These considerations provide a helpful framework for future experiments in synthetic biology, which will facilitate the buildup of a minimal living system and will assist in the establishment of the higher-level functions that characterize living systems of greater complexity than bacteria.

5.3 Defining Bacterial Organisms

To better understand how TDC operates in bacteria, a basic knowledge of how a cell functions is essential. By taking the example of bacteria, one can view this biological system as a single cell, a homogenous cellular colony or a more elaborate ecosystem involving a genetically heterogeneous cellular population. Controlling instances might be ascribed differently depending upon the context used as a reference. Herein, we consider the cell as a unit that is a complex intracellular network defined by chemical reactions, structural and functional parts, and the ensemble of their functional relationships with respect of one another.

5.3.1 *Hierarchical functional networks and master functions*

The understanding of the cell's functional organization is only possible through its understanding as a cellular network (e.g. [194-197]). While metabolic, genetic and regulatory sub-networks can be distinguished, it is the ensemble of these networks that constitute the fundamental system characterizing life. In a holistic way, it is the cell as an irreducible "controlling" unit that determines the set of informational feedback control loops necessary for its survival. In the literature, the metabolic regulatory network is often distinguished from the genetic regulatory network but it is clearly the interplay between the two that forms the controlling informational instances of the cell. The metabolite pools act as intracellular molecular signals that link the metabolic network to the genetic network through genetic regulations (e.g. [198]). Despite their apparent daunting complexity, these biological networks are themselves very modular in their overall organization with

subdomains and network motifs being easily distinguishable [195, 199]. In fact, mathematical models suggest that biological networks are inherently simple with modular units that could virtually perform independently [196, 199, 200]. Nevertheless, much remains to be done at an experimental level for understanding this modularity in term of functions, which is ultimately to comprehend what is the essential functional network that defines the invariance for life and how these functional modules are wired to create the fundamental living network.

Regardless of the above considerations, a living cell can be described at an organic level, as an autonomous functional entity while its constitutive parts cannot. The way we approach the cell as a network is to look at it as a hierarchical system of functions in informational relationships with one another where master functions are the functions essential for life. These master functions can be defined as higher order functional cyclic networks comprised of multiple subfunctions (Figure 5.1). They unify and integrate the system into a single autonomously behaving and responding, temporarily persistent, identifiably acting entity. A master function can be reduced to a minimal set of sub-functions, which itself is typically irreducible. Building up from von Neumann's idea of self-reproducing automata [201], Danchin [202] proposed that a bacterial cell could be described as a biological computer that is split into a machine (cellular machinery) and a program (genome). The cellular operating system (COS) is the link between the program and the cell machinery (Figure 5.1c). By involving multiple operations such as replication, transcription, translation and regulation, the COS is at the root of the two main master functions of bacterial life: the replication of the genome and reproduction of the cell machinery. The intimate

informational network that results from the symbiotic association of these two master functions is what characterizes cellular life. As such, "cellular life" could be seen as the higher-level master function. For a cell, its outcome is the formation of two daughter cells from one. Albeit reductive, this view is useful for describing the phenomenon of TDC by information control.

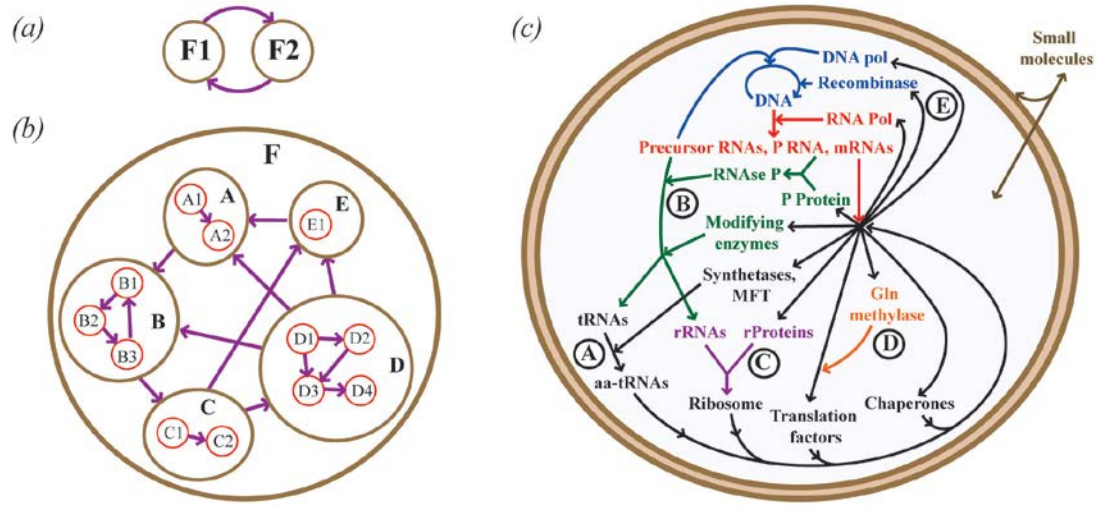


Figure 5.1 The concept of master functions and networks in bacterial cells.
(a) Two functions can be linked to one another by informational relationships (arrows) resulting from sharing, exchanging information or action on one another to form a functional cyclic network. **(b)** A master function F can be defined as a higher order functional network of lower functional entities (A to E) that can themselves involve multiple functional operations (e.g. A1, C2, D4, etc) in informational relationship with one another. (**c**) Example of functional cyclic network: the minimal cellular operating system (COS) network: it comprises biological macromolecules and pathways supposed necessary and sufficient for replication (and reproduction) from small molecule nutrients diffusing through the bilayer lipid vesicle (in brown) (Figure adapted from reference [203]). Biomolecules and chemical reactions are colored according to the biochemical pathway they belong to: DNA replication (blue), RNA transcription (red), RNA processing (green), ribosome assembly (violet), protein translation (black) and post-translational processing (orange). MFT: methionyl-tRNA^{fMet} formyltransferase. Circled letters correspond to steps subjected to complementation experiments described Table 5.1 (A: [204, 205]; B: [206, 207]; C:[208, 209]; D:[210, 211]; E:[212]).

5.3.2 *The cell as a flow of information*

While there are different kinds of information, it can generally be defined as that which brings about a reduction of uncertainty or indeterminacy [213, 214]. An anticipated characteristic of all master functions is that they are functional cyclic networks. Note that the "chemoton" proto-cell described by Szathmary and Smith [215] fits the present description of a cell as a functional cyclic network [216, 217]. Because of their cyclic organization, these networks can also be seen as informational networks with inbuilt informational links acting as feedback control loops. As such, an informational link can be broadly defined as any type of physical and/or chemical interaction existing between two functions. This link has both a functional and informational meaning by contrast to random molecular interactions that are part of the noise. Within the context of a cyclic network, this link can include a chemical reaction resulting in the transformation of one molecule into another, the molecular factors affecting this reaction such as the operation of an enzyme on a substrate and the operation of a molecular effector acting as a signal on another molecule.

The flow of information pertaining to the various biomolecular processes of the COS is schematically represented in Figure 5.2 within the context of the present "dogma" of molecular biology of the cell. It exemplifies the relations between the three major categories of informational polymers, DNA, RNA and proteins. The control of informational flows contributes to cellular homeostasis. Within the COS, informational flows mediated by protein, RNA and metabolite functions locally form numerous cycles. The COS is therefore characterized by multiple feedback loops of

information. As such, the COS has an inbuilt control of the information it processes. It contributes to the selection of the "quality" of the information carried by the informational molecules and regulates the level of expression of the information.

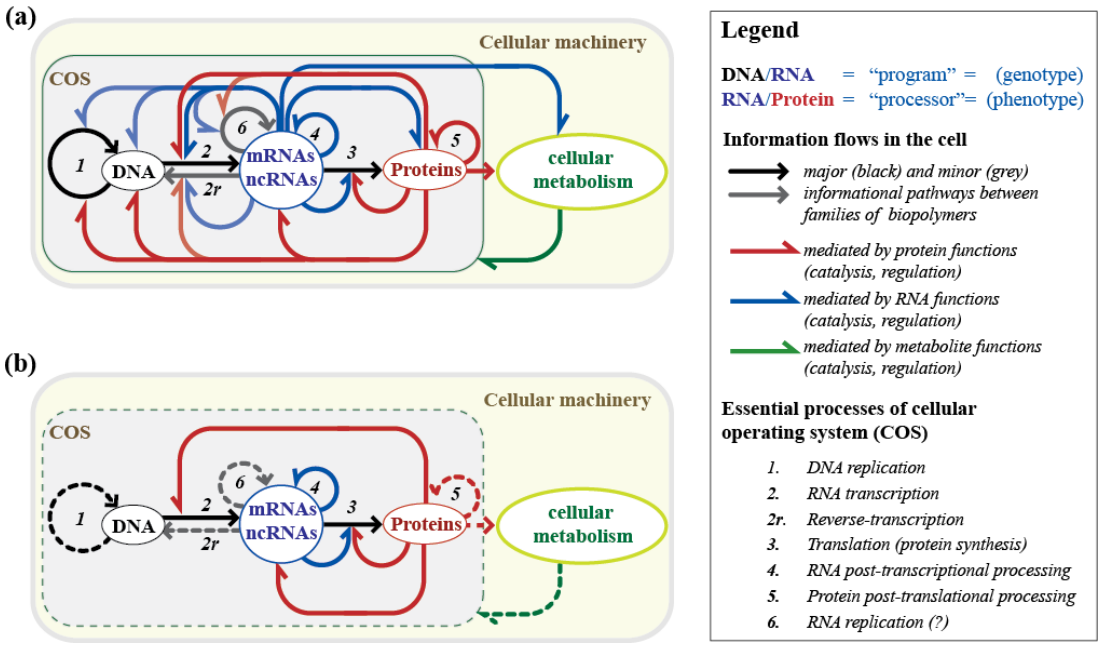


Figure 5.2 Flow of information within the central dogma of a cell.
The diagram in **(a)** conceptualizes and summarizes the different pathways and feedback loops that put in relation the three major categories of informational polymers involved in the COS. Abbreviations: ncRNAs, non-coding RNAs; mRNAs, messenger RNAs (coding for proteins). RNA replication (step 6) is a process that occurs in Eukaryotic cells [218, 219]. It might also occur in bacteria. The cellular metabolism can theoretically have an informational effect on any of the COS (this is indicated by a green arrow directed from the cellular metabolism to the COS). **(b)** Simplified diagram with information feedback loops involved in the processing of ncRNAs that directly participate to translation (e.g. rRNA, tRNA). Dashed arrows indicate steps that are not essential to the process. The gene corresponding to the ncRNA is transcribed (step 2) into a ncRNA precursor that might require post-transcriptional processing (step 4) so it can be used by the translational apparatus for synthesizing proteins (step 3). Translation (step 3) is possible only when the information encoded at the level of the ncRNA gene is valid. The quality of the information carried by the ncRNA is necessary for the production of the enzymes that are responsible of the production of the mature ncRNA (the RNA polymerase (step 2), the proteins of the translational apparatus (step 3) and the proteins involved in the maturation (step 4). This feedback loop controls the quality of the informational properties of the ncRNA.

In a functional network, it is possible that the flow of information forms a self-replicative cyclical operational network, also called a hypercycle [220]. However, functional cycles are not all self-replicative and could also take place at the level of lower constituent functional entities (Figure 5.1b).

5.4 The Cell and TDC by Information Control and Adaptive Selection

The sub-functions that constitute a master function emerge from the bottom-up. A master function as complex as the "modern day" replication of the cellular program is however extremely improbable to have emerged by mere "chance" (see Figure 5.1c). Even in the simplest modern cell, a master function "ought to come" from simpler molecular systems that nevertheless keep the intrinsic properties and characteristics of the master function conserved through the process of evolution. Therefore, a master function creates the functional constraints and boundaries within which the lower functions can evolve, diversify and eventually become more complex. It is important to realize here that it is the COS that operates the reproduction and replication of the cell. The various molecular components of the COS can interact, recombine and diversify to lead to further complexification and differentiation, offering new capabilities and potentialities. Thus, the COS does constitute, in the philosophical sense, the system formal cause [192]. This characterizes a top-down causal effect by the whole system (defined by master functions) on the molecular parts (defined by the functions of lower level that

operate the master functions). Considering the master functions of a cell, the functional constraints are informational constraints. Each sub-function that has evolved within the master function is constantly selected for the quality of its information. Therefore, it is the master function that ultimately "decides" whether the information will be kept or not. This is characterizing TDC by information control [192]. As the sub-functions are intimate parts of the master function, the acceptance or rejection of the new information carried by the new sub-functionalities leads either to the survival or death of the cellular system. However, within the context of evolution, death as an outcome is not a problem. There are typically millions of cells that can proofread the quality of the information and function in parallel (like parallel processors of information). To be fully operational, the cell therefore needs to perform within a space of possibilities. The process of selection for valid information is operated "in a blind way" by multiple copies of identical (or quasi-identical) cells. This is characteristic of TDC by adaptive selection. Bacteria cells can therefore be characterized by TDC by information control in conjunction with TDC by adaptive selection.

In summary, the TDC framework outlined above implies the cell can be operated by different modes of operations or functions of lower levels as long as the properties defined by the master functions are conserved [192] (see Figure 5.1). The direct philosophical implication of this framework is as follows: through TDC by information control, the master functions affect the outcome of their functional parts that contribute to the conservation of the properties of the master functions that define the whole system. Additionally, the master functions define the boundaries

within which the lower level functions evolve by natural selection. Aspects relative to the emergence of the functional properties for cellular life are discussed at a later stage in this paper. Within an evolutionary process (that takes advantage of adaptive selection by definition), it is because of information control from the top-down that the phenomenon of convergence of functions can possibly take place.

The concept of classes of functional equivalence can be extremely useful for providing experimental biological evidences in favor of this framework. Indeed, the demonstration that different modalities can operate similar or identical functions within the cell and that these functions are under information feedback loop control would constitute a rather eloquent proof of TDC by information control.

5.5 Classes of Functional Equivalence in Biology

In biology, a class of functional equivalence is defined by different modes of operations (typically defined by a set of molecules) that have the same function or lead to the same functional outcome [192]. The underlying concept of functions is therefore crucial for asserting without ambiguity whether two distinct modes of operations are equivalent or not in living organisms. Potential molecular candidates to a particular class of functional equivalence are often identified by comparative genomic sequence analysis of different species. This approach mainly allows identification of homologous functional molecules that result from evolutionary divergence of different organisms (Figure 5.3b) and that operate by similar conserved mechanisms (e.g. RNA polymerases, ribosomes or RNase P RNAs from different organisms). However, equivalent modes of operations are not necessarily

related by homology as they can result from ultimate evolutionary convergence (Figure 5.3c). Two functionally analogous sets of biomolecular operations may have completely different unrelated structural features, different mechanisms of catalysis, different modalities of recognizing substrates and may even involve a different number of molecular components. Still, they can have similar if not identical functional outcomes (Figure 5.3d-e). Consequently, analogous molecular candidates to a particular functional equivalence class require extensive experimental characterization of their function. This can partially explain why the number of examples of molecular functional convergence reported in the literature is still scarce.

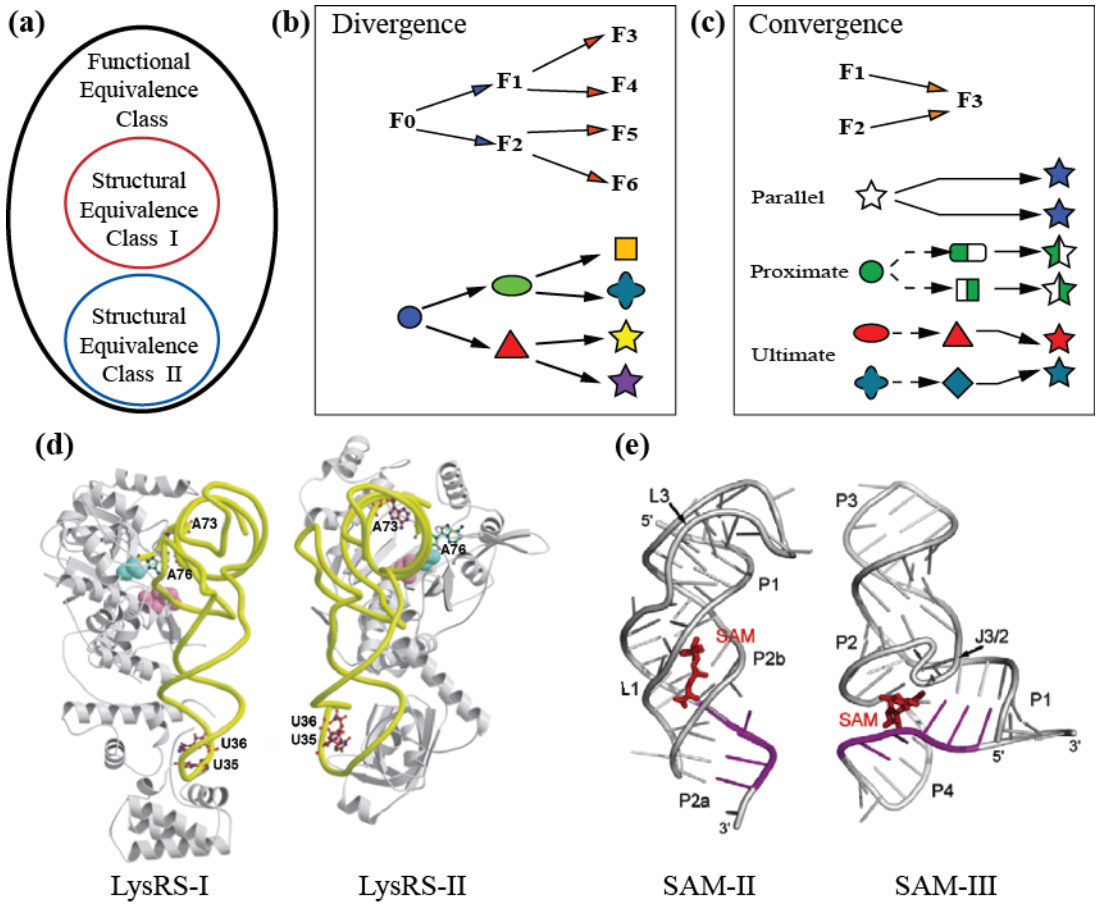


Figure 5.3 Evolutionary divergence, convergence and classes of functional equivalence.

(a) A functional equivalence class can be formed of different structural equivalence classes. (b) By evolutionary divergence, a functional molecular system can lead to new molecules with different functionalities (i) and significant sequence and structural variations (ii). Some of these molecules can still retain the same function and belong to the same functional class. Molecular divergence implies that the divergent molecules are evolutionary related as they have a common ancestor. (c) Two different unrelated molecules can evolve towards the same function through evolutionary convergence (i) (e.g. [221]). At a molecular level (ii), one can distinguish parallel convergence, where the same molecular system in two different organisms evolves independently the same way, proximate convergence, where a molecular system that has diverged significantly in two lineages converges towards the same features, and ultimate convergence, where two different evolutionarily unrelated molecular systems converge towards the same functional features (e.g. [222]). By contrast to ultimate convergence, parallel and proximate convergences are phenomena that pertain to evolutionarily related molecules. (d-e) Classes of

functional equivalence with molecules resulting from ultimate convergence: (d) Functional convergence of Class I and Class II Lysyl-tRNA synthetases. These two enzyme classes have different structural topologies with different modalities of recognition of Lysyl-tRNAs (adapted from [223]). They are distinct structural equivalence classes that belong to the same class of functional equivalence defined by Lysyl-tRNA aminoacylation functionalities. (e) Functional convergence of class SAM-II and class SAM-III of S-adenosyl-methionine (SAM) riboswitches [224]. These two distinct structural classes have different modalities of recognizing SAM [225, 226]. Figure adapted from reference [227].

The concept of functional equivalence classes allows the dismissal of evolutionary concerns such as homology or analogy because what really matters when comparing two modes of operation is their degree of structural and functional similarities, not their evolutionary relationship per se. Nevertheless, within a class of functional equivalence, evolutionary considerations are useful to identify its most compelling members, which should be structurally very different (different shapes and structure topologies) and most likely result from evolutionary convergence [221, 222].

Complementation experiments are the most straightforward methods for establishing the functional equivalence of different biomolecules or pathways in organisms (Figure 5.4 and 8.4.1). They are part of the tool kits of synthetic biology. By replacing a set of operations, defined by one or more molecules with a particular functional outcome, for another one, it is possible to verify whether the new set of operations is able to recover the initial function within an organism. Because of the remarkable functional modularity of living organisms, numerous classes of functional equivalence could potentially be defined that way. It is however important to keep in mind that the demonstration of functional equivalence by

complementation is dependent upon the initial experimental conditions. In some optimal conditions, two sets of operations might look functionally alike (equivalent) while in more discriminative conditions, they might show distinctive functional behavior. Grouping two different sets of operations into the same class of equivalence can therefore require a certain degree of coarse graining. Moreover, if the complex interconnectivity between components of a sub-function is not fully understood (i.e. not all of its operations have been characterized), then replacing only part of its operations might appear to have no functional equivalence.

From an evolutionary point of view, complementation experiments are good models for horizontal gene transfer (HGT), also called lateral gene transfer. In bacteria, it is not uncommon for genes to be transferred horizontally from one organism to another. HGT allows bacteria to acquire new functional modules and can therefore be seen as contributing to the innovation and potential emergence of new functions. It is also through this modality of genetic exchange that functional convergence of different modes of operations can potentially be established [228]. For instance, the quasi-universality of the genetic code shared by all living organisms might be resulting from such unifying processes [228].

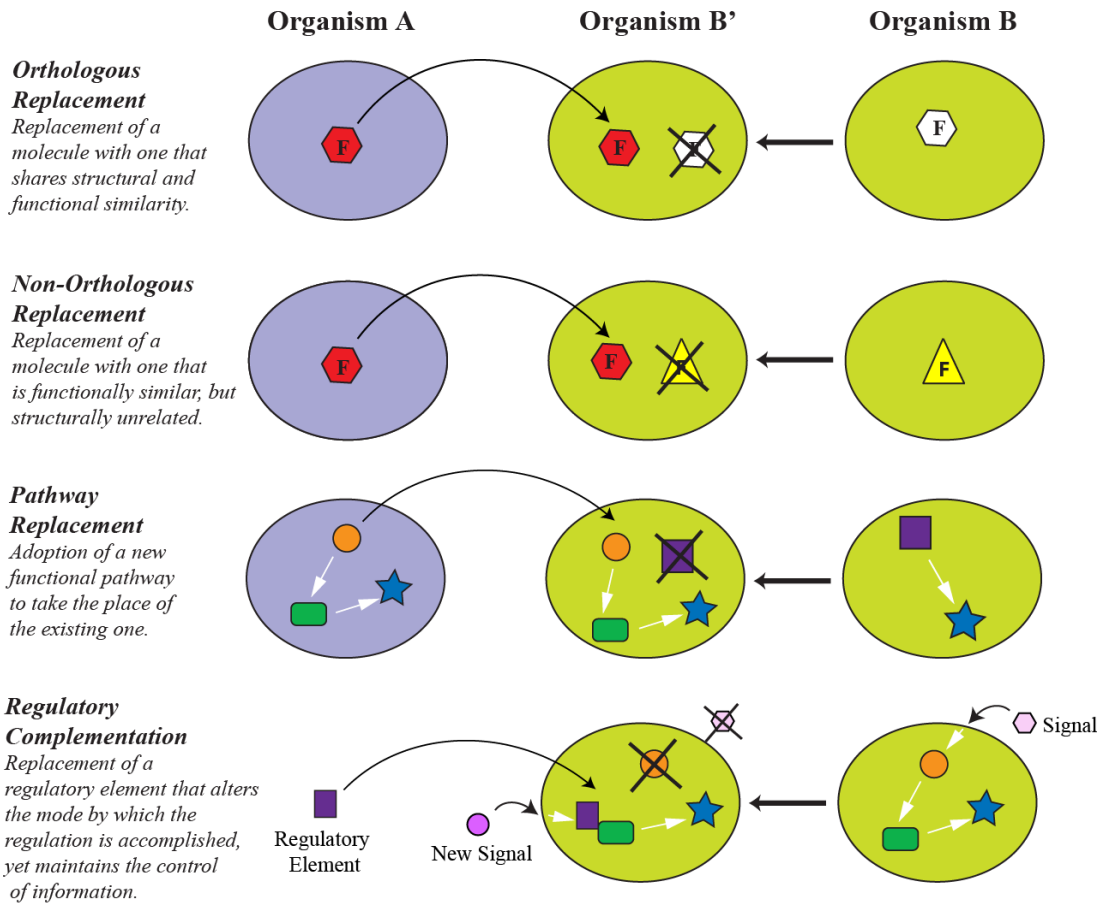


Figure 5.4 Examples of four different categories of complementation experiments.
Organism A is the donor of the operation of function F and organism B', resulting from organism B after nock out of the operation of function F in organism B, is the recipient. Molecular entities are represented by small colored geometric symbols. Selected references corresponding to each category of complementation: orthologous replacement [207], non-orthologous replacement [204, 206], pathway replacement [205, 229], regulatory complementation [230, 231]. Additional examples are provided Table 5.1.

5.6 Experimental Evidence for Classes of Functional Equivalence

Several examples of complementation experiments published in the literature demonstrate the functional interchangeability of different molecules or even different

molecular pathways that are involved in similar cellular processes (Table 5.1). They establish *in vivo* the existence of various classes of functional equivalence within the OS genetic network [204-206] and the metabolic networks [232, 233] in bacteria, Archaea and lower Eukaryotes (e.g. yeast [233]). Some of these complementation experiments demonstrate that functions from Eukaryotes could be similar or at least partially similar to the ones found in Bacteria, indicating strong conservation of functionalities throughout the three different branches of life [233]. With the exception of the work by Wegscheid et al [207], none of the reported complementation experiments has ever been interpreted in the literature within a conceptual TDC framework.

Function	Molecular systems	Type of complementation (evolutionary significance)		Reference
		Recipient(s)	Donor(s)	
Genetic OS network				
$tRNA^{lys}$ aminoacylation	Class I (*B. subtilis*) and class II *(B. burgdorferi*) Lysyl-tRNA synthetases	Non-orthologous replacement (convergence)		[204]
		B. subtilis (Bact.)	*B. burgdorferi (Bact.)*	
$tRNA^{gln}$ aminoacylation	Gln-$tRNA^{gln}$ direct (*E.coli*) and indirect (*B.subtilis*) aminoacylation pathways	Pathway replacement (convergence)		[205]
		E. coli (Bact.)	*B. subtilis (Bact.)*	
tRNA processing	Type A RNase P RNA (*E. coli*) and MRORP1 protein (*Arabidopsis*)	Non-orthologous replacement (convergence)		[206]
		E. coli (Bact.)	*Arabidopsis (Euk.)*	
TMP synthesis	ThyA (*E. coli)* and ThyX (*Borrelia burgdorferi)* Thymidylate synthases	Non-orthologous replacement (convergence)		[211]
		E. coli (Bact.)	*B. burgdorferi (Bact.)*	
protein folding	Rpl25 (S. *cerevisiae*) and TF (*E. coli*) ribosomal protein chaperones	Non-orthologous replacement (convergence)		[208]
		E. coli (Bact.)	*S. cerevisiae (Euk.)*	
tRNA processing	Type A (*E. coli*) and type B (*B. subtilis*) RNase P RNAs	Orthologous replacement (divergence)		[207]
		E. coli (Bact.) *B. subtilis (Bact.)*	*B. subtilis (Bact.)* *E. coli (Bact.)*	
Ribosome assembly	rRNA/r-protein operons	Orthologous replacement (divergence)		[209]
		E. coli (Bact.)	*S. typhimurium (Bact.)* *P. vulgaris (Bact.)*	
DNA recombination repair	RAD54 (*S. cerevisiae*) and AtRAD54 (*Arabipdopsis*) repair proteins	Orthologous replacement (divergence)		[210]
		S. *cerevisiae (Euk.)*	*Arabidopsis (Euk.)*	
Post-translational processing	alg7 (*S. cerevisiae*) and mv1751 *(M. voltae*) N-glycosylation proteins	Orthologous replacement (divergence)		[212]
		S. *cerevisiae (Euk.)*	*M. voltae (Arch.)*	
Metabolic network				
Lipid-linked oligosaccharides translocation	ABC type Wzx (*E. coli*) and non-ABC-type WlaB (*C. jejuni*) flippases	Non-orthologous replacement (convergence)		[232]
		E. coli (Bact.)	*C. jejuni (Bact.)*	
Inorganic pyrophosphate hydrolysis	Soluble (*S. cerevisiae)* and membrane-bound H^+-translocating (*Arabidopsis)* inorganic pyrophosphatases	Non-orthologous replacement (convergence)		[233]
		S. cerevisiae (Euk.)	*Arabidopsis (Euk.)* *C. aurantiacus (Bact.)*	
Antibiotic Resistance	Low Mg^{2+} (*S. enterica*) and high Mg^{2+} (*E. coli*) Polymyxin B resistance pathways	Pathway replacement (divergence)		[229]
		E. coli (Bact.)	*S. enterica (Bact.)*	
Molecular transport	YopB/YopD (*P. aeruginosai*) and PopB/PopD (*Y. pestis*) proteins	Orthologous replacement (divergence)		[234]
		P. aeruginosai (Bact.)	*Y. pestis (Bact.)*	
Regulatory network				
Bacteria motility	Pseudotaxis pathway (dependent on a theophylline-riboswitch) and natural chemotaxis pathway (*E.coli*)	Regulatory complementation (convergence)		[230]
		E. coli (Bact.)	synthetic parts	
Cellular and hormonal regulation	Lower and higher eukaryote calmodulins	Orthologous replacement (divergence)		[235]
		S. cerevisiae (Euk.)	*X. Laevis (Euk.)*	
Transcriptional regulation	piD261/Bud32 (*S. cerevisiae)* and PRPK (human) kinase proteins	Partial orthologous replacement (divergence)		[236]
		S. cerevisiae (Euk.)	*H. sapiens (Euk.)*	

Table 5.1 Complementation Experiments

Examples of complementation experiments demonstrating the existence of functional equivalence classes in the genetic, metabolic and regulatory networks. The listed examples only represent a small subset of experiments of complementation or gene replacements. They were identified in PubMed with the key words: complementation, gene, function replacement (or displacement), orthologous, parallel, nonorthologous (or non-orthologous), heterologous. The

experiment by Wegscheid et al [207] has been described in reference [192]. Experiments referenced [204-207, 230] are described in the text.

The complementation experiments involving functions carried out by amino-acyl tRNA synthetases (aaRS) [204, 205] (Figure 5.3d), RNase P [206], flippases [232] and pyrophosphatases [233] demonstrate without ambiguity that very different modes of operations at a molecular level could be substituted by one another as long as their function remains the same (Table 5.1 and Figure 5.5). In all cases, the molecular instructions that are exchanged have likely originated by ultimate convergent evolution [221, 222] because they do not share any apparent structural similarities. The best example of all is the complementation experiment in *E. coli* where the RNase P function, which is essentially based on an RNA in bacteria, is replaced by a purely proteinaceous RNase P from plant organelles [206]. By substituting a function carried by an RNA with one carried by a protein, one cannot argue here that the two molecular systems are swappable simply because they share conserved structural features. In this experiment and others [204-206, 232, 233], the only feature that the exchanged molecular systems share is their functional outcome. These experiments are therefore much stronger support of TDC than the one [207] originally mentioned by Auletta, Ellis and Jaeger [192]. Indeed, in Wegscheid et al [207], the RNase P RNAs that were exchanged are still phylogenetically related despite their significant biophysical differences.

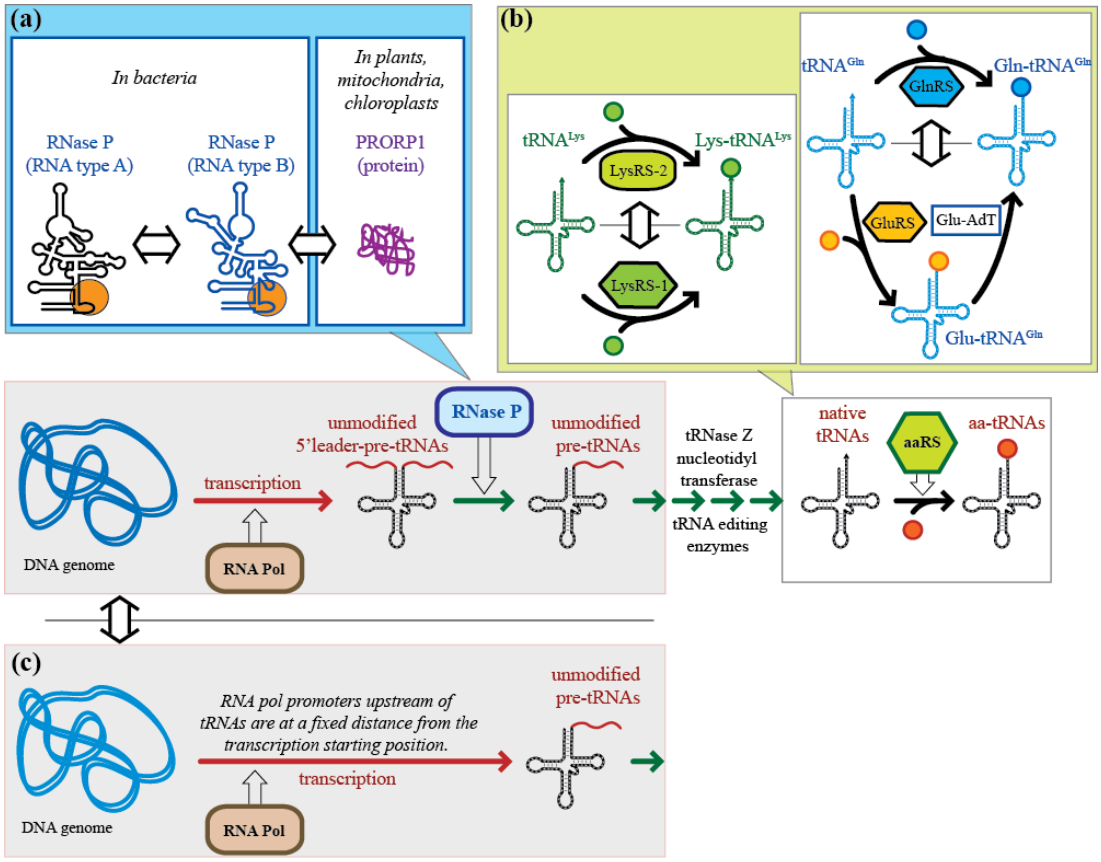


Figure 5.5 Examples of functional equivalence classes within the process of tRNA synthesis and activation from the bacterial operating system.
(a) In E. coli, the endogenous type A RNase P RNA can be substituted by a type B RNase P RNA [207]. It can also be substituted by PROP1, a proteinaceous RNase P from plant mitochondria (Arabidopsis thaliana)[206]. (b) Left: Lysyl-tRNA synthesases (LysRS) have been found in different organisms to be either of class 1 or class 2. Substituting a class 2 bacterial LysRS by a class 1 archaeal LysRS still allows the bacteria to operate the OS [204]. Right: The direct pathway of amino acylation of glutaminyl tRNA found in E. coli can potentially be substituted by an indirect pathway involving first mischarging of tRNAgln with Glu by glutamyl-tRNA synthetase (GluRS) to form glutamyl tRNAGln. Then, glutamyl tRNAGln is converted into glutaminyl tRNAgln by Glu-tRNAGln amidotransferase (Glu-AdT) [205]. (c) The process of maturation of 5'-leader pretRNA by RNase P does not exist anymore in Nanoarchaeum equitans. Instead, the tRNA genes code for tRNA precursors without 5' leader because transcription starts now at the level of the mature 5' end [237]. With respect of mature tRNAs production for translation, the Nanoarchaeum process is expected equivalent to the RNase P dependent's one. However, this remains to be tested in vivo. Color code for reaction processes is the same as in Figure 5.1c.

Based on the cellular framework presented above (Figure 5.2), the classes of functional equivalence identified in the COS readily provide strong clues, if not proof of TDC by information control. As shown in Figure 5.5, both tRNA aminoacylation (aaRS) and tRNA maturation (RNase P) are essential processes for the proper implementation of the basic protocol of the genetic code. If tRNAs are not able to be properly amino-acylated or matured, the enzymes responsible for tRNA production (RNA polymerases, aaRS and RNase P protein) cannot be properly synthesized by the translational apparatus (involving the ribosome) (see also Figure 8.17). Therefore, the very existence of the multiple feedback control loops at the level of the COS offers several possibilities to verify the quality of the operations necessary for replication and cellular reproduction (see Figure 5.2b).

In these examples, we can clearly establish that the information selection defining the operational elements of the class is conserved despite the differences in lower level variables. This is demonstrating TDC by information control. A more philosophical framework is also shown Figure 5.6b.

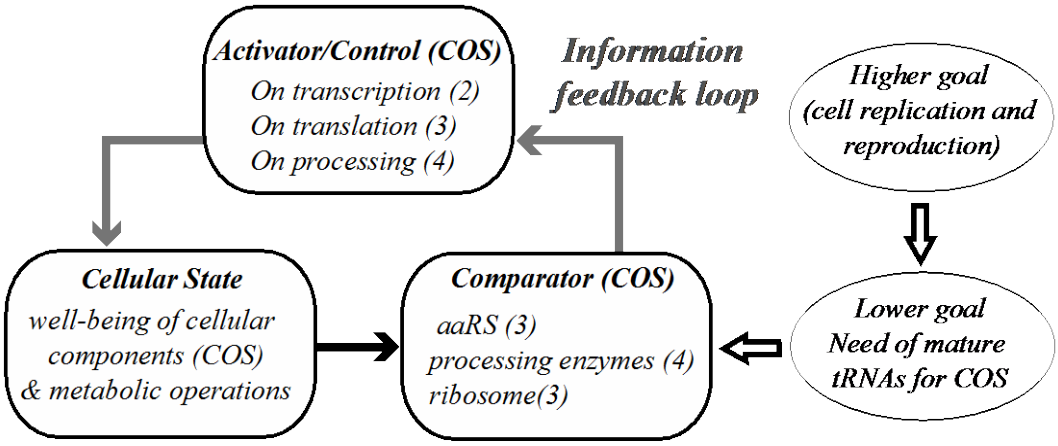


Figure 5.6 Philosophical interpretation of TDC by information control in the cellular operating system.
Numbers correspond to the steps indicated in Figure 5.2.

Presently, it becomes apparent that regulatory elements, which control the expression of gene operons in bacteria, can have very different modes of operations for regulating similar gene operons in different bacteria (e.g. [238, 239]). At least three different categories of attenuation mechanisms (ribosome mediated, protein mediated and uncharged-tRNA mediated) are involved in the regulation of the tryptophan operon in bacteria [238, 240]. Another remarkable example is given by the very distinct structural classes of S-Adenosyl-methionine (SAM) riboswitches that seem to be interchangeable for controlling via an attenuation mechanism the same type of genes in different bacteria species [224, 241] (Figure 5.3e). These examples are far from being isolated cases and would require a thorough investigation. At a molecular level, functional convergence is likely more common than it was initially thought to be. By looking at cellular functions that are linked to the intrinsic regulatory mechanisms of the bacterial cell as well as the extrinsic

physical and chemical environment in which the bacteria live [186], other interesting conserved functional features under TDC by information control are likely in existence, especially at the level of regulatory controlling elements. In the future, it will also be of prime importance to look at functional networks in an evolutionary context to ultimately understand their modularity and possibly how their parts came to be [199].

5.7 Discussion

One of the important questions relative to living systems is to identify the core functional properties that characterize life. TDC by information control and adaptive selection could be particularly insightful for explaining the emergence of novel functions in living systems. These considerations could also be useful for proposing new experiments of synthetic biology.

5.7.1 *The emergence of the core functional properties of life*

Among the definitions of life, many share the notion that cellular life is associated to the emerging properties of a replicating informational molecular system able to mutate [202, 242, 243]. While the informational template should be allowed to change for exploring a space of possibilities for new functions, it still needs to retain the ability to replicate effectively. Exploration of novel functions is possible because of imperfections in the replication process, which creates the constraints for this exploration. Like the pre-existing information, the new resulting molecular information requires selection for its viability within the system. In other

words, the new information is expressed and controlled for its ability to operate within the system. The control of the "quality" of the information (through feedback control) is embedded within the replicating system and can be seen as a necessary underlying property associated to the function of replication. However, the issue here is not merely to replicate but to also properly segregate the new information resulting from replication so that what is essential to the system is kept while what is deleterious (lethal) to the system is disregarded. Therefore, the replicating system requires compartmentalization with selective reproduction of the molecular sub-functions responsible for DNA replication and cell reproduction, as well as the novel functions that are not detrimental to the cellular master functions. While there is a drive here for perpetuating the informational master properties within the system through time (a drive for life that allows exploration of novelty), there is also a drive for eliminating parts of the molecular information that do not fulfill the master cellular functions (therefore a drive towards decay and death). The main selection drive behind perpetuation of the informational properties is to retain among others, the novel functions that allow intake of the sources of chemical energy, production of the molecular building blocks necessary for the synthesis of the biomolecules that support the master functions of the cell as well as their repair, etc. The main selection drive behind the elimination of undesirable information is to retain for example the functions that favor degradation processes and the segregation of undesirable molecules from the correct ones as well as elimination of the cellular systems that have aged, etc [244]. Because of TDC by information control and

adaptive selection, one can perceive how unrelated molecular systems could converge towards similar or identical functions during evolution.

The master functions of the cell (replication and reproduction) are the properties that characterize the cellular unit as a whole. They are defined at the molecular level by a set of relationships that abstractly define the COS network and absolutely need to be conserved (Figure 5.2). Through the COS, the cell selects the information that it can process, which is also the information that constitutes parts of the COS itself. The COS is by definition built up for working via feedback through information control. Any information that it processes, if disruptive, is ultimately eliminated because it induces the destruction of the whole cell. By contrast, if selected, the COS can operate with it and the replication of the information and the reproduction of the cell can occur. The cell therefore operates through TDC by information control; the higher-level master functions have causal power on the set of lower level functions or operations that are causally effective for controlling all the information processed by the cell in order to assure conservation of the master functions.

Interestingly, the constitutive lower level functions of replication and cellular reproduction can be carried out by different modalities of operations. This is particularly well exemplified by the COS of Bacteria, Archaea and Eukaryotes. Despite sharing similar overall mechanisms and numerous similar macromolecular machineries for replication, transcription and translation, (e.g. RNA polymerases, ribosomes, DNA replisomes) they can nevertheless present significant differences at the level of the molecular operations leading to these functional outcomes. As long as the master functions are conserved, the intracellular network of molecules and

their associated interactions can significantly vary as long as the overall process is maintained. For instance, different classes of enzymes are structurally highly divergent or even unrelated at the level of their structures (e.g. class I and class II lys-tRNA^{lys}, Figure 5.3d) but essentially carry the same function at the level of the COS (see above).

The phenomenon of TDC by information control through adaptive selection can explain how living systems could possibly have some of their parts evolving through time from an RNA/peptide world to a DNA/RNA/protein world [83] (Figure 5.7). As long as the basic master functions of the cell are conserved, the cellular system can increase in complexity by creating new functions that fit the goal of the master function. The optimization of the COS is essentially driven by the need to efficiently process a greater amount of information. The fitness of the cell is defined here in terms of being able to efficiently operate the master functions.

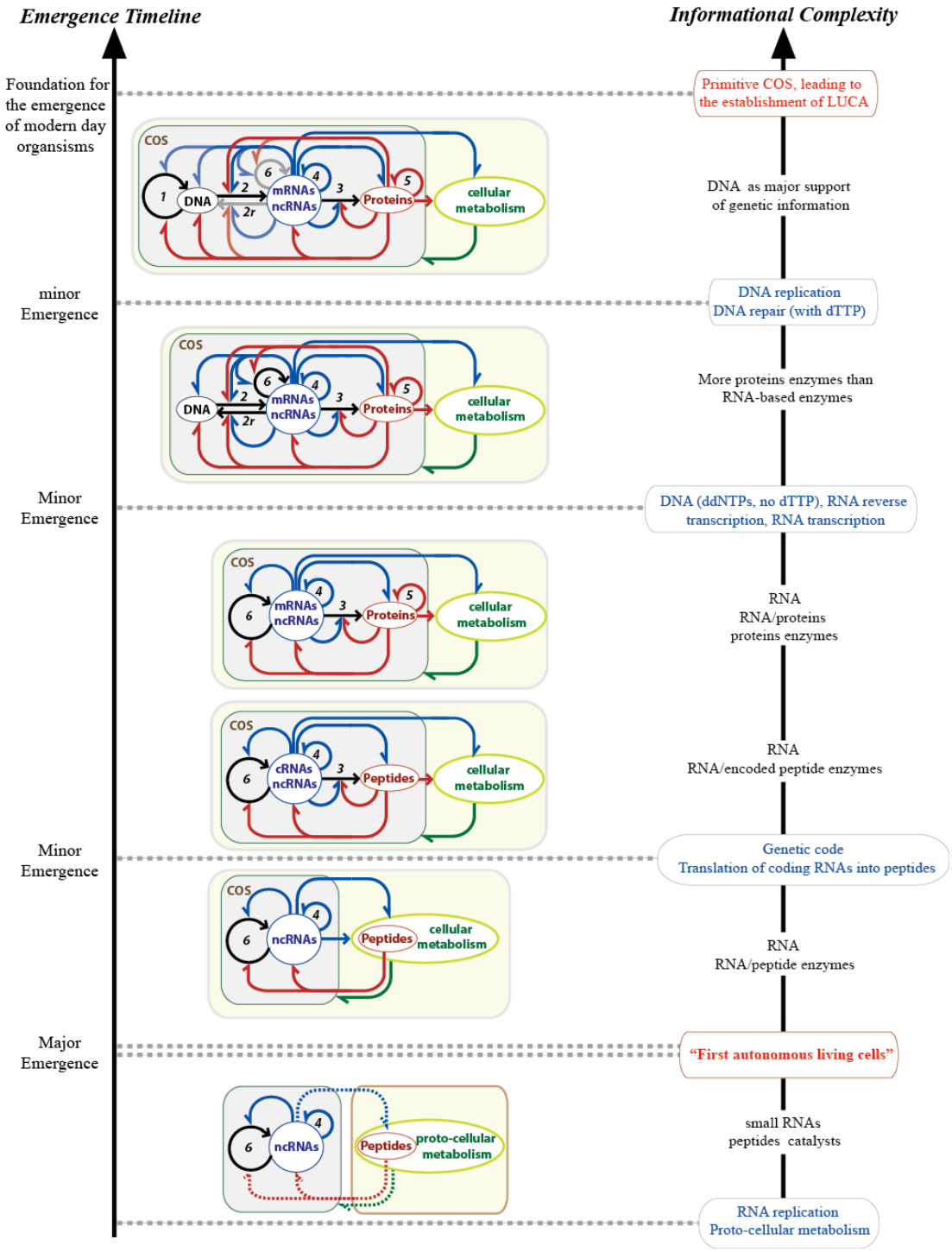


Figure 5.7 Possible evolutionary scenario for the optimization and increase in complexity of the cellular operating system through TDC by information control and adaptive selection.

This leads to the foundational "modern" core of information pathways belonging to the last universal common ancestor (LUCA) of all living organisms on Earth. While

this scenario delineates clear transition steps, some of the optimization steps could overlap in time. The major emergence of the first autonomous living cells corresponds to the emergence of the regulatory network linking RNA replication with cellular reproduction. See also Figure 5.2 legend.

Within the TDC framework outlined above, we propose a scenario explaining how simple replicating and reproducing cellular systems, based on COSs involving a limited set of functionalities could have developed into modern COSs (Figure 5.7). What is making this scenario possible is the existence of classes of functional equivalence with modes of operations able to be interchanged between different cellular entities through horizontal gene transfer [245, 246]. It is therefore with TDC by information control and adaptive selection that the phenomenon of convergence towards a unique COS can take place. Several important emergence events might have occurred. First, coupling between RNA replication and cellular metabolic reproduction led to the first autonomous living cells. This coupling corresponded to the emergence of the first regulatory network associated to the COS. Then, the natural drive of informational molecules to increase in size led to an increase of complexity that required improvement in the processing of this information. This could have taken place in multiple steps, which likely overlapped in time, each of these steps improving the accuracy and speed of the COS. Key emergence events were (i) the invention of the universal genetic code with translation of coding RNA into encoded peptides allowing the COS to rely more and more on proteins to carry important catalytic functions, followed by (ii) the invention of DNA as a superior support of the genetic information. As such, DNA genomes offered the advantage to be more easily repaired than RNA genomes. Additionally, other essential functions,

which paved the way towards efficient production of chemical energy, communication between cells and their environments as well as between different cells, emerged very early in conjunction with the emergence of the common "modern" COS of all living organisms [245]. This point in the history of cellular evolution, which can be traced back from modern day organisms, gave rise to the Last Universal Common Ancestor (LUCA) (Figure 5.7). It is only from that point that cellular life was ready for the next evolutionary step leading to speciation through division of labor and multi-cellular organization [245]. This point, called the Darwinian threshold, is therefore at the origin of the major division between eukaryotic, bacterial and archaeal cells. Note however that the process of optimization of the COS has continued to take place to a minor extent after that point of history.

Our discussion here pertains only to the core master functions of cellular life. During evolution, new master functions can emerge with selection drive towards higher order functions (such as those pertaining to differentiated multi-cellular living systems). However, any new set of emerging master functions cannot work against the master properties of lower levels but have to be built upon them. Good examples are parasitic or symbiotic living cellular systems. The genomes of these obligate systems might have lost some of the lower-level operations necessary for their autonomy outside their host. However, they rely on the fundamental lower-level properties provided by their host to operate their master functions. In all cases, any new master functions will create new functional boundaries for TDC by information control and adaptive selection. Therefore, TDC can enable new classes of

equivalence to appear through an explosion of diverse life forms and species [247] that are themselves subjected to new phenomena of convergence [248].

5.7.2 ***Future Perspectives for synthetic biology***

With synthetic biology, new biological functions and systems not found in nature can be created through a combination of *in vivo*, *in vitro* and *in silico* techniques. One can envision the engineering of new living organisms by modification of their genomes from a top-down approach [249, 250] or by integration of artificial molecular parts from the bottom–up [249, 251, 252]. For the past few years, bootstrapping experiments with synthetic bacterial genomes derived from natural ones have been underway for recreating minimal living bacterial organisms from *Mycoplasma* [13, 253, 254] and very recently, some of the technical challenges have been overcome [255]. For instance, the genesis of new *M. mycoides* cells has been demonstrated by transplantation of a synthetic genome into a *M. capricolum* recipient cell [255].

We believe that our present TDC framework could be useful for planning future experiments of synthetic biology in order to unravel the minimal set of functions that characterize living systems [203, 256] and providing insight about the way some of the essential cellular functions came to be. For instance, using the concept of classes of equivalence, it might be possible to substitute some of the existing cellular processes by simpler ones once a class of functional equivalence is established. As it has been shown in the archaeon *Nanoarchaeum equitan* [237], the production of mature tRNA does not require RNase P as long as the tRNA genes are organized at

the level of the genome such that all pre-tRNA sequences tRNAs transcripts are produced without 5'-leader sequences (Figure 5.5c). A deeper understanding of the different classes of functional equivalence and their modularity could be essential for the integration of new functionalities that would be more deeply rooted in the cellular regulatory network of the cell. Key factors will be the proper identification of the modality of regulations and flow of information at the level of these minimal living systems that are likely to present some degree of hierarchical organization. The minimal set of cellular functions identified in *mycoplasma* organisms is likely to be an interesting starting point for investigating the process of emergence of new functions and properties at the level of these organisms.

While it is unlikely that the new *M. mycoides* strain will be able to be (de)-evolved in a form of life that does not require the translational apparatus, it is however anticipated that the determination of the important characteristics at the root of the master functions of life through the top-down approach will be of some help to those interested by the bottom-up approach. It might be possible to substitute a certain number of operations of lower functions naturally carried out by proteins into some carried by artificial RNA molecules, creating a new cellular "RNA world". These experiments could be seen as "retro-synthetic" biology. In this exercise, several hypothetical evolutionary pathways could be investigated for their potential to have led to modern day cellular systems. A better understanding of the necessary emerging requirements for cellular life might come from giant viruses like mimiviruses [257-259]. These obligate parasitic systems offer the intriguing possibility to be engineered into biological systems that could express a COS (with a

fully operational translational apparatus) that might allow them to function independently from their cellular host. However, all these experiments come with possible ethical issues that should not be neglected.

Synthetic biology has presently demonstrated that it is possible to design novel metabolic pathways with new specific synthetic goals (systemically “needed” outcome or products) in bacteria [250, 260, 261] or to express orthologous pathways that can work in parallel with existing pathways in order to express new genes allowing the possibility to create artificial metabolic systems [262, 263]. For instance, it has been demonstrated that a synthetically evolved orthogonal ribosome system can function within *E. coli* to generate protein with unnatural amino acids in parallel to the "normal" ribosomal system [263]. It is possible to create new regulatory pathways by triggering specific “behavior” (e.g. bacterial motility) in response to specific molecular signals (e. g: theophylline or the herbicide atrazine) [231, 264]. In other words, at a microbial level, an outcome can be reached by “re-programming” the cell with artificial genes that could lead to this outcome with different pathways and different molecular triggers. Most of these approaches are based on the present understanding of how natural regulatory pathways work in bacteria [198, 199, 250]. Because of the high modularity and rather simple regulatory pathways in bacteria (especially when compared to those from Eukaryotes), artificial regulatory pathways can be built and existing functions and molecular parts can be easily exchanged with new ones (http://partsregistry.org) [265].

Within the context of synthetic biology, complementation experiments are extremely useful tools for experimentally unraveling the real nature of the modularity behind the various functions characterizing living organisms. Determination of this functional modularity allows identification of classes of functional equivalence that could potentially be hierarchically organized based on their degree of importance for the bacteria in well-defined experimental conditions.

5.8 Conclusion

In conclusion, TDC by information control and adaptive selection are at the root of converging forces that shape the evolution of living biosystems from the simplest to the most complex levels. Living systems could therefore be defined as self-reproducing systems that function via TDC by information control and adaptive selection. The functions of the COS that control cellular reproduction and DNA replication are maintained through TDC by information control leading to a converging driving force. This is apparent within the COS as it can have different modes of operations with the same functions. It is anticipated that functional convergence at a molecular level might not be as rare as initially thought. With the development of synthetic biology, it is expected that functional testing by complementation of newly identified molecules could lead to the discovery of a greater number of examples of ultimate functional convergence that are indicative of TDC. As long as the fundamental functions of reproduction and replication are kept, emergence of novel functions from the bottom-up is possible. This is however under the dependency of TDC by information control and adaptive selection. Indeed,

whatever emerges from the bottom-up has still to work within the context of the living system. Darwinian evolutionary processes in living systems are therefore not only ruled from the bottom-up but also by fundamental emerging organizational principles that are hierarchically built up and impose necessary constraints from the top-down. These principles are key for defining organic life.

Chapter 6 Conclusion

6.1 Introduction

RNA synthetic engineering research involving methodologies, design strategies and computational approaches is currently limited to small molecular RNA assemblies. Producing large RNA with self-assembling properties and multifunctional characteristics is beyond our scope and capabilities at this present time. Typically, these small RNA structures are symmetrical, regular shapes with very little in terms of complex structure involving multidimensional attributes and multifaceted functionality. Producing RNA molecules with a compacted, dense nucleotide population containing many variable and malleable interactions would prove extremely beneficial to the design of novel RNA architectures [35, 40]. Additionally, it is not fully understood how complex, structured RNA, such as the ribosome, evolved and grew increasing larger and more complex. Based on our current research it is clear that designability of RNA motifs and characterization of the constraints imposed on proper folding and function are key to our present understanding of synthetic biology and biochemical evolution [266].

6.2 RNA Motifs: Modular units for the Design of Complex RNA architectures

In the framework of this study, we have provided further evidence for functionally equivalent motifs that utilize identical folding properties irrespective of sequence. This designability feature allows for the precise control of the self-assembly of large RNA structures. The operation of folding requires that sequences

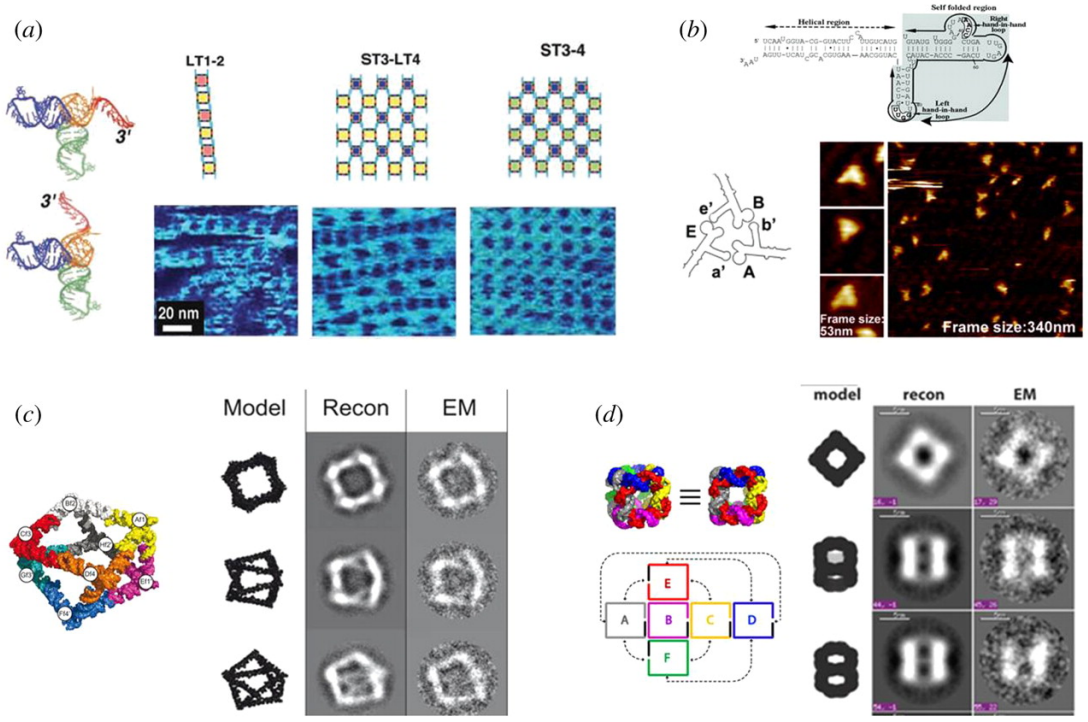
adopt their intended structure without alternative folds interfering with the final native state. “Folding traps”, as they are called, can be avoided by employing functionally equivalent structural motifs that maintain similar roles (such as a GUAA binding motif), yet contain a variety of primary sequences to choose from. Motif databases such as RNAmotif [178], RNA bricks, [267], RNA 3D Motif Atlas [268] and RNA Frabase [269] have a vast repertoire of RNA 2D and 3D structural motifs and have proven useful for structural comparisons. Building up these motifs into larger, grander structures proves to be more difficult than just putting the pieces together in the “pot”. Currently, 3-D structure prediction programs such as NAST, BARNACLE and FARFAR are tackling the problem of manual manipulation of variables and structure size. These programs, along with experimental evidence and additional crystal structures offer further knowledge regarding the sequence/structure relationship and have proven to be extremely valuable tools for designing more complex RNA structures. Expanding the repertoire of motifs to be utilized for RNA design, in addition to elucidating larger motif interactions is a key component to the future of this field. It is clear, that even with these modern advancements, further developments and refinements of existing models is needed to improve the predictability of the sequence/structure relationship.

6.3 Natural RNA as a Structural Guide for Developing Design Principles

One of the main focal points of current and future research is the generation of novel self-assembling motifs. Tertiary interactions are the key component to the building up of larger, more integrated RNA architectures. Because of this, much attention has been given to the discovery and generation of novel RNA tertiary motifs. Currently, there are several strategies being employed to tackle such a task. One such strategy employs nature's tool kit for utilizing adaptable building blocks. Figure 6.1 outlines some key design strategies that utilize natural RNA motifs to build up tectoRNA structures. Motifs utilized in these studies include the right angle motif, kissing loops [65], three-way junctions[35], pRNA [270], sticky-end cohesions [40]. The architectures constructed from such motifs are simple and include squares, triangles, anti-prisms and cubes. It is clear that with the increased efforts in obtaining more crystal structures of natural RNA interactions, the building block repertoire can expand, thus providing the tools needed to generate more complex RNA architectures.

6.4 Selection Pressures on RNA Evolution

GNRA tetraloop receptors are one of the most widely used RNA tertiary motifs in structured RNA. The GNRA tetraloop is versatile in that it is capable of binding to both helices and internal loops. GNRA tetraloop/receptor interactions found in nature can generally be classified as either belonging to the 11nt/IC3-like class or a



GYRA helical receptor. It has been demonstrated through experimentation, that synthetic

Figure 6.1 Nanoengineering of RNA-based Novel Architectures
(a) Nano-square construction and design patterns. (b) Trimers assembled from phi29 pRNA. (c) Anti-prism nanoparticles made from tRNA (d) Sticky-end cohesion RNA-nanocube assemblies [1].

GNRA receptors have comparable properties as well as functionalities [72, 73, 158], yet nature has not utilized other classes of GNRA receptors. So the question remains, why do we see so few receptor classes in nature?

From the analysis of GNRA receptor motifs and their respective networks, it is clear that interacting modules are hierarchically organized. While studying the details of the interactions at a local level is an all important aspect of characterization, this is just one small part. It is important to recognize that each

level of organization is not acting autonomously. In other words, other factors than just the hydrogen bonding patterns, stacking interactions, etc. have a role on the native fold of large RNAs. A key component to understanding how tertiary interactions generate large complex structures is first to understand how the global folding and interactions play a part and have an effect on the lower levels of organization. Future research in this arena will not only assist in improving structure prediction, but also help in gaining a further understanding of the evolutionary constraints that drive the change of these molecules over time.

Comparison of classes of GNRA receptors found in nature versus *in vitro* selection experiments has helped to lay the foundation for identifying important constraints dictating RNA structure for large self-assemblies. It has been shown that artificial receptors can be created *in vitro* that contain predictable folds and have high binding affinity for their cognate loop. However, binding affinity is contextual and the natural receptors that have been characterized range in K_d for their cognate loop from 3nM (11nt/GAAA), 800nM (IC3/GUAA), and 400-950 nM (Helix/GYRA). This range of affinity indicates that the strength of the interaction is predicated upon the context of the molecule as a whole and the role that the interaction plays. In support of this notion, an *in vitro* study was conducted utilizing the *Td* bacteriophage ribozyme for an *in vitro* selection of GNRA tetraloop/receptor interactions [73]. The scaffold used was constructed from an interaction between the L2 loop and the P8 helix, which is required for ribozyme function, cleavage in the P1 stem. Interestingly, the resulting pool winners from this study showed much more efficient cleavage than the wild type. This would indicate that the L2-P8 interaction

is less significant in the context of the ribozyme as a whole, which could also explain why strong interactions (such as 11nt/GAAA) are rarely found in this position. In other words, there is no selection pressure for adapting strong binding receptors in this position.

Zakrevsky (2015) developed a fitness landscape that links many different classes of both *in vitro* selected and naturally occurring receptors by point mutation. Each receptor was characterized and this study points out that the fitness landscape is able to be traversed while maintaining functionality (or binding). When comparing sequences of selected receptors versus naturally occurring ones (class I di-GMP riboswitches), it was noted that selected receptors have a markedly high G/C content over A/U Afonin (2012) and Zakrevsky (2015). Since *in vitro* selection is a highly utilized technique for "fishing" out RNA aptamers, this finding could prove useful in augmenting current selection criteria for these types of experiments. Additionally, it offers some insight as to the effect of higher level of organization. As presented herein, local GNRA interactions are highly robust in nature and the number of sequences that adopt similar folds with similar phenotypic properties is likely immense. It is clear that local interactions do not solely dictate the selection of sequence on tertiary interactions, but rather it is a compilation of higher order levels of organization that have an impact as well.

This notion of the pressure of selection being imposed on lower-levels of organization by the whole RNA structure can be explained through top-down causation. The work that has been done highlighting fitness landscapes, neutral walks through sequence space and a clear bias between natural receptor sequences

and artificial ones suggests that it's the RNA context as a whole that is creating the constraints within which loop-receptor interactions are selected by nature. It is for this reason that global folding and the avoidance of alternative structures is at the core of adaptation through natural selection. Clearly, many receptor sequences (developed *in vitro*) are capable of binding GNRA tetraloops, yet nature has a utilized a limited range of sequences for this purpose. Natural sequences have evolved with this additional selection pressure, which current scaffold design and selection criteria for *in vitro* selected receptors has yet to address for large RNA architectures. This a clear example of top-down causation where the global context affect the parts. As such, both bottom-up and top-down effects are causally effective during RNA structural evolution.

Chapter 7 Materials and Methods

7.1 Protocols and Methods for Chapter 2

7.1.1 *3-dimensional models for new GUAA/S8-like receptor interactions*

The tectoRNA scaffolds were based on a previously studied hairpin dimer that take advantage of two GNRA loop/receptor interactions [72, 93] (Figure 2.1D and E). Atomic model structures have been previously developed [39, 123]. Models of receptor/loop interactions were created using Swiss-Pdb Viewer [271]. ModeRNA was used to generate specific, single nucleobase replacements [272] and image rendering was performed in PyMOL [273].

7.1.2 *Synthesis of tectoRNA*

The sequence of the RNA monomer is as follows; 5'-GGGAAAGCGAUUUACUC-(N6-N8)-UAGUUCCGGG-(GRNA)-CUUGGUUCUA-(N6-N9)-GGGUAGGUCGUUUUUCUCU-3', where N is the receptor region being mutated. See Table 8.1 for a complete list of receptors tested. In addition to the 8 GNRA tetraloops (GAAA, GGAA, GUAA, GCAA, GAGA, GGGA, GUGA, GCGA), the loop regions were also replaced by the Archaea L39 loop (5'-GGUAAGC-3') or the *E. coli* L39 loop (5'-GGAUAAGC-3'). DNA templates, antisense and primer sequences were acquired from Integrated DNA Technologies (IDT). PCR amplification of the desired sequence, which included a T7 promoter sequence, was used to generate the tectoRNA via *in vitro* run-off transcription. PCR reactions were covered with mineral oil and a hot start using Taq

polymerase at 94°C initiated the reaction. The thermocycler was programmed for 25 cycles (94°C for 75 seconds, 56°C for 75 seconds, 72°C for 75 seconds). PCR reactions contained 30 ul 5X PCR buffer (50 mM Tris pH 8.9, 250 mM KCl, 2.5% NP40, 5 mg/ml gelatin), 2mM $MgCl_2$, 0.5 mM of each dNTP, 150 pmol of each primer and 0.3 pmol template. Purification of PCR reactions was performed with QiaQuick PCR purification kit (Qiagen). Run-off T7 transcription was performed on each of the purified PCR reactions as described [39, 43, 274]. The resulting RNA samples were purified on an 8% polyacrylamide denaturing gel. After purification the RNA concentrations were calculated with UV spectrophotometry at 260 nm.

7.1.3 *Association Assay*

Heterodimer interactions were characterized via binding assays utilizing native PAGE. Equimolar concentrations of monomers (ranging from 1nM-20uM) were mixed in water, including 1nM of 3'-end [^{32}P]pCp-labeled RNA monomer containing the GNRA probe with variable loops [72]. RNA was denatured at 95°C for 2 minutes, then placed on ice for 3 minutes and finally allowed to anneal at 30°C for 5 minutes prior to assembly. For the assembly step Mg^{2+} buffer was added, (final concentration-15mM $Mg(OAc)_2$, 89mM Tris-borate pH 8.3) and incubated at 30°C for 20 minutes. Prior to running constructs on native PAGE, assemblies were allowed to equilibrate at 10°C for at least 30 minutes. Visualization of the interaction between monomers was performed on native 7% polyacrylamide electrophoresis at 10°C with 15mM Mg^{2+}. Gels were dried and developed on phosphor screens, then scanned using the Typhoon 9410.

7.1.4 *K_d determination*

Titration experiments were performed (as described above) in order to determine the dissociation constants (K_d's) of each interaction. Monomers (Probe [P]) and heterodimers (Receptor [R] + Probe [P]) were quantified using ImageQuant software [40, 43, 72, 124] (Figure 8.1C). A non-linear fit from the experimental equation: $f = [2\beta M0 + K_d - (4M0\beta K_d + K_d2)0.5]/2M0$, was used to determine K_d's for the equilibrium reaction $P + R \leftrightarrow PR$, where f is the fraction of RNA heterodimer, defined as the weight-in-weight (w/w) ratio of the dimer (PR) to the total RNA species (P + R+PR) [275]. M0 is the total concentration of the probe. β is the maximum fraction of RNA able to dimerize. With β typically equal to 1 for most molecules tested, the K_d's equation is: $K_d = [(M0)(1 - f)2]/f$. Therefore, K_d's correspond to M0/2 when 50% of bi-molecular assemblies are formed [39, 72]. For each set of molecules, K_d values correspond to the average calculated from three independent experiments. The corresponding free energy of dimerization (ΔG) between tectoRNA receptors and RNA tetraloop probes are determined from the equation, $\Delta G = RT\ln K_d$, where R is the gas constant (1.985 cal/K/mol) and T is the temperature (283°K). The apparent free energy variation of dimerization at 10°C (ΔΔG) can be derived from the equation, $\Delta\Delta G = \Delta G(PR) - \Delta G(reference)$. For the reference we chose a general concentration that represented a threshold for optimal biological activity of 4000nM.

7.1.5 ***Receptor specificity***

Specificity profiles (utilizing color coded bars) for each receptor were created to identify binding patterns that could link sequence information and structure (Figure 3, 5 and 6). To create these profiles, associations, as described above, were performed with final concentrations ranging from 1nM to 20uM for each receptor mixed with GNRA tetraloop (with the exception of GAAA), of which 1nM was 3'-end [^{32}P]pCp-labeled GNRA tetraloop probe for detection of dimerization. K_d's and $\Delta\Delta G$'s were calculated as described above. Side-by-side comparisons were used to assess the affinity of each receptor and its target discrimination. The loop UUCG acted as a negative control in the specificity profile experiments.

7.2 Protocols and Methods for Chapter 3

7.2.1 *In vitro selection*

The DNA library was created via PCR amplification using 6 different partially randomized templates (5'-AGAGAAAAACGACCTACCCA-(Random)-GGATGAACCAAGTTTCCCCGGAACATCC-(Random)-TGAGTAAATCGCTTTCCCGCATCATAGCATCCACATCG-3'), where the random regions are 3N-3N, 3N-4N, 4N-3N, 4N-4N, 4N-5N, 5N-4N, respectively, forward primer (5'-TTCTAATACGACTCACTATAGGGAAAGCGATTTACTC-3')- and reverse primer (5'-GCATCATAGCATCCACATCGAGAGAAAAACGACCTACCC-3'). All DNA was ordered from IDT. The sequence of the DNA templates reflects the antisense sequence of the final RNA product. PCR was accomplished by annealing the forward primer containing a T7 RNA polymerase promoter, followed by the reverse primer annealing to the resulting complimentary strand of the template. PCR reactions were covered with mineral oil and a hot start using Taq polymerase at 94°C initiated the reaction. The thermocycler was programmed for 25 cycles (94°C for 75 seconds, 56°C for 75 seconds, 72°C for 75 seconds). PCR reactions contained 30 ul 5X PCR buffer (50 mM Tris pH 8.9, 250 mM KCl, 2.5% NP40, 5 mg/ml gelatin), 2mM $MgCl_2$, 0.5 mM of each dNTP, 150 pmol of each primer and 0.3 pmol template. Purification of PCR reactions was performed with QiaQuick PCR purification kit (Qiagen). Run-off T7 transcription was performed on each of the purified PCR reactions as described [39, 43, 274]. The resulting ^{32}P body-labeled

RNA samples were purified on an 8% polyacrylamide denaturing gel. After purification the RNA concentrations were calculated with UV spectrophotometry at 260 nm. The 'bank' RNA pool was created by calculating the possible number of variants and adjusting the amounts of the mutant pool so that the 'bank' pool statistically has one of every mutant.

The selection process begins with a counter selection removing any molecules capable of forming dimers or other complexes in the absence of a GUAA loop target (Figure 3.1B). The bank RNA pool (15.12 uM) was supplemented with the reverse transcription primer (1:2), designed to bind to the single stranded reverse transcription tail (5'-GCATCATAGCATCCACATCG-3') to eliminate any single stranded region interfering with the association of receptors to their target. The first round began with denaturation at 90°C for 1 minute, ice for 3 minutes, pre-incubation at 30°C for 2 minutes before association buffer (89 mM Tris-borate pH 8.3, 15 mM $Mg(OAc)_2$, 50 mM KCl) was added and incubated at 30°C for 30 minutes. A 1/10 volume of gel loading buffer (association buffer with 55% glycerol, 0.05% bromophenol blue and 0.01% xylene cyanol) was added to the sample and run on a native 7% (29:1) polyacrylamide gel in 15 mM Mg^{2+} buffer. The resulting monomer band was purified, eluted and ethanol precipitated to move on to the positive selection step.

The monomer bands of all counter selection steps continued to the next step of selection, whereas positive selection steps resulting in dimer formation moved on to the next round of selection. There were a total of six rounds of selection, where the

positive selection step of each round was followed by reverse-transcription PCR (RT-PCR) in order to amplify the 'winners' of that round. Dimer bands were extracted from the gel and ethanol precipitated, resulting in a new pool of RNA molecules capable of forming dimer complexes. The Improm II RT system from VWR was used to reverse transcribe the 'winning' RNA pool and PCR was performed as previously stated. After purification with QiaQuick PCR Purification Kit (Qiagen), the enriched DNA pool was then transcribed as described previously and the resulting RNA was subjected to additional rounds of both negative and positive selection (Figure 3.1C). Following six rounds of selection, the resulting DNA was cloned using Invitrogen's TA Cloning Kit (INVαF' competent *E. coli* cells) and 35 clones were sequenced (Table 8.5).

7.2.2 ***RNA self-assembly on native PAGE***

For determination of thermodynamic dissociation constants of heterodimeric constructs (K_d), equimolar concentrations of an RNA tetraloop receptor and GUAA tetraloop were dissolved in water (0.5 nM – 20 μM final), heated at 90°C, placed on ice, and incubated at 30°C for 20-30 minutes upon addition of Mg^{++} buffer (1X Tris borate (TB), 15 mM Mg^{++}, pH 8.2 final). Prior to incubation, a constant amount (0.5 nM – 3 nM) of 3' labeled radioactive GUAA tetraloop molecule was added to each sample. Following the incubation and subsequent addition of loading buffer (1X TB, 15 mM Mg^{++}, 55% glycerol, 0.05% bromophenol blue, 0.05% xylene cyanol) samples were immediately placed on ice and loaded on a 7% native 1X TB 15 mM Mg^{++} polyacrylamide gel and migrated at 30 W in 10-15°C.

Additionally, RNA receptors were screened against 7 different GNRA tetraloops (utilizing the GAAA/11nt interaction as an anchor, Ra/La) at equimolar concentrations, ranging from 1nM to 10,000 nM to probe the tetraloop specificity of each hetero-dimer construct. If self-dimers were evident during these experiments, this was an indication that the receptor may bind to GAAA. In these cases, the anchor was then changed to either the R5.16/GUAA or the R1/GGAA interaction to calculate the K_d for the GAAA tetraloop. Specificity gels were conducted with a constant amount of radioactive RNA receptor monomer (0.5 – 3 nM). All self-assembly gels were conducted in an identical fashion, with the exception of the radioactive monomer. Upon completion of gel migration, each gel was isolated, dried under heat and vacuum, and exposed overnight on a phosphorscreen. The fraction of heterodimer was calculated either by direct quantitation with ImageQuant or by determining the R_f of the dimer using Photoshop. The fraction of dimer was plotted against total RNA concentration and the K_d was calculated according to Afonin (2010) (Figure S1). All K_d values for GUAA are reported as an average of at least 2 separate experiments (Table S1).

For the purpose of representing data, plotted values were listed as ΔΔG's with a base K_d value of 4,000 nM. ΔΔG's below this value are considered biologically relevant as their thermodynamic stability affords a reasonable affinity in a natural context. ΔΔG's above this value are generally thought of as interactions that mostly likely will not occur in nature as structured RNA. The profiles for each receptor are represented as both color-coded bars, as well as line graphs for all the GNRA

tetraloops so that specificity profiles could be used as a phenotypic parameter for characterization.

7.2.3 ***Chemical probing with dimethylsulfate (DMS)***

RNA receptors with an elongated 20 nucleotide 5' tail for primer extension were assembled as monomer or dimer (dimer contains equimolar amounts of GUAA tetraloop and receptor) as described above at a single concentration (50 nM or 2 μM) using a DMS assembly buffer (50 mM HEPES, 100 mM KCl, 15 mM Mg^{++} final). Following the incubation, the samples were directly treated to a solution of DMS diluted in 100% EtOH (60 mM final) and reacted at room temperature for 4 min (monomer) or 8 min (dimer). 2 μL of DMS quench buffer (2.8 M NaOAc, 1 M β-mercaptoethanol) and 200 μL of cold 100% EtOH was added upon completion of reaction and samples were cooled at –80°C for 15 min and precipitated at 4°C, rinsed twice with 90% EtOH, and dried under vacuum. Samples were resuspended in water and subjected to primer extension using Superscript III RT. Each reaction contained approximately 0.5 pmol of RNA and 1.0 pmol of 5' labeled radioactive DNA primer (total of 3μL), at which point, the samples were heated at 90°C and 7 μL of reverse transcription mix was added (~50 nM RNA, ~100 nM primer, 1.25 mM dNTP, 5mM DDT, 1X first strand buffer, 30 units of Superscript III reverse transcriptase final**)**. Sequencing reactions were performed in an identical fashion in the presence of 0.313 mM (40%) of the appropriate ddNTP. Following primer extension, RNA was degraded with treatment of 2 N NaOH, heated at 95°C for 3 min, neutralized with identical amount of 2 N HCl. 3.75 μL of 3 M NaOAc and 200

μL of cold 100% EtOH were added and the samples were precipitated as described above. Samples were resuspended in urea blue buffer (8 M urea, 0.05% bromophenol blue, 0.05% xylene cyanol) and the radioactivity was equilibrated using a Geiger counter. Samples were heated at 90°C for 3 min and were finally loaded on a 0.35 mm denaturing 8M urea, 1X TTE, 8% polyacrylamide gel and migrated at 80 W for 1.5 hrs. Upon completion of gel migration, the gel was rinsed with 10% EtOH 10% HOAc solution, dried under heat and vacuum, and exposed overnight on a phosphorimager screen. DMS experiments were conducted in duplicate (Figure 3.5).

7.2.4 ***Molecular Modeling***

The molecular model for R5.58 was composed using the software Assemble [34]. The crystal structure of the S8 like secondary structure from *E. coli* 16S rRNA (PDB: 2AW7) was used to define the motif of interest. Further energy minimization was conducted using Sander. Models of receptor/loop interactions were created using Swiss-Pdb Viewer [271]. Figures of the molecular structure were rendered with Pymol [273].

7.3 Protocols and Methods for Chapter 4

7.3.1 *TectoRNA design and 3D modeling*

3D atomic models were manually constructed using the program Swiss-Pdb Viewer [271] following the RNA architectonics guidelines [63]. All tectoRNA attenuators contain a heterodimer-forming module that assembles with a probe through two intermolecular receptor/GNRA interactions (Figure 4.1 and Figure 4.2). This module was modeled after the tectoRNA heterodimer (HD) [39, 43, 153] for which atomic model structures are presently available [PDB_ID: 2adt][123, 276]. The 5' and 3' PK forming modules leading to the formation of 5' and 3' intramolecular pseudoknots (PK), respectively, were modeled after the NMR structure of the Box H/ACA snoRNA bound to its rRNA target [PDB code: 2p89, 2pcv][169, 170] (Figure 5.1B-C and Figure 8.17). In order to form the PK, the PK-forming module with a 10 bp stem apical stem, is linked through 4 nucleotides to the HD-forming module, which includes a 3 bp stem (Figure 4.1 and Figure 4.2). TectoRNA sequences (listed Table 5.1) were checked for proper folding with the program Mfold [52, 54] to maximize the stability of their secondary structure while minimizing the occurrence of alternative secondary structure folds. All tectoRNA attenuators are predicted to fold into a unique secondary structure prone to assemble with the probe. PKs with one or more GC bp are accurately predicted with Kinefold [277] (Table 8.12).

7.3.2 ***TectoRNA synthesis and assembly***

TectoRNAs were synthesized by in vitro T7 run-off transcription from PCR generate templates, purified by denaturing polyacrylamide gel electrophoresis (PAGE) and labeled at their 3' end using 3'-[^{32}P]pCp as previously described [39, 153]. For the determination of equilibrium constants of dissociation (K_d) by titration experiments, RNA samples were typically prepared by mixing equimolar amounts of each tectoRNA at various concentrations (1 nM to 50 µM) in water. After denaturation (2 min, 95° C; 2 min, 4° C; 2 min, 30° C), samples were renatured by addition of magnesium buffer (89 mM Tris– borate pH 8.3 (TB), 50 mM KCl, 2 or 15 mM $Mg(OAc)_2$ final concentration) at 30°C for 20 min before incubation at 10°C. For tri-molecular competition experiments, attenuator tectoRNA molecules were first assembled with linear switching (SW) RNA molecules at 15 mM $Mg(OAc)_2$ for 15 minutes, before further incubation with the RNA probe for 30 min. One of the tectoRNAs used in the self-assembly mix (usually the probe) contained a fixed amount of 3'-end [^{32}P] pCp-labeled RNA (1–10 nM final) for visual monitoring on native 10% (29:1) PAGE gels. Samples were cooled on ice before addition of blue loading buffer (magnesium buffer, 0.01% bromophenol blue, 0.01% xylene cyanol, 50% glycerol) and migration at a maximum temperature of 10°C for 3h on PAGE gels with 2 or 15 mM mM $Mg(OAc)_2$ and running buffer (89 mM Tris-borate, pH 8.3, 2 or 15 mM $Mg(OAc)_2$).

7.3.3 ***Dissociation constants (K_d) and free energy (ΔG) calculations***

K_d's were experimentally derived from titration experiments at 10°C performed as described above. Monomers (Probe (*P*), HD-forming module (M_{HD}) or attenuator tectoRNA (M_{AT})) and heterodimers (P^*M_{HD} or P^*M_{AT}) were quantified using ImageQuant software [39, 40, 43, 153]. K_ds for the equilibrium reaction $P + M \rightarrow P^*M$ (with $M = M_{HD}$ *or* M_{AT}), were determined from a nonlinear fit of the experimental data to equation: $f=(2\beta M_0 + K_d - (4M_0\beta K_d + K_d^2)^{0.5})/2M_0$, where *f* is the fraction of RNA heterodimer, defined as the weight-in-weight (w/w) ratio of the dimer (P^*M) to the total RNA species ($P + M + P^*M$) [278]. M_0 is the total concentration of the probe (or attenuator tectoRNA). β is the maximum fraction of RNA able to dimerize. With β typically equal to 1 for most molecules tested, the K_d's equation is: $K_d = ((M_0)(1 - f)^2)/f$. Therefore, K_d's correspond to $M_0/2$ when 50% of bi-molecular assemblies are formed [39, 153]. For each set of molecules, K_d's values correspond to the average calculated from three independent experiments. The corresponding free energy variations of dimerization (ΔG_{HD}) between tectoRNA attenuators and cognate RNA probes are determined from the equation, $\Delta G_{HD} = RTlnK_d$, where *R* is the gas constant (1.985 $cal.K^{-1}.mol^{-1}$) and *T* is the temperature (283°K). The apparent free energy variation of attenuation at 10°C ($\Delta\Delta G_{AT}$) can be derived from the equation, $\Delta\Delta G_{AT} = \Delta G_{HD}(M_{AT} + P) - \Delta G_{HD}(M_{HD} + P)$, where, $\Delta G_{HD}(M_{AT} + P)$ is the free energy of dimerization between the tectoRNA attenuator (comprising attenuator PK-forming and HD-forming modules) and its cognate RNA probe, and $\Delta G_{HD}(M_{HD} + P)$ is the free energy variation of dimerization between the

corresponding HD-forming module alone and its cognate probe. All K_ds and associated ΔG_{HD} and $\Delta\Delta G_{AT}$ are reported in Tables S2, S3 and S4 in the online Supplementary Data.

7.3.4 *Co-transcriptional assembly*

PCR-generated DNA templates coding for a tectoRNA attenuator and its cognate RNA probe (GAAA2 or GGAA2 (Table 8.11)) were mixed at equimolar concentrations in presence of the transcription mixture (50 mM Tris pH 7.5, 10 mM $MgCl_2$, 2 mM spermidine, 2.5 mM NTPs, 10 mM DTT, $\alpha[^{32}P]$-ATP (10 mCi/ml)). Transcription was initiated by addition of home-made T7 RNA polymerase (10 units/µl final) at 37°C. Small aliquots of the transcription mix were taken at 15, 30, 45, and 60 minutes time intervals and quenched by incubation with RQ1 RNase-free DNase (0.3 units/µl final) for 15 min at 37°C, just before native PAGE analysis at 10°C in presence of 10 mM $Mg(OAc)_2$ as described above.

7.3.5 *Lead Pb(II)-Induced Cleavage*

RNA samples (4 µM final with 10 nM of 3'-end labeled RNA) assembled as described above, were incubated in presence of $Pb(OAc)_2$ (8mM final) for 2 min before addition of 50 mM EDTA and ethanol precipitation [39, 153]. Lead-induced cleavage patterns were visualized on 8M Urea/ 20% PAGE (See also Supporting Information). Cleaved positions were identified using RNA samples treated by RNase T1 digestion and alkaline hydrolysis.

Chapter 8 Appendix

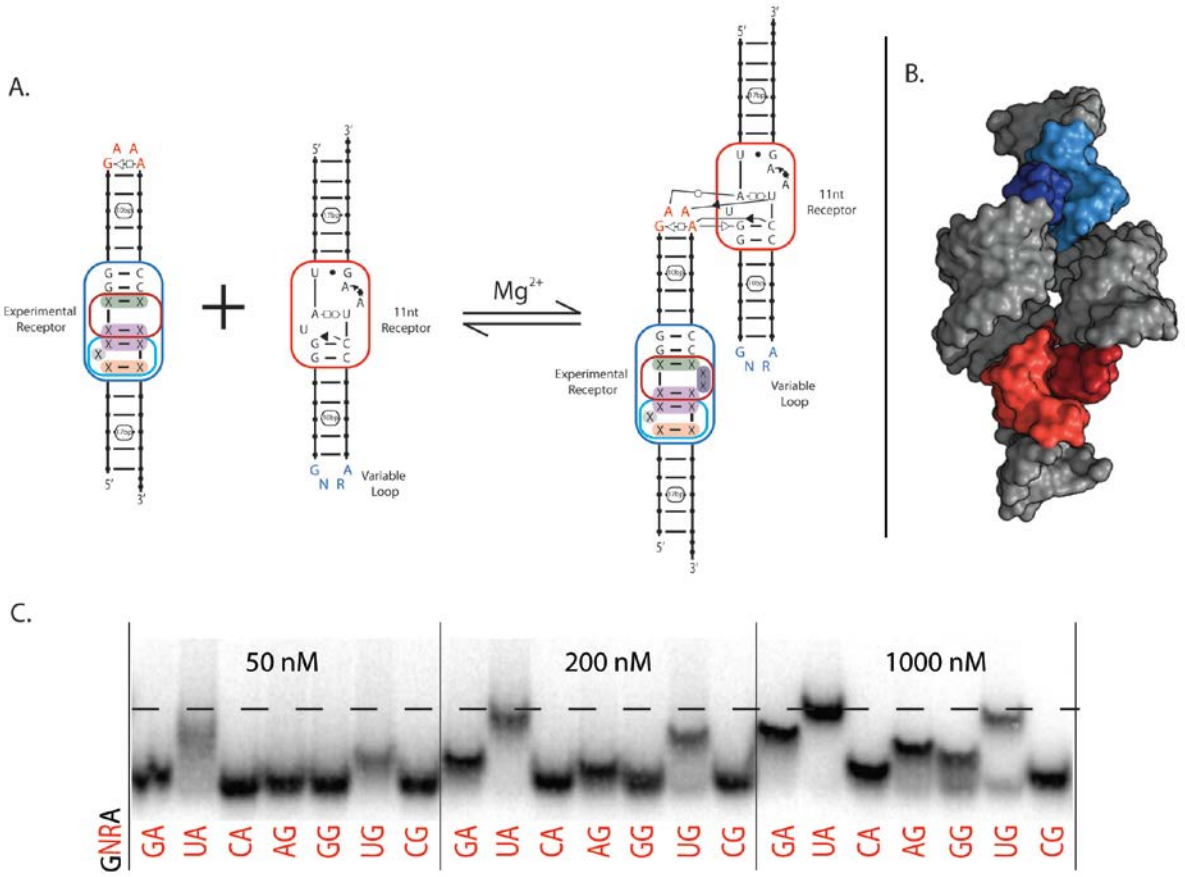


Figure 8.1 TectoRNA dimer assay
(A) Heterodimer assay consisting of two hairpin monomers interacting via two separate loop/receptor interactions. The 11nt receptor and its cognate GAAA loop, act to anchor the interaction (in red). The experimental receptor is inserted into the turquoise region and the GNRA tetraloop on the other monomer consists of one of the other GNRA (excluding GAAA). Upon the addition of Mg2+, dimerization is able to occur. (B) shows a space filling model of the dimer interaction (PDB_ID: 2JYJ). (C) Gel shift assay on native PAGE of tectodimer formation at three different concentrations (50 nM, 200 nM, and 1000 nM) shows the extent of binding to each loop for the experimental receptor.

8.1 Supplemental Information for Chapter 2

Molecule	RNA Sequence (5' to 3')
H89arch	GGGAAAGCGAUUUACUCGGCUUGCUUAGUUCCGGGGAAACUUGGUUCUA-AGUC-GUCGGGUAGGUCGUUUUUCUCU
H89bact	...ACUCUGUUUGGCUA...GAAA...UA-GCUG-AUAGGGU...
H89arch.1	...ACUCGGCUUGGGUA...GAAA...UA-CCUC-GUCGGGU...
H89arch.2	...ACUCGGCUUGGGUA...GAAA...UA-CCCC-GUCGGGU...
H89bact.1	...ACUCUGUUUGGGUA...GAAA...UA-CCUG-AUAGGGU...
H89bact.2	...ACUCUGUUUGGGUA...GAAA...UA-CCCG-AUAGGGU...
B7.8	...ACUCUAAGG-CCUA...GAAA...UA-GGCUACUGGGGU...
S8	...ACUCUA-AGGGGUA...GAAA...UA-CCCUACUGGGGU...
S8.2	...ACUCUA-AGGGGUA...GAAA...UA-CCCCACUGGGGU...
S8.3	...ACUCUA-AGCGGUA...GAAA...UA-CCCUACUGGGGU...
S8.4	...ACUCUA-ACGGGUA...GAAA...UA-CCCCAGUGGGGU...
S8.5	...ACUCUA-ACGGGUA...GAAA...UA-CCCUAGUGGGGU...
S8.7	...ACUCAA-ACGGGUA...GAAA...UA-CCCUAGUUGGGU...
S8.8	...ACUCAA-CGGGGUA...GAAA...UA-CCCUACGUGGGU...
S8.9	...ACUCAA-AGGGGUA...GAAA...UA-CCCUACUUGGGU...
S8.11	...ACUCUA-AGGGGUA...GAAA...UA-CCUUACUGGGGU...
S8.12	...ACUCUA-AGGGGUA...GAAA...UA-CCCUGCUGGGGU...
S8.13	...ACUCUA-AGGGGUA...GAAA...UA-CCCGUCUGGGGU...
S8.14	...ACUCUA-AGGGGUA...GAAA...UA-CCCAACUGGGGU...
S8.15	...ACUCUA-AGGGGUA...GAAA...UA-CCCACCUGGGGU...
S8.16	...ACUCUA-AGGGGUA...GAAA...UA-CCCUCCUGGGGU...
S8.17	...ACUCUA-AGGGGUA...GAAA...UA-CCCAUCUGGGGU...
S8.18	...ACUCUA-AGGGGUA...GAAA...UA-CCCAGCUGGGGU...
S8.19	...ACUCU--AGGGGUA...GAAA...UA-CCCUACUGGGGU...
S8.20	...ACUCUA-AGGGGUA...GAAA...UA-CCC--CUGGGGU...
S8.21	...ACUCUA-AGGGGUA...GAAA...UA-CCUCACUGGGGU...
S8.22	...ACUCUA-AGGGGUA...GAAA...UA-CCUAACUGGGGU...
S8.23	...ACUCUAAGGGGGUA...GAAA...UACCCCUACUGGGGU...
S8.24	...ACUCUA-AG-GGUA...GAAA...UA-CC-UACUGGGGU...
S8.25	...ACUCUA-AGGGGUA...GAAA...UA-CCUACCUGGGGU...
S8.26	...ACUCU--AGGGGUA...GAAA...UA-CCUUACUGGGGU...
S8.27	...ACUCU--AGGGGUA...GAAA...UA-CCCCACUGGGGU...
S8.28	...ACUCU--AGGGGUA...GAAA...UA-CCUCACUGGGGU...
S8.a1	...ACUCCUACUGGGUA...GAAA...UA-CCCA-AGGGGGU...
S8.a2	...ACUCCUACUGGGUA...GAAA...UA-CCUA-AGGGGGU...

H.GYaA	...ACUCU--AGUGGUA...GAAA...UA-CCC--CUGGGGU...
H.GYgA	...ACUCU--AGUGGUA...GAAA...UA-CCU--CUGGGGU...
H.GYgA.1	...ACUCU--AGAGGUA...GAAA...UA-CCU--CUGGGGU...
S8Tt1	...ACUCGG-CGGGGUA...GAAA...UA-CCCUACGCGGGU...
S8Tt2	...ACUCCG-CGGGGUA...GAAA...UA-CCCUACGGGGGU...
S8Tt3	...ACUCCG-CGGGGUA...GAAA...UA-CCUUACGGGGGU...
S8Tt4	...ACUCUG-CGGGGUA...GAAA...UA-CCCUACGGGGGU...
S8Tt5	...ACUCUG-CGGGGUA...GAAA...UA-CCCAACGGGGGU...
S8Tt6	...ACUCUG-CGGGGUA...GAAA...UA-CCUUACGGGGGU...
S8Tt7	...ACUCUG-CGGGGUA...GAAA...UA-CCUAACGGGGGU...
S8Tt8	...ACUCUG-CGGGGUA...GAAA...UA-CCCCACGGGGGU...
S8Tt9	...ACUCUG-CGGGGUA...GAAA...UA-CCUCACGGGGGU...
S8Tt.a	...ACUCCUACGGGGUA...GAAA...UA-CCCG-CGGGGGU...
S8Tt4.a	...ACAGUGC-GGGGUA...GAAA...UA-CCCUACGGCUGU...
S8Tt8.a	...ACAGUGC-GGGGUA...GAAA...UA-CCCCACGGCUGU...
H68	...ACUCUAAACGGGUA...GAAA...UA-CCCCAGUGGGGU...
H68.1	...ACUCUAAACGGGUA...GAAA...UA-CCCUAGUGGGGU...
H68.2	...ACUCUAAACCGGUA...GAAA...UA-CCCCAGUGGGGU...
H68.5	...ACUCUAAACGGGUA...GAAA...UA-CCCAAGUGGGGU...
H68.6	...ACUCUAAACGGGUA...GAAA...UA-CCCACGUGGGGU...
H68.7	...ACUCUAAACGGGUA...GAAA...UA-CCUAAGUGGGGU...
H68.8	...ACUCUAAACGGGUA...GAAA...UA-CCUCAGUGGGGU...
H68.9	...ACUCUAAACGGGUA...GAAA...UA-CCCAGGUGGGGU...
H68.a1	...ACUCCCAGUGGGUA...GAAA...UA-CCUAAACGGGGU...
H68.a2	...ACUCCCAGUGGGUA...GAAA...UA-CCCAAACGGGGU...
P12	...ACUCUG-CGGGGUA...GAAA...UA-CCCAGCGGGGGU...
P12a	...ACUCUG-CG-GGUA...GAAA...UA-CC-AGCGGGGGU...
P12b	...ACUCUG-CGCGGUA...GAAA...UA-CCCAGCGGGGGU...
P12c	...ACUCU--CGGGGUA...GAAA...UA-CCCAGCGGGGGU...
P12d	...ACUCUGCGGGGGUA...GAAA...UACCCCAGCGGGGGU...
P12g	...ACUCUG-CG-GGUA...GAAA...UA-CC-ACCGGGGGU...
P12h	...ACUCUGCGGGGGUA...GAAA...UACCCACCGGGGGGU...
P12n	...ACUCUG-CGGGAUA...GAAA...UA-UCCAGCGGGGGU...
P12.14	...ACUCUG-CGGGGUA...GAAA...UA-CCCACCGGGGGU...
P12.a1	...ACUCCAGCGGGGUA...GAAA...UA-CCUG-CGGGGGU...
P12.a2	...ACUCCAGCGGGGUA...GAAA...UA-CCCG-CGGGGGU...

Table 8.1 List of Receptor Names and Sequences
Nucleotides in black were constant in all constructs, nucleotides in red were variable. All hairpin loops contained a GAAA tetraloop, which interacted with the 11nt receptor on the probe monomer (see below).

Molecule	RNA Sequence (5' to 3')
Probe _GGAA	GGGAAAGCGAUUUACUCAUAUGGUAGUUCCGGG-GGAA-CUUGGUUCUA CCUAAGUGGGUAGGUCGUUUUUCUCU
Probe _GUAA	...ACUCAUAUGGUA...GUAA...UACCUAAGUGGGU...
Probe_GCAA	...ACUCAUAUGGUA...GCAA...UACCUAAGUGGGU...
Probe_GAGA	...ACUCAUAUGGUA...GAGA...UACCUAAGUGGGU...
Probe_GGGA	...ACUCAUAUGGUA...GAAA...UACCUAAGUGGGU...
Probe_GUGA	...ACUCAUAUGGUA...GUGA...UACCUAAGUGGGU...
Probe_GCGA	...ACUCAUAUGGUA...GCGA...UACCUAAGUGGGU...

Table 8.2 11nt Probe Sequences
Sequence in red represents the natural 11nt receptor, the loop sequence is variable in each molecule and represents GNRA tetraloops.

Molecule	GGAA	GUAA	GCAA	GAGA	GGGA	GUGA	GCGA	GUAAG	GAUAAG
	Kd (nM)	Kd(nM)	Kd (nM)	Kd (nM)	Kd (nM)	Kd (nM)	Kd (nM)	Kd (nM)	Kd (nM)
H89arch	150000	55000	55000	200000	100000	25000	75000	50000	200000
H89bact	150000	40000	100000	200000	200000	200000	200000	200000	200000
H89arch.1	150000	27000	100000	200000	200000	200000	200000		
H89arch.2	23000	570	1003	1975	23000	697	1029		
H89bact.1	150000	40000	150000	200000	200000	200000	200000		
H89bact.2	10000	2073	3711	5059	30000	6112	17000		
B7.8	404	31	3356	906	1730	110	9314		
S8	138	19	786	228	376	81	1462	10000	35000
S8.2	123	209	1389	887	223	736	3006		
S8.3	510	487	2386	617	426	1205	10509		
S8.4	88	147	1140	455	173	340	4652		
S8.5	200	41	866	543	360	385	3466		
S8.7	217	60	896	210	578	245	1365		
S8.8	331	73	1697	1027	752	467	4550		
S8.9	663	111	1650	1611	2593	811	8606		
S8.11	1747	21	395	1459	2512	1794	1377	314	100000
S8.12	3622	746	3415	6829	6174	8583	200000		
S8.13	226	117	467	232	416	177	503		
S8.14	54	7	311	106	74	28	352	122	7500
S8.15	513	26	212	280	898	125	788	222	35000
S8.16	9669	38	314	201	10913	205	9663		
S8.17	316	175	1142	347	981	228	479		
S8.18	1264	357	1671	2000	720	462	720		
S8.19	58	1	360	162	204	62	1202	265	4303
S8.20	40000	530	775	4110	40000	991	4000		
S8.21	2150	1350	40000	15000	15000	40000	200000		
S8.22	677	42	950	1500	2225	710	3500		
S8.23	2248	133	1000	1660	8000	360	9500		
S8.24	50000	631	2850	200000	200000	50000	50000		
S8.25	3029	644	50000	10000	15000	15000	50000		
S8.26	900	55	1731	10139	4572	3636	50000		
S8.27	1311	1237	28000	45000	25000	40000	100000		
S8.28	372	477	2056	1576	1576	1576	15000		
S8.a1	3285	358	20000	40000	20000	20000	40000	12000	200000
S8.a2	293	220	1041	1427	961	498	2548	7831	200000
H.GYaA	40000	404	750	1412	40000	1251	3158	30000	200000
H.GYgA	2740	950	2010	1825	8240	1825	7115		
H.GYgA.1	100000	8805	5320	200000	200000	8800	9100		

S8Tt1	64	70	673	76	114	117	1407		
S8Tt2	100	16	400	200	250	65	530	519	15000
S8Tt3	1362	121	644	4906	7385	3918	954		
S8Tt4	44	5	223	80	80	34	500	214	9000
S8Tt5	75	37	75	125	37	125	650	250	909
S8Tt6	1550	90	1550	13318	10500	10500	10500		
S8Tt7	1980	887	1980	1980	5795	2000	12690		
S8Tt8	12315	2060	3314	9508	30000	25000	20000		
S8Tt9	20000	10000	100000	200000	100000	200000	200000		
S8Tt.a	2350	750	2120	2120	4400	2300	4400		
S8Tt4.a	45	10	235	103	120	10	360		
S8Tt8.a	529	370	1057	1191	865	619	1961		
H68	487	688	1479	2512	732	1539	4058		
H68.1	469	96	1718	6021	1095	2966	8101		
H68.2	3679	25000	25000	3067	2168	25000	25000		
H68.5	22	132	1130	300	46	250	2600	413	50000
H68.6	58	84	390	110	55	156	736		
H68.7	1385	1575	5636	3862	3536	30000	200000		
H68.8	1569	1604	2672	3220	3045	2608	9330		
H68.9	6152	2054	150000	200000	150000	200000	200000		
H68.a1	25000	7900	40000	50000	50000	150000	150000		
H68.a2	1600	1600	6200	7750	5675	6200	10000		
P12	287	275	315	330	191	275	367		
P12a	50000	15000	30000	100000	25000	25000	30000		
P12b	527	10000	7500	1950	775	15000	15000		
P12c	1850	890	1794	1759	1831	1486	6800		
P12d	15000	734	1684	2000	6900	1900	20000		
P12g	20000	10000	20000	100000	20000	20000	30000		
P12h	100	200	2000	1150	250	800	3300		
P12n	701	465	1577	2102	712	2430	2400		
P12.14	315	48	232	494	261	415	1000		
P12.a1	50000	30000	100000	200000	200000	200000	200000		
P12.a2	2100	1250	3400	4300	13000	4300	6700		

Table 8.3 K_d values
Values are reported in nM and were calculated based off of native gel shift assays.

Molecule	GGAA	GUAA	GCAA	GAGA	GGGA	GUGA	GCGA	GUAAG	GAUAAG
	ΔΔG	ΔΔG	ΔΔG	ΔΔG	ΔΔG	ΔΔG	ΔΔG	ΔΔG	ΔΔG
H89arch	2.04	1.47	1.47	2.20	1.81	1.03	1.65	1.42	2.20
H89bact	2.04	1.30	1.81	2.20	2.20	2.20	2.20	2.20	2.20
H89arch.1	2.04	1.07	1.81	2.20	2.20	2.20	2.20		
H89arch.2	0.98	-1.09	-0.78	-0.40	0.98	-0.98	-0.76		
H89bact.1	2.04	1.30	2.04	2.20	2.20	2.20	2.20		
H89bact.2	0.52	-0.37	-0.04	0.13	1.13	0.24	0.81		
B7.8	-1.29	-2.73	-0.10	-0.83	-0.47	-2.02	0.48		
S8	-1.89	-3.01	-0.91	-1.61	-1.33	-2.19	-0.57	-1.78	1.36
S8.2	-1.96	-1.66	-0.59	-0.85	-1.62	-0.95	-0.16		
S8.3	-1.16	-1.18	-0.29	-1.05	-1.26	-0.67	0.54		
S8.4	-2.15	-1.86	-0.71	-1.22	-1.77	-1.39	0.09		
S8.5	-1.68	-2.57	-0.86	-1.12	-1.35	-1.32	-0.08		
S8.7	-1.64	-2.36	-0.84	-1.66	-1.09	-1.57	-0.60		
S8.8	-1.40	-2.25	-0.48	-0.76	-0.94	-1.21	0.07		
S8.9	-1.01	-2.01	-0.50	-0.51	-0.24	-0.90	0.43	0.52	1.22
S8.11	-0.47	-2.95	-1.30	-0.57	-0.26	-0.45	-0.60	-1.43	1.81
S8.12	-0.06	-0.94	-0.09	0.30	0.24	0.43	2.20		
S8.13	-1.62	-1.99	-1.21	-1.60	-1.27	-1.75	-1.17		
S8.14	-2.42	-3.57	-1.44	-2.04	-2.24	-2.79	-1.37	-1.96	0.35
S8.15	-1.15	-2.83	-1.65	-1.49	-0.84	-1.95	-0.91	-1.63	1.22
S8.16	0.50	-2.62	-1.43	-1.68	0.57	-1.67	0.50		
S8.17	-1.43	-1.76	-0.70	-1.37	-0.79	-1.61	-1.19		
S8.18	-0.65	-1.36	-0.49	-0.39	-0.96	-1.21	-0.96		
S8.19	-2.38	-4.66	-1.35	-1.80	-1.67	-2.34	-0.68	-1.53	0.04
S8.20	1.30	-1.14	-0.92	0.02	1.30	-0.78	0.00		
S8.21	-0.35	-0.61	1.30	0.74	0.74	1.30	2.20		
S8.22	-1.00	-2.56	-0.81	-0.55	-0.33	-0.97	-0.07		
S8.23	-0.32	-1.91	-0.78	-0.49	0.39	-1.35	0.49		
S8.24	1.42	-1.04	-0.19	2.20	2.20	1.42	1.42		
S8.25	-0.16	-1.03	1.42	0.52	0.74	0.74	1.42		
S8.26	-0.84	-2.41	-0.47	0.52	0.08	-0.05	1.42		
S8.27	-0.63	-0.66	1.10	1.36	1.03	1.30	1.81		
S8.28	-1.33	-1.20	-0.37	-0.52	-0.52	-0.52	0.74		
S8.a1	-0.11	-1.36	0.91	1.30	0.91	0.91	1.30	0.62	2.20
S8.a2	-1.47	-1.63	-0.76	-0.58	-0.80	-1.17	-0.25	0.38	2.20
H.GYaA	1.30	-1.29	-0.94	-0.58	1.30	-0.65	-0.13	1.13	2.20
H.GYgA	-0.21	-0.81	-0.39	-0.44	0.41	-0.44	0.32		
H.GYgA.1	1.81	0.44	0.16	2.20	2.20	0.44	0.46		

S8Tt1	-2.32	-2.27	-1.00	-2.23	-2.00	-1.99	-0.59		
S8Tt2	-2.07	-3.10	-1.29	-1.68	-1.56	-2.32	-1.14	-1.15	0.74
S8Tt3	-0.61	-1.97	-1.03	0.12	0.35	-0.01	-0.81		
S8Tt4	-2.54	-3.76	-1.62	-2.20	-2.20	-2.68	-1.17		
S8Tt5	-2.24	-2.63	-2.24	-1.95	-2.63	-1.95	-1.02		
S8Tt6	-0.53	-2.13	-0.53	0.68	0.54	0.54	0.54		
S8Tt7	-0.39	-0.85	-0.39	-0.39	0.21	-0.39	0.65		
S8Tt8	0.63	-0.37	-0.11	0.49	1.13	1.03	0.91		
S8Tt9	0.91	0.52	1.81	2.20	1.81	2.20	2.20		
S8Tt.a	-0.30	-0.94	-0.36	-0.36	0.05	-0.31	0.05		
S8Tt4.a	-1.14	-1.34	-0.75	-0.68	-0.86	-1.05	-0.40		
S8Tt8.a	-2.52	-3.37	-1.59	-2.06	-1.97	-3.37	-1.35		
H68	-1.18	-0.99	-0.56	-0.26	-0.95	-0.54	0.01		
H68.1	-1.20	-2.10	-0.47	0.23	-0.73	-0.17	0.40		
H68.2	-0.05	1.03	1.03	-0.15	-0.34	1.03	1.03		
H68.5	-2.92	-1.92	-0.71	-1.46	-2.51	-1.56	-0.24	-1.28	1.42
H68.6	-2.38	-2.17	-1.31	-2.02	-2.41	-1.82	-0.95		
H68.7	-0.60	-0.52	0.19	-0.02	-0.07	1.13	2.20		
H68.8	-0.53	-0.51	-0.23	-0.12	-0.15	-0.24	0.48		
H68.9	0.24	-0.37	2.04	2.20	2.04	2.20	2.20		
H68.a1	1.03	0.38	1.30	1.42	1.42	2.04	2.04		
H68.a2	-0.51	-0.51	0.25	0.37	0.20	0.25	0.52		
P12	-1.48	-1.50	-1.43	-1.40	-1.71	-1.50	-1.34		
P12a	1.42	0.74	1.13	1.81	1.03	1.03	1.13		
P12b	-1.14	0.52	0.35	-0.40	-0.92	0.74	0.74		
P12c	-0.43	-0.84	-0.45	-0.46	-0.44	-0.56	0.30		
P12d	0.74	-0.95	-0.49	-0.39	0.31	-0.42	0.91		
P12g	0.91	0.52	0.91	1.81	0.91	0.91	1.13		
P12h	-2.07	-1.68	-0.39	-0.70	-1.56	-0.90	-0.11		
P12n	-0.98	-1.21	-0.52	-0.36	-0.97	-0.28	-0.29		
P12.14	-1.43	-2.49	-1.60	-1.18	-1.53	-1.27	-0.78		
P12.a1	1.42	1.13	1.81	2.20	2.20	2.20	2.20		
P12.a2	-0.36	-0.65	-0.09	0.04	0.66	0.04	0.29		

Table 8.4 ΔΔG Values
Reference for ΔΔG values is 4000 nM. Each value was calculated from the Kd in Table 8.3.

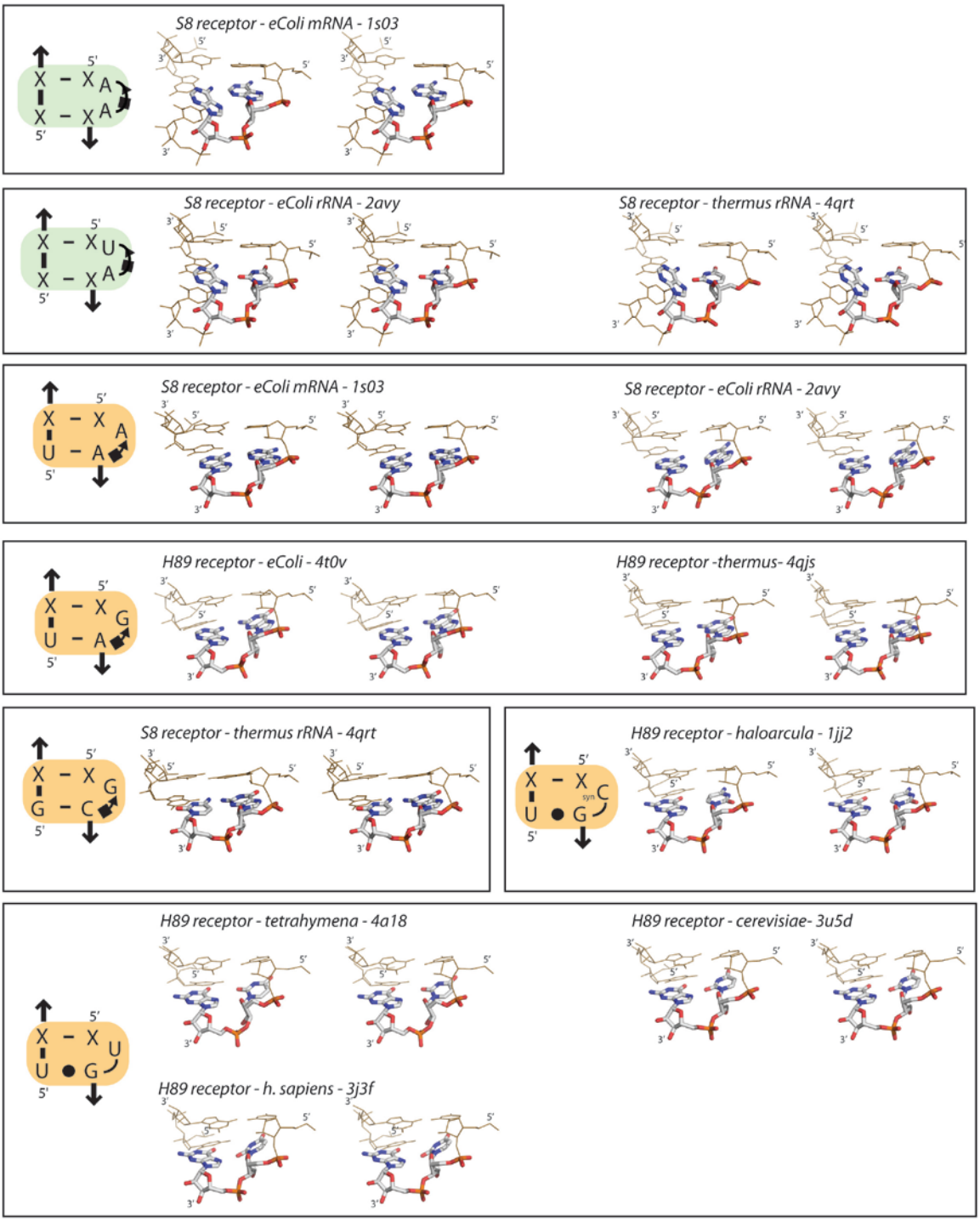


Figure 8.2 Crystal Structures of platforms
Crystal structure files are labeled on figure. S8 module in green is upper platform that directly interacts with the S8 protein, orange module is lower platform. H89 orange module is the platform that directly interacts with terminal loop.

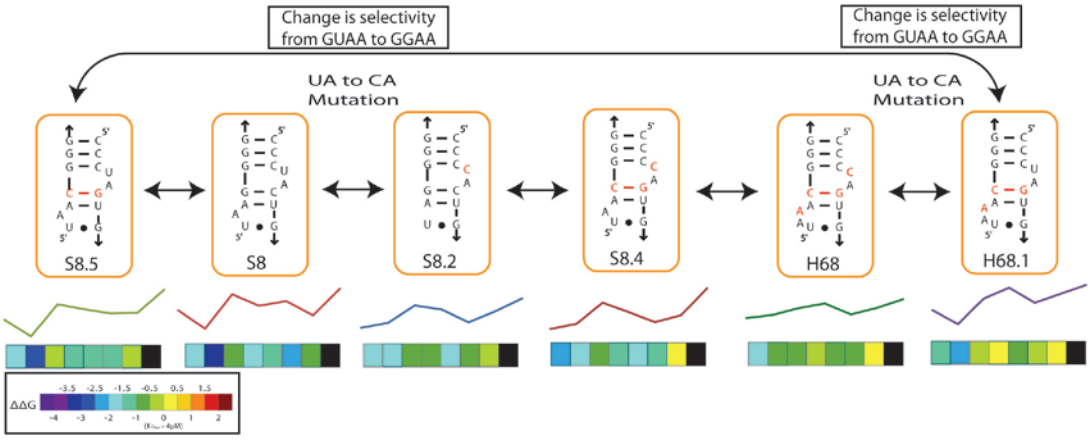
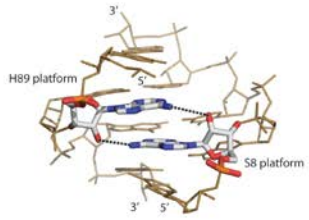
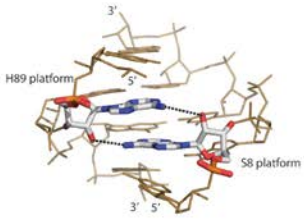


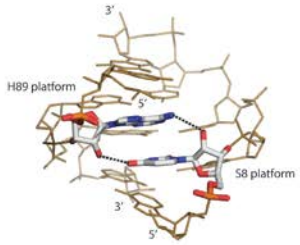
Figure 8.3 Mutational Pathway from S8 to H68
Two separate pathways from S8 to H68 via single point mutations. Specificity profiles highlight the change in selectivity from GUAA to GGAA. In each pathway, selectivity switches when the interacting platform changes to a CA (S8 to S8.2 and H68 to H68.1).

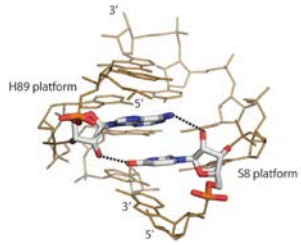
S8 eColi (spc openron mRNA)
- 1S03



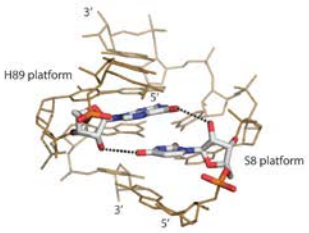


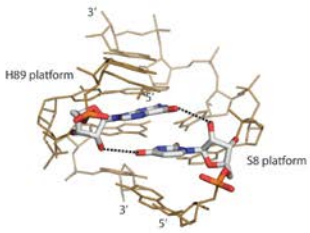
S8 eColi (16s rRNA)
- 2AVY



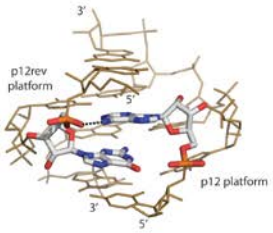


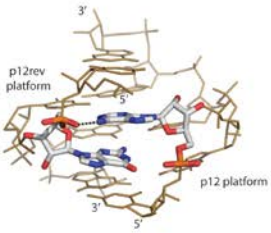
S8 thermus (16s rRNA)
- 4QJT



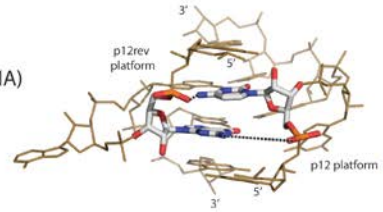


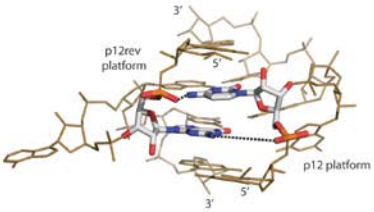
P12 thermus (RNase P)
- 1U9S



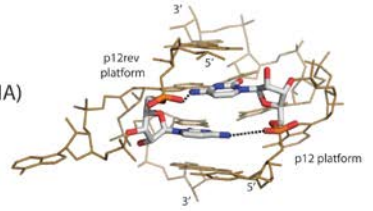


H68 thermus (23s rRNA)
- 4QJS





H68 haloarcula (23s rRNA)
- 1JJ2



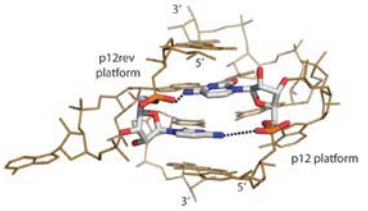


Figure 8.4 Crystal structures of stacking modules from S8, P12 and H68
Crystal structure files are labeled in figure at left. Figure highlights the hydrogen bond contacts between each module. The S8 stacks demonstrate a classic 2x_bulge that is locked in by two hydrogen bonds. It is important to note that in the P12 and H68 stacking modules, the orientation of the interacting nucleotides switches from top to bottom, when compared to the S8 orientation.

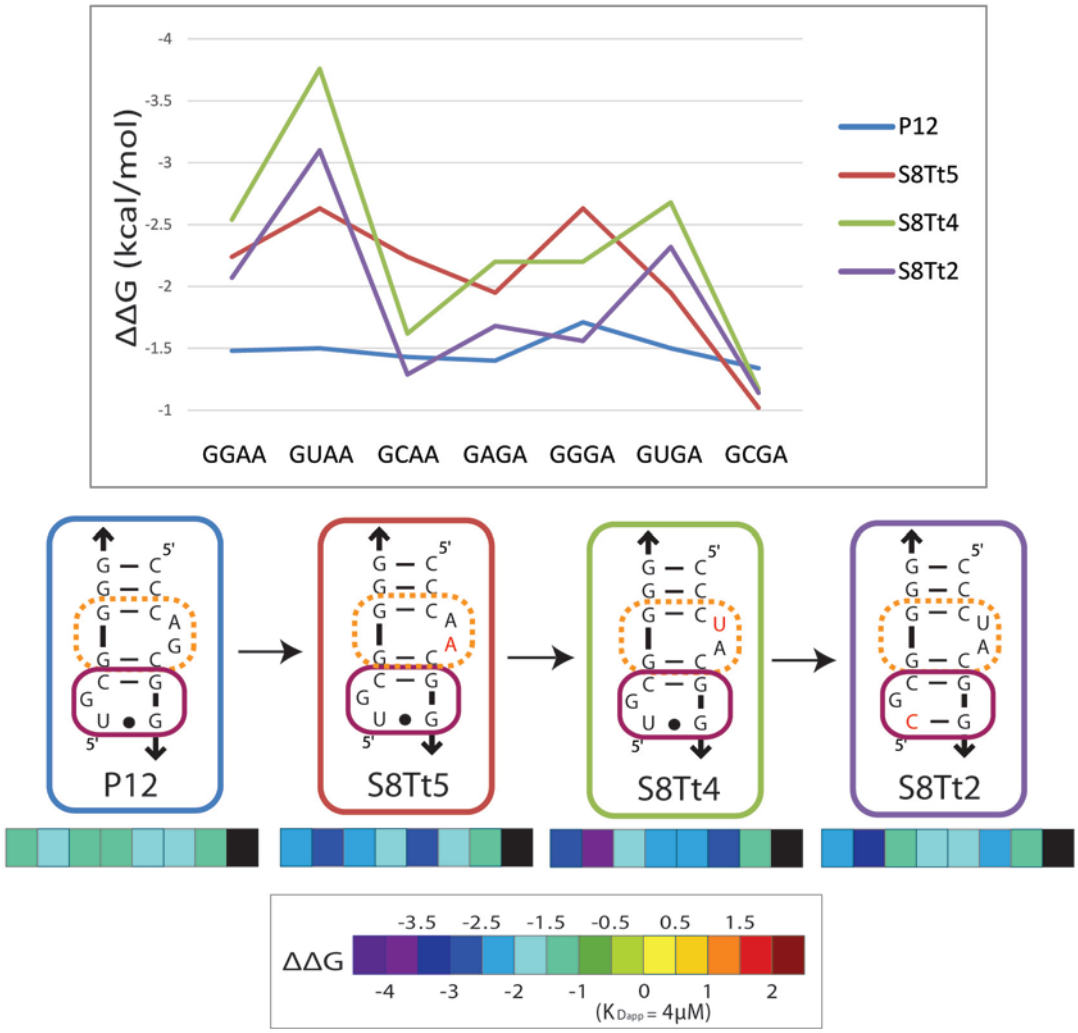


Figure 8.5 Mutational Pathway from P12 to S8Tt2
P12 and S8Tt2 have very different phenotypic behavior, yet only 3 point mutations separate the two classes. S8Tt4 increases affinity for all loops while maintain an identical profile to that of S8Tt2. S8Tt5 is one point mutation from P12 and from S8Tt4 and seems to have a hybrid of the two profiles.

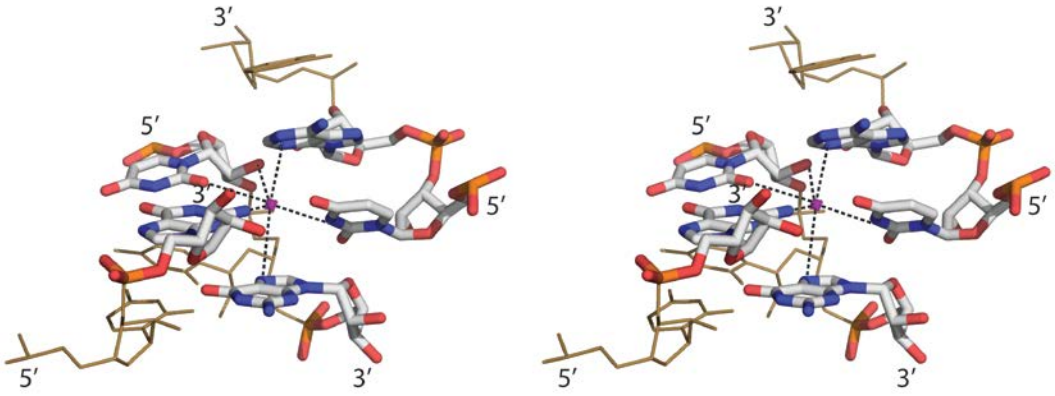


Figure 8.6 Crystal structure of H89 coordinating Mg^{2+} with terminal loop
Coordination of a Mg^{2+} ion in the binding site of the H89 receptor. Potential reasoning for why H89/S9-like receptor are not as selective as other GNRA tetraloops, like 11nt and R1 [72]

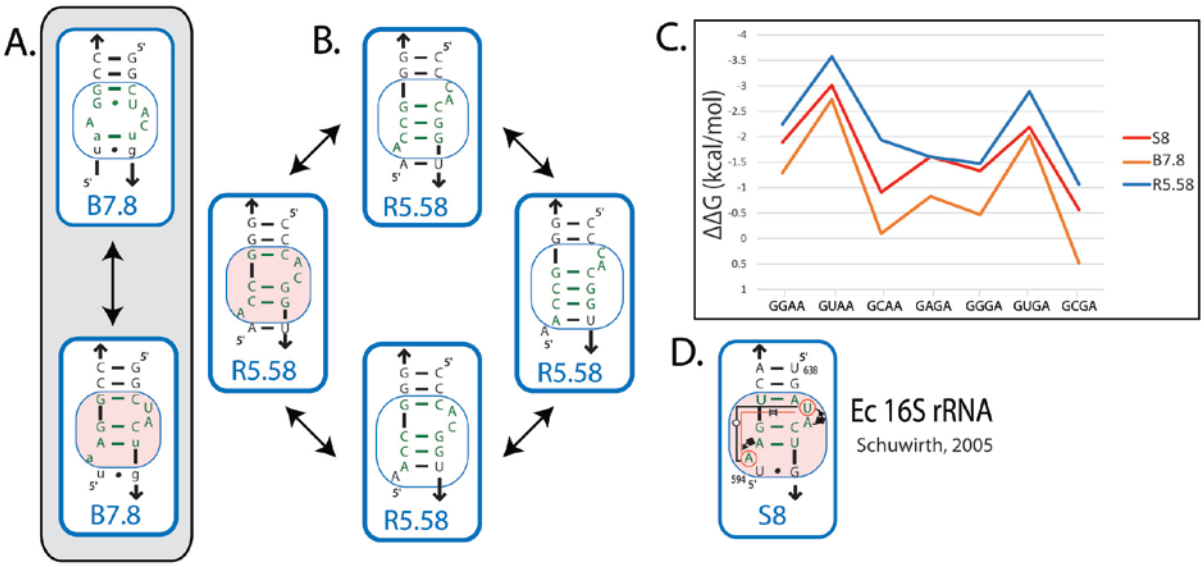


Figure 8.7 Alternative folds of other selected 2x_bulge motifs
(A-B) B7.8 and R5.58 are both artificially selected receptors that selectively bind GUAA. The selectivity profiles each have very similar loop specificity profiles to that of S8 (C) (Calkins 2015-in preparation). S8 adopts a 2X-bulge structure (D) and has a near identical sequence to that of B7.8. Alternative pairing demonstrate that both B7.8 and R5.58 are capable of adopting a 2x-bulge structure. This along with similar specificity profiles, suggests a possible structural context for the artificially selected receptors.

8.2 Supplemental Information for Chapter 3

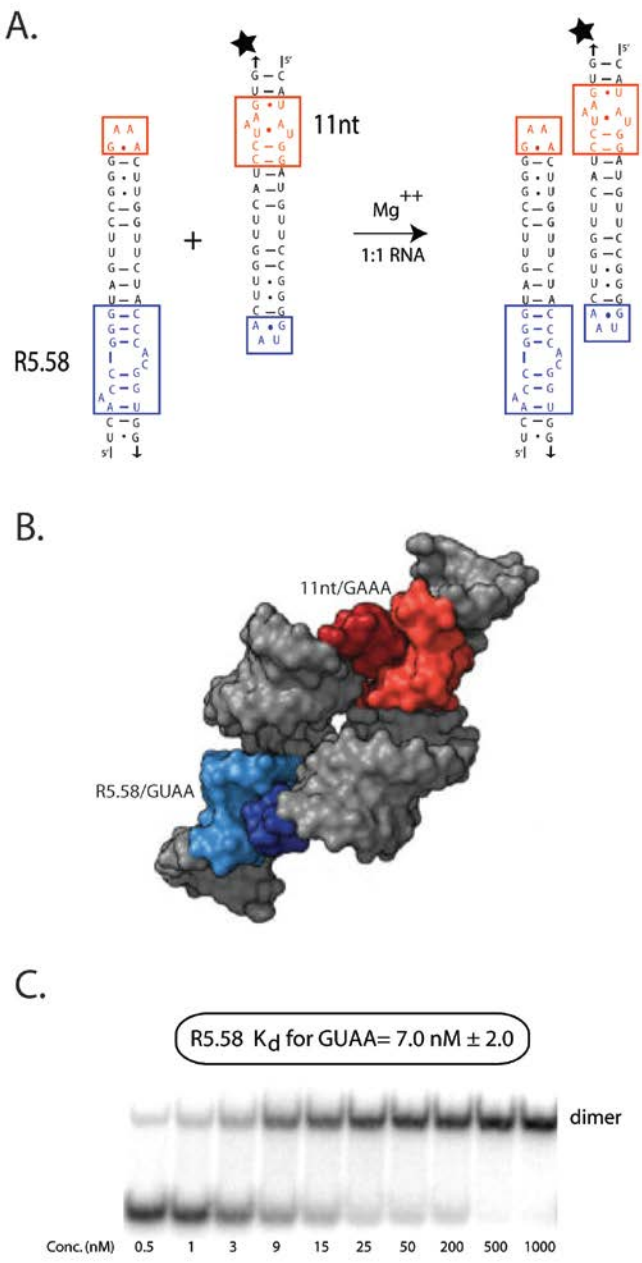


Figure 8.8 Tectodimer Self-Assembling Assay
(A) Heterodimer self-assembling assay utilizing two different tetraloop/receptor interactions. The synergistic binding of the llnt/GAAA interaction and the experimental loop/receptor interaction (shown as R5.58 and GUAA) create dimers upon the addition of Mg^{2+}. (B) Shows a space filling model of the dimer interaction (PDB_ID: 2JYJ). (C) Degree of dimerization and K_d values can be determined from through titration assays on native PAGE. Quantitation of monomer and dimer can be calculated using ImageQuant software and plotted against concentration to determine K_d values for each loop/receptor interaction.

Selection Receptors	# times selected	Sequence
R5.34	Selex (1)	GGGAAAGCGAUUUACUCAUAAC---GGUAGUUCCGGG...GAAA...CUUGGUUCUACC-AUGUG-UGGGUAGGUCGUUUUUCUCU
R5.16	Selex (1)	GGGAAAGCGAUUUACUCAUAAC---GGUAGUUCCGGG...GAAA...CUUGGUUCUACC-UGCGC-UGGGUAGGUCGUUUUUCUCU
R5.58	Selex (15)	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.8b	Selex (1)	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AUGG-UGGGUAGGUCGUUUUUCUCU
R5.15b	Selex (2)	GGGAAAGCGAUUUACUCA--ACCU-GGUAGUUCCGGG...GAAA...CUUGGUUCUACC-GUCGG-UGGGUAGGUCGUUUUUCUCU
R5.15	Selex (1)	GGGAAAGCGAUUUACUCA--A-CU-GGUAGUUCCGGG...GAAA...CUUGGUUCUACC-GUCG--UGGGUAGGUCGUUUUUCUCU
R5.13	Selex (1)	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC-GUCGG-UGGGUAGGUCGUUUUUCUCU
R5.33	Selex (1)	GGGAAAGCGAUUUACUCAUAACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGG-AGGGUAGGUCGUUUUUCUCU
R5.14	Selex (1)	GGGAAAGCGAUUUACUCA-AACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGGUUGGGUAGGUCGUUUUUCUCU
R5.12	Selex (9)	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGGUUGGGUAGGUCGUUUUUCUCU
R5.12b	Selex (2)	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGGCUGGGUAGGUCGUUUUUCUCU
Mutant Receptors	**Mutation**	**Sequence**
R5.58	Wild Type	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M1	G13U:C1G	GGGAAAGCGAUUUACUCA--ACC-GGUUAGUUCCGGG...GAAA...CUUGGUUCUAGCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M2	G12A:C2U	GGGAAAGCGAUUUACUCA--ACC-GAGUAGUUCCGGG...GAAA...CUUGGUUCUACUC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M3	G12C:C2G	GGGAAAGCGAUUUACUCA--ACC-GCGUAGUUCCGGG...GAAA...CUUGGUUCUACGC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M4	G12U:C2A	GGGAAAGCGAUUUACUCA--ACC-GUGUAGUUCCGGG...GAAA...CUUGGUUCUACAC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M5	G6C	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCG-UGGGUAGGUCGUUUUUCUCU
R5.58M6	C10G	GGGAAAGCGAUUUACUCA--ACG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M7	C10G:G6C	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCG-UGGGUAGGUCGUUUUUCUCU
R5.58M8	A7.2C:U7.1G	GGGAAAGCGAUUUACUCC--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-GGGGUAGGUCGUUUUUCUCU
R5.58M9	G11U:C3U	GGGAAAGCGAUUUACUCA--ACC-UGGUAGUUCCGGG...GAAA...CUUGGUUCUACCU-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M10	G11C:C3U	GGGAAAGCGAUUUACUCA--ACC-CGGUAGUUCCGGG...GAAA...CUUGGUUCUACCU-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M11	G11C:C3G	GGGAAAGCGAUUUACUCA--ACC-CGGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M12	G11C	GGGAAAGCGAUUUACUCA--ACC-CGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M12d1	G11del, C3del	GGGAAAGCGAUUUACUCA--ACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M13	G11A:C3U	GGGAAAGCGAUUUACUCA--ACC-AGGUAGUUCCGGG...GAAA...CUUGGUUCUACCU-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M14	U7.1G:A7.2U, A8G, C10G:C6G	GGGAAAGCGAUUUACUCU--GCG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCG-GGGGUAGGUCGUUUUUCUCU
R5.58M15	U7.1G:A7.2U, A8G	GGGAAAGCGAUUUACUCU--GCC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-GGGGUAGGUCGUUUUUCUCU
R5.58M16	G13C:C1G	GGGAAAGCGAUUUACUCA--ACC-GGCUAGUUCCGGG...GAAA...CUUGGUUCUAGCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M17	G13U:C1A	GGGAAAGCGAUUUACUCA--ACC-GGUUAGUUCCGGG...GAAA...CUUGGUUCUAACC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M18	G13C:C1G, G12C:C2G	GGGAAAGCGAUUUACUCA--ACC-GCCUAGUUCCGGG...GAAA...CUUGGUUCUAGGC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M19	G13A:C1U	GGGAAAGCGAUUUACUCA--ACC-GGAUAGUUCCGGG...GAAA...CUUGGUUCUAUCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M20	G11U:C3A	GGGAAAGCGAUUUACUCA--ACC-UGGUAGUUCCGGG...GAAA...CUUGGUUCUACCA-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M21	G11U	GGGAAAGCGAUUUACUCA--ACC-UGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M22	C5G	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AGGG-UGGGUAGGUCGUUUUUCUCU

R5.58M23	C9G:G7C	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M24	C9A:G7U	GGGAAAGCGAUUUACUCA--AAC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGU-UGGGUAGGUCGUUUUUCUCU
R5.58M25	A8del	GGGAAAGCGAUUUACUCA---CC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M27	C10G:G6C, A7.2U:U7.1G	GGGAAAGCGAUUUACUCU--ACG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCG-GGGGUAGGUCGUUUUUCUCU
R5.58M28	C9A:G7U, A7.2U:U7.1G	GGGAAAGCGAUUUACUCU--AAC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGU-GGGGUAGGUCGUUUUUCUCU
R5.58M29	A7.2U:U7.1G	GGGAAAGCGAUUUACUCU--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-GGGGUAGGUCGUUUUUCUCU
R5.58M30	C9U:G7A, A8G, A7.2U:U7.1G	GGGAAAGCGAUUUACUCU--GUC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGA-GGGGUAGGUCGUUUUUCUCU
R5.58M31	G7U	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGU-UGGGUAGGUCGUUUUUCUCU
R5.58M32	C9G	GGGAAAGCGAUUUACUCA--AGC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M33	C9A	GGGAAAGCGAUUUACUCA--AAC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58M34	C5A	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AAGG-UGGGUAGGUCGUUUUUCUCU
R5.58M35	C9U:G7A	GGGAAAGCGAUUUACUCA--AUC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGA-UGGGUAGGUCGUUUUUCUCU
R5.58M36	A5U: C6A	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UAGG-UGGGUAGGUCGUUUUUCUCU
R5.58M37	A4U:C5G	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UGGG-UGGGUAGGUCGUUUUUCUCU
R5.58M38	A4G:C5U	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-GUGG-UGGGUAGGUCGUUUUUCUCU
R5.58M39	A4C:C5A	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-CAGG-UGGGUAGGUCGUUUUUCUCU
R5.58M40	A4U	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UCGG-UGGGUAGGUCGUUUUUCUCU
R5.58M41	A4del, C5del	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC---GG-UGGGUAGGUCGUUUUUCUCU
R5.58M42	C3U	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCU-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58A4G	A4G	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-GCGG-UGGGUAGGUCGUUUUUCUCU
R5.58A4C	A4C	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-CCGG-UGGGUAGGUCGUUUUUCUCU
R5.58A4del	A4del	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC--CGG-UGGGUAGGUCGUUUUUCUCU
R5.58C5del	C5del	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-A-GG-UGGGUAGGUCGUUUUUCUCU
R5.58C3A	C3A	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCA-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58C3G	C3G	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58C3del	C3del	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58G11A	G11A	GGGAAAGCGAUUUACUCA--ACC-AGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58G11del	G11del	GGGAAAGCGAUUUACUCA--ACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58G6A/C10U	G6A;C10U	GGGAAAGCGAUUUACUCA--ACU-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACAG-UGGGUAGGUCGUUUUUCUCU
R5.58G6U/C10A	G6U:C10A	GGGAAAGCGAUUUACUCA--ACA-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACUG-UGGGUAGGUCGUUUUUCUCU
R5.58G6U	G6U	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACUG-UGGGUAGGUCGUUUUUCUCU
R5.58C10U	C10U	GGGAAAGCGAUUUACUCA--ACU-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58C10A	C10A	GGGAAAGCGAUUUACUCA--ACA-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58G6A	G6A	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACAG-UGGGUAGGUCGUUUUUCUCU
R5.58C10del	C10del/same as C9del	GGGAAAGCGAUUUACUCA--AC--GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58G7A	G7A	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGA-UGGGUAGGUCGUUUUUCUCU
R5.58G7C	G7C	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGC-UGGGUAGGUCGUUUUUCUCU
R5.58C9U	C9U	GGGAAAGCGAUUUACUCA--AUC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58A8U	A8U	GGGAAAGCGAUUUACUCA--UCC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU

R5.58A8G	A8G	GGGAAAGCGAUUUACUCA--GCC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58A8C	A8C	GGGAAAGCGAUUUACUCA--CCC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58rev	reverse orientaion of R5.58	GGGAAAGCGAUUUACUCA-ACCC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACC--UGGGUAGGUCGUUUUUCUCU
R5.58M7i1	G6C:C10G, U7.1G	GGGAAAGCGAUUUACUCA--ACG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCG-GGGGUAGGUCGUUUUUCUCU
R5.58M24i1	G7U:C9A, U7.1G	GGGAAAGCGAUUUACUCA--AAC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGU-GGGGUAGGUCGUUUUUCUCU
R5.58M27i1	G6C:C10G, U7.1G:A7.2U, C9A	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCG-GGGGUAGGUCGUUUUUCUCU
R5.58M28i1	G6C, G7U:C9A, U7.1G:A7.2U	GGGAAAGCGAUUUACUCA--ACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG-UGGGUAGGUCGUUUUUCUCU
R5.58c	C3G:G11U	GGGAAAGCGAUUUACUCA--ACC-UGGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-ACGG-UGGGUAGGUCGUUUUUCUCU

Table 8.5 Selected Receptors and Mutants with Sequences
Table showing name of each selected receptor and the number of times the sequence appeared in our results

Selection Receptors	Kd (nM) GGAA	Kd (nM) GUAA	Kd (nM) GAGA	Kd (nM) GCAA	Kd (nM) GGGA	Kd (nM) GUGA	Kd (nM) GCGA
R5.34	200000	5800	200000	200000	200000	200000	200000
R5.16	8665	12	30000	5908	150000	687	30000
R5.58	74	7	129	232	290	24	600
R5.8b	163	20	764	733	648	10	1500
R5.15b	348	50	66	382	3180	51	383
R5.15	1600	50	648	1665	10000	406	6326
R5.13	10000	5000	10000	10000	100000	10000	15000
R5.33	30000	23000	150000	200000	200000	50000	50000
R5.14	15000	601	30000	200000	200000	45000	200000
R5.12	23000	810	30000	200000	200000	25000	200000
R5.12b	23000	600	20000	200000	200000	27000	200000

Mutant Receptors	Kd (nM) GGAA	Kd (nM) GUAA	Kd (nM) GAGA	Kd (nM) GCAA	Kd (nM) GGGA	Kd (nM) GUGA	Kd (nM) GCGA	Kd (nM) GAAA
R5.58	74	7	129	232	290	24	600	
R5.58M1	168	7	165	2400	4671	191	4801	
R5.58M2	7729	106	2393	200000	200000	1620	200000	
R5.58M3	2695	44	1818	1828	1686	179	6766	
R5.58M4	5436	125	2374	200000	200000	2462	200000	
R5.58M5	364	50	666	728	671	370	1816	
R5.58M6	6712	1106	200000	9250	200000	200000	200000	
R5.58M7	378	97	799	525	636	544	2170	
R5.58M8	196	9	294	755	681	122	1523	
R5.58M9	838	455	1318	4884	4162	3303	6355	
R5.58M10	1143	620	3175	6730	7040	3591	200000	
R5.58M11	803	19	256	493	875	275	945	
R5.58M12	71	1285	3007	892	45	1301	5331	
R5.58M12d1	17800	1722	28300	200000	200000	49700	200000	
R5.58M13	117	104	474	2781	7632	924	3076	
R5.58M14	315	48	232	494	261	415	1000	
R5.58M15	1231	82	569	1322	2719	521	2398	
R5.58M16	233	23	339	2003	3199	147	3585	
R5.58M17	190	22	254	1696	2589	148	3587	
R5.58M18	1209	38	1245	3697	2197	216	8946	
R5.58M19	244	37	192	690	1205	99	1431	
R5.58M20	1449	239	1262	2892	5000	2281	7093	
R5.58M21	491	274	1007	1568	1321	1568	2825	
R5.58M22	687	1132	4016	2410	1308	1609	4222	

R5.58M23	140	126	1130	820	273	808	2453
R5.58M24	266	49	311	179	541	135	763
R5.58M25	623	72	717	745	1242	335	1999
R5.58M27	121	82	196	116	153	107	432
R5.58M28	146	3	46	221	307	37	380
R5.58M29	147	13	160	208	320	85	345
R5.58M30	930	70	690	1011	1646	279	1913
R5.58M31	1536	637	1254	929	2752	1225	3411
R5.58M32	1578	286	717	865	3739	764	1715
R5.58M33	182	75	337	320	260	222	1025
R5.58M34	44	15	353	121	48	26	504
R5.58M35	56	39	455	296	88	194	1668
R5.58M36	111	79	1651	1025	254	706	4003
R5.58M37	1260	1573	8270	7045	1854	2944	10000
R5.58M38	2278	381	1346	957	2993	636	4445
R5.58M39	167	255	2279	1232	340	483	3544
R5.58M40	173	62	220	346	358	335	2305
R5.58M41	8244	366	711	2090	9811	1021	2042
R5.58M42	2150	233	2150	6326	6650	6650	200000
R5.58A4G	4185	767	4566	2159	6784	1445	3310
R5.58A4C	1293	736	1796	1156	1129	1264	3250
R5.58A4del	8228	2305	8616	4033	12000	4218	4743
R5.58C5del	6	391	1932	899	153	801	2387
R5.58C3A	15271	23911	105000	200000	200000	120000	180000
R5.58C3G	1168	3990	15572	200000	200000	50000	85000
R5.58C3del	200000	30000	170000	100000	200000	30000	60000
R5.58G11A	30000	35000	200000	200000	50000	200000	200000
R5.58G11del	471	3163	3016	1150	621	3060	2439
R5.58G6A/C10U	64	52	491	231	149	326	1482
R5.58G6U/C10A	551	336	1322	1319	1362	1135	4956
R5.58G6U	421	233	801	751	927	656	2048
R5.58C10U	938	305	1550	1549	2648	901	2856
R5.58C10A	696	171	1074	1257	951	579	2331
R5.58G6A		199	1102	506	3266	625	8398
R5.58C10del	290	156	813	604	530	646	2037
R5.58G7A	852	639	20000	1130	10000	1255	25000
R5.58G7C	2063	805	1918	1553	2937	2404	5529
R5.58C9U	1536	637	1254	929	2752	1225	3411
R5.58A8U	717	270	877	845	2033	696	2227
R5.58A8G	3280	114	2311	3400	6225	1344	6411

115

R5.58A8C	1879	801	2207	2336	11801	1900	9540
R5.58rev	156	618	3590	512	2804	3492	6271
R5.58M7i1	772	21400	5990	588	481	11950	5324
R5.58M24i1	72	1	114	173	86	1	381
R5.58M27i1	2132	624	1663	2055	2949	1328	4059
R5.58M28i1	795	188	1403	798	1502	467	3148
R5.58c	425	27	193	460	8302	147	973

Table 8.6 Selected Receptors and Mutants with K_ds
K_ds are in nM and were calculate from native gel shift assays

Selection Receptors	ΔΔG GGAA	ΔΔG GUAA	ΔΔG GAGA	ΔΔG GCAA	ΔΔG GGGA	ΔΔG GUGA	ΔΔG GCGA
R5.34	1.421	-0.659	2.201	2.201	2.201	1.22	2.201
R5.16	0.435	-3.293	1.134	0.22	2.039	-0.991	1.134
R5.58	-2.247	-3.572	-1.935	-1.602	-1.477	-2.885	-1.067
R5.8b	-1.801	-2.981	-0.931	-0.955	-1.024	-3.371	-0.553
R5.15b	-1.373	-2.466	-2.313	-1.321	-0.129	-2.45	-1.32
R5.15	-0.516	-2.466	-1.024	-0.493	0.516	-1.287	0.258
R5.13	0.516	0.126	0.516	0.516	1.811	0.516	0.744
R5.33	1.134	0.984	2.039	2.201	2.201	1.421	1.421
R5.14	0.744	-1.067	1.134	2.201	2.201	1.362	2.201
R5.12	0.984	-0.899	1.134	2.201	2.201	1.031	2.201
R5.12b	0.984	-1.067	0.906	2.201	2.201	1.074	2.201

Mutant Receptors	ΔΔG GGAA	ΔΔG GUAA	ΔΔG GAGA	ΔΔG GCAA	ΔΔG GGGA	ΔΔG GUGA	ΔΔG GCGA	ΔΔG GAAA
R5.58	-2.247	-3.572	-1.935	-1.602	-1.477	-2.885	-1.067	
R5.58M1	-1.784	-3.58	-1.794	-0.287	0.087	-1.712	0.103	
R5.58M2	0.371	-2.043	-0.289	1.421	1.421	-0.509	1.421	
R5.58M3	-0.222	-2.541	-0.444	-0.441	-0.486	-1.748	0.296	
R5.58M4	0.173	-1.948	-0.294	1.421	1.421	-0.273	1.421	
R5.58M5	-1.349	-2.465	-1.009	-0.959	-1.005	-1.339	-0.444	
R5.58M6	0.291	-0.723	1.421	0.472	1.421	1.421	1.421	
R5.58M7	-1.327	-2.095	-0.906	-1.143	-1.035	-1.123	-0.344	
R5.58M8	-1.697	-3.412	-1.469	-0.938	-0.996	-1.964	-0.543	
R5.58M9	-0.879	-1.223	-0.625	0.112	0.022	-0.108	0.26	
R5.58M10	-0.705	-1.049	-0.13	0.293	0.318	-0.061	1.421	
R5.58M11	-0.903	-3.025	-1.547	-1.178	-0.855	-1.506	-0.812	
R5.58M12	-2.268	-0.639	-0.161	-0.844	-2.525	-0.632	0.162	
R5.58M12d1	0.84	-0.474	1.101	2.201	2.201	1.418	2.201	
R5.58M13	-1.987	-2.051	-1.2	-0.205	0.364	-0.825	-0.148	
R5.58M14	-1.43	-2.489	-1.602	-1.177	-1.536	-1.275	-0.78	
R5.58M15	-0.663	-2.191	-1.097	-0.623	-0.217	-1.147	-0.288	
R5.58M16	-1.6	-2.903	-1.389	-0.389	-0.126	-1.859	-0.062	
R5.58M17	-1.714	-2.928	-1.551	-0.483	-0.245	-1.855	-0.061	
R5.58M18	-0.673	-2.62	-0.657	-1.3	-0.337	-1.642	0.453	
R5.58M19	-1.574	-2.635	-1.709	-0.989	-0.675	-2.081	-0.578	
R5.58M20	-0.571	-1.584	-0.649	-0.183	0.126	-0.316	0.322	
R5.58M21	-1.18	-1.509	-0.776	-0.527	-0.623	-0.527	-0.196	
R5.58M22	-0.991	-0.71	0.002	-0.285	-0.629	-0.512	0.03	

R5.58M23	-1.886	-1.947	-0.711	-0.892	-1.511	-0.9	-0.275
R5.58M24	-1.525	-2.475	-1.437	-1.748	-1.126	-1.907	-0.932
R5.58M25	-1.046	-2.261	-0.967	-0.946	-0.658	-1.395	-0.39
R5.58M27	-1.968	-2.19	-1.697	-1.992	-1.836	-2.038	-1.252
R5.58M28	-1.863	-4.129	-2.513	-1.629	-1.445	-2.635	-1.324
R5.58M29	-1.859	-3.237	-1.811	-1.664	-1.421	-2.167	-1.379
R5.58M30	-0.821	-2.278	-0.989	-0.774	-0.5	-1.498	-0.415
R5.58M31	-0.539	-1.034	-0.653	-0.821	-0.21	-0.666	-0.09
R5.58M32	-0.523	-1.485	-0.967	-0.862	-0.038	-0.932	-0.477
R5.58M33	-1.739	-2.239	-1.392	-1.421	-1.538	-1.627	-0.766
R5.58M34	-2.538	-3.143	-1.366	-1.968	-2.489	-2.834	-1.166
R5.58M35	-2.402	-2.613	-1.223	-1.465	-2.148	-1.703	-0.492
R5.58M36	-2.017	-2.207	-0.498	-0.766	-1.551	-0.976	0
R5.58M37	-0.65	-0.525	0.409	0.318	-0.433	-0.172	0.516
R5.58M38	-0.317	-1.323	-0.613	-0.805	-0.163	-1.035	0.059
R5.58M39	-1.787	-1.549	-0.317	-0.663	-1.387	-1.19	-0.068
R5.58M40	-1.767	-2.345	-1.632	-1.377	-1.358	-1.395	-0.31
R5.58M41	0.407	-1.346	-0.972	-0.365	0.505	-0.768	-0.378
R5.58M42	-0.349	-1.6	-0.349	0.258	0.286	0.286	2.201
R5.58A4G	0.025	-0.929	0.074	-0.347	0.297	-0.573	-0.107
R5.58A4C	-0.635	-0.953	-0.451	-0.698	-0.712	-0.648	-0.117
R5.58A4del	0.406	-0.31	0.432	0.005	0.618	0.03	0.096
R5.58C5del	-3.64	-1.308	-0.409	-0.84	-1.836	-0.905	-0.29
R5.58C3A	0.754	1.006	1.839	2.201	2.201	1.914	2.142
R5.58C3G	-0.693	-0.001	0.765	2.201	2.201	1.421	1.72
R5.58C3del	2.201	1.134	2.11	1.811	2.201	1.134	1.524
R5.58G11A	1.134	1.22	2.201	2.201	1.421	2.201	2.201
R5.58G11del	-1.204	-0.132	-0.159	-0.701	-1.048	-0.151	-0.278
R5.58G6A/C10U	-2.327	-2.444	-1.18	-1.605	-1.851	-1.411	-0.559
R5.58G6U/C10A	-1.115	-1.394	-0.623	-0.624	-0.606	-0.709	0.121
R5.58G6U	-1.267	-1.6	-0.905	-0.941	-0.823	-1.017	-0.377
R5.58C10U	-0.816	-1.448	-0.533	-0.534	-0.232	-0.839	-0.19
R5.58C10A	-0.984	-1.774	-0.74	-0.651	-0.808	-1.088	-0.304
R5.58G6A		-1.688	-0.725	-1.163	-0.114	-1.044	0.417
R5.58C10del	-1.477	-1.825	-0.897	-1.064	-1.137	-1.026	-0.38
R5.58G7A	-0.87	-1.032	0.906	-0.711	0.516	-0.652	1.031
R5.58G7C	-0.373	-0.902	-0.414	-0.532	-0.174	-0.286	0.182
R5.58C9U	-0.986	-3.182	-0.955	-0.978	-0.821	-1.35	-0.389
R5.58A8U	-0.967	-1.517	-0.854	-0.875	-0.381	-0.984	-0.33
R5.58A8G	-0.112	-2.002	-0.309	-0.091	0.249	-0.614	0.265

-1.997

R5.58A8C	-0.425	-0.905	-0.335	-0.303	0.609	-0.419	0.489
R5.58rev	-1.825	-1.051	-0.061	-1.157	-0.2	-0.076	0.253
R5.58M7i1	-0.926	0.944	0.227	-1.079	-1.192	0.616	0.161
R5.58M24i1	-2.26	-4.667	-2.002	-1.767	-2.161	-4.667	-1.323
R5.58M27i1	-0.354	-1.045	-0.494	-0.375	-0.172	-0.62	0.008
R5.58M28i1	-0.909	-1.72	-0.59	-0.907	-0.551	-1.208	-0.135
R5.58c	-1.262	-2.804	-1.705	-1.217	0.411	-1.857	-0.795

Table 8.7 Selected Receptors and Mutants with ΔΔG Values
Reference for ΔΔG values is 4000 nM. Each value was calculated from the Kd in Table 8.6

Network Intermediates	Sequence
B7.8	GGGAAAGCGAUUUACUCU--AAG-GCCUAGUUCCGGG...GAAA...CUUGGUUCUAGGC-UACU--GGGGUAGGUCGUUUUUCUCU
S8	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UACU--GGGGUAGGUCGUUUUUCUCU
S8.2	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-CACU--GGGGUAGGUCGUUUUUCUCU
S8.5	GGGAAAGCGAUUUACUCU--AAC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UAGU--GGGGUAGGUCGUUUUUCUCU
S8.7	GGGAAAGCGAUUUACUCA--AAC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UAGU--UGGGUAGGUCGUUUUUCUCU
S8.8	GGGAAAGCGAUUUACUCA--ACG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UACG--UGGGUAGGUCGUUUUUCUCU
S8.9	GGGAAAGCGAUUUACUCA--ACG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UACG--UGGGUAGGUCGUUUUUCUCU
S8.11	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCU-UACU--GGGGUAGGUCGUUUUUCUCU
S8.12	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UGCU--GGGGUAGGUCGUUUUUCUCU
S8.13	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-GUCU--GGGGUAGGUCGUUUUUCUCU
S8.14	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AACU--GGGGUAGGUCGUUUUUCUCU
S8.15	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACCU--GGGGUAGGUCGUUUUUCUCU
S8.16	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UCCU--GGGGUAGGUCGUUUUUCUCU
S8.17	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AUCU--GGGGUAGGUCGUUUUUCUCU
S8.18	GGGAAAGCGAUUUACUCU--AAG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AGCU--GGGGUAGGUCGUUUUUCUCU
S8Tt4	GGGAAAGCGAUUUACUCU--GCG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-UACG--GGGGUAGGUCGUUUUUCUCU

R4.3c2	GGGAAAGCGAUUUACUCU--GCG-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCU-AGUG--UGGGUAGGUCGUUUUUCUCU
R5.4b	GGGAAAGCGAUUUACUCA-UACC-GGGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AUGG--UGGGUAGGUCGUUUUUCUCU
R5.4c	GGGAAAGCGAUUUACUCA-UACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AUGG--UGGGUAGGUCGUUUUUCUCU
R5.8d	GGGAAAGCGAUUUACUCA-UACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AUGG--UGGGUAGGUCGUUUUUCUCU
R5.8f	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-AUGG--UGGGUAGGUCGUUUUUCUCU
R5.12c	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGG--UGGGUAGGUCGUUUUUCUCU
R5.13b	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCA-UCGG--UGGGUAGGUCGUUUUUCUCU
R5.15c	GGGAAAGCGAUUUACUCA-AUCU--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-UCGG--UGGGUAGGUCGUUUUUCUCU
R5.15d	GGGAAAGCGAUUUACUCA-AUCU--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-UCG---UGGGUAGGUCGUUUUUCUCU
R5.15e	GGGAAAGCGAUUUACUCAGACCU--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-UCGG--UGGGUAGGUCGUUUUUCUCU
R5.15f	GGGAAAGCGAUUUACUCA--ACU--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCG-UCGG--UGGGUAGGUCGUUUUUCUCU
R5.16b	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGCG--UGGGUAGGUCGUUUUUCUCU
R5.16e	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGCGG-UGGGUAGGUCGUUUUUCUCU
R5.16f	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGCAG-UGGGUAGGUCGUUUUUCUCU
R5.16h	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGCAC-UGGGUAGGUCGUUUUUCUCU
R5.16i	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGGGC-UGGGUAGGUCGUUUUUCUCU
R5.16j	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UACGC-UGGGUAGGUCGUUUUUCUCU
R5.16L	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UACAC-UGGGUAGGUCGUUUUUCUCU
R5.16n	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGGAC-UGGGUAGGUCGUUUUUCUCU
R5.16o	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGGGG-UGGGUAGGUCGUUUUUCUCU
R5.16c2	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC-AUGCGC-UGGGUAGGUCGUUUUUCUCU
R5.21b	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--AUGG--UGGGUAGGUCGUUUUUCUCU
R5.21c	GGGAAAGCGAUUUACUCA-GACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--AUGG--UGGGUAGGUCGUUUUUCUCU
R5.21d	GGGAAAGCGAUUUACUCA-UACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--ACGG--UGGGUAGGUCGUUUUUCUCU
R5.33b	GGGAAAGCGAUUUACUCA-UAACC-GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--

	AUGG--UGGGUAGGUCGUUUUUCUCU
R5.34b	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--UGUG--UGGGUAGGUCGUUUUUCUCU
R5.34c	GGGAAAGCGAUUUACUCA-UACC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--AUGUG-UGGGUAGGUCGUUUUUCUCU
R5.34d	GGGAAAGCGAUUUACUCA-UAC---GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--AUGUG-UGGGUAGGUCGUUUUUCUCU
R5.34b2	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--AAGUG-UGGGUAGGUCGUUUUUCUCU
R5.34b3	GGGAAAGCGAUUUACUCA-UAAC--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--AUGCG-UGGGUAGGUCGUUUUUCUCU
R5.58b	GGGAAAGCGAUUUACUCA-ACCU--GGUAGUUCCGGG...GAAA...CUUGGUUCUACCC-ACGG--UGGGUAGGUCGUUUUUCUCU
R5.58c	GGGAAAGCGAUUUACUCA-ACCU--GGUAGUUCCGGG...GAAA...CUUGGUUCUACC--GACGG-UGGGUAGGUCGUUUUUCUCU

Table 8.8 Network intermediate mutants and sequences
Sequences for all point mutation receptors that connect all the receptors in the selected network.

Network Intermediates	**Kd (nM) GGAA**	**Kd (nM) GUAA**	**Kd (nM) GAGA**	**Kd (nM) GCAA**	**Kd (nM) GGGA**	**Kd (nM) GUGA**	**Kd (nM) GCGA**
B7.8	404	31	3356	906	1730	110	9314
S8	138	19	786	228	376	81	1462
S8.2	123	209	1389	887	223	736	3006
S8.5	200	41	866	543	360	385	3466
S8.7	217	60	896	210	578	245	1365
S8.8	331	73	1697	1027	752	467	4550
S8.9	663	111	1650	1611	2593	811	8606
S8.11	1747	21	395	1459	2512	1794	1377
S8.12	3622	749	3415	6829	6174	8583	200000
S8.13	226	117	467	232	416	177	503
S8.14	54	7	311	106	74	28	352
S8.15	513	26	212	280	898	125	788

S8.16	9669	38	314	201	10913	205	9663
S8.17	316	175	1142	347	981	228	479
S8.18	1264	357	1671	2000	720	462	720
S8Tt4	44	5	223	80	80	34	500
R4.3c2	7513	13510	50000	40000	50000	200000	200000
R5.4b	1	1118	5000	5000	198	7500	35000
R5.4c	3979	100000	100000	15000	2509	50000	50000
R5.8d	186	3	3979	1165	547	145	5365
R5.8f	846	12500	14750	1346	784	3134	5271
R5.12c	14600	1362	43000	200000	160000	36000	71000
R5.13b	2583	6121	10508	40000	100000	15000	40000
R5.15c	1190	25	312	726	20000	187	1361
R5.15d	30000	7250	30000	200000	200000	20000	50000
R5.15e	440	1	110	944	4130	286	640
R5.15f	611	116	272	509	1774	302	1026
R5.16b	200000	100000	200000	200000	200000	27000	50000
R5.16e	30500	277	200000	46600	57000	3448	30500
R5.16f	41000	4162	66000	200000	200000	43800	41000
R5.16h	27000	1117	46000	38000	49000	16000	29500
R5.16i	31000	151	50000	25000	43000	6300	38000
R5.16j	2295	2052	12000	4388	4872	12000	11200
R5.16L	9975	1632	15000	12600	12600	20500	17000
R5.16n	27000	5145	110000	57000	91000	57000	110000
R5.16o	5911	282	22500	12500	14000	6537	17000
R5.16c2	2612	13147	50000	11486	6064	30000	200000
R5.21b	50000	56785	55000	200000	200000	50000	50000
R5.21c	168	50000	150000	200000	200000	50000	50000

R5.21d	21600	4605	40500	48500	28300	41000	41000
R5.33b	1050	28300	200000	40500	27100	95000	120000
R5.34b	60000	60000	60000	200000	200000	60000	70000
R5.34c	100000	100000	200000	200000	25000	100000	100000
R5.34d	20000	25000	50000	200000	200000	50000	100000
R5.34b2	404	17000	175000	30000	10000	200000	200000
R5.34b3	200000	50000	100000	200000	200000	50000	50000
R5.58b	277	382	584	1517	1679	1189	5551
R5.58c	425	27	193	460	8302	147	973

Table 8.9 K_ds for Selected Receptor Network intermediates
K_ds were calculated from native gel shift assays.

Receptor	**ΔΔG GGAA**	**ΔΔG GUAA**	**ΔΔG GAGA**	**ΔΔG GCAA**	**ΔΔG GGGA**	**ΔΔG GUGA**	**ΔΔG GCGA**
B7.8	-1.29	-2.73	-0.1	-0.83	-0.47	-2.02	0.48
S8	-1.89	-3.01	-0.91	-1.61	-1.33	-2.19	-0.57
S8.2	-1.96	-1.66	-0.59	-0.85	-1.62	-0.95	-0.16
S8.5	-1.68	-2.57	-0.86	-1.12	-1.35	-1.32	-0.08
S8.7	-1.64	-2.36	-0.84	-1.66	-1.09	-1.57	-0.6
S8.8	-1.4	-2.25	-0.48	-0.76	-0.94	-1.21	0.07
S8.9	-1.01	-2.01	-0.5	-0.51	-0.24	-0.9	0.43
S8.11	-0.47	-2.95	-1.3	-0.57	-0.26	-0.45	-0.6
S8.12	-0.06	-0.94	-0.09	0.3	0.24	0.43	2.2
S8.13	-1.62	-1.99	-1.21	-1.6	-1.27	-1.75	-1.17
S8.14	-2.42	-3.57	-1.44	-2.04	-2.24	-2.79	-1.37
S8.15	-1.15	-2.83	-1.65	-1.49	-0.84	-1.95	-0.91
S8.16	0.5	-2.62	-1.43	-1.68	0.57	-1.67	0.5
S8.17	-1.43	-1.76	-0.7	-1.37	-0.79	-1.61	-1.19
S8.18	-0.65	-1.36	-0.49	-0.39	-0.96	-1.21	-0.96

S8Tt4	-2.54	-3.76	-1.62	-2.2	-2.2	-2.68	-1.17
R4.3c2	0.355	0.685	1.421	1.296	1.421	2.201	2.201
R5.4b	-4.667	-0.717	0.126	0.126	-1.691	0.354	1.22
R5.4c	-0.003	1.811	1.811	0.744	-0.262	1.421	1.421
R5.8d	-1.726	-4.108	-0.003	-0.694	-1.12	-1.867	0.165
R5.8f	-0.874	0.641	0.734	-0.613	-0.917	-0.137	0.155
R5.12c	0.729	-0.606	1.336	2.201	2.076	1.236	1.618
R5.13b	-0.246	0.239	0.543	1.296	1.811	0.744	1.296
R5.15c	-0.682	-2.867	-1.435	-0.96	0.906	-1.723	-0.607
R5.15d	1.134	0.335	1.134	2.201	2.201	0.906	1.421
R5.15e	-1.242	-4.667	-2.022	-0.812	0.018	-1.484	-1.031
R5.15f	-1.057	-1.992	-1.513	-1.16	-0.457	-1.454	-0.766
R5.16b	2.201	1.811	2.201	2.201	2.201	1.074	1.421
R5.16e	1.143	-1.502	2.201	1.382	1.495	-0.084	1.143
R5.16f	1.31	0.022	1.577	2.201	2.201	1.347	1.31
R5.16h	1.074	-0.718	1.374	1.267	1.41	0.78	1.124
R5.16i	1.152	-1.844	1.421	1.031	1.336	0.256	1.267
R5.16j	-0.313	-0.376	0.618	0.052	0.111	0.618	0.579
R5.16L	0.514	-0.504	0.744	0.646	0.646	0.919	0.814
R5.16n	1.074	0.142	1.865	1.495	1.758	1.495	1.865
R5.16o	0.22	-1.492	0.972	0.641	0.705	0.276	0.814
R5.16c2	-0.24	0.67	1.421	0.594	0.234	1.134	2.201
R5.21b	1.421	1.493	1.475	2.201	2.201	1.421	1.421
R5.21c	-1.783	1.421	2.039	2.201	2.201	1.421	1.421
R5.21d	0.949	0.079	1.303	1.404	1.101	1.31	1.31
R5.33b	-0.753	1.101	2.201	1.303	1.077	1.782	1.914
R5.34b	1.524	1.524	1.524	2.201	2.201	1.524	1.61

R5.34c	1.811	1.811	2.201	2.201	1.031	1.811	1.811
R5.34d	0.906	1.031	1.421	2.201	2.201	1.421	1.811
R5.34b2	-1.29	0.814	2.126	1.134	0.516	2.201	2.201
R5.34b3	2.201	1.421	1.811	2.201	2.201	1.421	1.421
R5.58b	-1.503	-1.322	-1.083	-0.546	-0.488	-0.682	0.184
R5.58c	-1.262	-2.804	-1.705	-1.217	0.411	-1.857	-0.795

Table 8.10 **ΔΔG values for Selected Receptor Network Intermediates**
Reference for ΔΔG values is 4000 nM. Each value was calculated from the Kd in Table 8.9

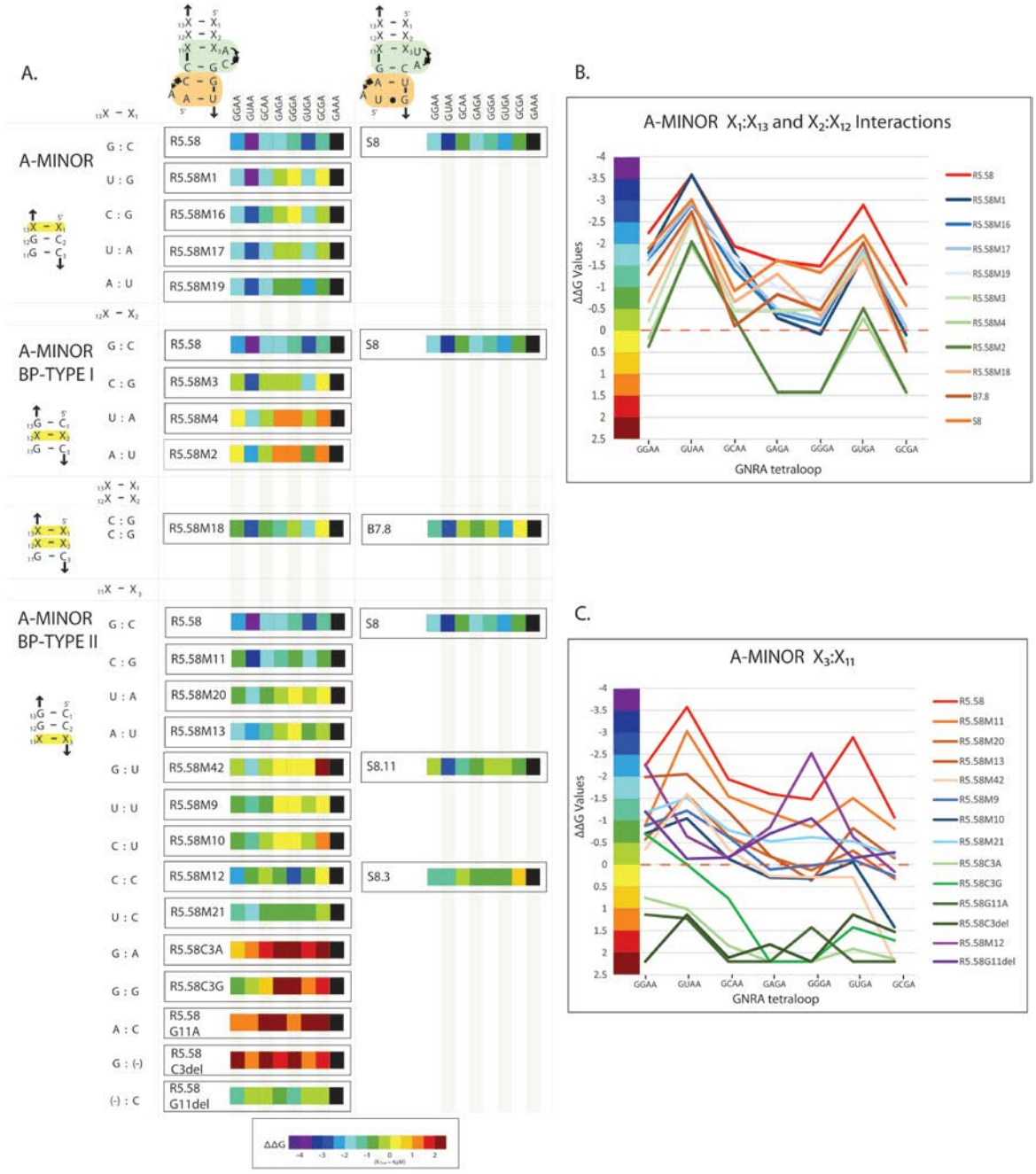


Figure 8.9 A-minor Mutational Analysis

(A) Mutations in the base pair stem of the A-minor (X_{13}:X_1, X_{12}:X_2 and X_{11}:X_3) were tested for optimization. Base pairs in the A-minor (X_{13}:X_1, X_{12}:X_2) were mutated to maintain Watson Crick interactions for the integrity of the stem. In position X_{11}:X_3, mismatches were tested to understand the effects of this on the A-minor interaction. R5.58 mutations were made, as well as some S8 mutations to see if the effects in both contexts was comparable. Bar codes were used as a measurement for comparability. (B) Specificity profiles for mutations made within the A-minor stem (X_{13}:X_1, X_{12}:X_2). (C) Specificity profiles for mutations made within the A-minor stem (X_{11}:X_3).

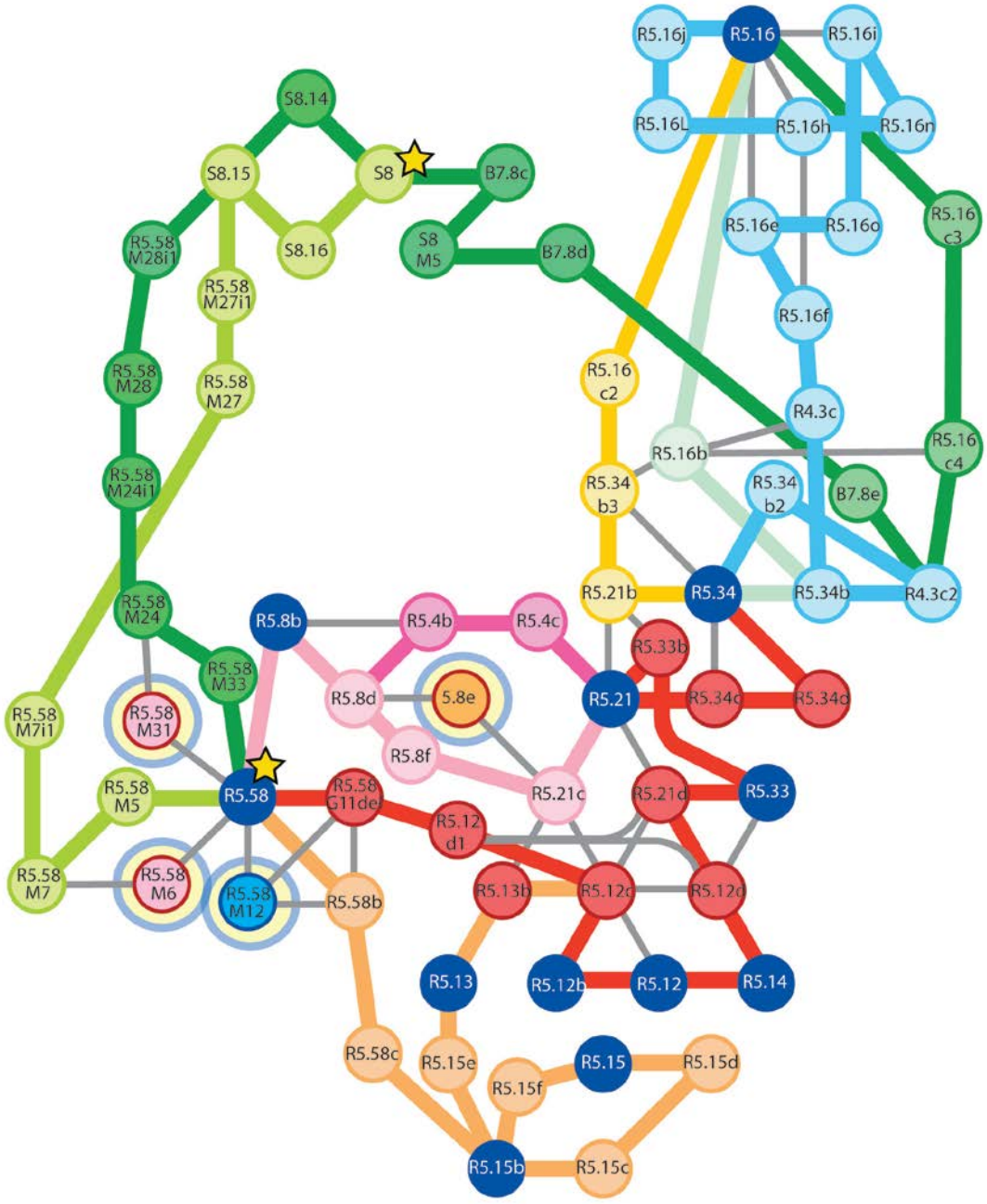


Figure 8.10 Selected Receptor Point Mutation Network
Point mutation network connecting R5.58, S8 and all other GUAA receptors from our selection. Color coded pathways denote different receptor classes (light green=S8-like, light pink=R5a class, red=R5b class, orange=R5c class, blue=R5d, other colors are representative of similar receptors with comparable phenotypic behavior). Receptors that are circled were not tested. All Kds and ΔΔG for receptors can be found in Table 8.9 and Table 8.10. Graphical representation of network can be found in Figure 3.7

8.3 Supplemental Information for Chapter 4

Name	Molecular modularity	Sequence alignment
Probes		(((((((..(((((..(((((....)))))..))))).)))))))).
pGGAA	GGAA_probe	gggauuau_ggauguuccgggggaacuugguucauccuaaguccuuu
pGAAA	GAAA_probe	gggauaucuggauguuccgggggaaacuugguucauccugugguccuuu
	Probes for co-tx assembly	((((((((((((((((((..(((((..(((((....)))))..))))).)))))))))))))))))).
pGGAA2	GGAA_probe_extended-helix	gggaaagcgauuuacucuau_ggauguuccgggggaacuugguucauccuaagggguaggucguuuuucucu
pGAAA2	GAAA_probe_extended-helix	gggaaagcgauuuacucaucuggauguuccgggggaaacuugguucauccugugggguaggucguuuuucucu
Dimer-forming controls		(((((((((.....(((((..(((((....)))))..))))).....))))))))).
HD_1	11nt_AA	gccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaauggcu
HD_4	11nt_AC	ggaauuacuauau_ggauguuccgggggaauuugguucauccuacguuguaauuccu
HD_7	11nt_GU	gccauuacuauau_ggauguuccgggggaauuugguucauccugguuguaauggcu
HD_14	11nt_A/CC	ggaauuacuacac_ggauguuccgggggaauuugguucauccuaaauuguaauuccu
HD_1"	(11nt_AA)c	gccguuacuauau_ggauguuccgggggaacuugguucauccuaaguuguaauggcu
HD_7"	(11nt_GU)c	gccguuacuauau_ggauguuccgggggaacuugguucauccugguuguaauggcu
HD_16	R1	gccguuacuaaucuggauguuccgggggaaauuugguucauccugguuguaauggcu
HD_16"	(R1)c	gccguuacuaaucuggauguuccgggggaaacuugguucauccugguuguaauggcu
Controls with helices		.....((((((((((((((....))))))))))))))....(((((((((.....(((((..(((((....)))))..))))).....)))))))))).
HD_1b	(11nt_AA):(helix)	gggaacgcaacuuacaacuucgguuguaaguugcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
HD_7b	(11nt_GU):(helix)	gggaacgcaacacacaacuucgguuguguuugcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
		Attenuators with gGRAAu
	3′ PKL attenuators	.....(((.....((((.....)))).....)))....(((((((((.....(((((..(((((....)))))..))))).....)))))))))).
1	(11nt_AA):(3`KL_AA)	gggaacgcaacuuacaacuucgguugaaa___gcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
1a	(11nt_AA):(3`KL_AA-5A)	Gggaacgcaacuuacaacuucgguugaaaaa_gcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
2	(11nt_AA):(3`KL_AC)	gggaaggcaacguacaacuucgguugaaa___gccaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
3	(11nt_AA):(3`KL_GU)	gggaacgcaacacaacaacuucgguugaaa___gcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
4	(11nt_AC):(3`KL_AC)	gggaaggcaacguacaacuucgguugaaa___gccaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuacguuguaaugcu
5	(11nt_AC):(3`KL_AA)	gggaacgcaacuuacaacuucgguugaaa___gcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuacguuguaaugcu
6	(11nt_AC):(3`KL_GU)	gggaacgcaacacaacaacuucgguugaaa___gcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuacguuguaaugcu
7	(11nt_GU):(3`KL_GU)	gggaacgcaacacaacaacuucgguugaaa___gcgaaaugccguuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
7a	(11nt_GU):(3`KL_GU-5A)	Gggaacgcaacacaacaacuucgguugaaaaa_gcgaaaugccguuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
8	(11nt_GU):(3`KL_AC)	gggaaggcaacguacaacuucgguugaaa___gcCaaaugccauuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
9	(11nt_GU):(3`KL_AA)	gggaacgcaacuuacaacuucgguugaaa___gcGaaaugccauuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
16	(R1):(3`KL_GU)	gggaacgcaacacaacaacuucgguugaaa___gcgaaaugccguuacuaaucuggauguuccgggggaauuugguucauccugguuguaaugcu
16a	(R1):(3`KL_GU-5A)	Gggaacgcaacacaacaacuucgguugaaaaa_gcgaaaugccguuacuaaucuggauguuccgggggaauuugguucauccugguuguaaugcu
	5′ PKL attenuators	.....(((.....((((....)))).....)))....(((((((((.....(((((..(((((....)))))..))))).....)))))))))).
10	(11nt_AA):(5`KL)	gggaacgcaaa___caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
10a	(11nt_AA):(5`KL-5A)	Gggaacgcaaaaa_caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
11	(11nt_AA):(5`KL_A/CC)	gggaacccaaa___caacuucgguuggugua_gggaaaugccauuacuauau_ggauguuccgggggaauuugguucauccuaaguuguaaugcu
12	(11nt_GU):(5`KL-)	gggaacgcaaa___caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
12a	(11nt_GU):(5`KL-5A)	Gggaacgcaaaaa_caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
13	(11nt_GU):(5`KL_A/CC)	gggaacgcaaa___caacuucgguuggugua_gcgaaaugccauuacuauau_ggauguuccgggggaauuugguucauccugguuguaaugcu
14	(11nt_A/CC):(5`KL_A/CC)	gggaacgcaaa___caacuucgguuggugua_gcgaaaugccauuacuacac_ggauguuccgggggaauuugguucauccuaaauuguaaugcu
15	(11nt_A/CC):(5	gggaacgcaaa___caacuucgguugauaua_gcgaaaugccauuacuacac_ggauguuccgggggaauuugguucauccuaaauuguaaug

	`KL)	gcu
17	(R1):(5`KL_R1)	gggaacgcaaa___caacuucgguugagauuagcgaaaugccguuacuaaucuggauguuccggggaaauuuggguucauccuguguuguaaaug gcu
17a	(R1):(5`KL_R1-5A)	Gggaacgcaaaaa_caacuucgguugagauuagcgaaaugccguuacuaaucuggauguuccggggaaauuuggguucauccuguguuguaaaug gcu
		Attenuators with gGRAAc
	3′ PKL attenuators	.....(((......((((....)))).....))).....(((((((((.....(((((..(((((....)))))..)))))).....)))))))).
1"	(11nt_AA):(3`KL_AA)c	gggaacgcaacuuacaacuucgguugaaa___gcgaaaugccauuacuauau_ggauguuccggggaacuuggguucauccuaaguuguaaaug gcu
1a"	(11nt_AA):(3`KL_AA-5A)c	Gggaacgcaacuuacaacuucgguugaaaaa_gcgaaaugccauuacuauau_ggauguuccggggaacuuggguucauccuaaguuguaaaug gcu
7"	(11nt_GU):(3`KL_GU)c	gggaacgcaacacacaacuucgguugaaa___gcgaaaugccguuacuauau_ggauguuccggggaacuuggguucauccuguguuguaaaug gcu
7a"	(11nt_GU):(3`KL_GU-5A)c	Gggaacgcaacacacaacuucgguugaaaaa_gcgaaaugccguuacuauau_ggauguuccggggaauuuggguucauccuguguuguaaaug gcu
16"	(R1):(3`KL_GU)c	gggaacgcaacacacaacuucgguugaaa___gcgaaaugccguuacuaaucuggauguuccggggaacuuggguucauccuguguuguaaaug gcu
16a"	(R1):(3`KL_GU-5A)c	Gggaacgcaacacacaacuucgguugaaaaa_gcgaaaugccguuacuaaucuggauguuccggggaacuuggguucauccuguguuguaaaug gcu
	5′ PKL attenuators	.....(((......((((....)))).....))).....(((((((((.....(((((..(((((....)))))..)))))).....)))))))).
10"	(11nt_AA):(5`KL)c	gggaacgcaaa___caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccggggaacuuggguucauccuaaguuguaaaug gcu
10a"	(11nt_AA):(5`KL-5A)c	Gggaacgcaaaaa_caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccggggaacuuggguucauccuaaguuguaaaug gcu
12"	(11nt_GU):(5`KL)c	gggaacgcaaa___caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccggggaacuuggguucauccuguguuguaaaug gcu
12a"	(11nt_GU):(5`KL-5A)c	Gggaacgcaaaaa_caacuucgguugauaua_gcgaaaugccguuacuauau_ggauguuccggggaacuuggguucauccuguguuguaaaug gcu
17"	(R1):(5`KL_R1)c	gggaacgcaaa___caacuucgguugagauuagcgaaaugccguuacuaaucuggauguuccggggaacuuggguucauccuguguuguaaaug gcu
17a"	(R1):(5`KL_R1-5A)c	Gggaacgcaaaaa_caacuucgguugagauuagcgaaaugccguuacuaaucuggauguuccggggaacuuggguucauccuguguuguaaaug gcu
		Switches (SWI & SWII)
18	(11nt_GU):(3`KL_GU-P)	Gggaacgcaacacacaacuucgguuggccgaagcgaaaugccguuacuauau-ggauguuccgggggaauuuggguucauccuguguuguaauggcu
SW_I	SW(I)	gggaaguucggcucuguguuaagucug
SW_II	SW(II)	gggaaguguguuucuucggcaagucuc

Table 8.11 List of sequences designed and used in this work.

Each tectoRNA attenuator comprises a dimer-forming (HD) and attenuator modules. Their modularity is indicated as follow: (receptor_name):(PKL_name). Molecule name followed by “a” indicates when the PK-forming internal loop has 5 As (instead 3 as depicted Figure 5.2 Flow of information within the central dogma of a cell.B). At the level of the modular description, it is indicated as (PKL_name-5A). Molecule name followed by “ indicates when the GNRA loop of the dimer-forming module is gGRAAc (instead gGRAAu): the modular description is followed by a ”c“. For example, molecule 1a” or (11nt_AA):(3`KL_AA-5A)c corresponds to the attenuator with a PK-forming loop containing 5 As opposite to the KL sequence, which pairs to the 3` side of the 11nt_AA receptor. It also contains a gGAAAc loop. Pink and light blue nucleotides (nts) correspond to the R1 and 11nt variant receptors, respectively. Green nts from the attenuator module are involved in PK formation.

Name	Receptor	PKL	PK pairing stability and PK prediction	Kd (nM)	ΔG (kcal/mol)	ΔΔG (kcal/mol)
Controls						**$\Delta\Delta G_{HD}$ = ΔG(HD_x)-ΔG(HD_1)**
HD_1	11nt_AA	None	-	198±34	-8.68	**-**
HD_4	11nt_AC	None	-	259±50	-8.53	**0.15±0.05**
HD_7	11nt_GU	None	-	252±29	-8.54	**0.14±0.05**
HD_14	11nt_A/CC	None	-	425±23	-8.25	**0.43±0.10**
HD_16	R1	None	-	112±29	-9	**-0.32±0.05**
HD_1b	11nt_AA	helix	-	339±71	-8.38	**0.3±0.12**
HD_7b	11nt_GU	helix	-	379±63	-8.31	**0.23±0.1**
3' PK attenuators						**$\Delta\Delta G_{AT}$ = ΔG(x)-ΔG(HD_x)**
1	11nt_AA	3`KL_AA	-3.3(yes)	221±19	-8.62	**0.06±0.03**
2	11nt_AA	3`KL_AC	-(no)	323±31	-8.44	**0.24±0.1**
3	11nt_AA	3`KL_GU	-(no)	240±44	-8.58	**0.1±0.05**
4	11nt_AC	3`KL_AC	-4.8(yes)	1356±263	-7.53	**1±0.06**
5	11nt_AC	3`KL_AA	-(no)	250±22	-8.55	**-0.02±0.01**
6	11nt_AC	3`KL_GU	-(no)	385±33	-8.31	**0.22±0.04**
7	11nt_GU	3`KL_GU	-5.3(yes)	>80000	(-5.3)<	**>3.24**
8	11nt_GU	3`KL_AC	-(no)	479±25	-8.18	**0.36±0.01**
9	11nt_GU	3`KL_AA	-(no)	359±72	-8.34	**0.2±0.03**
16	R1	3`KL_GU	-5.3(yes)	5276±219	-6.83	**2.17±0.1**
1_a	11nt_AA	3`KL_AA -5A	-3.3(no)	485±128	-8.18	**0.5±0.03**
7_a	11nt_GU	3`KL_GU -5A	-5.3(yes)	>80000	(-5.3)<	**>3.24**
16_a	R1	3`KL_GU -5A	-5.3(yes)	5321±517	-6.83	**2.17±0.11**
5' PK attenuators						**$\Delta\Delta G_{AT}$ = ΔG(x)-ΔG(HD_x)**
10	11nt_AA	5`KL	-0.6(no)	110±18	-9.01	**-0.33±0.04**
11	11nt_AA	5`KL_A/CC	-(no)	134±22	-8.9	**-0.22±0.01**
12	11nt_GU	5`KL	-0.6(no)	298±39	-8.45	**0.09±0.01**
13	11nt_GU	5`KL_A/CC	-(no)	366±43	-8.33	**0.21±0.05**
14	11nt_A/CC	5`KL_A/CC	-3.7(yes)	1772±135	-7.45	**0.8±0.05**
15	11nt_A/CC	5`KL	-(no)	568±89	-8.09	**0.16±0.03**
17	R1	5`KL_R1	-3.5(yes)	3341±379	-7.09	**1.91±0.12**
10_a	11nt_AA	5`KL-5A	-0.6(no)	631±280	-8.03	**0.65±0.06**
12_a	11nt_GU	5`KL-5A	-0.6(no)	642±169	-8.02	**0.52±0.03**
17_a	R1	5`KL_R1 -5A	-3.5(yes)	7854±646	-6.61	**2.39±0.16**

Table 8.12 Dissociation constants (Kd's), free energies (ΔG) and difference in free energies (ΔΔG).
Dissociation constants (Kd's), free energies (ΔG) and difference in free energies (ΔΔG) determined for attenuators with gGRAAu tetraloops at 15 mM Mg(OAc)2 and 10°C. The anticipated secondary structure for each molecule is also the one predicted to be the most stable by Mfold [52]. We have indicated the calculated free energy change of PK pairing formation (in kcal/mol at 37°C) according to Freier and colleagues [166]: between brackets, is indicated whether the PK_forming conformer is predicted (yes) or not (no) by KineFold [277] or/and Cylofold [144, 279].

Control name	Receptor	GRAA with gu or gc closing bp	Kd (nM)	ΔG (kcal/mol)	$\Delta\Delta G_{HD}$ (kcal/mol) ΔG(HD_x)- ΔG(HD_x")
HD_1 HD_1"	11nt_AA	gu gc	198±34 23±3	-8.7 -9.9	- **1.2±0.05**
HD_7 HD_7"	11nt_GU	gu gc	252±29 75±13	-8.54 -9.23	- **0.69±0.1**
HD_16 HD_16"	R1	gu gc	112±29 49±9	-9 -9.47	- **0.47±0.09**

Table 8.13 Dissociation constants (K_d`s), free energies (ΔG) and difference in free energies (ΔΔG) determined for heterodimer.
Heterodimer assembly with HD forming modules comprising gGRAAu (gu) or gGRAAc (gc) tetraloops, at 15 mM Mg(OAc)2 and 10°C (see in Figure 5.2 Flow of information within the central dogma of a cell.B).

Name	Receptor	PKL	Kd (nM)	ΔG (kcal/mol)	ΔΔG (kcal/mol)
15 mM Mg2+					
controls					$\Delta\Delta G_{HD}$ = ΔG(HD_x")- ΔG(HD_1")
HD_1"	11nt_AA	None	23±3	-9.9	-
HD_7"	11nt_GU	None	75±13	-9.23	**0.67±0.05**
HD_16"	R1	None	49±9	-9.47	**0.43±0.05**
3' PK attenuators					$\Delta\Delta G_{AT}$ = ΔG(x")-ΔG(HD_x")
1"	11nt_AA	3`KL_AA	21±3	-9.94	**-0.05±0.08**
7"	11nt_GU	3`KL_GU	2480±424	-7.26	**1.96±0.1**
16"	R1	3`KL_GU	152±23	-9.09	**0.65±0.1**
1_a"	11nt_AA	3`KL_AA-5A	56±13	-9.39	**0.49±0.16**
7_a"	11nt_GU	3`KL_GU-5A	3951±771	-7	**2.03±0.03**
16_a"	R1	3`KL_GU-5A	96±12	-8.91	**0.38±0.22**
5' PK attenuators					$\Delta\Delta G_{AT}$ = ΔG(x")-ΔG(HD_x")
10"	11nt_AA	5`KL_AA	12±3	-10.28	**-0.38±0.14**
12"	11nt_GU	5`KL_GU	99±6	-9.07	**-0.16±0.03**
17"	R1	5`KL_R1	95±6	-8.83	**0.38±0.04**
10_a"	11nt_AA	5`KL_AA-5A	49±5	-9.46	**-0.44±0.06**
12_a"	11nt_GU	5`KL_GU-5A	53±5	-9.42	**-0.11±0.05**
17_a"	R1	5`KL_R1-5A	131±	-9.06	**0.41±0.04**
2 mM Mg2+					
controls					ΔG(HD_x"@2mM)- ΔG(HD_x"@15mM)
HD_1"	11nt_AA	None	771±89	-7.91	**-1.99±0.1**
HD_7"	11nt_GU	None	1240±168	-7.65	**-1.58±0.04**
HD_16"	R1	None	995±161	-7.77	**-1.7±0.07**
3' PK attenuators					$\Delta\Delta G_{AT}$ = ΔG(x")-ΔG(HD_x")
1"	11nt_AA	3`KL_AA	1487±198	-7.55	**0.36±0.04**
7"	11nt_GU	3`KL_GU	>100000	(-5.3)<	**>3.2**
16"	R1	3`KL_GU	6797±936	-6.84	**0.93±0.04**
1_a"	11nt_AA	3`KL_AA-5A	843±170	-7.86	**0.05±0.01**
7_a"	11nt_GU	3`KL_GU-5A	>100000	(-5.3)<	**>3.2**
16_a"	R1	3`KL_GU-5A	6055±870	-6.7	**1.01±0.07**
5' PK attenuators					$\Delta\Delta G_{AT}$ = ΔG(x")-ΔG(HD_x")
10"	11nt_AA	5`KL_AA	1577±107	-7.51	**0.4±0.4**
12"	11nt_GU	5`KL_GU	4203±830	-6.96	**0.68±0.11**
17"	R1	5`KL_R1	6742±824	-6.69	**1.08±0.07**
10_a"	11nt_AA	5`KL_AA-5A	2062±211	-7.36	**0.55±0.06**
12_a"	11nt_GU	5`KL_GU-5A	2430±495	-7.27	**0.37±0.12**
17_a"	R1	5`KL_R1-5A	5198±823	-6.76	**1.02±0.09**

Table 8.14 Dissociation constants (Kd`s), free energies (ΔG) and difference in free energies (ΔΔG) determined for attenuators.
Attenuators with gGRAAc tetraloops at 15 mM and 2 mM Mg(OAc)2 and 10°C.

A.

Type of tetraloops	Number of GNRA loops	Number of GNRA/11nt(or 11nt-like) receptor interactions	Number of GNRA/helix receptor interactions
GYRA Loops	***177 (35.5%)***	***7 (1.4 %)***	***170 (34.1%)***
GUAA	36 (7%)	5 (1.0%)	31 (6.2%)
GUGA	41 (8%)	0 (0%)	41 (8.2%)
GCAA	18 (4.5%)	0 (0%)	18 (3.6%)
GCGA	82 (16%)	2 (0.4%)	80 (16.1%)
GRRA Loops	***309 (62%)***	***287 (57.6%)***	***22 (4.4%)***
GAAA	297 (61%)	281 (56.4%)	16 (3.2%)
GAGA	7 (2%)	5 (1.0%)	2 (0.4%)
GGAA	2 (0.4%)	1 (0.2%)	1 (0.2%)
GGGA	3 (0.6%)	0 (0%)	3 (0.6%)
Other Loops	***12 (2.5%)***	***12 (2.4%)***	***0 (0%)***
AAAA	2 (0.4%)	2 (0.4%)	0 (0%)
UAAA	9 (1.9%)	9 (1.8%)	0 (0%)
GGAC	1 (0.2%)	1 (0.2%)	0 (0%)
TOTAL	**498 (100%)**	**306 (61%)**	**192 (39%)**

B.

	11nt/11nt-like receptors
G/C content (not including A-minor interaction)	
0 G or C	12%
1 G or C	52%
2 G or C	24%
3 G or C	11%
4 G or C	1.3%
Platform sequence	
AA	71%
GU	8.5%
AC	7.5%
UC	5.2%
GA	2.3%
GC	0.65%
Unclassified	5.2%

Table 8.15 Sequence analysis of GNRA/receptor interactions.
Sequence analysis of GNRA/receptor interactions in class I cyclic di-GMP riboswitches based on the alignment performed by Sudarsan and colleagues [157]. A) Distribution of GNRA/helix and GNRA/11nt (or 11nt-like) receptor interactions: the values (calculated percentages) are based on a total of 498 sequences. B) Analysis of the G/C content and di-nucleotide platform sequence (see also Figure 5.2 based on 306 class I cyclic di-GMP riboswitch sequences that contain an 11nt or 11nt-like receptor.

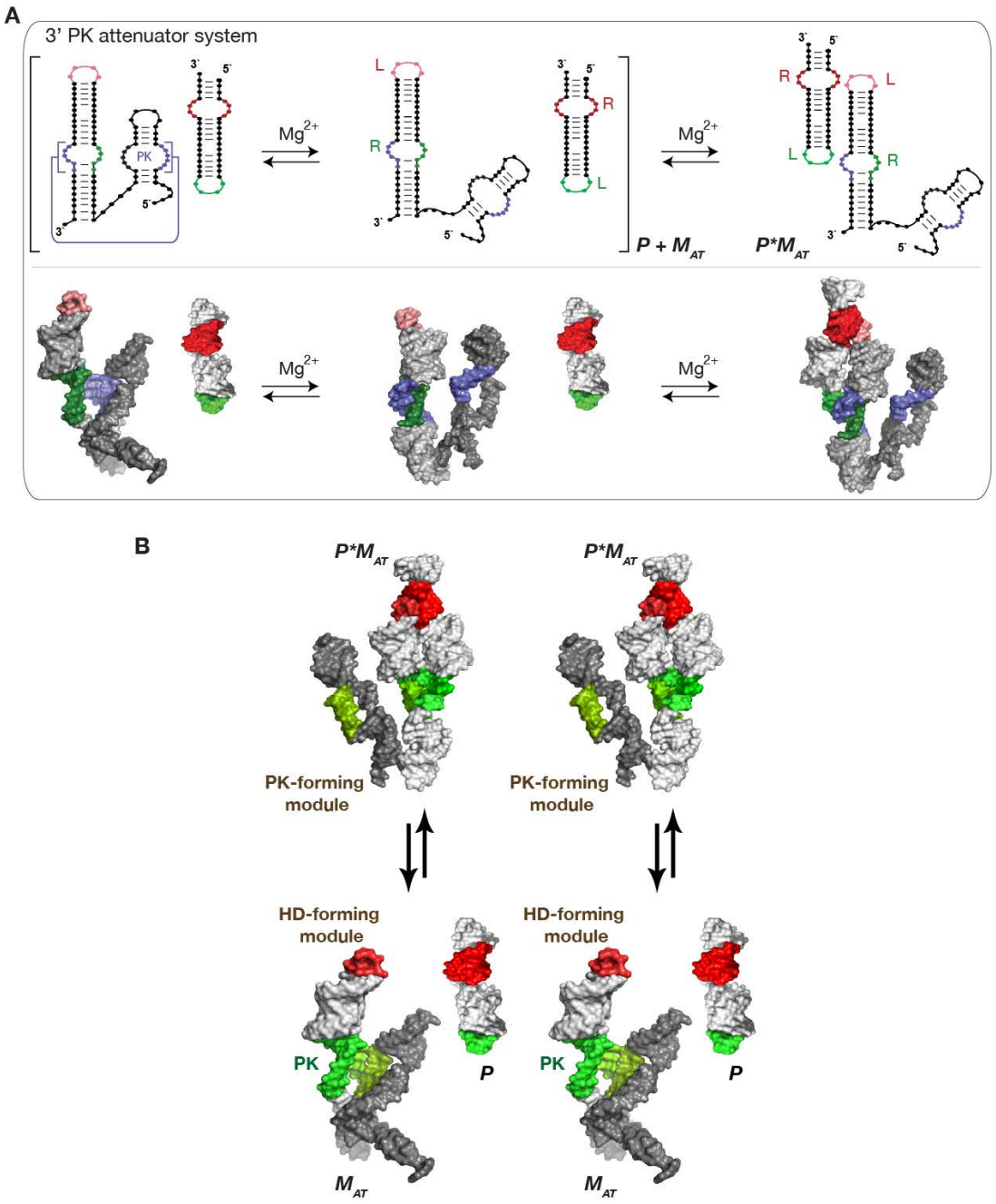


Figure 8.11 Self-assembly equilibrium reactions for the 3' PK attenuator system.

(A) 3' PK attenuator system: the tectoRNA attenuator (MAT), consisting of a HD-forming module linked to a PK-forming module, can assemble with a probe (P) through their HD-forming module (reaction on the right) to form the heterodimer (P*MAT). Attenuation of intermolecular self-assembly between the tectoRNA attenuator and the probe occurs when the PK-forming module interacts with the 3' side (in blue) of the receptor of the HD-forming module to form a 3' PK

(equilibrium reaction between brackets). Interacting receptor (R) and loop (L) motifs as well as pseudoknot (PK) are indicated. (B) 3D stereo view scheme of the self-assembly equilibrium reaction for the tectoRNA 3'PK attenuator system. The formation of a PK between the PK-forming module and the 3' side of the receptor from the HD-forming module can prevent assembly of the HD-forming module with the GRAA probe.

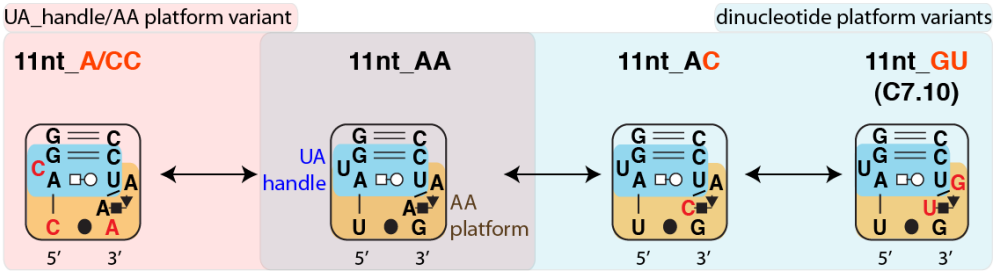


Figure 8.12 Structural modularity of the 11nt receptor variants.
These receptors are formed of two conserved G:C bps that form A minor interactions with the two last As from GAAA tetraloops (not shown) [153]. The second module is a UA_handle motif (in blue) that recognizes the second A from the GAAA [144, 279]. The U in bulge can theoretically be any nucleotide. The third module is a dinucleotide platform that stabilizes the second A of the GAAA through stacking interactions [135]. While not as abundant than AA platforms, the AC and GU platform variants are both found in nature, with the AC platform being more abundant than the GU one in group I and group II introns [73, 94]. The G:U closing base pair can sometimes be substituted by the isosteric C:A(+) bp [73, 94]. Base-pairings are indicated according to Leontis and Westhof annotation [9].

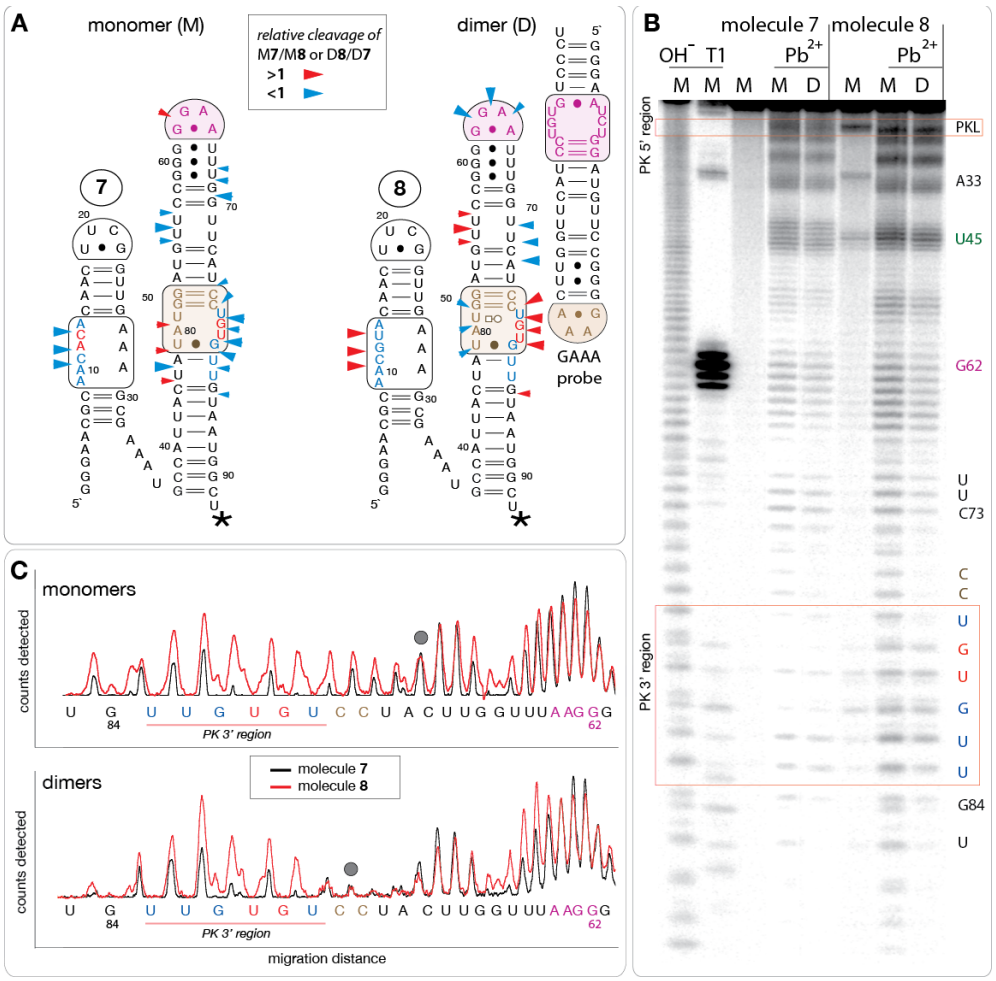


Figure 8.13 Lead(II)-induced cleavage patterns for tectoRNA attenuators 7 and 8. Lead(II)-induced cleavage patterns for tectoRNA attenuators 7 and 8 in their monomeric and heterodimeric states. (A) 2D diagrams of tectoRNA attenuators with reported differential Pb(II) cleavage patterns in the monomeric (M) and heterodimeric (D) states. Phosphate positions in monomer 7 (M7) that show enhanced or reduced Pb(II) cleavage with respect to monomer 8 (M8) are indicated by red or blue arrows on the 2D diagram of 7, respectively. Phosphate positions in heterodimer 8 (D8) that show enhanced or reduced Pb(II) cleavage with respect to heterodimer 7 (D7) are indicated by red or blue arrows on the 2D diagram of 8, respectively. The size of the arrows is roughly proportionate to the difference in cleavage for M7 versus M8 or D8 versus D7. A star indicates the radiolabeled RNA 3' end. (B) Pb(II) cleavage patterns of 32P radiolabeled molecules 7 and 8 either alone or bound to their non-radioactive cognate GAAA probe (as shown in (A)). M and D correspond to monomer and dimer lanes, respectively. Cleavage experiments (indicated by Pb2+) were carried out as described in the Materials and Methods section; OH- indicates alkaline hydrolysis ladder; T1 indicates RNaseT1 digestion.

(C) Superposed lead cleavage profiles for monomers 7 and 8 (top) and for the corresponding heterodimers in presence of GAAA probe (bottom). Black dots indicate positions used for normalization.

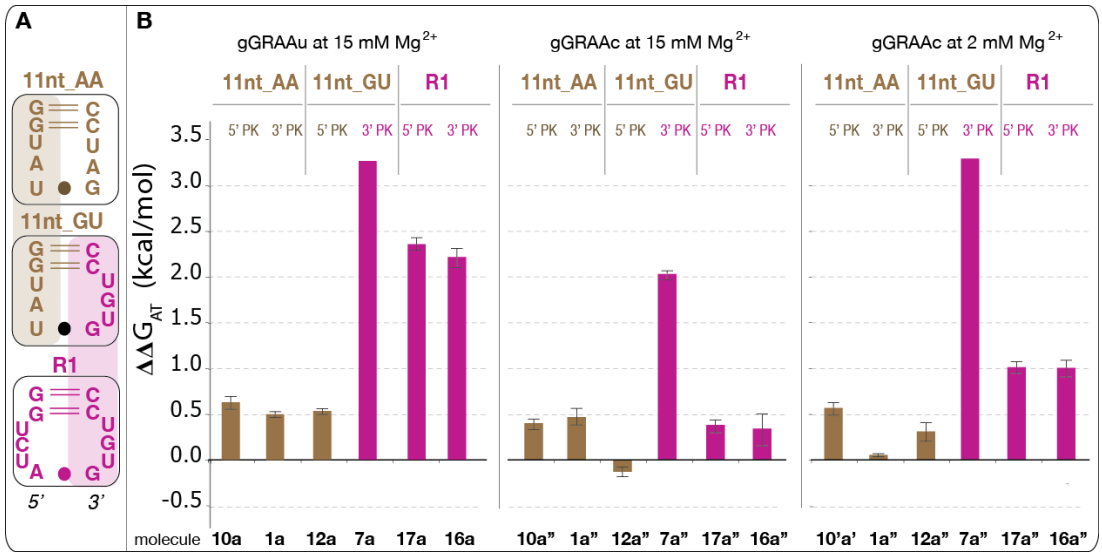


Figure 8.14 Thermodynamic analysis of tectoRNA attenuators based on the 11nt and R1 receptors.
This tectoRNA series has PK internal loops (PKL) with five As instead three (see Table S1). (A) Sequence relationships between the R1 [153], 11nt_GU (or C7.10 [73, 153]) and 11nt receptors. (B) Free energies of attenuation of heterodimer formation for all attenuator constructs based on the 11nt and R1 receptors (see also Table S4). The free energies of attenuation (ΔΔGAT) were estimated at 10°C and 2 or 15 mM Mg(OAc)2 as described in Materials and Methods. Attenuators 1a", 7a", 10a", 12a", 16a" and 17a" differ from molecules 1a, 7a, 10a, 12a, 16a and 17a by the presence of gGRAAc terminal loops (instead gGRAAu). This single nucleotide variation increases heterodimer stability. No major differences are observed between these constructs, with PKL of 5 As, and those with PKL of 3 As (see Figure 4.5).

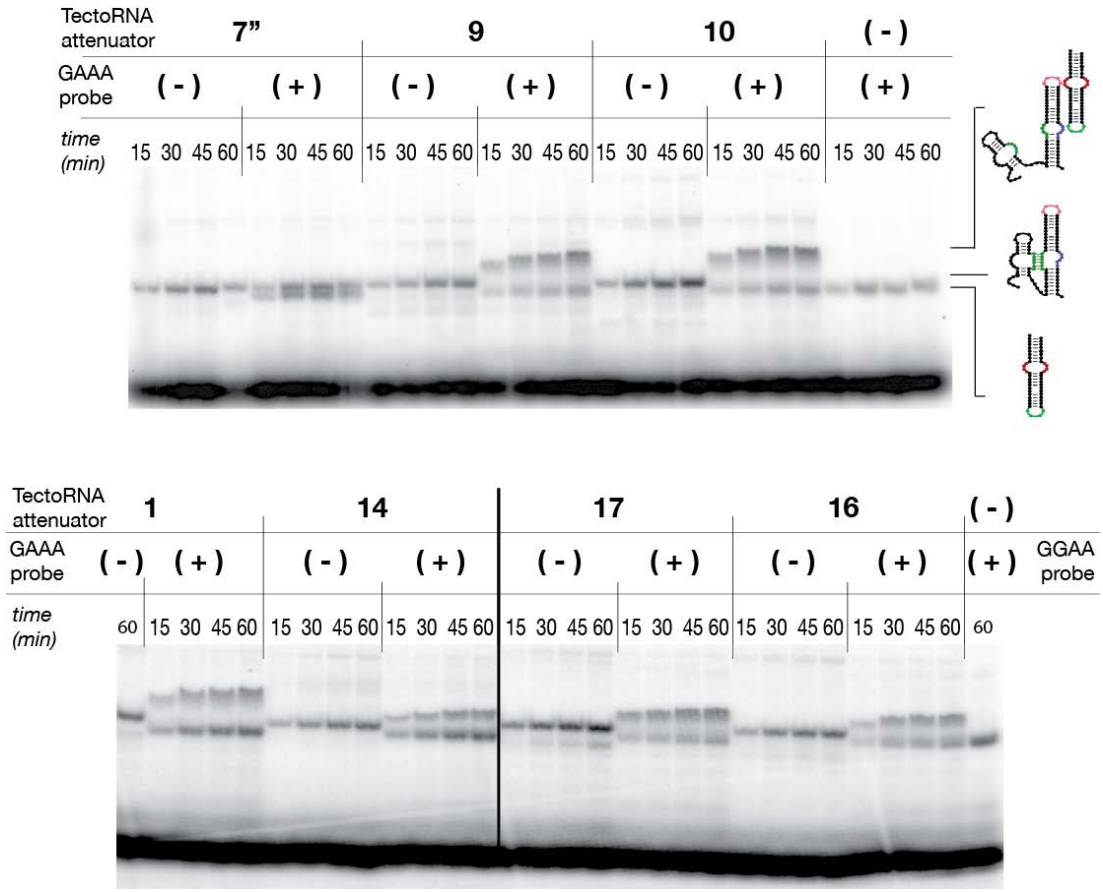
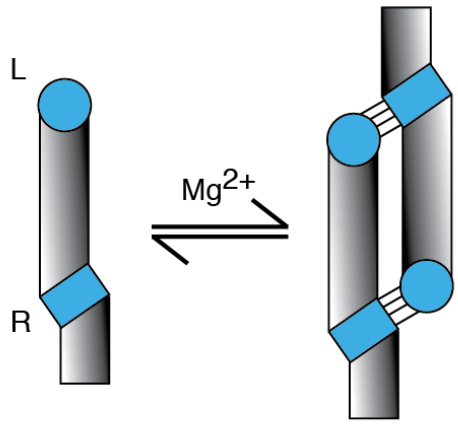


Figure 8.15 Co-transcriptional assemblies.
TectoRNA attenuators 1, 7’’, 9, 10, 14, 16 and 17 in presence (or absence) of their cognate GAAA or GGAA probe. Native PAGE autoradiograms of different tectoRNA attenuator transcription mixtures at various times in presence (+) or absence (-) of probe are shown. Co-transcriptional assembly is monitored by RNA body-labeling with α[P32]ATP and native PAGE is performed at 10°C and 10 mM $Mg(OAc)_2$ after quenching the transcription with DNase as described in Materials and Methods.

A



B

Receptor name	Receptor sequence	GAAA	GGAA	UAAA	AAAA
11nt_AA	CCUAAG[ ]UAU-GG	3 nM*+	>100000 nM	50 nM+	140 nM+
C7.2	CCUGUA[ ]-AU-GG	4 nM*	>100000 nM	n.d.	n.d.
11nt_GU	CCUGUG[ ]UAU-GG	20 nM*	>100000 nM	n.d.	n.d.
11nt_GUb	CCUGUC[ ]UAG-GG	30000 nM+	>100000 nM+	>100000 nM+	n.d.
R1	CCUGUG[ ]AUCUGG	>100000 nM	40 nM	n.d.	n.d.

Figure 8.16 *In vitro* tectoRNA self-assembly into homodimers.
Most of these tectoRNAs are based on receptors identified in class 1 di-GMP riboswitches [157]. These results indicate that the GNRA/receptor interaction within this natural molecule can span a broad range of binding affinities. During evolution, it is unlikely that the bias towards "U/A-rich" receptors is therefore due to a selection pressure for an intrinsic stability or/and loop selectivity of the receptor. Indeed, this GNRA/receptor interaction is tunable so that the riboswitch can function in different gene regulatory contexts more as a dimmer or rheostat than a binary on of switch. (**A**) Self-assembly scheme. (**B**) Table with apparent equilibrium dissociation constants (Kd in nM) for various homodimer constructs of the type: R(receptor):(LGAAA), R(receptor:(LGGAA), R(receptor):(LUAAA), R(receptor):(LAAAA). *: Indicates values previously reported in [153]; +: Indicates that the loop/receptor combination is found in the context of the di-GMP riboswitch [157]. The combination **GAAA** loop/ **11nt_A/CC** is also found in the di-GMP riboswitch and has been characterized in the HD heterodimer context (Table S2). Note that for homodimers, K_ds were determined from a nonlinear fit of the experimental data to equation: $f=(4\beta M_0 + K_d - (8M_0\beta K_d + K_d^2)^{0.5})/4M_0$. With β typically equal to 1, the K_d's equation is: $K_d = ((2M_0)(1 - f)^2)/f$. Therefore, Kds correspond to M_0 when 50% of bi-molecular assemblies are formed [39, 153].

8.4 Supplemental Information for Chapter 5

8.4.1 ***Box 1***

Experiment of complementation or gene replacement refers to the substitution of one gene for another, while maintaining the function of the gene. Typically, it consists in the transfer of DNA molecules from a population of identical copies of DNA molecules into bacteria cells from a population of identical bacteria cells (they are all clones) through a process called transformation. Statistically one bacteria cell do not receive more than one copy of the DNA molecules that codes for a new gene or ensemble of genes. This bacteria cell then divides into a progeny of identical cells that will all have the new gene. From the point of view of the bacterial population, several identical cells receive the same DNA molecule, each of them leading to new colonies that are identical to one another. While there is a lot of different methods to perform complementation at a technical level, it is possible to complement the bacteria by a new DNA gene that will either be directly incorporated into the DNA bacterial chromosome, or carried by a DNA vector (plasmid), which is not physically linked to the single bacterial chromosome molecule inside the cell. The endogenous bacterial gene that needs to be replaced by the new one can either be removed (or knocked out) before or after introduction of the new gene of similar function. The choice of the strategies of complementation essentially depends upon the function of the gene of interest that is replaced. In their finality, these strategies amount to the replacement of one gene by another in a single cell that after division

will lead to a progeny of new identical cells that will all contain the same new gene of interest.

The experiments of complementation that have been previously demonstrated can be performed in at least four different ways (see Figure 5.4). Two types of gene replacement experiments can be distinguished: orthologous replacement and non-orthologous replacement. Orthologous replacement refers to experiments where a gene is replaced by another gene that codes for a functionally and structurally similar product (e.g. [215]). These types of genes are said to be homologous and are derived from a divergent evolutionary event (Figure 4.3b). In contrast, non-orthologous gene replacement is the replacement of one gene for another that is structurally different, yet functionally similar (e.g. [188, 200]). These types of genes are said to be non-homologous and are derived from evolutionary convergence (Figure 5.3c). Non-orthologous replacement experiments offer the most compelling evidence for the existence of classes of functional equivalence. In addition, there are at least two other types of complementation that can involve the replacement of an entire metabolic pathway [199] or the insertion of alterative regulatory elements or pathways to program new functionalities leading to identical outcomes (e.g. [233, 235]). By altering "lower level" modes of operations at the level of metabolic pathways, one can demonstrate that an artificial set of molecular operations encoded at the level of DNA can be equivalent to a natural one as far as it concerns their outcome.

Most experiments of complementation have been initially designed to address questions of evolutionary significance. It is however important to see that the

concept of classes of functional equivalence allows to bypass these evolutionary concerns and can be used in Synthetic Biology to unravel the functions of cellular entities. For instance, newly discovered cellular regulatory circuits can potentially be tested through this concept.

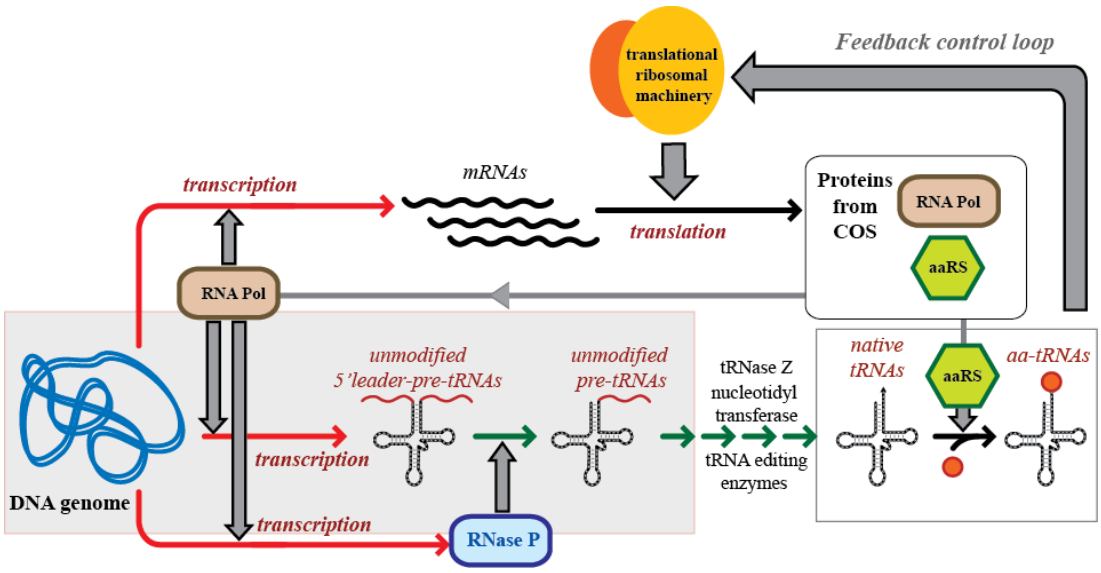


Figure 8.17 TDC by information control in the cellular operating system
Schematic describing the feedback informational control loops (in grey) involving protein translation. If tRNAs are not properly amino-acylated or matured, the enzymes (RNA polymerases [RNA Pol], amino-acyl transferases [aaRS] and ribonuclease P [RNase P]), which are responsible for tRNA production, cannot be properly synthesized by the translational apparatus (involving the ribosome).

References

1. Li, H., T.H. Labean, and K.W. Leong, *Nucleic acid-based nanoengineering: novel structures for biomedical applications.* Interface Focus, 2011. **1**(5): p. 702-24.
2. Staehelin, T., D.M. Maglott, and R.E. Monro, *On the catalytic center of peptidyl transfer: a part of the 50 S ribosome structure.* Cold Spring Harb Symp Quant Biol, 1969. **34**: p. 39-48.
3. Grimm, D., K. Pandey, and M.A. Kay, *Adeno-associated virus vectors for short hairpin RNA expression.* Methods Enzymol, 2005. **392**: p. 381-405.
4. Brantl, S. and E.G. Wagner, *Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181.* Mol Microbiol, 2000. **35**(6): p. 1469-82.
5. Yuan, J., C.M. Dutton, and S.P. Scully, *RNAi mediated MMP-1 silencing inhibits human chondrosarcoma invasion.* J Orthop Res, 2005. **23**(6): p. 1467-74.
6. Allain, F.H. and G. Varani, *Structure of the P1 helix from group I self-splicing introns.* J Mol Biol, 1995. **250**(3): p. 333-53.
7. Woodson, S.A., *Metal ions and RNA folding: a highly charged topic with a dynamic future.* Curr Opin Chem Biol, 2005. **9**(2): p. 104-9.
8. Woodson, S.A., *Recent insights on RNA folding mechanisms from catalytic RNA.* Cell Mol Life Sci, 2000. **57**(5): p. 796-808.
9. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs.* RNA, 2001. **7**(4): p. 499-512.
10. Leontis, N.B., J. Stombaugh, and E. Westhof, *The non-Watson-Crick base pairs and their associated isostericity matrices.* Nucleic Acids Res, 2002. **30**(16): p. 3497-531.
11. Benenson, Y., *Synthetic biology with RNA: progress report.* Curr Opin Chem Biol, 2012. **16**(3-4): p. 278-84.
12. McDaniel, R. and R. Weiss, *Advances in synthetic biology: on the path from prototypes to applications.* Curr Opin Biotechnol, 2005. **16**(4): p. 476-83.
13. Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science, 2008. **319**(5867): p. 1215-20.
14. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nat Rev Genet, 2010. **11**(5): p. 367-79.
15. Bachellerie, J.P., J. Cavaille, and A. Huttenhofer, *The expanding snoRNA world.* Biochimie, 2002. **84**(8): p. 775-90.
16. Blumenberg, M. and C. Yanofsky, *Regulatory region of the Klebsiella aerogenes tryptophan operon.* J Bacteriol, 1982. **152**(1): p. 49-56.
17. Ferapontova, E.E., E.M. Olsen, and K.V. Gothelf, *An RNA aptamer-based electrochemical biosensor for detection of theophylline in serum.* J Am Chem Soc, 2008. **130**(13): p. 4256-8.
18. Hamilton, A.J. and D.C. Baulcombe, *A species of small antisense RNA in posttranscriptional gene silencing in plants.* Science, 1999. **286**(5441): p. 950-2.

19. Lilley, D.M., *The origins of RNA catalysis in ribozymes.* Trends Biochem Sci, 2003. **28**(9): p. 495-501.
20. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 resolution.* Science, 2012. **289**(5481): p. 1678-85.
21. Cate, J.H., et al., *Crystal structure of a group I ribozyme domain: principles of RNA packing.* Science, 1996. **273**(5282): p. 1678-85.
22. Yusupov, M.M., et al., *Crystal structure of the ribosome at 5.5 A resolution.* Science, 2001. **292**(5518): p. 883-96.
23. Afonin, K.A., et al., *Design and self-assembly of siRNA-functionalized RNA nanoparticles for use in automated nanomedicine.* Nat Protoc, 2011. **6**(12): p. 2022-34.
24. Zhang, Y., et al., *Combinational delivery of c-myc siRNA and nucleoside analogs in a single, synthetic nanocarrier for targeted cancer therapy.* Biomaterials, 2013. **34**(33): p. 8459-68.
25. Doty, P., et al., *Configurational studies of polynucleotides and ribonucleic acid.* Ann N Y Acad Sci, 1959. **81**: p. 693-708.
26. Moore, P.B., *Structural motifs in RNA.* Annu Rev Biochem, 1999. **68**: p. 287-300.
27. Leontis, N.B. and E. Westhof, *Analysis of RNA motifs.* Curr Opin Struct Biol, 2003. **13**(3): p. 300-8.
28. Tamura, M. and S.R. Holbrook, *Sequence and structural conservation in RNA ribose zippers.* J Mol Biol, 2002. **320**(3): p. 455-74.
29. Holbrook, S.R., *RNA structure: the long and the short of it.* Curr Opin Struct Biol, 2005. **15**(3): p. 302-8.
30. Qin, H., T.R. Sosnick, and T. Pan, *Modular construction of a tertiary RNA structure: the specificity domain of the Bacillus subtilis RNase P RNA.* Biochemistry, 2001. **40**(37): p. 11202-10.
31. Tinoco, I., Jr. and C. Bustamante, *How RNA folds.* J Mol Biol, 1999. **293**(2): p. 271-81.
32. Greenleaf, W.J., et al., *Direct observation of hierarchical folding in single riboswitch aptamers.* Science, 2008. **319**(5863): p. 630-3.
33. Bindewald, E., et al., *Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler.* J Mol Graph Model, 2008. **27**(3): p. 299-308.
34. Jossinet, F., T.E. Ludwig, and E. Westhof, *Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels.* Bioinformatics, 2010. **26**(16): p. 2057-9.
35. Severcan, I., et al., *A polyhedron made of tRNAs.* Nat Chem, 2010. **2**(9): p. 772-9.
36. Grabow, W.W., et al., *Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes.* Nano Lett, 2011. **11**(2): p. 878-87.
37. Binning, J.M., et al., *Development of RNA aptamers targeting Ebola virus VP35.* Biochemistry, 2013.

38. Duconge, F. and J.J. Toulme, *In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1.* RNA, 1999. **5**(12): p. 1605-14.
39. Jaeger, L., E. Westhof, and N.B. Leontis, *TectoRNA: modular assembly units for the construction of RNA nano-objects.* Nucleic Acids Res, 2001. **29**(2): p. 455-63.
40. Afonin, K.A., et al., *In vitro assembly of cubic RNA-based scaffolds designed in silico.* Nat Nanotechnol, 2010. **5**(9): p. 676-82.
41. Afonin, K.A., D.J. Cieply, and N.B. Leontis, *Specific RNA self-assembly with minimal paranemic motifs.* J Am Chem Soc, 2008. **130**(1): p. 93-102.
42. Nasalean, L., et al., *Controlling RNA self-assembly to form filaments.* Nucleic Acids Res, 2006. **34**(5): p. 1381-92.
43. Jaeger, L. and N.B. Leontis, *Tecto-RNA: One-Dimensional Self-Assembly through Tertiary Interactions This work was carried out in Strasbourg with the support of grants to N.B.L. from the NIH (1R15 GM55898) and the NIH Fogarty Institute (1-F06-TW02251-01) and the support of the CNRS to L.J. The authors wish to thank Eric Westhof for his support and encouragement of this work.* Angew Chem Int Ed Engl, 2000. **39**(14): p. 2521-2524.
44. Grabow, W.W. and L. Jaeger, *RNA self-assembly and RNA nanotechnology.* Acc Chem Res, 2014. **47**(6): p. 1871-80.
45. Seeman, N.C., *Nucleic acid junctions and lattices.* J Theor Biol, 1982. **99**(2): p. 237-47.
46. Cao, H., et al., *RNA interference by nanofiber-based siRNA delivery system.* J Control Release, 2010. **144**(2): p. 203-12.
47. Yao, Z., Z. Weinberg, and W.L. Ruzzo, *CMfinder--a covariance model based RNA motif finding algorithm.* Bioinformatics, 2006. **22**(4): p. 445-52.
48. Riccitelli, N.J. and A. Luptak, *Computational discovery of folded RNA domains in genomes and in vitro selected libraries.* Methods, 2010. **52**(2): p. 133-40.
49. Freyhult, E., V. Moulton, and P. Gardner, *Predicting RNA structure using mutual information.* Appl Bioinformatics, 2005. **4**(1): p. 53-9.
50. Aguirre-Hernandez, R., H.H. Hoos, and A. Condon, *Computational RNA secondary structure design: empirical complexity and improved methods.* BMC Bioinformatics, 2007. **8**: p. 34.
51. Delisi, C. and D.M. Crothers, *Prediction of RNA secondary structure.* Proc Natl Acad Sci U S A, 1971. **68**(11): p. 2682-5.
52. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.
53. Zuker, M., *On finding all suboptimal foldings of an RNA molecule.* Science, 1989. **244**(4900): p. 48-52.
54. Mathews, D.H., et al., *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.* J Mol Biol, 1999. **288**(5): p. 911-40.
55. Schuster, P., et al., *From sequences to shapes and back: a case study in RNA secondary structures.* Proc Biol Sci, 1994. **255**(1344): p. 279-84.

56. Yingling, Y.G. and B.A. Shapiro, *Computational design of an RNA hexagonal nanoring and an RNA nanotube.* Nano Lett, 2007. **7**(8): p. 2328-34.
57. Mokdad, A. and A.D. Frankel, *ISFOLD: structure prediction of base pairs in non-helical RNA motifs from isostericity signatures in their sequence alignments.* J Biomol Struct Dyn, 2008. **25**(5): p. 467-72.
58. Robertson, D.L. and G.F. Joyce, *Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA.* Nature, 1990. **344**(6265): p. 467-8.
59. Laing, C. and T. Schlick, *Computational approaches to 3D modeling of RNA.* J Phys Condens Matter, 2010. **22**(28): p. 283101.
60. Parsch, J., J.M. Braverman, and W. Stephan, *Comparative sequence analysis and patterns of covariation in RNA secondary structures.* Genetics, 2000. **154**(2): p. 909-21.
61. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-10.
62. Hermann, T., *Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition.* Biochimie, 2002. **84**(9): p. 869-75.
63. Jaeger, L. and A. Chworos, *The architectonics of programmable RNA and DNA nanostructures.* Curr Opin Struct Biol, 2006. **16**(4): p. 531-43.
64. Grabow, W.W., et al., *The right angle (RA) motif: a prevalent ribosomal RNA structural pattern found in group I introns.* J Mol Biol, 2012. **424**(1-2): p. 54-67.
65. Chworos, A., et al., *Building programmable jigsaw puzzles with RNA.* Science, 2004. **306**(5704): p. 2068-72.
66. Tinoco, I., Jr., *From RNA hairpins to kisses to pseudoknots.* Nucleic Acids Symp Ser, 1997(36): p. 49-51.
67. Severcan, I., et al., *Square-shaped RNA particles from different RNA folds.* Nano Lett, 2009. **9**(3): p. 1270-7.
68. Aquino-Jarquin, G. and J.D. Toscano-Garibay, *RNA aptamer evolution: two decades of SELEction.* Int J Mol Sci, 2011. **12**(12): p. 9155-71.
69. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. **346**(6287): p. 818-22.
70. Seelig, B. and A. Jaschke, *A small catalytic RNA motif with Diels-Alderase activity.* Chem Biol, 1999. **6**(3): p. 167-76.
71. Wilson, C. and J.W. Szostak, *In vitro evolution of a self-alkylating ribozyme.* Nature, 1995. **374**(6525): p. 777-82.
72. Geary, C., S. Baudrey, and L. Jaeger, *Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors.* Nucleic Acids Res, 2008. **36**(4): p. 1138-52.
73. Costa, M. and F. Michel, *Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution.* EMBO J, 1997. **16**(11): p. 3289-302.

74. Barlow, M. and B.G. Hall, *Predicting evolutionary potential: in vitro evolution accurately reproduces natural evolution of the tem beta-lactamase.* Genetics, 2002. **160**(3): p. 823-32.
75. Hall, B.G., *Predicting evolution by in vitro evolution requires determining evolutionary pathways.* Antimicrob Agents Chemother, 2002. **46**(9): p. 3035-8.
76. Diaz Arenas, C. and N. Lehman, *Darwin's concepts in a test tube: parallels between organismal and in vitro evolution.* Int J Biochem Cell Biol, 2009. **41**(2): p. 266-73.
77. Johnston, W.K., et al., *RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension.* Science, 2001. **292**(5520): p. 1319-25.
78. Joyce, G.F., *Directed evolution of nucleic acid enzymes.* Annu Rev Biochem, 2004. **73**: p. 791-836.
79. Jaeger, L., M.C. Wright, and G.F. Joyce, *A complex ligase ribozyme evolved in vitro from a group I ribozyme domain.* Proc Natl Acad Sci U S A, 1999. **96**(26): p. 14712-7.
80. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world.* Crit Rev Biochem Mol Biol, 2004. **39**(2): p. 99-123.
81. Joyce, G.F., *Evolution in an RNA world.* Cold Spring Harb Symp Quant Biol, 2009. **74**: p. 17-23.
82. Jimenez, J.I., et al., *Comprehensive experimental fitness landscape and evolutionary network for small RNA.* Proc Natl Acad Sci U S A, 2013. **110**(37): p. 14984-9.
83. Ellington, A.D., et al., *Evolutionary origins and directed evolution of RNA.* Int J Biochem Cell Biol, 2009. **41**(2): p. 254-65.
84. Kauffman, S. and S. Levin, *Towards a general theory of adaptive walks on rugged landscapes.* J Theor Biol, 1987. **128**(1): p. 11-45.
85. Stich, M., E. Lazaro, and S.C. Manrubia, *Phenotypic effect of mutations in evolving populations of RNA molecules.* BMC Evol Biol, 2010. **10**: p. 46.
86. Woese, C.R., S. Winker, and R.R. Gutell, *Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops".* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8467-71.
87. Cheong, C., G. Varani, and I. Tinoco, Jr., *Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC.* Nature, 1990. **346**(6285): p. 680-2.
88. Heus, H.A. and A. Pardi, *Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops.* Science, 1991. **253**(5016): p. 191-4.
89. SantaLucia, J., Jr., R. Kierzek, and D.H. Turner, *Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin.* Science, 1992. **256**(5054): p. 217-9.
90. Michel, F. and E. Westhof, *Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.* J Mol Biol, 1990. **216**(3): p. 585-610.

91. Uhlenbeck, O.C., *Tetraloops and RNA folding.* Nature, 1990. **346**(6285): p. 613-4.
92. Woese, C.R., et al., *Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids.* Microbiol Rev, 1983. **47**(4): p. 621-69.
93. Jaeger, L., F. Michel, and E. Westhof, *Involvement of a GNRA tetraloop in long-range RNA tertiary interactions.* J Mol Biol, 1994. **236**(5): p. 1271-6.
94. Costa, M. and F. Michel, *Frequent use of the same tertiary motif by self-folding RNAs.* EMBO J, 1995. **14**(6): p. 1276-85.
95. Massire, C., L. Jaeger, and E. Westhof, *Phylogenetic evidence for a new tertiary interaction in bacterial RNase P RNAs.* RNA, 1997. **3**(6): p. 553-6.
96. Torres-Larios, A., et al., *Structure of ribonuclease P--a universal ribozyme.* Curr Opin Struct Biol, 2006. **16**(3): p. 327-35.
97. Noller, H.F., *RNA structure: reading the ribosome.* Science, 2005. **309**(5740): p. 1508-14.
98. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 A resolution.* Science, 2000. **289**(5481): p. 905-20.
99. Wimberly, B.T., et al., *Structure of the 30S ribosomal subunit.* Nature, 2000. **407**(6802): p. 327-39.
100. Ramakrishnan, V., *Ribosome structure and the mechanism of translation.* Cell, 2002. **108**(4): p. 557-72.
101. Pley, H.W., K.M. Flaherty, and D.B. McKay, *Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix.* Nature, 1994. **372**(6501): p. 111-3.
102. Doherty, E.A., et al., *A universal mode of helix packing in RNA.* Nat Struct Biol, 2001. **8**(4): p. 339-43.
103. Nissen, P., et al., *RNA tertiary interactions in the large ribosomal subunit: the A-minor motif.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 4899-903.
104. Adams, P.L., et al., *Crystal structure of a self-splicing group I intron with both exons.* Nature, 2004. **430**(6995): p. 45-50.
105. Regulski, E.E., et al., *A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism.* Mol Microbiol, 2008. **68**(4): p. 918-32.
106. Weinberg, Z., et al., *Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline.* Nucleic Acids Res, 2007. **35**(14): p. 4809-19.
107. Geary, C., A. Chworos, and L. Jaeger, *Promoting RNA helical stacking via A-minor junctions.* Nucleic Acids Res, 2011. **39**(3): p. 1066-80.
108. Tishchenko, S.V., et al., *Isolation, crystallization, and investigation of ribosomal protein S8 complexed with specific fragments of rRNA of bacterial or archaeal origin.* Biochemistry (Mosc), 2001. **66**(9): p. 948-53.
109. Holbrook, S.R., *Structural principles from large RNAs.* Annu Rev Biophys, 2008. **37**: p. 445-64.
110. Muto, Y. and S. Yokoyama, *Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems.* Wiley Interdiscip Rev RNA, 2012. **3**(2): p. 229-46.

111. Hendrix, D.K., S.E. Brenner, and S.R. Holbrook, *RNA structural motifs: building blocks of a modular biomolecule.* Q Rev Biophys, 2005. **38**(3): p. 221-43.
112. Leontis, N.B., A. Lescoute, and E. Westhof, *The building blocks and motifs of RNA architecture.* Curr Opin Struct Biol, 2006. **16**(3): p. 279-87.
113. Butcher, S.E. and A.M. Pyle, *The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks.* Acc Chem Res, 2011. **44**(12): p. 1302-11.
114. Jaeger, L., *Defining the syntax for self-assembling RNA tertiary architectures.* Nucleic Acids Symp Ser (Oxf), 2009(53): p. 83-4.
115. Grabow, W. and L. Jaeger, *RNA modularity for synthetic biology.* F1000Prime Rep, 2013. **5**: p. 46.
116. Rupert, P.B. and A.R. Ferre-D'Amare, *Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis.* Nature, 2001. **410**(6830): p. 780-6.
117. Ecker, D.J., et al., *Pseudo--half-knot formation with RNA.* Science, 1992. **257**(5072): p. 958-61.
118. Vijayasarathy, C., et al., *Identification of a stable RNA encoded by the H-strand of the mouse mitochondrial D-loop region and a conserved sequence motif immediately upstream of its polyadenylation site.* Gene Expr, 1995. **4**(3): p. 125-41.
119. Abramovitz, D.L. and A.M. Pyle, *Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction.* J Mol Biol, 1997. **266**(3): p. 493-506.
120. Fiore, J.L. and D.J. Nesbitt, *An RNA folding motif: GNRA tetraloop-receptor interactions.* Q Rev Biophys, 2013. **46**(3): p. 223-64.
121. Xin, Y., et al., *Annotation of tertiary interactions in RNA structures reveals variations and correlations.* RNA, 2008. **14**(12): p. 2465-77.
122. Butcher, S.E., T. Dieckmann, and J. Feigon, *Solution structure of a GAAA tetraloop receptor RNA.* EMBO J, 1997. **16**(24): p. 7490-9.
123. Davis, J.H., et al., *RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex.* J Mol Biol, 2005. **351**(2): p. 371-82.
124. Ikawa, Y., et al., *A comparative study on two GNRA-tetraloop receptors: 11-nt and IC3 motifs.* J Biochem, 2001. **130**(2): p. 251-5.
125. Chow, C.S., T.N. Lamichhane, and S.K. Mahto, *Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications.* ACS Chem Biol, 2007. **2**(9): p. 610-9.
126. Hsiao, C., et al., *Peeling the onion: ribosomes are ancient molecular fossils.* Mol Biol Evol, 2009. **26**(11): p. 2415-25.
127. Johnsson, P., et al., *Evolutionary conservation of long non-coding RNAs; sequence, structure, function.* Biochim Biophys Acta, 2014. **1840**(3): p. 1063-71.
128. Hsiao, C., et al., *Molecular paleontology: a biochemical model of the ancestral ribosome.* Nucleic Acids Res, 2013. **41**(5): p. 3373-85.

129. Gu, W., et al., *The impact of RNA structure on coding sequence evolution in both bacteria and eukaryotes.* BMC Evol Biol, 2014. **14**: p. 87.
130. Wilson, D.S. and J.W. Szostak, *In vitro selection of functional nucleic acids.* Annu Rev Biochem, 1999. **68**: p. 611-47.
131. Zhou, M., et al., *A light-driven DNA nanomachine for the efficient photoswitching of RNA digestion.* Angew Chem Int Ed Engl, 2010. **49**(12): p. 2167-70.
132. Leconte, A.M., et al., *Directed evolution of DNA polymerases for next-generation sequencing.* Angew Chem Int Ed Engl, 2010. **49**(34): p. 5921-4.
133. Sheehy, J.P., A.R. Davis, and B.M. Znosko, *Thermodynamic characterization of naturally occurring RNA tetraloops.* RNA, 2010. **16**(2): p. 417-29.
134. Afonin, K.A. and N.B. Leontis, *Generating new specific RNA interaction interfaces using C-loops.* J Am Chem Soc, 2006. **128**(50): p. 16131-7.
135. Cate, J.H., et al., *RNA tertiary structure mediation by adenosine platforms.* Science, 1996. **273**(5282): p. 1696-9.
136. Tijerina, P., S. Mohr, and R. Russell, *DMS footprinting of structured RNAs and RNA-protein complexes.* Nat Protoc, 2007. **2**(10): p. 2608-23.
137. Chauhan, S. and S.A. Woodson, *Tertiary interactions determine the accuracy of RNA folding.* J Am Chem Soc, 2008. **130**(4): p. 1296-303.
138. Ishikawa, J., et al., *Rational optimization of the DSL ligase ribozyme with GNRA/receptor interacting modules.* Arch Biochem Biophys, 2009. **490**(2): p. 163-70.
139. Davies, C., V. Ramakrishnan, and S.W. White, *Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from Bacillus stearothermophilus at 1.9 A resolution.* Structure, 1996. **4**(9): p. 1093-104.
140. Schuwirth, B.S., et al., *Structures of the bacterial ribosome at 3.5 A resolution.* Science, 2005. **310**(5749): p. 827-34.
141. Moine, H., et al., *The RNA binding site of S8 ribosomal protein of Escherichia coli: Selex and hydroxyl radical probing studies.* RNA, 1997. **3**(3): p. 255-68.
142. Tishchenko, S., et al., *Detailed analysis of RNA-protein interactions within the ribosomal protein S8-rRNA complex from the archaeon Methanococcus jannaschii.* J Mol Biol, 2001. **311**(2): p. 311-24.
143. Merianos, H.J., J. Wang, and P.B. Moore, *The structure of a ribosomal protein S8/spc operon mRNA complex.* RNA, 2004. **10**(6): p. 954-64.
144. Jaeger, L., E.J. Verzemnieks, and C. Geary, *The UA_handle: a versatile submotif in stable RNA architectures.* Nucleic Acids Res, 2009. **37**(1): p. 215-30.
145. Ramakrishnan, V. and S.W. White, *Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome.* Trends Biochem Sci, 1998. **23**(6): p. 208-12.
146. Agalarov, S.C., et al., *Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain.* Science, 2000. **288**(5463): p. 107-13.

147. Aguirre, J., et al., *Topological structure of the space of phenotypes: the case of RNA neutral networks.* PLoS One, 2011. **6**(10): p. e26324.
148. Manrubia, S.C. and C. Briones, *Modular evolution and increase of functional complexity in replicating RNA molecules.* RNA, 2007. **13**(1): p. 97-107.
149. Held, D.M., et al., *Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers.* J Mol Evol, 2003. **57**(3): p. 299-308.
150. Michel, F., et al., *Activation of the catalytic core of a group I intron by a remote 3' splice junction.* Genes Dev, 1992. **6**(8): p. 1373-85.
151. Adams, P.L., et al., *Crystal structure of a group I intron splicing intermediate.* Rna, 2004. **10**(12): p. 1867-87.
152. Ikawa, Y., et al., *A conserved motif in group IC3 introns is a new class of GNRA receptor.* Nucleic acids research, 1999. **27**(8): p. 1859-65.
153. Geary, C., S. Baudrey, and L. Jaeger, *Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors.* Nucleic Acids Res, 2008. **36**(4): p. 1138--1152.
154. Michel, F. and E. Westhof, *Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.* Journal of molecular biology, 1990. **216**(3): p. 585-610.
155. Lehnert, V., et al., *New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme.* Chemistry & biology, 1996. **3**(12): p. 993-1009.
156. Massire, C., L. Jaeger, and E. Westhof, *Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis.* J Mol Biol, 1998. **279**(4): p. 773-93.
157. Sudarsan, N., et al., *Riboswitches in eubacteria sense the second messenger cyclic di-GMP.* Science, 2008. **321**(5887): p. 411-3.
158. Ishikawa, J., et al., *GNRA/receptor interacting modules: versatile modular units for natural and artificial RNA architectures.* Methods, 2011. **54**(2): p. 226-38.
159. Dawid, A., B. Cayrol, and H. Isambert, *RNA synthetic biology inspired from bacteria: construction of transcription attenuators under antisense regulation.* Phys Biol, 2009. **6**(2): p. 025007.
160. Gutierrez-Preciado, A., et al., *Biochemical features and functional implications of the RNA-based T-box regulatory mechanism.* Microbiol Mol Biol Rev, 2009. **73**(1): p. 36-61.
161. Wachter, A., *Riboswitch-mediated control of gene expression in eukaryotes.* RNA Biol, 2011. **7**(1): p. 67-76.
162. Breaker, R.R., *Riboswitches and the RNA World.* Cold Spring Harb Perspect Biol, 2011.
163. Hess, H. and L. Jaeger, *Nanobiotechnology.* Curr Opin Biotechnol, 2010. **21**(4): p. 373-5.
164. Topp, S. and J.P. Gallivan, *Emerging applications of riboswitches in chemical biology.* ACS Chem Biol, 2011. **5**(1): p. 139-48.
165. de Las Heras, A., et al., *Engineering input/output nodes in prokaryotic regulatory circuits.* FEMS Microbiol Rev, 2011. **34**(5): p. 842-65.

166. Freier, S.M., et al., *Improved free-energy parameters for predictions of RNA duplex stability.* Proc Natl Acad Sci U S A, 1986. **83**(24): p. 9373-7.
167. Mahen, E.M., et al., *Kinetics and thermodynamics make different contributions to RNA folding in vitro and in yeast.* Molecular cell, 2005. **19**(1): p. 27-37.
168. Mahen, E.M., et al., *mRNA secondary structures fold sequentially but exchange rapidly in vivo.* PLoS Biol, 2010. **8**(2): p. e1000307.
169. Wu, H. and J. Feigon, *H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6655-60.
170. Jin, H., J.P. Loria, and P.B. Moore, *Solution Structure of an rRNA Substrate Bound to the Pseudouridylation Pocket of a Box H/ACA snoRNA.* Mol.Cell, 2007. **26**: p. 205-215.
171. Ikawa, Y., et al., *A conserved motif in group IC3 introns is a new class of GNRA receptor.* Nucleic Acids Res, 1999. **27**(8): p. 1859-65.
172. Smith, K.D., et al., *Structural basis of ligand binding by a c-di-GMP riboswitch.* Nat Struct Mol Biol, 2009. **16**(12): p. 1218-23.
173. Kulshina, N., N.J. Baird, and A.R. Ferre-D'Amare, *Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch.* Nat Struct Mol Biol, 2009. **16**(12): p. 1212-7.
174. Lehnert, V., et al., *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, 1996. **3**(12): p. 993-1009.
175. Toor, N., et al., *Tertiary architecture of the Oceanobacillus iheyensis group II intron.* RNA, 2010. **16**(1): p. 57-69.
176. Baird, N.J., N. Kulshina, and A.R. Ferre-D'Amare, *Riboswitch function: flipping the switch or tuning the dimmer?* RNA Biol, 2010. **7**(3): p. 328-32.
177. Prathiba, J. and R. Malathi, *Group I introns and GNRA tetraloops: remnants of 'The RNA world'?* Mol Biol Rep, 2008. **35**(2): p. 239-49.
178. Macke, T.J., et al., *RNAMotif, an RNA secondary structure definition and search algorithm.* Nucleic Acids Res, 2001. **29**(22): p. 4724-35.
179. Gardner, P.P., et al., *Rfam: Wikipedia, clans and the "decimal" release.* Nucleic Acids Res, 2011. **39**(Database issue): p. D141-5.
180. Jaeger, L. and E.R. Calkins, *Downward causation by information control in micro-organisms.* Interface Focus, 2012. **2**(1): p. 26-41.
181. Plotkin, J.B. and J. Dushoff, *Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 7152-7.
182. Bard, J., *A systems biology view of evolutionary genetics: network-driven processes incorporate much more variation than evolutionary genetics can handle. This variation is hard to formalise but allows fast change.* Bioessays, 2010. **32**(7): p. 559-63.
183. Martins dos Santos, V.A. and J. Damborsky, *Systems biology at work.* Current opinion in biotechnology, 2010. **21**(4): p. 498-501.

184. de Lorenzo, V., *Synthetic biology: something old, something new.* Bioessays, 2010. **32**(4): p. 267-70.
185. Danchin, A., *Information of the chassis and information of the program in synthetic cells.* Syst Synth Biol, 2009. **3**(1-4): p. 125-34.
186. Young, E. and H. Alper, *Synthetic biology: tools to design, build, and optimize cellular processes.* J Biomed Biotechnol, 2010. **2010**: p. 130781.
187. Luisi, P.L., *Toward the engineering of minimal living cells.* Anat Rec, 2002. **268**(3): p. 208-14.
188. Luisi, P.L., *The emergence of life: from chemical origins to synthetic biology*. 2006, Cambridge: Cambridge University Press, Cambridge. 315.
189. Clayton, P. and P.C.W. Davies, eds. *The re-emergence of emergence*. 2006, Oxford University Press: Oxford, UK.
190. Murphy, N. and W.R. Stoeger, *Evolution and emergence: systems, organisms, persons*. 2007, Oxford, UK: Oxford University Press.
191. Ellis, G.F., *On the nature of causation in complex systems.* trans. R. Soc. S. Afr., 2008. **63**: p. 1-16.
192. Auletta, G., G.F. Ellis, and L. Jaeger, *Top-down causation by information control: from a philosophical problem to a scientific research programme.* J R Soc Interface, 2008. **5**(27): p. 1159-72.
193. Ellis, G.F., *Top-Down Causation and the Human Brain*, in *Downward causation and the Neurobiology of Free*
Will, N. Murphy, G.F. Ellis, and T. O'Connor, Editors. 2009, Springer-Verlag: Berlin Heidelberg.
194. Barabasi, A.L. and Z.N. Oltvai, *Network biology: understanding the cell's functional organization.* Nat Rev Genet, 2004. **5**(2): p. 101-13.
195. Kashtan, N. and U. Alon, *Spontaneous evolution of modularity and network motifs.* Proc Natl Acad Sci U S A, 2005. **102**(39): p. 13773-8.
196. Alon, U., *Network motifs: theory and experimental approaches.* Nat Rev Genet, 2007. **8**(6): p. 450-61.
197. Silva-Rocha, R. and V. de Lorenzo, *Noise and robustness in prokaryotic regulatory networks.* Annu Rev Microbiol, 2010. **64**: p. 257-75.
198. Goelzer, A., et al., *Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of Bacillus subtilis.* BMC Syst Biol, 2008. **2**: p. 20.
199. Spirin, V., et al., *A metabolic network in the evolutionary context: multiscale structure and modularity*. Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8774-9.
200. Harmer, R., et al., *Intrinsic information carriers in combinatorial dynamical systems.* Chaos, 2010. **20**(3): p. 037108.
201. von Neumann, J., *The Computer and the Brain*. 1958, New Haven: Yale University Press.
202. Danchin, A., *Bacteria as computers making computers.* FEMS Microbiol Rev, 2009. **33**(1): p. 3-26.
203. Forster, A.C. and G.M. Church, *Towards synthesis of a minimal cell.* Mol Syst Biol, 2006. **2**: p. 45.

204. Jester, B.C., et al., *Nonorthologous replacement of lysyl-tRNA synthetase prevents addition of lysine analogues to the genetic code.* Proc Natl Acad Sci U S A, 2003. **100**(24): p. 14351-6.
205. Baick, J.W., et al., *Growth inhibition of Escherichia coli during heterologous expression of Bacillus subtilis glutamyl-tRNA synthetase that catalyzes the formation of mischarged glutamyl-tRNA1 Gln.* J Microbiol, 2004. **42**(2): p. 111-6.
206. Gobert, A., et al., *A single Arabidopsis organellar protein has RNase P activity.* Nat Struct Mol Biol, 2010. **17**(6): p. 740-4.
207. Wegscheid, B., C. Condon, and R.K. Hartmann, *Type A and B RNase P RNAs are interchangeable in vivo despite substantial biophysical differences.* EMBO Rep, 2006. **7**(4): p. 411-7.
208. Rauch, T., et al., *Dissecting functional similarities of ribosome-associated chaperones from Saccharomyces cerevisiae and Escherichia coli.* Mol Microbiol, 2005. **57**(2): p. 357-65.
209. Asai, T., et al., *An Escherichia coli strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1971-6.
210. Klutstein, M., et al., *Functional conservation of the yeast and Arabidopsis RAD54-like genes.* Genetics, 2008. **178**(4): p. 2389-97.
211. Zhong, J., et al., *Function and evolution of plasmid-borne genes for pyrimidine biosynthesis in Borrelia spp.* J Bacteriol, 2006. **188**(3): p. 909-18.
212. Shams-Eldin, H., et al., *Identification of the archaeal alg7 gene homolog (encoding N-acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in Saccharomyces cerevisiae.* J Bacteriol, 2008. **190**(6): p. 2217-20.
213. Coffman, J.A., *Information as a manifestation of development.* Information, 2011. **2**: p. 102-116.
214. Coffman, J.A., *On causality in nonlinear complex systems: the developmentalist perspective*, in *Handbook of the Philosophy of Science: Philosophy of Complex Systems*, C. Hooker, Editor. 2011, Elsevier: Amsterdam, The Netherlands. p. 214-221.
215. Szathmary, E. and J.M. Smith, *The major evolutionary transitions.* Nature, 1995. **374**(6519): p. 227-32.
216. Fontana, W. and L.W. Buss, *What would be conserved if "the tape were played twice"?* Proc Natl Acad Sci U S A, 1994. **91**(2): p. 757-61.
217. Silvestre, D.A. and J.F. Fontanari, *The information capacity of hypercycles.* J Theor Biol, 2008. **254**(4): p. 804-6.
218. Kapranov, P., et al., *New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism.* Nature, 2010. **466**(7306): p. 642-6.
219. Maida, Y. and K. Masutomi, *RNA-dependent RNA polymerases in RNA silencing.* Biological chemistry, 2011. **392**(4): p. 299-304.

220. Eigen, M., et al., *The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage.* Biochemistry, 1991. **30**(46): p. 11005-18.
221. Arendt, J. and D. Reznick, *Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation?* Trends Ecol Evol, 2008. **23**(1): p. 26-32.
222. Leander, B.S., *A hierarchical view of convergent evolution in microbial eukaryotes.* J Eukaryot Microbiol, 2008. **55**(2): p. 59-68.
223. Terada, T., et al., *Functional convergence of two lysyl-tRNA synthetases with unrelated topologies.* Nat Struct Biol, 2002. **9**(4): p. 257-62.
224. Roth, A. and R.R. Breaker, *The structural and functional diversity of metabolite-binding riboswitches.* Annu Rev Biochem, 2009. **78**: p. 305-34.
225. Gilbert, S.D., et al., *Structure of the SAM-II riboswitch bound to S-adenosylmethionine.* Nat Struct Mol Biol, 2008. **15**(2): p. 177-82.
226. Lu, C., et al., *Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism.* Nat Struct Mol Biol, 2008. **15**(10): p. 1076-83.
227. Serganov, A., *The long and the short of riboswitches.* Curr Opin Struct Biol, 2009. **19**(3): p. 251-9.
228. Vetsigian, K., C. Woese, and N. Goldenfeld, *Collective evolution and the genetic code.* Proc Natl Acad Sci U S A, 2006. **103**(28): p. 10696-701.
229. Winfield, M.D. and E.A. Groisman, *Phenotypic differences between Salmonella and Escherichia coli resulting from the disparate regulation of homologous genes.* Proc Natl Acad Sci U S A, 2004. **101**(49): p. 17162-7.
230. Topp, S. and J.P. Gallivan, *Guiding bacteria with small molecules and RNA.* J Am Chem Soc, 2007. **129**(21): p. 6807-11.
231. Sinha, J., S.J. Reyes, and J.P. Gallivan, *Reprogramming bacteria to seek and destroy an herbicide.* Nat Chem Biol, 2010. **6**(6): p. 464-70.
232. Alaimo, C., et al., *Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides.* Embo J, 2006. **25**(5): p. 967-76.
233. Perez-Castineira, J.R., et al., *Functional complementation of yeast cytosolic pyrophosphatase by bacterial and plant H+-translocating pyrophosphatases.* Proc Natl Acad Sci U S A, 2002. **99**(25): p. 15914-9.
234. Frithz-Lindsten, E., et al., *Functional conservation of the effector protein translocators PopB/YopB and PopD/YopD of Pseudomonas aeruginosa and Yersinia pseudotuberculosis.* Mol Microbiol, 1998. **29**(5): p. 1155-65.
235. Davis, T.N. and J. Thorner, *Vertebrate and yeast calmodulin, despite significant sequence divergence, are functionally interchangeable.* Proc Natl Acad Sci U S A, 1989. **86**(20): p. 7909-13.
236. Facchin, S., et al., *Functional homology between yeast piD261/Bud32 and human PRPK: both phosphorylate p53 and PRPK partially complements piD261/Bud32 deficiency.* FEBS Lett, 2003. **549**(1-3): p. 63-6.
237. Randau, L., I. Schroder, and D. Soll, *Life without RNase P.* Nature, 2008. **453**(7191): p. 120-3.

238. Merino, E., R.A. Jensen, and C. Yanofsky, *Evolution of bacterial trp operons and their regulation.* Curr Opin Microbiol, 2008. **11**(2): p. 78-86.
239. Vitreschak, A.G., et al., *Comparative genomic analysis of T-box regulatory systems in bacteria.* Rna, 2008. **14**(4): p. 717-35.
240. Yanofsky, C., *RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria.* Rna, 2007. **13**(8): p. 1141-54.
241. Poiata, E., et al., *A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria.* Rna, 2009. **15**(11): p. 2046-56.
242. Luisi, P.L., *About various definitions of life.* Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life, 1998. **28**(4-6): p. 613-22.
243. Joyce, G.F., *RNA evolution and the origins of life.* Nature, 1989. **338**(6212): p. 217-24.
244. Danchin, A., *Natural selection and immortality.* Biogerontology, 2009. **10**(4): p. 503-16.
245. Woese, C.R., *On the evolution of cells.* Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8742-7.
246. Vogan, A.A. and P.G. Higgs, *The advantages and disadvantages of horizontal gene transfer and the emergence of the first species.* Biol Direct, 2011. **6**: p. 1.
247. Gould, S.J., *Wonderful Life: The Burgess Shale and the Nature of History*. 1989: W. W. Norton & Company. 347.
248. Conway Morris, S., *The Crucible of Creation: The Burgess Shale and the Rise of Animals*. 1998: Oxford University Press. 235.
249. Pereto, J. and J. catala, *The Renaissance of Synthetic Biology.* Biological Theory, 2007. **2**((2)): p. 128-130.
250. Keasling, J.D., *Synthetic biology for synthetic chemistry.* ACS Chem Biol, 2008. **3**(1): p. 64-76.
251. Szostak, J.W., D.P. Bartel, and P.L. Luisi, *Synthesizing life.* Nature, 2001. **409**(6818): p. 387-90.
252. Schrum, J.P., T.F. Zhu, and J.W. Szostak, *The origins of cellular life.* Cold Spring Harbor perspectives in biology, 2010. **2**(9): p. a002212.
253. Lartigue, C., et al., *Genome transplantation in bacteria: changing one species to another.* Science, 2007. **317**(5838): p. 632-8.
254. Kaiser, J., *Synthetic biology. Attempt to patent artificial organism draws a protest.* Science, 2007. **316**(5831): p. 1557.
255. Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome.* Science, 2010. **329**(5987): p. 52-6.
256. Jewett, M.C. and A.C. Forster, *Update on designing and building minimal cells.* Curr Opin Biotechnol, 2010. **5**: p. 697-703.
257. Claverie, J.M. and C. Abergel, *Mimivirus: the emerging paradox of quasi-autonomous viruses.* Trends Genet, 2010. **26**(10): p. 431-7.
258. Forterre, P., *Giant viruses: conflicts in revisiting the virus concept.* Intervirology, 2010. **53**(5): p. 362-78.

259. Boyer, M., et al., *Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21848-53.
260. Steen, E.J., et al., *Microbial production of fatty-acid-derived fuels and chemicals from plant biomass.* Nature, 2010. **463**(7280): p. 559-62.
261. Peralta-Yahya, P.P. and J.D. Keasling, *Advanced biofuel production in microbes.* Biotechnol J, 2010. **5**(2): p. 147-62.
262. An, W. and J.W. Chin, *Synthesis of orthogonal transcription-translation networks.* Proc Natl Acad Sci U S A, 2009. **106**(21): p. 8477-82.
263. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.* Nature, 2010. **464**(7287): p. 441-4.
264. Gallivan, J.P., *Toward reprogramming bacteria with small molecules and RNA.* Curr Opin Chem Biol, 2007. **11**(6): p. 612-9.
265. Anderson, J.C., et al., *BglBricks: A flexible standard for biological part assembly.* J Biol Eng, 2010. **4**(1): p. 1.
266. Afonin, K.A., et al., *Attenuation of loop-receptor interactions with pseudoknot formation.* Nucleic Acids Res, 2012. **40**(5): p. 2168-80.
267. Chojnowski, G., T. Walen, and J.M. Bujnicki, *RNA Bricks--a database of RNA 3D motifs and their interactions.* Nucleic Acids Res, 2014. **42**(Database issue): p. D123-31.
268. Petrov, A.I., C.L. Zirbel, and N.B. Leontis, *Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas.* RNA, 2013. **19**(10): p. 1327-40.
269. Popenda, M., et al., *RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures.* BMC Bioinformatics, 2010. **11**: p. 231.
270. Chen, C., et al., *A dimer as a building block in assembling RNA. A hexamer that gears bacterial virus phi29 DNA-translocating machinery.* J Biol Chem, 2000. **275**(23): p. 17510-6.
271. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-23.
272. Rother, M., et al., *ModeRNA: a tool for comparative modeling of RNA 3D structure.* Nucleic Acids Res, 2011. **39**(10): p. 4007-22.
273. Delano, W.L., *The PyMOL Moleular Graphics System.* 2002.
274. Liu, B., et al., *Characterization of tectoRNA assembly with cationic conjugated polymers.* J Am Chem Soc, 2004. **126**(13): p. 4076-7.
275. Paillart, J.C., et al., *A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.* Proc Natl Acad Sci U S A, 1996. **93**(11): p. 5572-7.
276. Davis, J.H., et al., *Role of metal ions in the tetraloop-receptor complex as analyzed by NMR.* RNA, 2007. **13**(1): p. 76-86.
277. Xayaphoummine, A., T. Bucher, and H. Isambert, *Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots.* Nucleic acids research, 2005. **33**(Web Server issue): p. W605-10.

278. Paillart, J.C., et al., *A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(11): p. 5572-7.
279. Bindewald, E., T. Kluth, and B.A. Shapiro, *CyloFold: secondary structure prediction including pseudoknots.* Nucleic acids research, 2010. **38**(Web Server issue): p. W368-72.