The Dissertation Committee for Brant Richard Gracia Certifies that this is the approved version of the following dissertation:

Quantitative Dissection of RNA Structure Formation Reveals a Cooperative and Modular Folding and Assembly Landscape

Committee:

Rick Russell, Supervisor

Lydia Contreras

Ilya Finkelstein

Kenneth A. Johnson

Andreas Matouschek

Quantitative Dissection of RNA Structure Formation Reveals a Cooperative and Modular Folding and Assembly Landscape

by

Brant Richard Gracia, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Dedication

For my Parents and Sister who always supported me.

Acknowledgements

First and foremost, I would like to thank my advisor Rick Russell for his steady guidance and friendship during my research studies. Rick went above and beyond the role of research mentor on many occasions, and I cannot thank him enough for his help with the problems and challenges I faced inside and outside of the lab. I also want to thank Pilar Tijerina for all the work she has done to make the lab an effective place to work, for unconditional support inside and outside the lab, and for her help with ordering reagents and supplies. Inga Jarmoskaite was one of the best graduate student mentors I had, and I want to thank her for teaching me so many critical aspects of being a scientist. I have so many fond memories of our friendship: working in Argonne National Laboratory, eating pizza and visiting art museums in downtown Chicago, and working late into the night in the lab. I also want to give special thanks to my close colleague and friend Namita Bisaria who taught me a huge deal about how to ask a scientific question. I also have to thank her along with Breena Stoner, Joseph Yesselman, and Dan and Peggy Herschlag for giving me a place to stay during research trips and being wonderful friends.

I want to thank Rhiju Das for his mentorship and for giving me a place to research in his lab at Stanford. Anne Wipapat Kladwang taught me the high throughput footprinting technique and a good deal about experimental design, I cannot thank her enough. Yi Xue and Hashim Al-Hashimi were wonderful research scientists to work with and learn from. I thoroughly enjoyed the countless hours of stimulating conversation we had over skype. I want to thank Dan Herschlag for being a fantastic mentor, a leader, and a friend. I learned a great deal about science and life from the interactions I had with Dan, and the rest of the Program Project Grant group at Stanford University, thank you.

Philip Yangyuoru has been a wonderful friend to me. I will always remember the fun we had watching different sporting events while still managing to discuss the impending experiments to be done on the next day. Travis Middleton has been a great friend as well. I have so many fun memories of our times traversing the streets of Austin and San Francisco, finding secret bars, and going to concerts. Collaborative meetings at Stanford were much better with fellow graduate students Isabel Strohkendl and Ian Hamilton. I have no doubt all four of them will succeed. I also want to thank Woongsoon Choi, Jeff Potratz, Cynthia Pan, and Luke Ward for being good friends and mentors during their time in the lab. Lastly, I have to thank Josh Russell, Greg Clark, and Pallavi Kompella for their work in collaboration with the Present Your Ph.D. Outreach project.

Quantitative Dissection of RNA Structure Formation Reveals a Cooperative and Modular Folding and Assembly Landscape

Brant Richard Gracia, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Rick Russell

Structured RNAs are pervasive in biology with ubiquitous roles in gene expression and regulation. RNAs must fold from a linear chain of nucleotide sequence to attain three-dimensional structure. RNA folding can be described as modular and hierarchical with tiers of structure that form independently: secondary structure forms first and defines the helices followed by formation of tertiary structure. The separation between secondary and tertiary structure is not absolute. Many biological RNAs couple secondary structure changes to RNA tertiary structure formation and link these changes to downstream functional consequences. To predict how these biological RNAs fold requires a deep understanding of the structural intermediates, folding pathways, and mechanisms of cooperativity that promote folding. To test the modularity and predictability of secondary and tertiary RNA folding and assembly, we have investigated the folding and assembly of the P5abc subdomain from the Tetrahymena thermophila Group I intron ribozyme. P5abc folds cooperatively in isolation, binding Mg^{2+} ions and adopting tertiary structure. Mg^{2+} binding is linked to a shift in the secondary structure of seventeen nucleotides and prior work concluded that there is a high degree of cooperativity for this seemingly concerted transition. With the already established principles of RNA modularity in mind, we develop a reconstitution hypothesis to test if

cooperative secondary and tertiary folding and assembly of P5abc can be understood from the component pieces. By using rational mutagenesis, we find that higher order folding of P5abc is modular, and we elucidate the physical origins of cooperativity (Chapter 2). With our knowledge of isolated P5abc folding, we demonstrate that the local folding transition within P5abc controls the rate and pathway of assembly with the P5abc-deleted ribozyme core ($E^{\Delta P5abc}$), further highlighting the modularity of RNA structure (Chapter 3). Lastly, we show that the kinetics of assembly can be attributed to specific tertiary contacts that form in the assembly transition state such that the rate of a particular folding pathway is dictated by the properties of an individual tertiary contact (Chapter 4). The modularity of RNA structure makes it a reasonable molecule for the origins of life and an adaptable tool for bioengineering applications.

Table of Contents

List of Tables

List of Figures

Chapter 1: Versatility and modularity of RNA structure

INTRODUCTION

Structured RNAs have been ubiquitous dating back to the earliest stages of life on Earth, and since the discovery that RNA can act as an enzyme catalyst, or ribozyme, our knowledge of structured RNAs has greatly expanded (1). Nearly half a century of research has uncovered the intimate connection between RNA structure and cellular metabolism and developed a new world of RNA bioengineering. Clearly, RNAs are versatile biomolecules with endless applications, necessitating a clear understanding of how it forms functional structures.

In this chapter, I outline a body of prior work that leads to a testable quantitative folding model to further establish that RNA structure is modular, and more importantly, can be rationally controlled. First, I recall the fundamental principles of RNA structure formation that led to a hierarchical RNA folding model. Then, I explore previous observations that reinforced the independence of structure formation within each tier of the hierarchy. I then discuss complexities that arise for a class of biological RNAs that violate the RNA folding hierarchy, riboswitches, as well as non-canonical designer structured RNAs currently used in research laboratories and clinics. There is evidence that RNA structure is built from smaller modules of structure, exemplified by the reoccurrence of a common tetraloop tetraloop-receptor tertiary contact (TL/TLR) in almost all kingdoms of life. To further test this concept, I utilized the P5abc subdomain from *Tetrahymena thermophila*, a strong model system to study RNA folding thanks to the large body of preexisting biochemical and biophysical measurements. Central to my studies is the cooperative folding behavior observed for tertiary structure formation in P5abc and other RNAs. To quantitatively predict cooperative RNA structure formation, I develop and test a reconstitution hypothesis that hinges on the notion that RNA structure

can be rationally separated into smaller energetically independent components or modules of structure.

Fundamental studies of structured RNAs

RNAs fold from a negatively charged linear chain of nucleotides to structured RNAs with three-dimensional architecture. The energetics of folding suggest that folding proceeds through discrete steps, or tiers, of structure formation. The independence of structure formation has been further exemplified by in-depth investigations of each energetic tier (secondary and tertiary structure).

The RNA folding hierarchy

The RNA folding hierarchy can be separated into three energetic tiers of structure (Figure 1.1). The primary structure is the nucleic acid sequence contained within a given RNA. In the presence of positively charged cations, RNA folds to secondary structure in the form of helical segments, and these helices can form when separated from the full length RNA (2). These results suggest that the RNA secondary structure tier is made of helical elements that are independent modules of structure. As the concentration of positively charged cations is further increased, long range interactions form at the tertiary structure level and the three-dimensional architecture is attained (3). The independence of energetic tiers afforded by the RNA folding hierarchy has led to astounding discoveries at the level of secondary and tertiary structure alone.

RNA secondary structure and covariation

An RNA secondary structure can be represented as a two-dimensional map of the helices and base-pairs. From thermodynamic measurements of different RNA helices, the stability of individual RNA base pairs has been determined (4, 5). These results suggested that the energetic contribution of an RNA base pair can be calculated from the

identity of the base pairs immediately flanking. This 'nearest neighbor' rule has led to energetic parameters that can predict, with relatively high confidence, the secondary structure of an RNA (5-8). Indeed, when the ribosome crystal structures were solved, it was found that much of the secondary structure of the 16S and 23S rRNAs was accurately predicted (9-11).

Before elucidation of the ribosome structures, secondary structure models for the 16S and 23S ribosomal RNAs were proposed using comparative methods of RNA sequences from many different species (12, 13). Covariation analysis, a particularly powerful subset of comparative analysis, infers the participation of individual bases in a helix by analyzing pairs of sequence changes between organisms that would serve to maintain Watson-Crick base pairing, e.g. if a C->A mutation is observed near a G->U mutation, then it is likely that these two nucleotides pair as part of an RNA helix. Thus, a mutation that disrupts a Watson-Crick base pair is often counterbalanced by a compensatory mutation in the opposite strand that restores the base pair. It would appear that some RNA helices are critical for function and selective pressures are present to maintain helical elements. This novel finding further suggests that helical elements may act as independent structure modules.

RNA tertiary structure and isolated folding motifs

Considering the pre-formation of secondary structure in the RNA folding hierarchy, an initial inspection of RNA tertiary structure begins with nucleotides that do not participate in helices. RNA hairpins are capped by loops of unpaired nucleotides, and unpaired nucleotides can be present within an otherwise continuous helix when nucleotides that cannot form Watson-Crick base pairs are inserted on one strand (bulges) or on both strands (internal loop) (14). Helices can also converge at one junction, e.g. a

three-way junction connects three helices by linking them together with unpaired nucleotides at the junction (Figure 1.1B). The structured RNA studied herein adopts a three-way junction topology and has a bulge that is critical for folding.

Bulges in an RNA helix create junctions that serve to bend and orient helices at angles relative to each other. A bend in the 5' UTR of the Hepatitis C viral RNA is critical for hijacking the host translation machinery through an internal ribosome entry site (15, 16). The sequence conservation of the 5' UTR has led to the development of various strategies for targeting this structure to treat the virus (17). Comparative sequence analysis and biophysical characterization has revealed a widespread bulge motif that tightly bends the helix at a 50° angle introducing a kink-turn in the RNA helix (18-20). Bends in the helix are critical to promote long-range interactions between distant regions of the RNA.

The unpaired loops of two helices can 'kiss' to form long-range base pairs (21) (Figure 1.2A). A detailed analysis of assembly rates and equilibria of different kissing loop structures showed that kissing loop stability is determined primarily by the dissociation rate constant (off rate) with similar on rates observed for different kissing loop sequences (22). From these results, one could speculate that the kinetics of assembly for kissing loops are dictated primarily by the probability of a successful collision, resulting in formation of the Watson-Crick base pairs within the contact. This notion is quite powerful because it raises the possibility that the folding kinetics of structured RNAs can be predicted from the properties of the individual tertiary contacts, at least those structured RNAs that involve kissing loops.

Another common long-range interaction is the tetraloop-tetraloop receptor (TL/TLR) motif, in which the loop of one helix docks into the minor groove of the receptor helix (Figure 1.2B). By comparing the solution structure of the tetraloop

obtained by NMR with RNA crystal structures containing a TL/TLR motif, it's clear that the tetraloop is a preformed structure module while the receptor undergoes significant conformational changes upon binding (23-25). NMR measurements using a *trans* TL/TLR assembly system and smFRET measurements using a *cis* assembly system showed that, despite some conformational heterogeneity, the tertiary contact is able to form in isolation (26). This result further implies modularity to the structural connections in RNA.

A-minor tertiary contacts are long-range interactions involving docking of a distant helix. These contacts involve the docking of an adenosine residue into the minor groove of a Watson-Crick base pair and are critical for biological functions such as decoding in the ribosome (27). One type of a-minor motif discussed in more detail below is the metal core metal core-receptor (MC/MCR) tertiary interaction, where a 5 nucleotide bulge docks into the minor groove of an adjacent helix. Only within the last few years was it found that this complex element of tertiary structure could also readily fold in an isolated context, as observed with smFRET (Figure 1.2C) (28). These results strongly suggest that RNA tertiary structure is defined by modules of structure that form readily in isolation in support of the hierarchical folding model. A comparison of the measured rates for different modules of RNA structure is shown in Table 1.1. However, the energetic independence between secondary and tertiary structures hardly holds for larger more complex structured RNAs that may undergo subtle or even dramatic changes in base pairing when tertiary structure forms.

Structured RNAs in biology and bioengineering

RNAs begin to fold during transcription in the cell, so regulatory mechanisms that are linked to RNA folding begin with the biogenesis of the RNA in the nucleus. Because

secondary structure forms first during folding, non-native base pairs can form and persist for biological RNAs. Nature has harnessed the switch between alternative and native RNA base pair conformations to regulate cellular processes. These switches violate the energetic independence between structured tiers in the RNA folding hierarchy, making it critical to understand how regulatory RNA switches fold. Additionally, selection techniques have uncovered RNA structures that are used as experimental tools (spinach RNA) and in the treatment of disease (pegaptanib).

RNA secondary structure switches

It has become clear that RNAs are quite dynamic (29). The most prevalent example of secondary structure switches in both bacteria and eukaryotes are riboswitches (30). These RNAs bind to metabolite ligands, resulting in conformational changes at the tertiary structure level that are linked to changes in the base pairing of the riboswitch (Figure 1.3A). Regulation through RNA structure is achieved by the change in secondary structure that results from ligand binding. The well characterized *S*-adenosylmethionine (SAM) riboswitch is co-transcriptionally regulated by SAM binding, resulting in the formation of a terminator hairpin and RNA degradation ('off' state) (31). In contrast, in the absence of SAM, an alternative secondary structure forms that sequesters nucleotides that would participate in the terminator hairpin structure, thus forming an anti-terminator hairpin and allowing RNA transcription to proceed ('on' state) (32-35).

Regulation can also be achieved at the translational level. The cobalamine B12 forms a complex with metabolite ligand that directly sequesters the Shine-Dalgarno sequence used for ribosome binding and translation initiation (36, 37). In eukaryotes, regulation is achieved at the level of alternative splicing, resulting in mRNA products containing internal stop codons (38, 39). Understanding the fundamental principles that

define the relative abundance of structured intermediates and folding pathway followed for linked secondary and tertiary folding transitions is critical for understanding RNA gene regulation.

Structured RNAs as bioengineering tools and therapeutics

The well-known green fluorescent protein (GFP) is commonly used for various molecular biology applications including, but not limited to, in cell imaging to fluorescent activated cell sorting (FACS) techniques. GFP is a practical tool for studying proteins that are fused to a GFP reporter and then translated in cell, but tracking RNAs in cell is not easily achieved with this approach.

To overcome this barrier, a green fluorescent RNA was developed using a SELEX approach that selects for RNA apatamers that bind tightly to a particular ligand (40-43). The ligands designed for these experiments were derivatives of 4-hydroxy-benzylidene-imidazolinone (HBI), the fluorophore at the center of GFP. The ligand (Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one (DFHBI) binds to the spinach RNA aptamer (K_d~500 nM) and is fluorescent upon radiation with UV light (365 nm) (44). Interestingly, when the sequence of the spinach RNA was determined, it was unclear how the DFHBI ligand might bind to the RNA based the predicted secondary structure (Figure 1.3B). These results suggest that ligand binding changes the tertiary structure of the spinach RNA in a way that completely reformats the secondary structure, a feature discussed previously for riboswitches. Indeed, crystal structures of the spinach RNA aptamer in complex with DFHBI show that the DFHBI stacks in the center of a G-quartet, stacking within a G-quadruplex structure at the heart of the spinach RNA (45-47).

I would like to briefly mention the RNA drug 'Pegaptanib' which is actively used for the treatment of neovascular age-related macular degeneration (AMD) (48-50). This RNA aptamer binds to the vascular endothelial growth factor (VEGF) and inhibits VEGF activity. The manipulability and adaptability to high throughput approaches makes RNA an attractive molecule for developing aptamers that can target disease at the molecular level.

Convergent evolution of RNA modules

After it was discovered that RNAs can possess catalytic activity similar to protein enzymes, there was a realization in the field that RNA could serve as both genetic material and catalyst. The proposal of an 'RNA World' preexisting DNA and proteins, which had come into existence decades earlier, was dramatically increased in momentum as a result of this discovery (51). Group I introns and other self-splicing autocatalytic RNAs with reversible chemistry were particularly important for this idea because they could drive evolution by mobilizing themselves out of and into different RNA sequences (52). Consistent with this model, group I introns are found in rRNA, tRNA, protein coding RNAs, mtRNAs, plant chloroplast DNA, and bacteriophages (53). Below, I compare and illustrate the similarities in structural architecture observed for three different introns from a bacteriophage (*Staphylococcus aureus*), a bacterium (*Azoarcus*), and a lower eukaryote (*Tetrahymena thermophila*). The common structural architecture of these group I introns from distantly related organisms and the ubiquitous distribution of the TL/TLR motif in RNA suggest that structured RNAs have propagated through horizontal gene transfer, and at the level of structural modules have evolved through convergent evolution.

A common architecture for Group I intron catalytic cores

A comparison between secondary and high-resolution tertiary structures of Group I introns from *Azoarcus*, *Tetrahymena thermophila*, and bacteriophage *Twort* illustrates the structural conservation of the Group I intron catalytic core (54-57). The conserved catalytic core is comprised of two coaxially stacked helices, P4-P6 and P3-P9, which surround the ribozyme active site (Figure 1.4A and B). There are structural differences between introns in the peripheral domains that flank the catalytic core (Figure 1.4C), but removal of these domains does not fully eliminate catalytic activity (58, 59), suggesting that only the core is required for catalysis. Indeed, ribozyme catalytic activity can be reconstituted from a two-piece trans assembly system with the P4-P6 and P3-P9 helices of the *Tet* ribozyme (60).

Additionally, the same tertiary contact motif is present in all three Group I introns, Group II introns, and RNase P, the previously mentioned TL/TLR (25). These long range tertiary contacts, and others, stabilize the ring of tertiary contacts that form around the intron active site (61-63). Even though the architecture of the folded group I introns are largely conserved and the content of tertiary contacts are similar, there are no direct correlations between observed folding rate and intron size or the order of tertiary contact formation, suggesting that introns fold through unique folding pathways to the folded structure (64). It's possible that the different peripheral elements substantially contribute to the folding pathway followed. A greater understanding of the fundamental factors that determine RNA folding flux through alternative assembly pathways is needed to address these models.

The Tetrahymena thermophila Group I intron as a model RNA

The *Tetrahymena thermophila* ribozyme is an attractive model system for investigating questions in RNA structure and RNA folding because of the wealth of

preexisting structural and biochemical data (55, 65, 66). While early folding studies followed folding of the entire intron or ribozyme to the native state, it was shown that the measurable steps in this process represent escape from misfolded kinetic traps that accumulate during folding (67). Thus, to directly probe folding steps that involve accumulation of structure during RNA folding and assembly, smaller, more tractable structured RNAs are valuable tools. Within the P4-P6 domain of the ribozyme is the P5abc subdomain, which can functionally separated from the *Tetrahymena thermophila* ribozyme core, folds in isolation, and can be reconstituted to form an active P5abc-ribozyme core complex. In my work, I have quantitatively dissected the folding and assembly of the P5abc subdomain with the ribozyme core and demystified literature confounding our understanding of cooperative RNA folding and assembly processes.

P5abc Folding: the P5c switch and metal core (MC)

The P5abc subdomain has three helices (P5a/P5b/P5c) that are connected by single stranded linkers that converge at a three-way junction (Figure 1.5A). From analysis of RNAs with three-way junction topology in the protein data bank, a classification system was developed based on the number of linking nucleotides connected the three helices (68). The P5abc subdomain was classified as 'Family C' because the 3' most junction has more nucleotides than the other junctions (Figure 1.5B). It is not known how the nucleotides in the junction contribute to folding.

P5abc folds in isolation when separated from the ribozyme (69). In the presence of modest concentration of monovalent cations, an alternative secondary structure is adopted and the three helices extend into solution (Figure 1.5A). Upon addition of Mg^{2+}, tertiary structure forms in the metal core (MC) of P5a and there is a shift in the base pairing of seventeen nucleotides primarily in the P5c helix (70, 71) (Figure 1.5C). The

secondary structure rearrangement in P5c reorganizes the nucleotides in the junction region and reformats the P5c loop (L5c) to include an additional uracil (GCAA→UGCAA). By making point mutants in P5c that stabilize the alternative secondary structure over the native secondary structure, it was demonstrated that tertiary structure formation is dependent on switching to the native P5c secondary structure, inconsistent with the separability of secondary and tertiary structure tiers in the RNA folding hierarchy (72). A more recent study in 2012 obtained confounding results when they found that mutations in P5c that stabilize either the alternative or native secondary structure had no effect on MC folding equilibrium or kinetics (73). The authors concluded that the transition state for folding did not depend on P5c and made no interpretation about the lack of an effect on equilibrium, and they did nothing to test if their mutants indeed had the anticipated effects. To develop a predictive and quantitative understanding of the RNA folding landscape requires a physical description of how tertiary structure formation in the MC is directly coupled to P5c switching.

Assembly of P5abc and the ribozyme core

When P5abc is removed from the full length *Tetrahymena* ribozyme, the P5abc-deleted ribozyme core ($E^{\Delta P5abc}$) can fold and still carry out activity (Figure 1.6). $E^{\Delta P5abc}$ folds in isolation and can robustly assemble with P5abc to form an 'exceptionally stable' complex (74, 75). To gain structural insight into the assembled complex, footprinting approaches were used, such as hydroxyl radical footprinting, an approach that preferentially cleaves the RNA phosphate backbone in regions that are exposed to solvent (76, 77). Cleaved RNA products can be resolved on a sequencing gel to determine with reasonable precision the regions that are protected from the solution probe. When P5abc was mixed with $E^{\Delta P5abc}$, protections were observed at the interface of the ribozyme core

and at tertiary contact locations within P5abc (Figure 1.6) (74). Additional measurements analyzing the onset of catalytic activity for the assembled $E^{\Delta P5abc}$-P5abc complex demonstrated that the two structures are preformed, and when P5abc assembles with the natively folded $E^{\Delta P5abc}$, neither partner must undergo extensive refolding rearrangements during assembly (78). (As a separate observation, approximately half of the $E^{\Delta P5abc}$ is misfolded at equilibrium, and the complex with P5abc remains misfolded before slowly refolding to the more stable, tighter binding native complex (75)).

The data strongly suggest that the two RNAs are structural modules that are preformed and rapidly assemble under conditions that support folding of P5abc and $E^{\Delta P5abc}$. As mentioned above, P5abc also undergoes a cooperative secondary and tertiary folding transition. It is not known how this cooperative folding transition contributes to the rate of folding or to the pathway followed during RNA assembly. What are the features of tertiary structure formation that dictate how RNAs assemble? I strive to answer these questions in my studies here.

Cooperativity in RNA tertiary structure formation

When this project was first started, there were few fundamental studies that directly probed cooperativity in RNA tertiary structure formation, and the aforementioned results on P5abc folding were confounding. Still, one study in particular probing the cooperativity between two tertiary contacts in P4-P6 provided a framework for quantitative predictions necessary to establish cooperativity in RNA folding. So, a reconstitution model is developed to test if a linked secondary and tertiary folding and assembly process can be understood from the properties of the individual structural modules.

Cooperativity of two tertiary contacts in the P4-P6 domain

The P4-P6 domain folds in isolation from the ribozyme and adopts a sharp 180° bend that accompanies the formation of two long-range tertiary contacts with the adjacent P5abc subdomain. The folding over of the P4-P6 domain is well suited to single molecule fluorescence resonance energy transfer (smFRET) measurements because the folding transition brings two helices into close proximity (79). The two tertiary contacts that form during P4-P6 folding can be knocked out directly by mutagenesis. P4-P6 variant RNAs were made with each of the two tertiary contacts knocked out and the folding equilibria were measured (80, 81). The authors observed that folding of both tertiary contacts was better than either one alone, amounting to 3.2 kcal/mol of cooperativity between the formation of these two tertiary contacts. It was proposed that the cooperativity might arise from paying an electrostatic penalty associated with bringing the two helices together, or by better orienting the second forming tertiary contact for more successful collision probability.

A reconstitution model to predict cooperative secondary and tertiary RNA folding thermodynamics and kinetics

From the reductionist measurements conducted on smaller domains and subdomains of RNA structure, it appears that RNA structure is composed of smaller units of structure that can fold in isolation (modular). It is less clear how secondary and tertiary structure change impacts RNA-RNA assembly.

To address these questions, a conformational search and capture model for assembly through one pathway was generated (Figure 1.7). If the conformational search step is in rapid equilibrium relative to the capture step, then the folding rate through this pathway can be predicted if the rate constant for capture is known (k_2) and the pre-equilibrium is measured (K_1). The predicted rate constants can then be directly tested

against measurements of the assembly kinetics through this folding pathway. In these studies (Chapter 2), the limits of predictability for the reconstitution hypothesis are tested, leading to a greater understanding of RNA folding and assembly and interpretation of kinetics data.

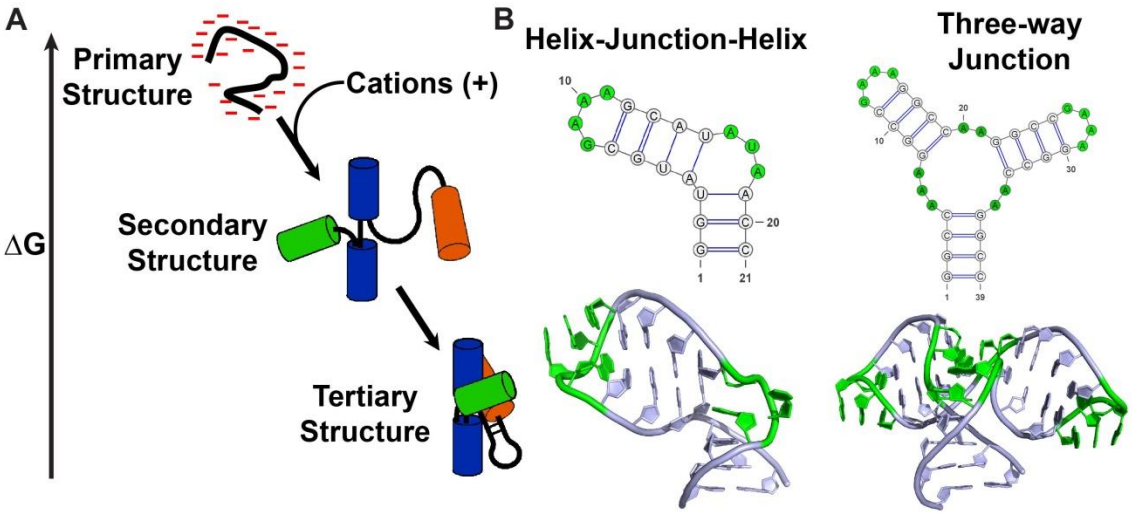

Figure 1.1: The energetic tiers of RNA structure.

(*A*) Simple free energy diagram of RNA structure formation and hierarchy. (*B*) RNA secondary structural motifs. Unpaired nucleotides are highlighted in green. RNA structures were generated using RNAComposer (82).

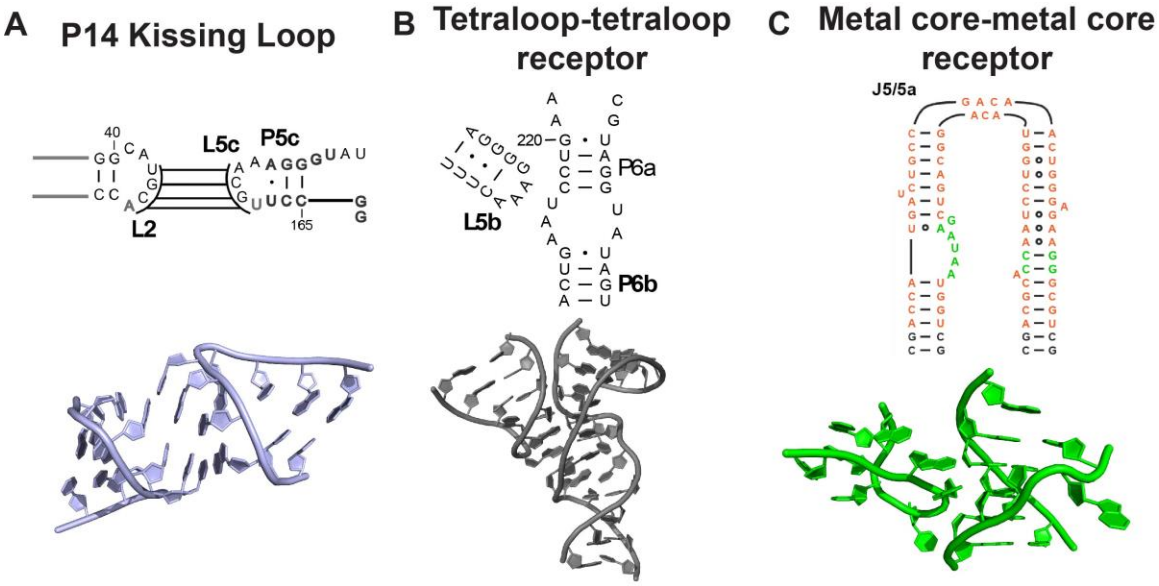

Figure 1.2: Isolated modules of tertiary structure.

(*A*-*C*) Structures shown are from the *Tetrahymena thermophila* ribozyme (63, 65).

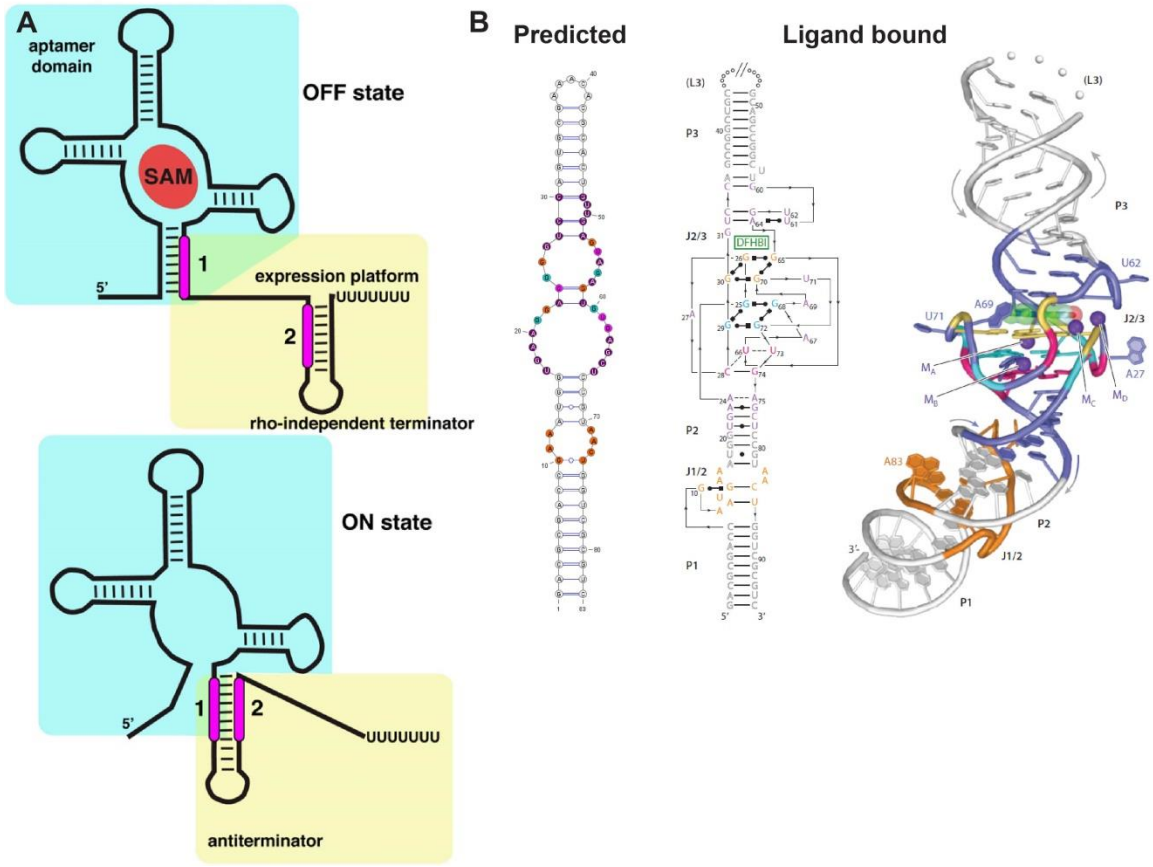

Figure 1.3: Examples of RNA secondary structure switches.

(*A*) Co-transcriptional folding of the SAM-I riboswitch. The aptamer domain is highlighted in teal, SAM ligand is red, the expression platform is yellow, and the relevant nucleotides in the base pair switch are magenta. (*B*) The predicted and ligand bound secondary structure of the Spinach RNA aptamer. The ligand DFHBI is highlighted in a green box. Figures are adapted from (83) and (47) with permissions from the publisher.

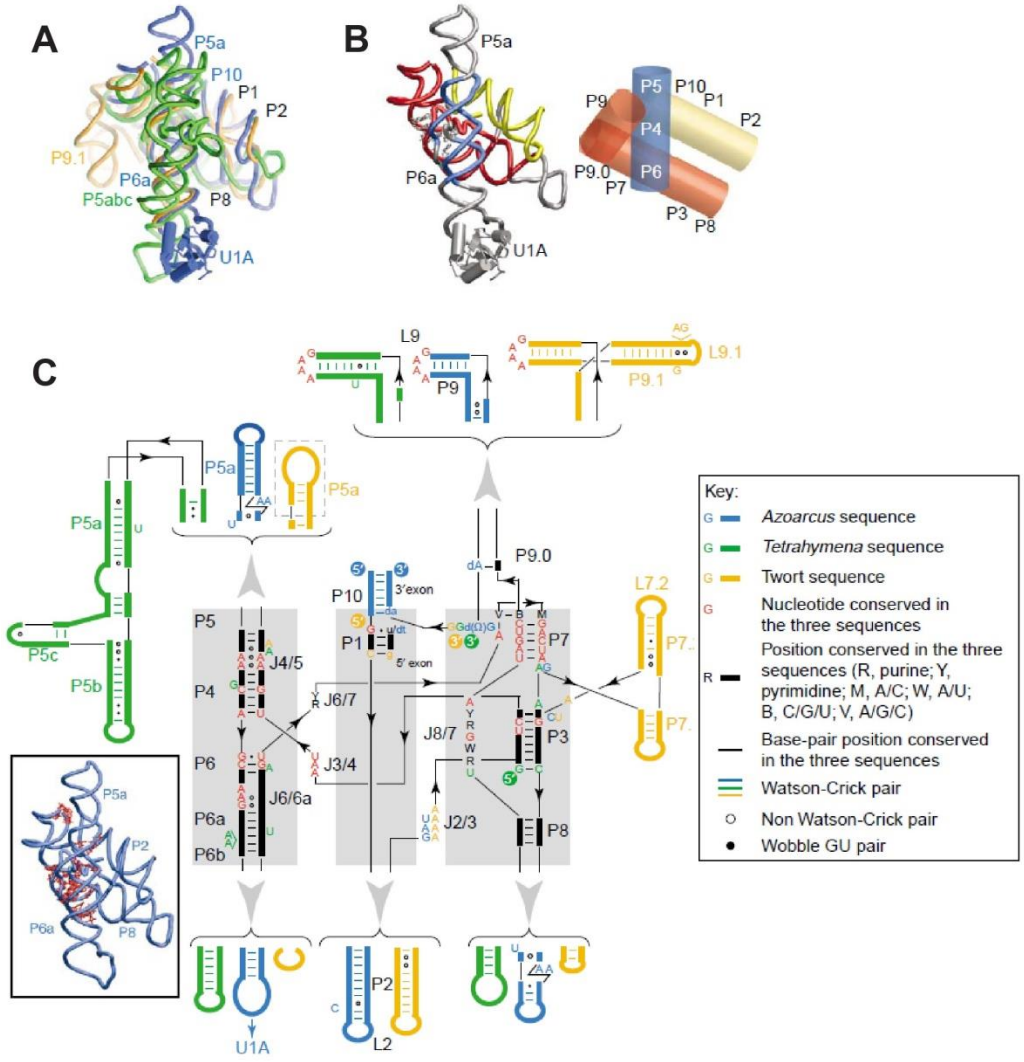

Figure 1.4: Structural similarities between group I introns.

(*A*) Superimposition of ribbon structures from three different group I introns. *Azoarcus* is blue (84), *Tetrahymena* is green (55), and *Twort* is yellow (57). (*B*) The conserved catalytic core of group I introns. (*C*) Peripheral domains that flank the conserved catalytic core. Figure is adapted from (54) with permissions from the publisher.

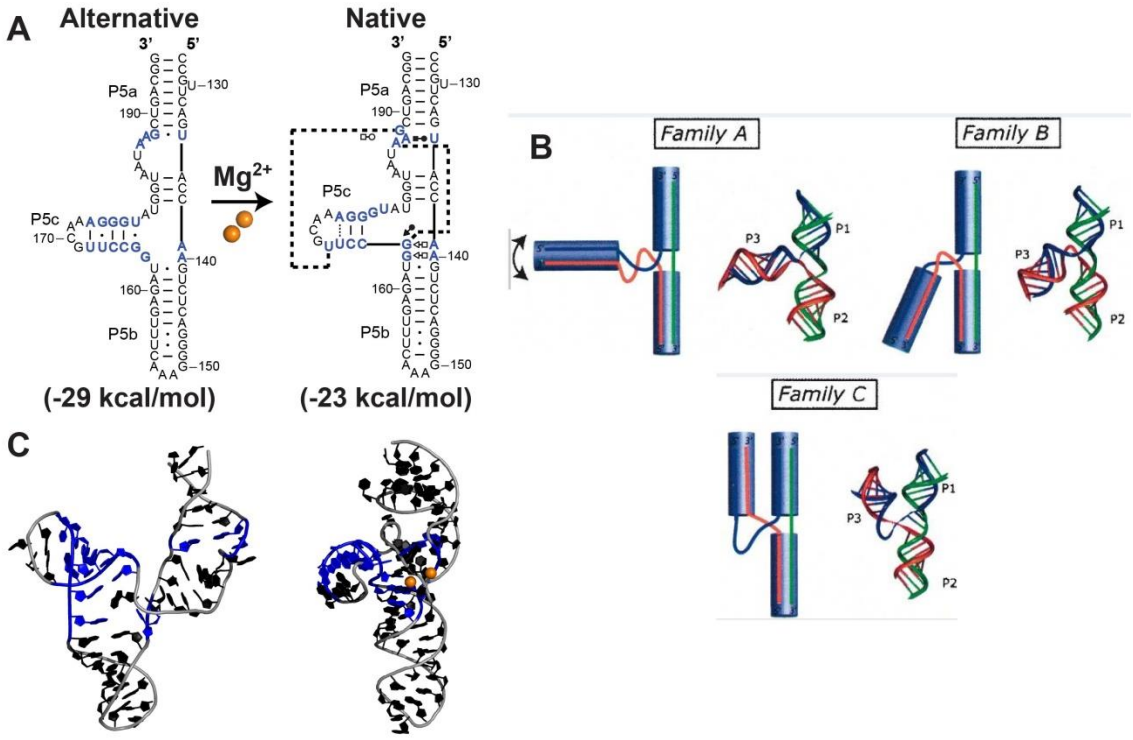

Figure 1.5: The P5abc subdomain changes secondary and tertiary structure during folding.

(*A*) Secondary structure diagrams of the alternative and native P5abc. The calculated stability of each structure was calculated with ViennaRNA (85). (*B*) Classification of different family types in folded RNAs with three-way junction topologies. P5abc is part of Family C. Figure is adapted from (68). (*C*) NMR structure of a truncated version of P5abc (tP5abc) in the absence of Mg^{2+} (70) and X-ray structure of P5abc within the P4-P6 domain (65).

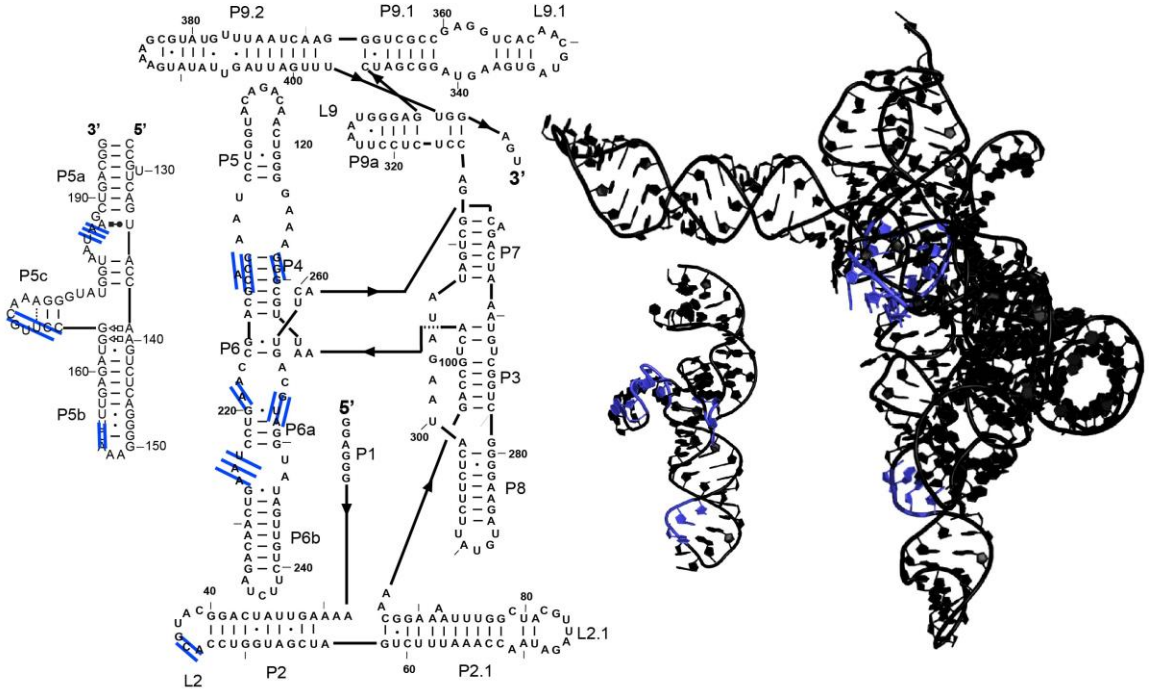

Figure 1.6: P5abc assembles with $E^{\Delta P5abc}$ (P5abc-deleted ribozyme core).

The secondary structure and a model three dimensional structure of $E^{\Delta P5abc}$ (63). Protections observed by hydroxyl radical footprinting in the presence of P5abc relative to the absence of P5abc are indicated with light blue lines (74).

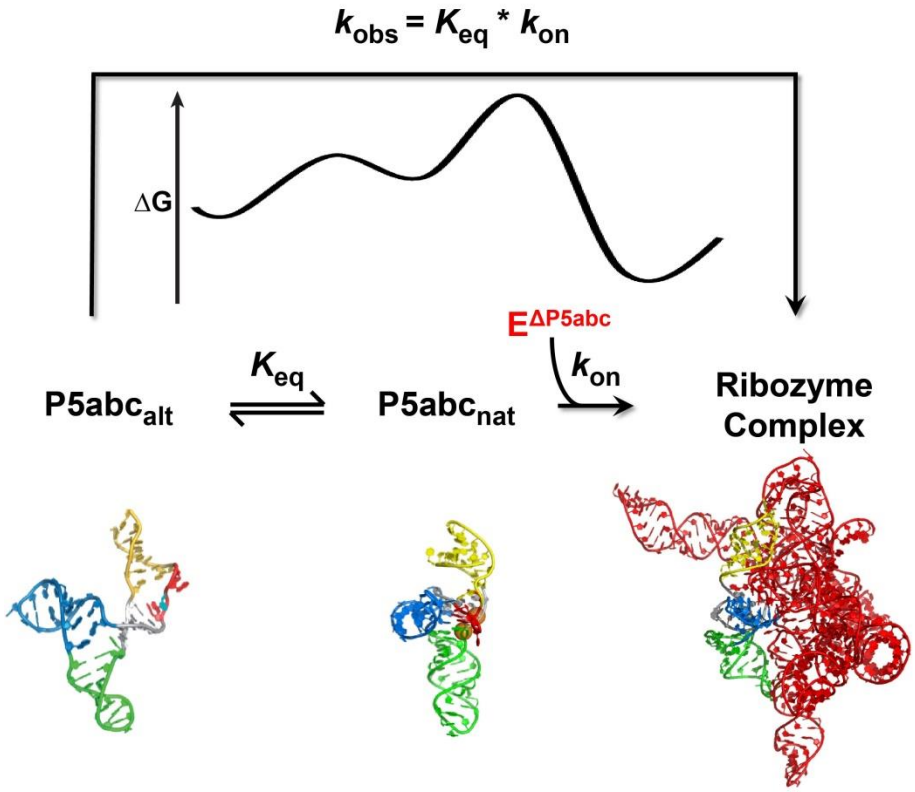

Figure 1.7: A reconstitution model to quantitatively predict RNA folding.

Structural Element	k_f (M^{-1} s^{-1})	k_u (s^{-1})
RNA Duplex	$10^5 – 10^7$	$10^{-6} – 10^{-8}$
Kissing loop (22)	3×10^6	0.17
Tetraloop / tetraloop-receptor (26, 86)	1.7×10^3	3.3
Metal core / metal core receptor (a-minor) (28)	1×10^4	37

Table 1.1: Kinetics measured for isolated RNA structural motifs.

Chapter 2: Dissection of a cooperative RNA secondary and tertiary conformational switch

INTRODUCTION

Structured RNAs function in a myriad of cellular processes that include protein production and RNA expression, and these RNAs are currently under active investigation as therapeutic targets (49, 50, 87-90). As structural information about functional RNAs has grown it has become apparent that complex RNA/protein complexes such as the spliceosome, the ribosome, and telomerase undergo orchestrated conformational transitions that enable them to switch between functional states as they carry out complex functions (29, 91-94).

A general feature of these RNA conformational steps is a coupling of structural changes in multiple regions of the complex, allowing transmission of information between these regions. For example, translocation by the ribosome involves coordinated movements that propagate to the interface of the large and small subunits and potential restructuring of the tertiary contacts between subunits (95, 96). Similarly, coupled rearrangements are observed in riboswitch signaling, where binding of a ligand results in an alternative structure being favored and a biological outcome through changes in transcription, translation, or splicing (30, 31, 97-99).

While in-depth functional and structural studies have allowed us to describe many of the conformational steps involved with function of RNA complexes (with still more to be uncovered), we know much less about the underlying energetics that drive these processes. An illustrative example comes from the P5abc subdomain of the Tetrahymena group I intron, which was identified as undergoing a conformational switch upon the addition of divalent cations (Figure 2.1A). Initial studies demonstrated a coupled rearrangement involving changes in the secondary and tertiary structure of the RNA, and

subsequent work and discussions implied an extremely high degree of cooperativity in this transition (70-73).

These observations raise the question of how these complex transitions are carried out. Despite the observation of a concerted transition, there is presumably a need for a kinetic pathway with a series of intermediates to provide a viable route to formation of the final state, as well as an evolutionary route through which the complex structure would have arisen. Indeed, recently we uncovered a 'partial' switch in P5abc, with a transient change to native secondary structure in the absence of the full native tertiary structure (100) and formation of a native tertiary structure element in the absence of the native secondary structure (101). A deep understanding of RNA structure and function requires knowledge of the physical and energetic basis of coupling between RNA structural elements.

Here, we built on the substantial prior work on P5abc to dissect the conformational steps in its concerted folding transition. Our results demystify the molecular events in this cooperative transition, quantify the underlying energetics, and unravel the structural connections. We find that this transition is multiple steps and modular, with modest levels of coupling between native secondary and tertiary structure formation at each step. The progressive increase in cooperativity likely functions to specify the native, functional state of the RNA, while the modular, independent folding of each element enables a kinetically accessible path to the native state.

Results

Structure and folding of P5abc

P5abc is composed of three helical elements (P5a/P5b/P5c) that converge at a three-way junction (3WJ) (Figure 2.1A) (68). In P5abc folding, two Mg^{2+} ions bind

specifically to sites in P5a at an A-rich bulge termed the metal core (MC), resulting in the formation of a 270° corkscrew in the RNA backbone built upon these two site-bound Mg^{2+} ions (Figure 2.1A, bottom right structure) (65, 69, 102-104). The formation of tertiary structure in the MC is accompanied by a change in the base pairing of 17 nucleotides, including a shift in the register of P5c and the formation of non-canonical base pairs in the 3WJ (Figure 2.1A).

Mg^{2+} concentration dependence of P5abc folding reveals multiple transitions

To monitor equilibrium folding, we incubated P5abc with various concentrations of Mg^{2+} (0.02-200 mM) and performed SHAPE and DMS footprinting (Figure 2.1B). In contrast to the simplest expectation from early results, the Mg^{2+} dependences were complex, with a range of Mg^{2+} midpoints and some nucleotides with multiple phases (Figure 2.3A and B, Table 2.1, and Table 2.2; see also Figure 2.1D for corresponding pseudo-gel image). These data indicate that P5abc folding is not a two-state process. To derive a minimal model, we fit all of the footprinting data with a global model using KinTek Global Kinetic Explorer (see Methods section: *Global modeling of SHAPE and DMS footprinting data*). The global fitting gave a minimal model with four discrete states and three Mg^{2+}-dependent transitions between them (Figure 2.4 and Table 2.3).

At sub-millimolar concentrations of Mg^{2+}, several nucleotides in the 3WJ and the MC display modest enhancements of reactivity with increasing Mg^{2+} concentration (Figure 2.5A, B, and C; U to U*). These enhancements are also observed when Mg^{2+} is replaced with Ba^{2+}, which does not support MC formation (Figure 2.5D) (105). Further, when divalent cations are added in the presence of 1 M NaCl, the enhancement is not observed (Figure 2.5E) (106). These results suggest that this transition reflects alleviation of electrostatic repulsion (107, 108), and it is not associated with secondary or tertiary

structure changes in P5c or the MC. Therefore, it is not discussed further and U and U* are collectively referred to as 'Alt'.

As the Mg^{2+} concentration is further increased, a second transition is observed, with protections in several regions of P5abc. Nucleotides within the MC and the 3WJ become protected from the SHAPE reagent (Figure 2.3C and D), indicating structure formation, and enhancements at key positions within P5c indicate the formation of native secondary structure (Figure 2.3E). These changes occur with near overlapping Mg^{2+} dependences and are cooperative with Mg^{2+} concentration ($n = 2.1 \pm 0.1$), as expected for a transition linked to cooperative binding of two Mg^{2+} ions in the MC (73). Thus, we conclude that in the second transition Mg^{2+} binds specifically to the MC and the MC forms tertiary structure, P5c forms the native secondary structure, and the 3WJ forms tertiary structure.

At even higher Mg^{2+} concentrations, a third transition is observed by SHAPE (Figure 2.3B, dashed boxes, Figure 2.3E and F, Table 2.1, and Table 2.2). These nucleotides are within loop L5c, the 3WJ, and the MC. The distributed nature of these changes and their positions suggest native tertiary packing of P5c into a groove formed by the MC and 3WJ to give the conformation observed in crystal structures of P4-P6 (Figure 2.1A, bottom right structure) (65, 66) and the ribozyme core (55).

Quantitating cooperativity between the P5c switch and the MC

The coincident formation of native P5c secondary structure and MC tertiary structure in the transition from Alt to I_F provided support for the previous conclusion that these structural elements fold cooperatively (70-73). To test this model further and to measure the cooperativity, we used mutagenesis to block one of these structural elements from folding, and then we measured folding of the other element by SHAPE footprinting.

A P5abc variant that stabilizes the alternative secondary structure of P5c (U167C) (72, 100, 101) displayed clear protections that indicated formation of the MC, but with an increased $[Mg^{2+}]_{1/2}$ value of 9.8 mM (Figure 2.6A and B, Table 2.1, Figure 2.7). The increased Mg^{2+} requirement indicates cooperativity between the P5c secondary structure change and MC formation.

Two additional results show that for this mutant, when MC forms at the higher Mg^{2+} concentration, P5c remains in the alternative secondary structure, and thus that this measurement reveals the full extent of cooperativity. First, there was no visible Mg^{2+}-dependent transition in P5c up to 50 mM Mg^{2+} (Figure 2.7C). Second, a double mutant in P5c (U167C/U177C) that further stabilizes the alternative P5c secondary structure (101) gave the same Mg^{2+} dependence for MC nucleotides as the single mutant (Figure 2.6A and B). Together, these results show that MC formation and P5c secondary structure switching are thermodynamically coupled, but that this coupling is sufficiently modest for folding of the MC to be readily observed when P5c is prevented from switching to its native secondary structure.

We also carried out the converse experiment, blocking MC formation by replacing A186 with U or the entire A-rich bulge with uridines (80, 106, 109) and monitoring the secondary structure transition of P5c by SHAPE (Figure 2.6C). We observed enhancements within P5c indicating native secondary structure formation (Figure 2.6D and Figure 2.8). Again, the Mg^{2+} requirement was increased, and the lack of a signal at MC nucleotides in these mutants showed that MC structure formation was indeed disrupted (Figure 2.8C). Thus, the P5c secondary structure rearrangement can occur without MC formation.

To evaluate the cooperativity quantitatively, we focused on a Mg^{2+} concentration in the transition range for both the wild-type P5abc and the mutants. At 3 mM Mg^{2+}, the

equilibrium values could be determined directly from the Mg^{2+} dependences without extrapolation. Comparison of these values shows that the coupling between P5c and the MC in this transition is modest, with a value of 1.3 kcal/mol (Figure 2.6E).

A strong prediction of our results is that a mutant P5abc that favors the native P5c secondary structure would decrease the Mg^{2+} requirement for MC formation, because the wild type must 'pay' to switch the P5c secondary structure, whereas the mutants would already have native P5c formed (Figure 2.6E, bottom right equilibrium K'_{MC}). Indeed, a mutant that stabilizes native P5c by inserting three additional Watson-Crick base pairs (Nat+3) displayed a reduced Mg^{2+} requirement for folding the MC (Figure 2.6A and B and Figure 2.9A). The equilibrium value at 3 mM Mg^{2+} was 7 ± 1, even a bit larger than the predicted value of 3 ± 1. Interestingly, protections in the 3WJ were also responsive to the stability of native P5c, as well as the MC, suggesting cooperativity between the MC and the 3WJ (Figure 2.6F).

Together, these results indicate that P5c switching and MC tertiary structure formation are modular and cooperative. Each transition occurs readily in the absence of the other, and the two transitions are energetically linked, such that each transition is more favorable in the presence of the other, with modest cooperativity comparable to the formation of 1-2 hydrogen bonds in many systems (110).

Structural origin of cooperativity in P5abc

P5c and the MC are connected by the 3WJ, making it a leading candidate for how 'information' on the folding of P5c could be transmitted to the MC and vice versa (see Appendix A: *Structural connections in the 3WJ of P5abc*). Further, three nucleotides that are liberated in the transition to native P5c form contacts in the 3WJ (Figure 2.1A, green nucleotides in Alt P5c). To test the role of the 3WJ in cooperative folding of P5c and the

MC, we followed folding of a construct in which equivalents of these three nucleotides were inserted and available for contacts in the 3WJ without the need for P5c native secondary structure (Figure 2.10A). In the background of mutations to ensure that P5c remained in the alternative secondary structure (U167C/U177C), we found that the Mg^{2+} requirement for MC formation was reduced for this 3WJ restoration mutant, giving approximately the same dependence as the wild-type P5abc (Figure 2.9B and Figure 2.10B). Thus, cooperativity between P5c secondary structure switching and MC tertiary structure formation is mediated through the intervening 3WJ.

In the fully folded structure, P5c packs against the MC, forming long-range contacts that provide another possible source of cooperativity (Figure 2.11). Indeed, this packing transition would be expected to be cooperative with native structure because a long-range base pair between P5c and the MC (U168-G188, Figure 2.10C) is not possible without native structure, as both nucleotides form alternative local contacts (Figure 2.1A). We first tested the role of this base pair by mutating U168 (73) and measuring P5abc folding (Figure 2.10C). As expected, the P5c packing transition was absent (Figure 2.10D), while the lower Mg^{2+} transition that results in MC formation and P5c secondary structure remained intact (Figure 2.10D, A186 in open symbols). The high Mg^{2+} transition was also absent for the Nat+3 mutant, which blocks the U168-G188 base pair by extending the P5c helix (Figure 2.10D, A184 in closed symbols). Thus, this tertiary folding transition depends on native structure in P5c and involves the folded MC, providing experimental support for additional cooperativity between these two structural elements.

We next quantitated the cooperativity by ablating MC folding with the A186U substitution and following the P5c packing transition. To separate effects of MC formation on P5c secondary structure switching and to ensure that U168 is available to

interact with the MC, we used a variant that locks P5c into the native secondary structure (Figure 2.10E). We found that when MC formation was ablated, signals for P5c packing were also perturbed, indicating cooperativity between MC formation and P5c packing (Figure 2.10F). From the partial transition observed at high Mg^{2+} concentrations, we calculated a coupling value of 2.6 kcal/mol between P5c packing and MC formation (Figure 2.12).

Discussion

Cooperativity is a hallmark of biological macromolecules as it enables the population of a specific functional structure or structures in preference to the enormous number of alternative structures (111, 112). Here we used quantitative, high-throughput footprinting to dissect a complex RNA folding transition to two separable transitions, and we measured the cooperativity in each step and elucidated its origins (Figure 2.13A). In one transition, tertiary folding in the MC is linked to a secondary structure change in P5c by structure formed in the intervening 3WJ. In the second transition, the folded MC and P5c are mutually stabilized by tertiary packing of P5c against the folded MC. Overall, this cooperativity amounts to approximately 4 kcal/mol of stabilization.

The multi-step, cooperative folding of P5abc

Although cooperativity is present in each step of P5abc folding and the total is substantial, the values generated at each step are modest. Previously, this process had been described as being strictly concerted, but we find that P5abc folding is highly modular (Figure 2.13B). P5c can switch to the native secondary structure without the folded MC and the MC can form its tertiary structure without P5c. This latter result is consistent with earlier single molecule studies showing that the isolated MC can form tertiary contacts with its receptor in the P4 helix (80). Our results underscore that even a

modest amount of cooperativity is enough to make structural rearrangements appear to be strictly concerted.

General properties of cooperativity in RNA folding

The incremental, step-wise cooperativity we observe here may be a general feature of structured RNAs. Overall, tertiary packing of structured RNAs is relatively sparse, with extensive penetration by water, such that the RNA structural elements engage in discrete tertiary contacts with each other. Cooperativity can arise between local elements of native structure and long-range contacts, as we have observed here, or between tertiary contacts that are distant from each other due to effects on positioning and overcoming electrostatic repulsion (80). Thus, there are multiple mechanisms for generating cooperativity in structured RNAs, and it may be a general property of RNA that cooperativity increases in increments as native secondary structure and tertiary structure form progressively.

Limited cooperativity as a virtue for RNA folding kinetics

The modularity and limited cooperativity in P5abc could be important in biology by allowing a facile kinetic route to the fully folded state (113). While cooperativity serves a critical role for the thermodynamics of the folded state by ensuring that partially structured intermediates do not dominate the population, high cooperativity presents a potential problem for the kinetics of the folding process (Figure 2.13B and C). A corollary of high cooperativity is that the individual structural modules are unstable individually. During folding, the two contacts are unlikely to form simultaneously by chance, necessitating formation of one or even both contacts individually, followed by their collision and mutual stabilization. If each contact is highly unstable alone, the probability of its individual formation will be so low that the kinetics of global structure

formation may be too slow to permit biological function. In contrast, the presence of structural modules with significant independent stability provides a viable path through these steps to produce a functional RNA.

Methods

RNA preparation

DNA templates were prepared by assembly PCR with four partially overlapping oligonucleotides (Integrated DNA Technologies). Primer extension reactions were carried out using Phusion High-Fidelity DNA Polymerase (New England Bio-labs) followed by purification of full-length, double-stranded DNA templates by Magnetic Bead PCR Clean-up (Axygen). PCR-amplified DNA template (0.2 μM) was used in transcription reactions with T7 RNA polymerase as previously described (114). RNA transcripts were isolated by affinity column (Qiagen). We used RNA constructs in which P5abc is flanked on both sides by a reference sequence that was used for internal normalization (Figure 2.1C and D). The RNA also includes a 3′ primer site. The labeled FAM primer used for reverse transcription includes a 5′ poly-A tail and a 3′ RNA binding site that bridges the RNA to an oligo-dT magnetic bead for purification. After correcting for signal saturation, overmodification correction, background, and normalization to the internal control (115), the 5′ and 3′ hairpin loops were comparably reactive at all Mg^{2+} concentrations probed (Figure 2.2A) and independent data sets were highly reproducible (Figure 2.2B), supporting the accuracy of the quantitative analysis. Signals were normalized to the single stranded linkers such that a reactivity value of '1' indicates that a given nucleotide has the same reactivity as nucleotides within the 3′ single-stranded linker. The results were unchanged when the data were normalized to the 3′ hairpin loop or the loop L5b within P5abc.

SHAPE and DMS footprinting

200 nM RNA was pre-folded in 19 μL at the desired concentration of Mg^{2+} for 15 min at 37 °C in a background of 100 mM KCl and 50 mM K-MOPS, pH 8.0, and then equilibrated at 25 °C for 10 min. Longer incubation times up to 120 min did not change the results, suggesting that the incubation time is sufficient to reach equilibrium. RNA was chemically modified by adding 1 μL of 1m7 SHAPE reagent (0.42 μg / μL, 5% DMSO final) or DMS (10 mM DMS, 5% DMSO, 0.45% ethanol final) and incubating for 15 min. DMSO was included in DMS footprinting reactions because we found that DMSO decreased the Mg^{2+} requirement for P5c switching (116). The presence of 0.45% ethanol has no effect on the SHAPE footprinting profile of wild-type P5abc. A fluorescently labeled DNA primer (FAM) with a 5′ poly-A tail was annealed to oligo-dT magnetic beads (Poly-A Purist kit, Life Technologies), resulting in a 3′ overhang on the DNA primer that is fully complementary to the 3′ end of the RNA (Figure 2.1C). To purify modified RNA, 10 μL of the primer-bead mixture was incubated with RNAs for 10 min, washed twice with 200 μL of 70% ethanol, and left to dry at room temperature for 20 min. RNAs were reverse transcribed with 5 μL of Superscript III reverse transcriptase mix (Life Technologies) at 42 °C for 60 min, followed by RNA degradation with 0.2 M NaOH at 90 °C for 3 min. The solution was neutralized with 0.3 M HCl, 0.4 M Na-acetate, pH 4.5, and 700 mM NaCl. cDNAs were purified by oligo-dT magnetic beads and ethanol washed, as described above, for analysis by capillary electrophoresis (ABI 3730). Electrophoretic traces were aligned using HiTRACE and then quantified as described (115, 117).

Data Processing and Analysis

The presence of a transition in Mg^{2+} titrations was assessed by two criterions. First, only nucleotides that undergo a change in reactivity of at least 0.15 reactivity units

were considered significant. Second, only changes in reactivity that exceed the average of the error bars within the change were considered. Positions that result in significant blockage to reverse transcription in the absence of footprinting were not interpreted due to the high background. Changes in reactivity were fit to a sigmoidal binding equation to extract midpoints, amplitudes, and hill coefficients.

Global modeling of SHAPE and DMS footprinting data

We imported the quantified SHAPE and DMS footprinting data from 29 nucleotides into Kintek Kinetic Explorer and fit all the data using one global model. We globally fit the data to a four state model, as described in Results. A clear feature of the data was steeper folding transitions than would be expected for a non-cooperative two-state model. To enable one of the transitions to be cooperative with respect to Mg^{2+} binding, which is not explicitly allowed in the Kinetic Explorer software, we incorporated an additional state into the model, making a total of five states (Scheme 1) (one of these states would never be populated significantly but allows a single transition to involve binding of two Mg^{2+} ions). The observed reactivity for each nucleotide is defined as a linear combination of parameter scaling factors associated with each state in the model (Equation 1).

$$A \leftrightarrow B \leftrightarrow C \leftrightarrow D \leftrightarrow E \qquad \text{(Scheme 1)}$$

$$\text{Reactivity} = sa1*A + sb1*B + sc1*C + sd1*D + se1*E \qquad \text{(Equation 1)}$$

In the model of Scheme 1, there are three possibilities for a single cooperative folding step: the transition from A to C, from B to D, or C to E. We enforced these three possibilities by systematically locking the scaling factors for states A and B (sa1=sb1), for states B and C (sb1=sc1), or for states C and D (sc1=sd1). These constraints ensure that during global fitting the corresponding folding intermediate does not substantially

populate and thus forms cooperatively. After global fitting under each constraint, the condition where state C does not significantly populate (five state B-C coop) produced a global fit that agreed best with all the data and included a cooperative transition (118) (Figure 2.4 and Table 2.3).

We contracted the four state model to only have three states, and then we systematically constrained each species as the cooperative folding intermediate.

$A \leftrightarrow B \leftrightarrow C \leftrightarrow D$ (Scheme 2)

React = sa1*A + sb1*B + sc1*C + sd1*D (Equation 2)

Global fitting under constraints of B not significantly populating converged on a model where the first two transitions were non-cooperative, i.e. $K_1 < K_2$. When state C was constrained to be the cooperative folding intermediate, global fitting did converge on a cooperative model, but the $\chi 2$ value was 2-fold higher because parts of the data deviated from the model (30.1 vs. 63.7, Table 2.3). Taken together, global fitting of a model with four states including a cooperative folding transition did not account for all the footprinting data.

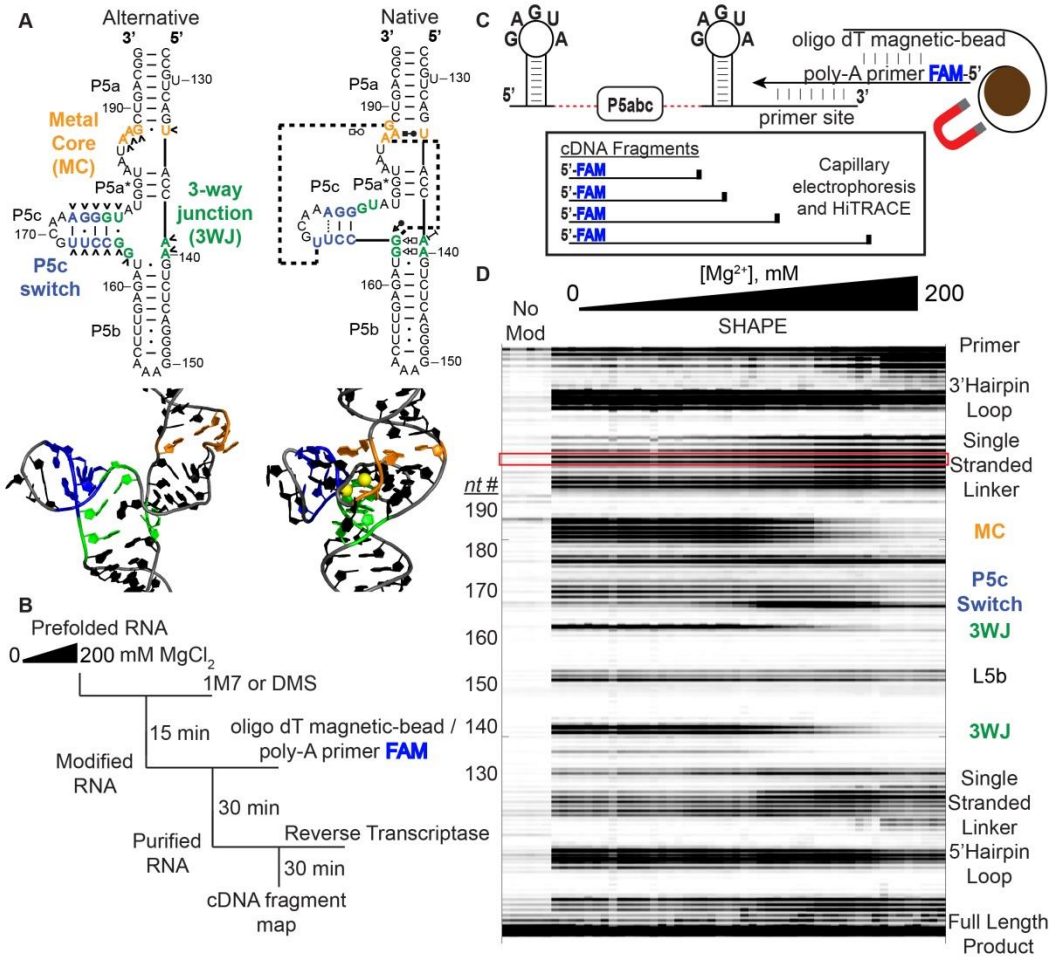

Figure 2.1: SHAPE and DMS footprinting approach.

Figure 2.1: continued.,

(*A*) Alternative and native structures of P5abc (left and right, respectively) (65, 70). Nucleotides that change base pairing during folding are marked by a '^' and colored (P5c, blue; 3WJ, green; MC, orange). Two site-bound Mg^{2+} ions are yellow in the native tertiary structure model (26, 31). Base pairs are shown here and elsewhere using Leontis/Westhof nomenclature (119). (*B*) Steps in the high-throughput footprinting approach. The FAM indicates that the primer is fluorescently labeled with fluorescein. (*C*) P5abc footprinting cassette. The RNA includes flanking hairpins linked by single-stranded segments used for internal normalization (red dashed lines). (*D*) Electrophoretic traces of reverse transcription products aligned and quantified by HiTRACE (117). The area under each peak is converted to gray scale and represented as a band (117). The red box indicates the nucleotides used for internal normalization. Eight control reactions without SHAPE reagent are shown at the left (No Mod) and span the Mg^{2+} concentrations used (0-200 mM Mg^{2+}). The nucleotide number (nt #) in the P5abc region is shown on the left edge.

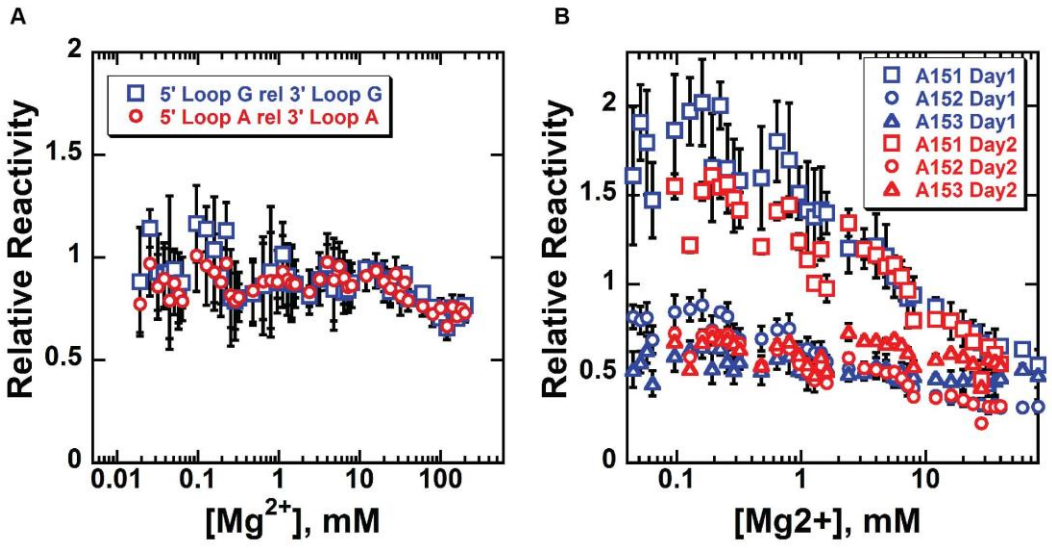

Figure 2.2: Quantitative analysis of footprinting data is reproducible.

(*A*) Reactivity values are calculated by dividing the reactivity of nucleotides within the 5' loop by the 3' loop. Two of the five loop nucleotides are shown in the plot (GAGUA). (*B*) Normalized reactivity values of L5b loop nucleotides measured on two different days. Error bars are propagated from the error associated with fitting peaks in the chromatogram.

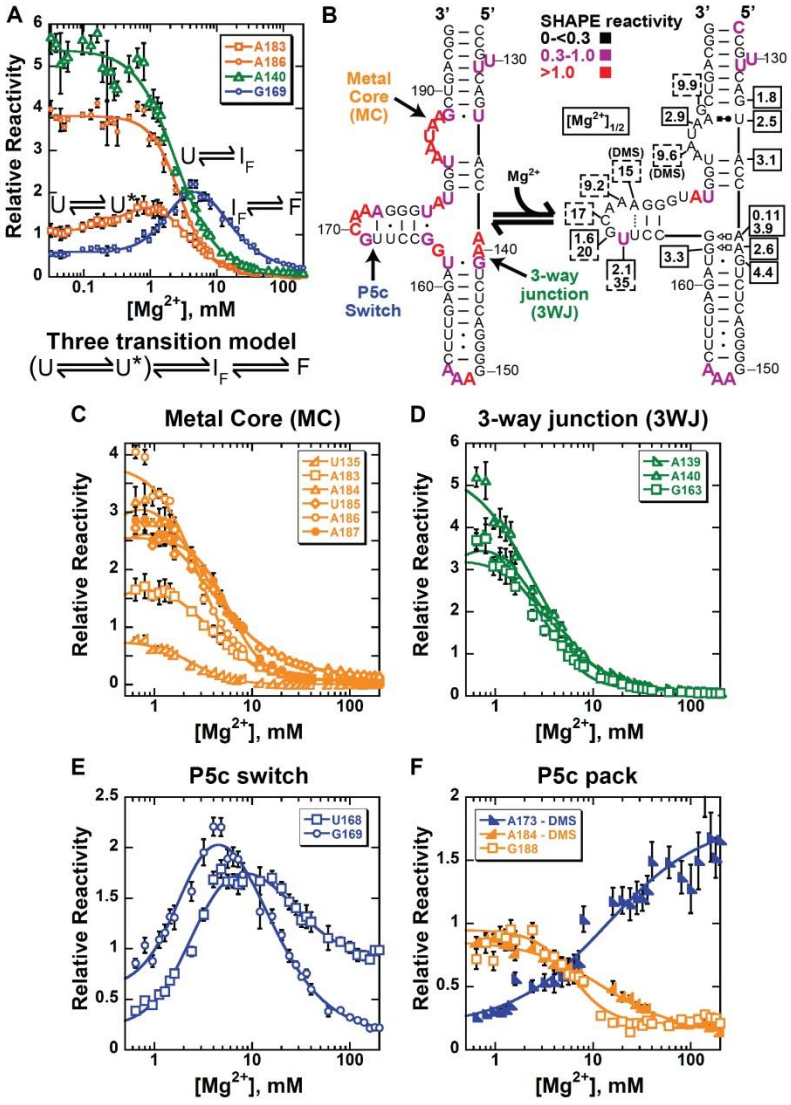

Figure 2.3: SHAPE and DMS footprinting of wild-type P5abc.

Figure 2.3: continued.,

(*A*) Mg^{2+} dependences of SHAPE reactivity for representative residues to illustrate the three transitions. Labels for each transition are displayed near the corresponding Mg^{2+} midpoint. The coloring is the same as in Figure 2.1. A minimal model is shown beneath the plot, with U to U* in parenthesis because this transition does not depend on the P5c switch or MC formation (Figure 2.5, below). (*B*) Normalized SHAPE reactivity values in the absence of Mg^{2+} and the presence of saturating Mg^{2+}. Mg^{2+} midpoints are indicated in boxes, with dashed boxes indicating higher Mg^{2+} concentrations required to fold. Not all measured midpoints are shown and two midpoints are shown for transitions measured by DMS (Table 2.2). (*C-E*) Mg^{2+} dependences of nucleotides that undergo changes in the U*→I_F transition within the MC (*C*), the three-way junction (3WJ, *D*), and P5c (*E*). (*F*) Mg^{2+} dependences of nucleotides that undergo changes in the I_F→F transition. In all panels, error bars reflect the residual values associated with peak fitting of the raw chromatograms.

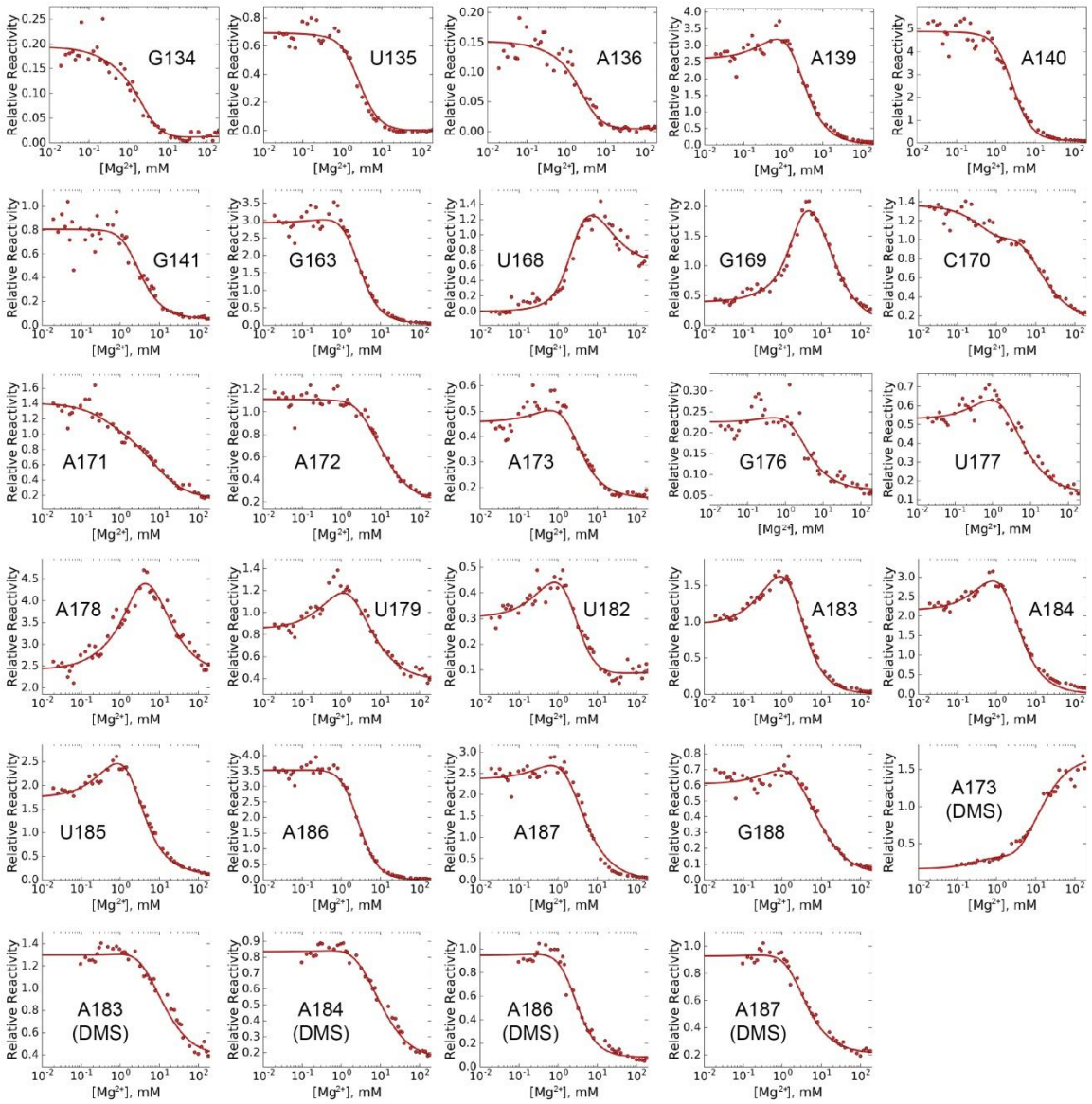

Figure 2.4: Global fitting of SHAPE and DMS footprinting data in Kintek Global Kinetic Explorer.

Red dots indicate the footprinting data and the curve shows the best fit by a global model, which included four states and three transitions. See Methods section: *Global modeling of SHAPE and DMS footprinting data* above for further details.

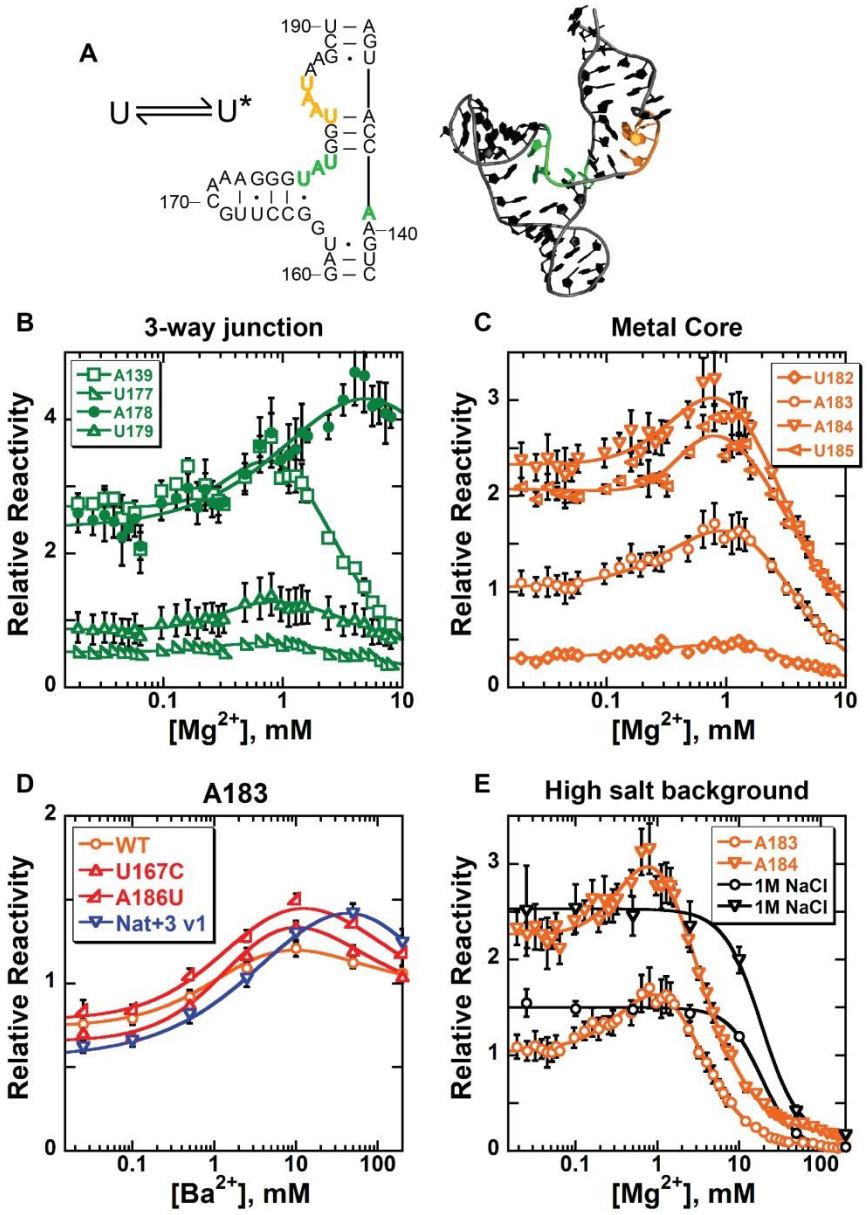

Figure 2.5: The first folding transition depends on electrostatic environment.

Figure 2.5: continued.,

(*A*) Nucleotides that underwent enhancements are in linking segments of the 3WJ (*B*) and in the MC (*C*). (*D*) SHAPE footprinting of mutants in Ba^{2+}, a divalent ion that does not support tertiary structure formation of the MC (105). (*E*) SHAPE footprinting of the MC in the presence of 1M NaCl (black) (106) eliminates the observed enhancement at sub-millimolar Mg^{2+} concentrations (shown in orange for comparison).

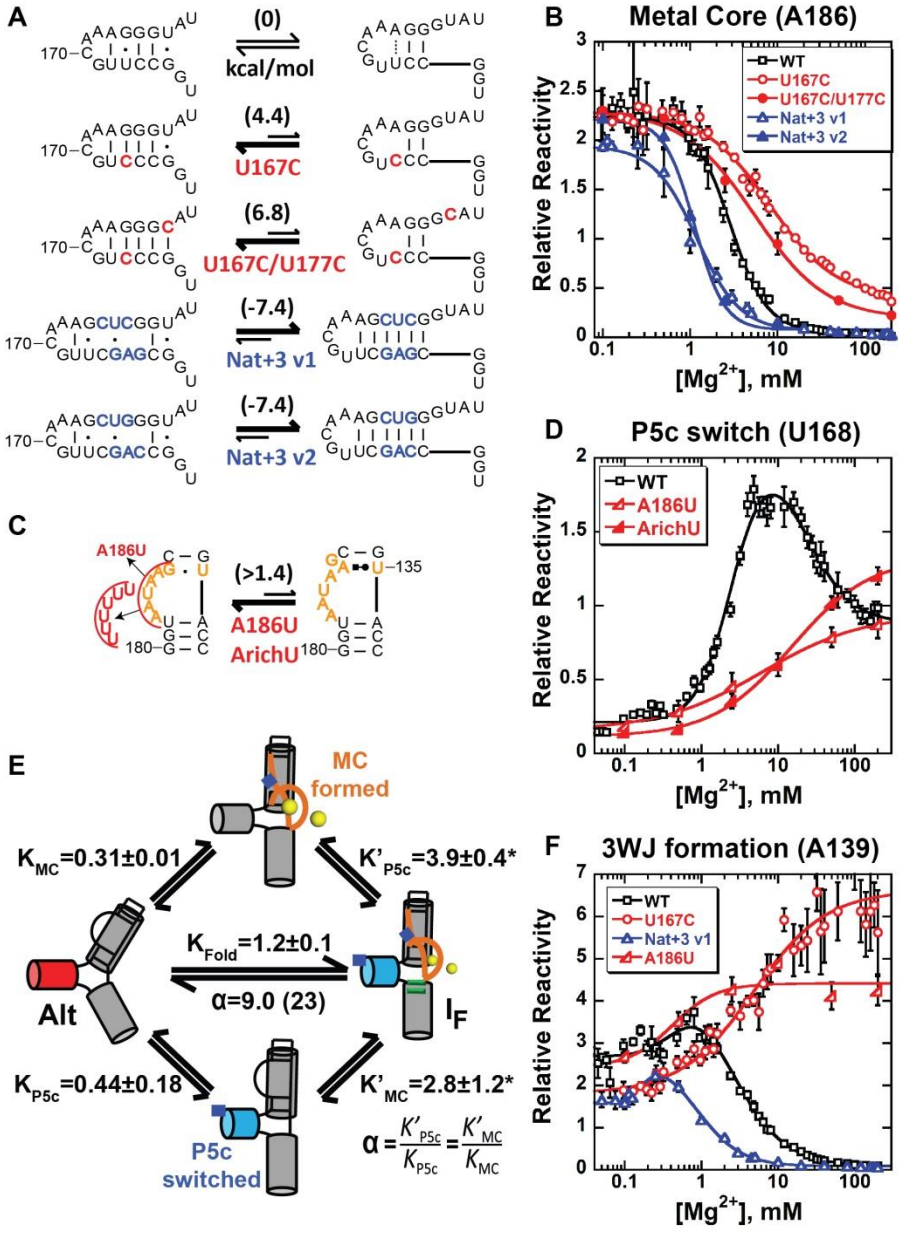

Figure 2.6: Cooperativity between P5c and the MC folding.

Figure 2.6: continued.,

(*A, B*) SHAPE footprinting of the MC at nucleotide A186 for mutants that shift P5c to the alternative or native secondary structure. Predicted effects of the mutations on P5c were calculated using ViennaRNA (85). (*C, D*) SHAPE footprinting of the P5c secondary structure switch at nucleotide U168 for mutants that block MC formation. Effects of the mutations on the MC are shown as a lower limit from the absence of a detectable transition herein. (*E*) Quantitative analysis of cooperativity between P5c switching and MC formation. Calculated values are labeled with an asterisk (*). The coupling constant of 9, equivalent to 1.3 kcal/mol, is obtained from completion of the thermodynamic cycles top and bottom, and the value of 23 in parenthesis is obtained from comparison of measurements for MC formation with altP5c and natP5c (*B*). (*F*) SHAPE footprinting of A139 in the 3WJ for mutants that affect P5abc folding.

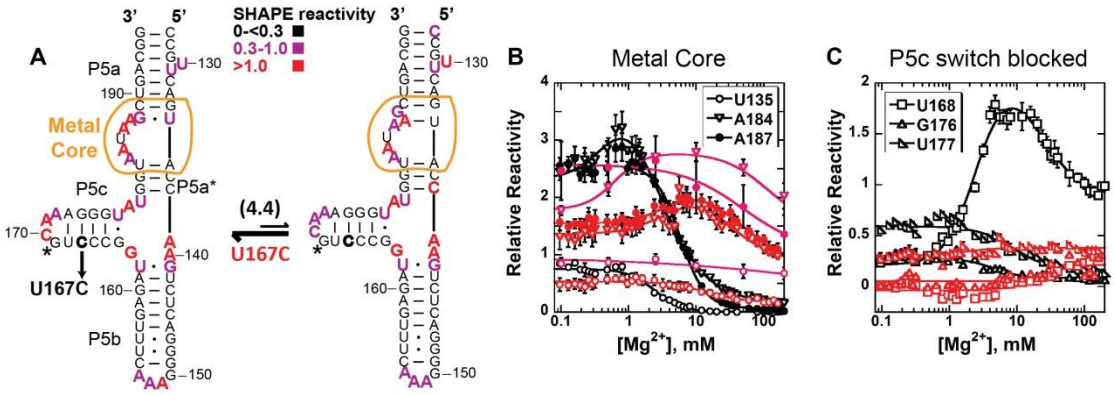

Figure 2.7: SHAPE reactivity for P5c variant U167C.

(*A*) SHAPE reactivity heat maps in the absence and presence of saturating Mg^{2+} is shown. Nucleotides that have values between 0.3-1.0 are colored purple and those with higher values are red. Here and below, an asterisk (*) is shown at position G169 for certain mutants because this position has high reverse transcription stopping precluding confident interpretation of the signal at this position for mutants. (*B*) The MC can still form at position U135 when P5c is maintained in the alternative secondary structure. The double mutant U167C/A186U is shown in magenta. (*C*) The P5c secondary structure rearrangement is effectively blocked by the U167C substitution.

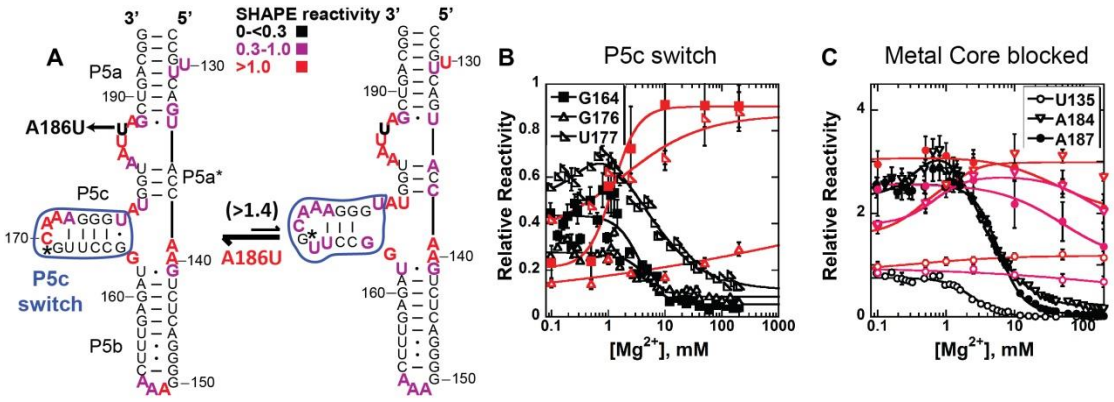

Figure 2.8: SHAPE reactivity for MC variant A186U.

(*A*) SHAPE reactivity heat maps in the absence and presence of saturating Mg^{2+} is shown. (*B*) The P5c switch can still occur when tertiary structure formation in the MC is knocked out by the A186U substitution. (*C*) Tertiary structure formation in the MC is effectively blocked by the A186U substitution. The double mutant U167C/A186U is shown in magenta.

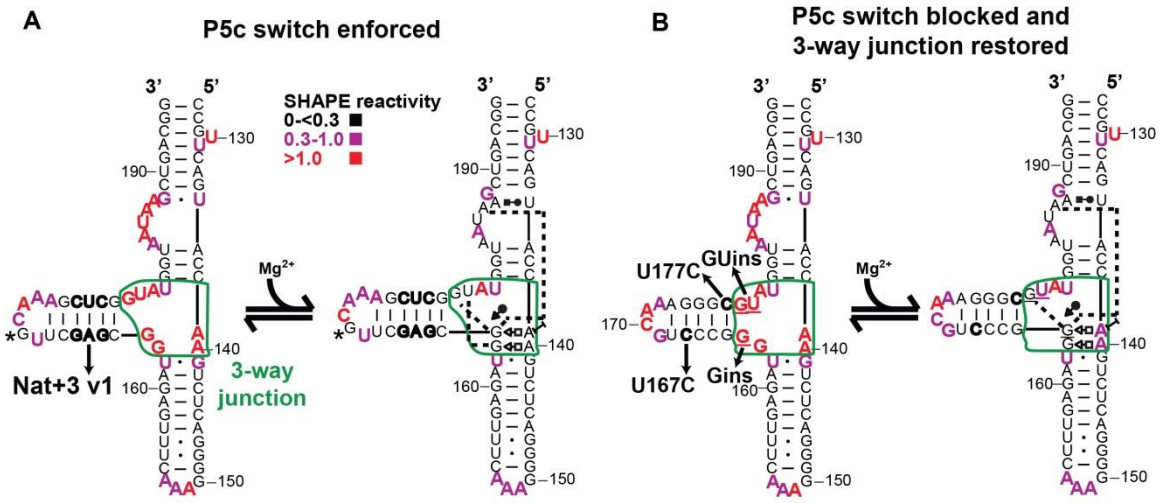

Figure 2.9: SHAPE reactivity for variants that impact the 3-way junction.

Mutations are indicated by bold black nucleotides with arrows, inserted nucleotides are underlined, long range interactions are shown as dashed lines.

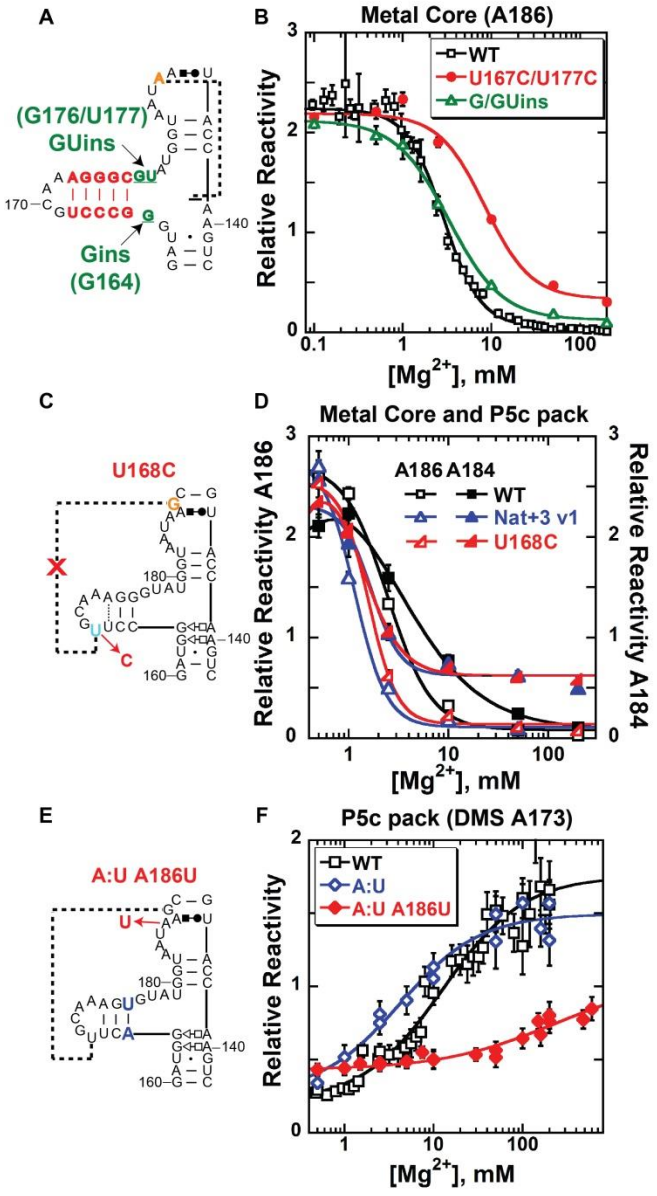

Figure 2.10: Dissecting the sources of folding cooperativity.

Figure 2.10: continued.,

(*A*) Gins-GUins mutant in the background of alternative P5c shown in red. (*B*) SHAPE footprinting of the MC at A186 for mutants that block P5c switching and restore native 3WJ nucleotides. (*C*) Perturbation of the U168-G188 base pair to knock out the P5c packing transition. (*D*) SHAPE footprinting of mutants that knock out the P5c packing transition. MC formation (A186) is shown in open symbols, and P5c packing (A184) is shown in closed symbols. (*E*) Construct designed to enforce the native P5c secondary structure (C165A/G175U, designated A:U) and knock out the MC (A186U). (*F*) DMS footprinting of mutants that knock out the MC.

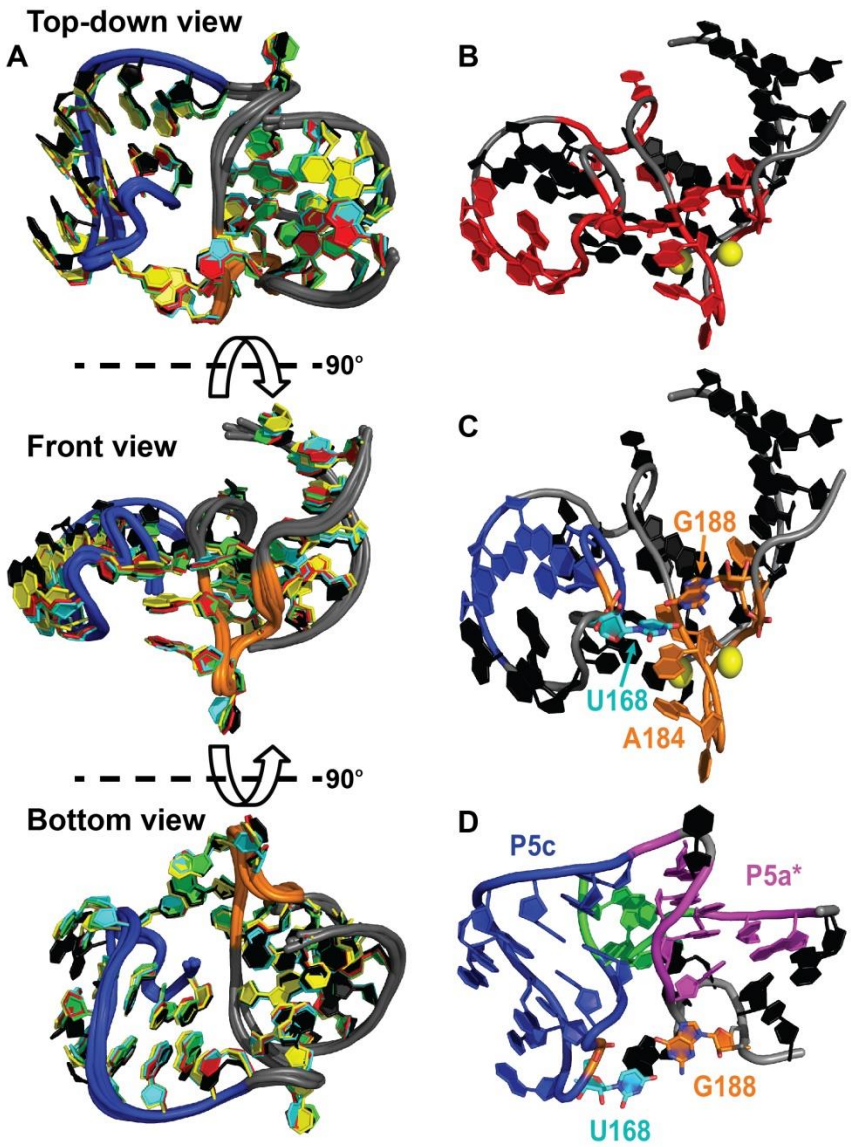

Figure 2.11: The packed P5abc conformation.

Figure 2.11: continued.,

(*A*) Structure alignment of the P5abc subdomain with front, top, and bottom views. The backbone of the P5c helix (blue) and MC (orange) are highlighted. P5abc regions where aligned from PDB files 1GID (green), 1L8V (cyan), 2R8S (black), 1GRZ (yellow), and 1HR2 (red). (*B*) Nucleotides with an observed Mg^{2+} midpoint higher than 5 mM are highlighted in red. (*C*) Labeled structure illustrating the long-range U168-G188 base pair. D. A non-canonical k-junction motif between P5c and P5a* orients P5c toward the MC (120).

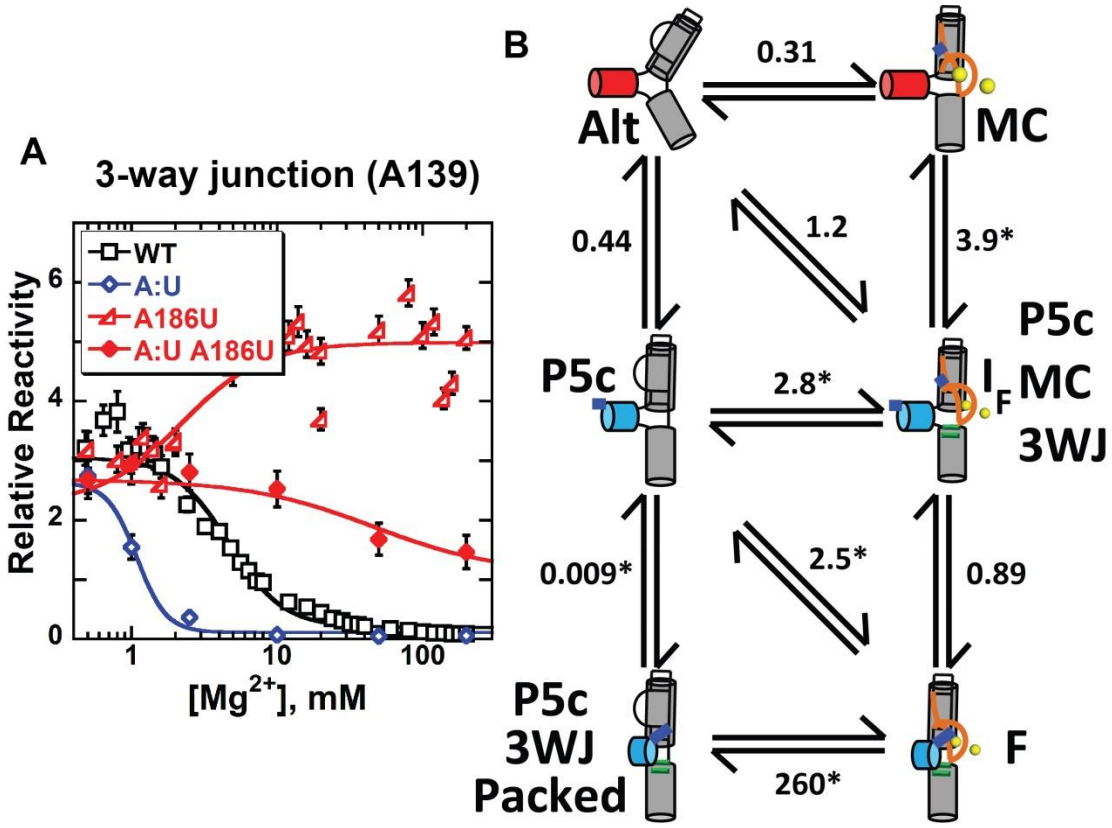

Figure 2.12: Quantitative analysis of cooperativity during folding.

(*A*) Partial 3WJ formation in the P5c packing mutant. (*B*) Thermodynamic framework for MC folding, P5c switching, and P5c packing. Values are measured at 3 mM $MgCl_2$. Calculated values are indicated with an asterisk (*).

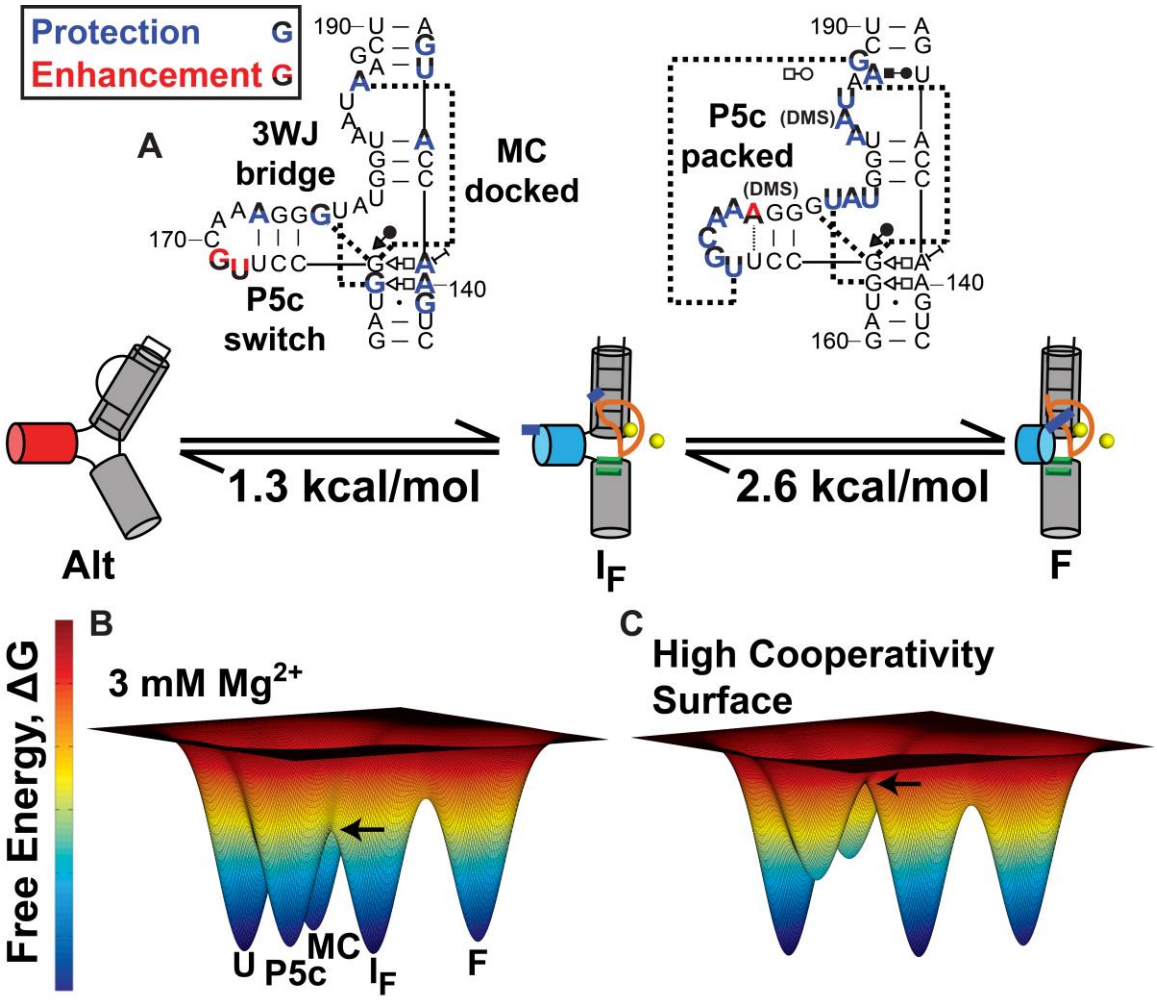

Figure 2.13: P5abc folding model and free energy landscapes.

(*A*) Compiled SHAPE and DMS footprinting signals associated with each transition are indicated with colored letters. The secondary structures illustrate the base pairs formed when the corresponding transition is completed. Cartoon RNA structures are shown as cylinders. Equilibria and cooperativities are calculated under conditions of 3 mM Mg^{2+}. (*B*) Free energy surface for folding P5abc illustrating a modular folding landscape at 3 mM Mg^{2+}. (*C*) A cartoon illustrating the increase in the transition state (T.S.) barrier for folding in a high cooperativity surface.

Nucleotide	Region	Change	$[Mg^{2+}]_{1/2}$	Hill, n	Amp, A
G134	MC	Protection	1.8 ± 0.2	1.7 ± 0.1	0.18 ± 0.02
U135	MC	Protection	2.5 ± 0.1	2.1 ± 0.2	0.68 ± 0.07
A136	MC	Protection	3.1 ± 0.4	2.2 ± 0.3	0.14 ± 0.01
A139	3HJ	Enhance-Protect	0.11-3.9	7.1-1.8	0.71-3.2
A140	3HJ	Protection	2.6 ± 0.2	1.5 ± 0.3	3.7 ± 0.7
G141	3HJ	Protection	4.4 ± 1.0	2.3 ± 0.6	0.42 ± 0.17
G163	3HJ	Protection	3.3 ± 0.1	2.1 ± 0.2	2.4 ± 0.3
G164	3HJ	Protection	NC	NC	NC
U168	P5c	Enhance-Protect	2.1-35	2.4-2.1	1.2-0.60
G169	P5c	Enhance-Protect	1.6-20	2.3-1.8	1.6-1.8
C170	P5c	Protection	17 ± 3	0.78 ± 0.14	1.3 ± 0.1
A171	P5c	Protection	4.5 ± 0.1	0.92 ± 0.14	1.1 ± 0.2
A172	P5c	Protection	9.2 ± 1.9	1.6 ± 0.1	0.76 ± 0.11
A173	P5c	Enhance-Protect	0.13-3.9	2.9-1.8	0.16-0.38
G176	P5c	Protection	3.7 ± 0.7	1.9 ± 0.4	0.20 ± 0.05
U177	P5c	Enhance-Protect	0.23-7.7	1.8-1.2	0.14-0.51
U182	MC	Enhance-Protect	0.12-4.0	1.1-2.2	0.17-0.36
A183	MC	Enhance-Protect	0.24-4.6	1.9-1.9	0.57-1.6
A184	MC	Enhance-Protect	0.27-4.8	2.4-1.9	0.70-2.6
U185	MC	Enhance-Protect	0.28-5.3	2.3-1.7	0.56-2.3
A186	MC	Protection	2.8 ± 0.1	1.9 ± 0.1	2.7 ± 0.5
A187	MC	Protection	6.2 ± 0.4	2.7 ± 0.3	1.7 ± 0.4
G188	MC	Protection	7.2 ± 1.4	2.4 ± 0.3	0.52 ± 0.04

Table 2.1: Mg^{2+} induced folding parameters from SHAPE footprinting. Nucleotides with biphasic changes in reactivity have two parameters shown separated by a hyphen. For these positions, the parameters are apparent midpoints because the first transition will influence the magnitude of subsequent transitions. Nucleotides that do not change are labeled as NC.

Nucleotide	Region	Change	$[Mg^{2+}]_{1/2}$	Hill, n	Amp, A
A173	P5c	Enhancement	15 ± 1	0.78 ± 0.10	1.4 ± 0.21
A183	MC	Protection	12 ± 5	0.96 ± 0.22	1.3 ± 0.3
A184	MC	Protection	9.6 ± 2.8	1.0 ± 0.2	0.80 ± 0.08
A186	MC	Protection	2.9 ± 0.3	1.6 ± 0.4	1.1 ± 0.2
A187	MC	Protection	3.9 ± 0.5	1.4 ± 0.1	0.78 ± 0.06

Table 2.2: Mg^{2+} induced folding parameters from DMS footprinting.

Model	K_1	K_2	K_3	K_4	$\chi 2$
4 state, A-B coop	0.042	11.3	0.98	NA	30.1 ± 0.2
4 state, B-C coop	0.28	2.7	1.99	NA	63.7 ± 0.2
5 state, A-B coop	0.04	1.22	2.1	11.7	16.0 ± 0.1
5 state, B-C coop	0.5	2.37	1.89	9.1	15.3 ± 0.1
5 state, C-D coop	2.8	27.7	0.097	3.03	30.3 ± 0.2

Table 2.3: Parameters from global modeling. See Methods section *Global modeling of SHAPE and DMS footprinting data* for the initial constraints of each global fit. NA means that the parameter is not applicable for the model.

Chapter 3: RNA structural modules control the rate and pathway of RNA folding and assembly[1]

INTRODUCTION

Structured RNAs are ubiquitous in biology and perform essential processes in all cells. They fold to functional forms by traversing complex energy landscapes along pathways that include multiple transitions and intermediates, and many RNAs must also assemble with proteins, other RNAs, or small molecules. While considerable progress has been made in predicting folded RNA structures (121), the complexity of RNA folding and assembly reactions and the myriad of potential pathways present formidable challenges to obtaining a deep and predictive understanding of these processes. Underscoring the need for such understanding, many RNAs are thought to function in the cell under kinetic control, with the RNA having a short period of time to respond to cellular cues before a commitment is made to a folding pathway and thereby to downstream signals (122, 123). Further, improper ribosome assembly and incorrect formation of splice junctions are linked to diseases, highlighting the essential nature of accurate and efficient RNA folding (124, 125).

A feature of RNA that has greatly aided understanding is the largely hierarchical nature of its folding (126). Local RNA secondary structure typically forms prior to tertiary structure –isolated helices form in microseconds– and secondary structure formation can be highly favorable even in the absence of enforcing tertiary contacts (127, 128). Typically, tertiary structure is thought to form as these pre-formed secondary structure elements are arranged relative to one another. Energetic separation of RNA secondary and tertiary structure formation is common, although not absolute, as limited

[1]Adapted and reproduced from Gracia B*, et al.* (2016) RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly. *J Mol Biol*.; all co-authors contributed equally to the work.

changes in secondary structure occur in tandem with the formation of tertiary structure, including around the catalytic sites of group I introns (129, 130) and in riboswitches as a consequence of ligand binding (123, 131-133). Indeed, even when higher-order folding is intertwined with secondary structure changes, a general view has emerged in which local secondary elements fold first to native structure, with pre-formed native local structure then reinforced as it forms tertiary contacts with other structural elements.

The prominence of local structure has led to the idea that RNA structures can be understood from the properties of its structural modules: RNA helices, junctions, and tertiary contacts (134). This idea can be traced back to observations that small components of large RNAs sometimes adopt essentially the same structures in isolation and in their corresponding biological RNAs (9, 55, 66, 135, 136), a clear indication that the structure is determined locally. Even for RNA elements that display structural differences in their biological RNAs relative to the same element in isolation, recent work indicates that the native structure can be present in the isolated element as a minor conformation among an ensemble –an excited state (137, 138)– that is then trapped by intra- or inter-molecular interactions in a process of conformational selection (33, 35, 139-143). Thus, understanding the excited states and their relative populations will likely give profound insights into the possible final structures adopted by a given structural module and the relative energetics of different possible structures.

Analogously, it is possible that the structural and dynamic properties of RNA modules can be used to understand and even predict the folding pathways and rates for complex RNAs. The Tetrahymena group I intron ribozyme and its component subdomains have been instrumental for studies of RNA folding, structure, and function. Rapid tertiary structure formation occurs in the peripheral subdomain P5abc in a transition that includes a secondary structure rearrangement from an alternative

conformation to the native conformation (70, 71, 144, 145). The pre-folded P5abc assembles with the rest of the intron by forming two long-range contacts with another helical subdomain to form the P4-P6 domain and a third long-range contact with another peripheral element of the intron. In light of the prior results, the simplest model for ribozyme folding is that P5abc forms its native local structure first and then forms the long-range tertiary contacts via conformational selection. However, it is not known whether this pathway is indeed followed, and if so, whether it is obligate, and it is not known whether properties of the structural modules within P5abc that form local and long-range contacts can be used to understand or predict the folding pathways and rates.

Here we used a two-piece system of the P5abc subdomain and the Tetrahymena group I intron core to probe whether perturbing one such structural module within P5abc would give a predictable effect on the overall assembly process. The rationale is that if P5abc is required to fold locally first and then assemble with the intron core via conformational selection, it should be possible to 'tune' the assembly kinetics by generating point mutations that modulate the local folding transition of P5abc. Indeed, we find that modulating the stability of the alternative secondary structure in P5abc results in tuning across a ~100-fold range of assembly rate constants, supporting the conformational selection pathway. Nevertheless, there is a limit of ~100-fold for the rate decrease, and we show that this limit arises because a distinct, induced fit pathway becomes dominant upon further destabilization of native P5abc structure, with assembly preceding the local rearrangement and likely nucleated by a tertiary contact that does not require the native P5abc structure. The presence of this pathway allows assembly of highly destabilized P5abc mutants to proceed orders of magnitude faster than would be calculated along the conformational selection pathway, thereby smoothing the complex and rugged RNA folding landscape. Our results highlight the pliability of RNA folding

pathways and bolster the idea that RNA folding processes can be understood at a quantitative level from an understanding of the properties of component modules.

Results

We used a two-piece system consisting of P5abc and the P5abc-deleted ribozyme ($E^{\Delta P5abc}$) (58, 74). In the absence of Mg^{2+}, P5abc forms an alternative conformation that lacks tertiary structure but has secondary structure, including a set of non-native base pairs in the helix P5c and one in P5a (Figure 3.1A). In the presence of millimolar Mg^{2+}, the formation of tertiary structure is linked to a one-nucleotide shift in base pairing in P5c, a one-nucleotide shift in P5a with formation of a bulge, and the formation of additional non-canonical base pairs in P5b (Figure 3.1A and B) (70-72). Upon folding to the native state, P5abc can robustly assemble with the largely pre-folded $E^{\Delta P5abc}$ core by forming two tertiary contacts with the other helical stack of the P4-P6 domain, a metal core/metal core receptor (MC/MCR) and a tetraloop/tetraloop receptor (TL/TLR) interaction, and a kissing loop tertiary contact between the loop of P5c and loop L2 of the core (P14) (Figure 3.1C) (63, 69, 74). Two of the specific structures within P5abc that form these tertiary contacts, the MC and the native sequence in the loop of P5c, form their native structures in the local P5abc folding transition (Figure 3.1A and B), contributing to the expectation that this transition precedes assembly. The overall structure of P5abc is the same in isolation and within the intact intron or with the rest of the P4-P6 domain (55, 66, 146). The two-piece system also adopts a native conformation, as it retains full catalytic activity and gives the same overall protection pattern as the full, one-piece ribozyme in chemical footprinting experiments (59).

Mutations tune assembly kinetics of P5abc with the intron core

If P5abc assembles with $E^{\Delta P5abc}$ via conformational selection, mutants for which the local native folding of P5abc is unfavorable would be predicted to assemble slower than wild-type P5abc because they would first have to undergo this unfavorable transition to the native state (Figure 3.2A). Further, if the secondary and tertiary transitions within P5abc are tightly coupled, mutations that stabilize the alternative secondary structure of P5abc relative to the native secondary structure are predicted to shift the equilibrium for the coupled transition. We tested three such point mutations (Figure 3.1A): U167C and G174A, which convert a G-U pair in the alternative P5abc structure to a Watson-Crick base pair and were shown previously to favor the alternative structure in the isolated P5abc (72), and U177C, which also changes a G-U to a G-C pair in the alternative structure and does not change the base pairing in the native structure. Notably, these three nucleotides do not form tertiary contacts in the native structure.

We monitored the assembly kinetics of radiolabeled wild-type and mutant versions of P5abc with the $E^{\Delta P5abc}$ core by using an electrophoretic mobility gel shift assay (EMSA) with a 'chase' of excess unlabeled P5abc (Figure 3.2B). Each P5abc mutant bound $E^{\Delta P5abc}$ with a pseudo-first-order rate constant that was linearly dependent on $E^{\Delta P5abc}$ concentration (Figure 3.2C and D). The wild-type P5abc assembled with a second-order rate constant of 1.1 (± 0.1) × 10^7 M^{-1} min^{-1} (Figure 3.2D and Table 3.1), as measured previously (75, 78). As predicted, all three mutants assembled slowly, giving rate constants similar to each other (~2 × 10^5 M^{-1} min^{-1}) and approximately 50-fold lower than that of the wild-type P5abc (Figure 3.2E).

We also tested a point mutation predicted to stabilize the native P5abc secondary structure relative to the alternative structure. The G176A mutation disrupts a C-G base pair in the alternative structure without changing base pairing in the native structure

(Figure 3.1A). In the background of the otherwise wild-type P5abc, this mutation was not expected to impact the assembly rate significantly because wild-type P5abc adopts the native structure as the predominant ground state under our experimental conditions (10 mM Mg^{2+}; (70) and data herein). Thus, further stabilization of the native structure by G176A would not increase the fraction of native P5abc significantly. Indeed, we found that the G176A P5abc assembled with $E^{\Delta P5abc}$ with the same rate constant within error as the wild-type P5abc (Fig. 2(d), Fig. S2, and Table 1).

A further prediction was that the G176A mutation would accelerate P5abc assembly in the background of the mutations that stabilize the alternative secondary structure – i.e., G176A would confer at least a partial rescue on these mutations. Indeed, in the background of each of the three mutations above, the rate constant for P5abc assembly with $E^{\Delta P5abc}$ was increased to within 10-fold of that for wild-type P5abc. We observed similar effects over a range of conditions, with rate decreases for the destabilizing mutations in the range of 100-fold and consistent rescue by G176A (data not shown).

Together, these results show that the assembly rate of P5abc with the intron core can be tuned by mutations that are predicted to change the equilibrium for a local secondary structure transition within P5abc. Mutations designed to stabilize the alternative conformation slowed assembly by as much as 100-fold, and this decrease could be rescued by a second mutation (G176A) predicted to shift the equilibrium back toward the native conformation. Thus, these data are consistent with a simple two-step, conformational selection pathway in which P5abc rearranges locally to form native P5abc and then assembles with the core.

Direct measurement of the folding transition within P5abc

To further test whether the changes in assembly kinetics reflect changes in the P5abc folding equilibrium, we used NMR to directly measure the equilibrium constant for the conformational transition for the isolated wild-type P5abc and mutants (70, 71). It was shown previously using a truncated, 56-nt version of P5abc (tP5abc) that the transition results in chemical shift changes for several imino protons (70). In particular, the Mg^{2+}-induced change from the alternative to the native structure results in the disappearance of four imino peaks and the appearance of four new peaks, assigned to be G141, U142, G160 and U162 (70). A particularly prominent signal is for U142, near the 'top' of P5b. This residue gives a sharp peak in the absence of Mg^{2+} with a chemical shift of 14.31 ppm and shifts to a new peak at 14.00 ppm in the presence of Mg^{2+} (70). The new peak is consistent with the X-ray crystal structure of native P4-P6 in which two G-A base pairs form above the U142-A161 pair and bind to a magnesium ion, presumably changing the environment of U142 to give the observed chemical shift change (70).

For the wild-type tP5abc, incubation in the absence of Mg^{2+} resulted in a spectrum with sharp, well-resolved imino proton resonances (10-14 ppm), and Mg^{2+} addition shifted the peaks as expected (Figure 3.3). We assigned the resonances both in the absence and presence of Mg^{2+} using 2D NMR experiments and ^{15}N-labeled tP5abc, and the results were in excellent agreement with assignments reported previously (70) (Figure 3.3A). To monitor the Mg^{2+}-dependence of the folding transition, we recorded 1D NMR spectra of tP5abc RNA at various Mg^{2+} concentrations. With 2 mM free Mg^{2+} the U142 peak (14.31 ppm) decreased in intensity and the 14.00 ppm peak, termed U142*, increased (Figure 3.3B). At higher Mg^{2+} concentrations substantial peak broadening and losses of signal were observed, as noted previously (70), likely due to RNA aggregation.

We performed analogous NMR measurements on tP5abc variants. U167C RNA gave a ^{1}H imino spectrum similar to that of the wild-type tP5abc in the absence of Mg^{2+}, but it did not give the characteristic shifts in peak positions even with 50 mM free Mg^{2+}, as expected for a mutation that stabilizes the alternative P5c secondary structure (Figure 3.3). In contrast, the G176A RNA, with its mutation that is expected to stabilize the native secondary structure, gave a prominent U142* peak in as little as 2 mM free Mg^{2+}, indicating stabilization of the native tP5abc conformation (Figure 3.3B).

A key test of predictions from assembly kinetics was the behavior of the double mutant, U167C/G176A. This mutant assembles more slowly than the wild-type P5abc, but the decrease is modest (Table 3.1), suggesting that the equilibrium between the alternative and native conformations is close to unity. Indeed, at 10 mM Mg^{2+} this RNA prominently displayed both the U142 and U142* peaks, indicating that both species are populated (Fig. 3(b)). We used the relative peak volumes of U142* and U142 to estimate the equilibrium value for the two conformations (K_{eq} = [Native]/[Alt]). With 2 mM free Mg^{2+}, K_{eq} was 0.31, and its value increased with increasing Mg^{2+} concentration to 0.79 at 10 mM Mg^{2+} (Table 3.2).

To obtain an independent measure of the effects of the mutations on the equilibrium between the alternative and native structures, we used quantitative DMS footprinting (115). It was shown previously that the transition to native P5abc enhances reactivity at A173, which forms a protected Watson-Crick base pair in the alternative conformation but a reactive, non-canonical base pair in the native conformation (72). As expected, the G176A mutant gave a high level of reactivity at 5 mM Mg^{2+}, consistent with the native structure, and U167C gave a low level of reactivity, consistent with the alternative structure (Table 3.3). Supporting the interpretations of the NMR results, the

wild-type P5abc and the U167C/G176A double mutant gave intermediate levels of reactivity, suggesting mixtures of the native and alternative structures at 5 mM Mg^{2+}.

While these results are in qualitative agreement with the model of a direct linkage between the conformational transition in P5abc and assembly with the ribozyme core, the assembly rate constant for the U167C/G176A and the U177C/G176A mutants were somewhat lower than would be predicted from the simplest model in which, upon rearrangement to the native state, the U167C/G176A mutant assembles with the same rate constant as the wild-type P5abc. Further work suggests that this deviation is an idiosyncratic effect of the G176A mutation in these backgrounds (see Appendix A: *Structural connections in the 3WJ of P5abc*).

Mutations switch flux from conformational selection to induced fit

The three point mutations, U167C, G174A, and U177C, were predicted to slow assembly relative to the wild-type P5abc and did so. However, the three mutations each slowed assembly by ~50-fold, despite much larger effects predicted for two of them from nearest neighbor rules (Table 3.1). To test the possibility that the predictions were incorrect, with the three mutations coincidentally giving the same effect on the P5abc equilibrium, we probed double mutants, noting that two independent substitutions would be expected to slow assembly by ~2,500-fold (50 × 50). The double mutants U167C/U177C and G174A/U177C have substitutions expected to produce independent effects on the secondary structure transition equilibrium as they alter non-adjacent base pairs in P5c (Figure 3.1A). In contrast to the empirically predicted 2,500-fold effect, and despite the nearest neighbor prediction of a shift in the equilibrium toward the alternative secondary structure by as much as 10^5-fold (Table 3.1), these mutants assembled with rate constants within 2-fold of the values for the single mutants (Figure 3.2E and Table

3.1). DMS footprinting results indicated that the second mutation stabilized the alternative secondary structure, as expected (data not shown). Thus, further weakening the native structure of P5abc relative to the alternative structure does not slow assembly beyond the decrease of ~100-fold.

The loss of dependence of the assembly rate on the equilibrium for P5abc structure in this regime suggested that, for the mutants that strongly stabilize the alternative structure, assembly with the intron core does not require rearrangement to the native state prior to the rate-limiting assembly step. Instead, the data strongly suggest that the rearrangement occurs after initial assembly in an induced fit process or, alternatively, not at all, with the alternative conformation remaining present in the complex.

To test whether these P5abc mutants (U167C, G174A, and U177C, and mutants with two of these substitutions) rearranged in their complexes with the core, we measured catalytic activity and performed DMS footprinting. Upon complex formation, mutants tested from this group displayed wild-type catalytic activity for cleavage of the standard, all-ribose substrate and a modified substrate that destabilizes tertiary docking of the substrate and would therefore reveal docking differences between the wild type and mutants (see Methods Section, *Ribozyme catalytic activity measurements* and Table 3.4). Prior results demonstrated that loss of the P14 tertiary contact, which is formed by P5c and not expected for the alternative conformation (see below), resulted in weakened docking (62). Thus, the observation of intact docking indicates formation of the native P5abc structure. Further, the DMS reactivity at A173 was enhanced for all three point mutants (Figure 3.4), although the G174A variant gave smaller increases, perhaps because the G-to-A substitution at the adjacent position impacts the DMS reactivity of A173. These results provide strong support for native complex formation with the P5abc mutants. Thus, together the results indicate that there are two pathways to the native

complex –a conformational selection pathway and an induced fit pathway– and the relative flux through each pathway depends on the stability of the native P5abc structure.

Probing the interactions formed along each pathway

Having established the two folding pathways, we set out to learn about the interactions formed in the assembly transition states along them. A strong candidate for a difference between the two pathways was P14, the long-range 'kissing loop' interaction of loop L5c with the complementary sequence in L2 (U43 – A46, Figure 3.5A). Assembly of the native P5abc could involve the early formation of P14, but the alternative conformation has a shift in the register of L5c that would prevent formation of the native P14 base pairs. Thus, it was possible that P14 would be formed in the transition state for assembly of the wild-type P5abc but not for mutants that use the alternative folding pathway (U167C, G174A, U177C, Figure 3.5A). Although some P14 base pairs or non-native interactions might be possible, U168 of L5c should not be able to form a P14 base pair with A46 because U168 forms an internal base pair with A173 in the alternative P5abc structure.

We tested for formation of the U168-A46 base pair by measuring P5abc assembly with a point mutant of $E^{\Delta P5abc}$ that eliminates the potential to form it (A46C; Figure 3.5B). This mutation decreased the assembly rate constant of the wild-type P5abc by 6-fold (Figure 3.5B, square vs. triangle), consistent with a base pair with A46 contributing to transition state stabilization. The double mutants that include the P5c rescuing mutation G176A (U167C/G176A, G174A/G176A, and U177C/G176A) also assembled slower with the A46C core than the wild-type core (2-5-fold). In striking contrast, the assembly rates of U167C and U177C were not affected and G174A was decreased less than 2-fold (Figure 3.5B and Table 3.1). These results suggest that the base pair between

A46 and U168 is formed in the transition state for assembly of the wild-type P5abc and the double rescue mutants, because eliminating the ability to form this base pair slowed assembly of these P5abc variants. Assembly rates of the P5abc mutants that strongly stabilize the alternative conformation and use the induced fit pathway were unaffected, as expected because U168 is base paired within the alternative P5c helix and unavailable to base pair (Figure 3.5A). We infer that the other three base pairs of P14 are also unlikely to be formed in the rate-limiting transition state along the induced fit pathway. Nucleation of complex assembly must involve formation of an alternative contact or contacts, likely including the TL/TLR interaction (see Chapter 4).

These results also provide additional support for the existence of the two pathways, as one involves P14 formation in the transition state and one does not appear to do so. As elaborated in the Discussion, the wild-type P5abc and variants in which the native structure remains moderately stable use primarily the conformational selection pathway, and mutations that stabilize the alternative P5abc structure beyond a critical threshold switch the bulk of the flux to the induced fit pathway.

Discussion

Building on prior evidence for structured RNAs as collections of structural modules whose overall structure can be understood from the properties of the modules, we tested whether the assembly kinetics of the P5abc peripheral element with the intron core could be rationally tuned by manipulating the native stability of P5abc. P5abc was thought to fold first and use a conformational selection assembly process.

We found that the rate of assembly of P5abc for the intron core can be tuned across a range of 100-fold in assembly rate constant. Further results indicated a second regime in which mutations that modulate the local structure in P5abc produce a switch to

an alternative assembly pathway that uses an induced fit mechanism instead of conformational selection. As described further below, these pathways involve the formation of distinct tertiary contacts in the assembly transition states and raise the possibility that the maximal rate constants are simple reflections of the type of tertiary contact that nucleates assembly along each pathway. This behavior suggests a high degree of modularity in folding and assembly, highlights the connections between the properties of RNA structural modules and the behaviors of complex folding processes, and demonstrates how manipulation of underlying structural modules can provide highly specific tests of folding pathways and models.

Folding then assembly: the conformational selection pathway

Figure 3.6 depicts the assembly pathway that dominates flux for a given equilibrium for folding P5abc. The wild-type P5abc and mutants that include the native-stabilizing G176A substitution (U167C/G176A, G174A/G176A, and U177C/G176A as well as the single mutant G176A) follow the conformational selection pathway. In this pathway, P5abc first rearranges to the native conformation and then assembles with the intron core via a rate-limiting transition state in which P5abc remains in the native conformation and P14 is formed (Figure 3.7, blue pathway). The maximal rate constant for this pathway is ~10^7 M^{-1} min^{-1}, the observed rate constant for the wild-type P5abc and the G176A mutant. These two versions of P5abc populate the native state favorably in the absence of the intron core, so this value, ~10^7 M^{-1} min^{-1}, represents the second-order rate constant for binding between native P5abc and the core. Most simply, this value may reflect formation of the P14 tertiary contact (cartoon along the blue pathway). Supporting this model, isolated kissing loops have been shown to associate with comparable rate constants (147, 148). The double mutants that include G176A display

overall rate constants that are 2-20-fold slower, yet are also affected by the A46C mutation within P14 and thus also appear to have P14 formed in the transition state. We conclude that these mutants use the same pathway as wild type and G176A. They must first rearrange unfavorably to the native P5abc conformation and then bind the core in a process of conformational selection.

Assembly then folding: the induced fit pathway

In contrast, U167C and other P5abc mutants that strongly stabilize the alternative conformation (G174A, U177C and combinations of these mutations) populate a distinct induced fit pathway (Figure 3.6). In this pathway, the assembly transition state that does not include P14 –at least not the U168-A46 base pair– and maintains P5c in the alternative structure (red pathway in Figure 3.7). To reach the native structure, P5abc must rearrange after forming an initial complex with the core. This rearrangement apparently permits formation of P14, as DMS footprinting revealed that P14 is formed in the final complex (Figure 3.5B). The re-ordering of events for this pathway, with the rearrangement occurring after binding to the core rather than before binding, requires a new pathway rather than a change in rate-limiting step along a common pathway.

In the induced fit pathway, the initial binding step occurs with P5c in the alternative structure and unavailable to form native P14 (Figure 3.5A). Thus, the initial complex presumably requires one or more of the other long-range tertiary contacts or formation of a partial P14. Because the tetraloop L5b is separated from P5c by the intervening helix P5b, the TL/TLR tertiary contact formed by L5b and its partner J6a/6b is a good candidate (red contact in Figure 3.7). Remarkably, the second-order rate constant along this pathway, ~10^5 M^{-1} min^{-1}, is close to that for formation of the same tetraloop-receptor pair in an isolated system (5×10^5 M^{-1} min^{-1}) (26, 86, 149), consistent

with this model (see Methods section: *Estimation of the rate constant for TL/TLR formation*). Further work will be necessary to test this hypothesis (see Chapter 4).

Interestingly, a complex in which only the TL/TLR contact is formed would be expected to dissociate rapidly, with a rate constant of at least 200 min^{-1} (26, 149). Both of the additional tertiary contacts, the MC/MCR and P14, are most simply expected to require the P5c rearrangement and native structure formation in P5abc. Thus, these results imply that the P5c rearrangement and subsequent tertiary contact formation occur faster than this complex dissociates, on the timescale of milliseconds, defying the general notion of RNA folding landscapes as being rugged (Figure 3.7). This interpretation is supported by recent results indicating that the secondary structure in P5c fluctuates rapidly (100).

The local P5abc folding transition controls the rate and pathway

Together, the results indicate that there are two pathways for P5abc assembly with the intron core, and the relative stability of the native and alternative conformations of P5abc determines which pathway receives the greatest flux (Figure 3.6). Mutations that result in the alternative structure being favored by a small amount decrease the folding rate but they do not change the dominant pathway. The alternative structure rearranges to the native structure reversibly, and then P5abc assembles with the core along the pathway that dominates for the wild-type P5abc. In this regime, mutations that alter the relative stabilities of the native and misfolded structures tune the assembly rate constant up and down. On the other hand, mutations that result in the alternative P5abc structure being favored by more than 100-fold (e.g. U167C) change the dominant pathway, with P5abc remaining in the alternative conformation until after formation of an initial complex. As the relative stability of the alternative conformation of P5abc reaches a threshold,

mutations that further stabilize this conformation switch the flux from conformational selection to induced fit without further impacting the assembly rate.

Methods

RNA preparation for assembly kinetics and NMR

RNAs were prepared by in vitro transcription essentially as described (114). The DNA template was plasmid DNA for $E^{\Delta P5abc}$ or a PCR product assembled from oligonucleotides for P5abc. P5abc constructs for DMS footprinting included 5′ and 3′ flanking hairpins for normalization and single-stranded regions to serve as markers or to anneal with a primer for reverse transcription and RNA isolation (115). $E^{\Delta P5abc}$ mutants were generated by a Quikchange mutagenesis protocol (Agilent). Oligonucleotides for PCR assembly and mutagenesis were from IDT (San Diego, CA). DNA templates and RNA were isolated by affinity column (Qiagen). P5abc RNA was 5'-end labeled by treating it with shrimp alkaline phosphatase (New England Biolabs) followed by polynucleotide kinase and [γ-^{32}P]-ATP. Labeled RNA was purified by polyacrylamide gel electrophoresis, eluted into TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA), and stored at –20 °C.

Measuring assembly kinetics

Reactions were performed in 50 mM Na-MOPS, pH 7.0, and 10 mM $MgCl_2$ unless indicated otherwise. The $E^{\Delta P5abc}$ ribozyme core (15 μL) was prefolded for 30 min at 50 °C in buffer containing the Mg^{2+} concentration of the measurement. Labeled P5abc was prefolded under the same solution conditions at the reaction temperature for at least 15 min for reactions at 25 °C or at least 60 min for reactions at 10 °C and 4 °C. $E^{\Delta P5abc}$ and radiolabeled P5abc (1 μL) were mixed and aliquots were quenched at various times thereafter by adding unlabeled P5abc (at least 5-fold excess over $E^{\Delta P5abc}$) and increasing

the Mg^{2+} concentration to 50 mM in a solution that also included 20% glycerol and 0.04% xylene cyanol for gel loading. A zero time point was collected by pre-incubating labeled and unlabeled P5abc together and then adding the mixture to $E^{\Delta P5abc}$. Quenched samples were placed on ice and then run on a 12% native polyacrylamide gel for 30 min at 200 V to separate bound and unbound P5abc (75). Bands were quantitated by using a Phosphorimager (GE Healthcare).

NMR measurements

All RNA samples were prepared by in vitro transcription using $^{13}C/^{15}N$ labeled nucleotide triphosphates (Cambridge Isotope Laboratories), T7 RNA polymerase (Fisher Scientific), and chemically synthesized DNA templates containing T7 promoter at the 5′ end (Integrated DNA Technologies). These samples were subsequently purified by anion-exchange chromatography (Protein-Pak Hi Res Q, Waters) using an HPLC (Alliance, Waters), and were exchanged into NMR buffer (50 mM MOPS, pH 7) using 3-kDa-cutoff centrifugal concentrators (Millipore Corp.). Prior to the final round of buffer-exchange, the samples were annealed by heating at 95 °C for 10 min and rapid cooling on ice.

DMS footprinting

Reactions were performed using a method developed previously (150) with extensive modifications to increase throughput (151). P5abc constructs including flanking regions (50 nM) were prefolded for 15 min at 37 °C and then equilibrated for at least 15 min at 25 °C or 60 min for measurements at 10 °C. For footprinting P5abc in complex with $E^{\Delta P5abc}$, 50 nM P5abc was incubated with 1 μM $E^{\Delta P5abc}$ for 60 min at 37 °C to allow binding, followed by equilibration at the desired temperature. DMS (Sigma-Aldrich) was added to 10 mM (with 0.45% ethanol) and incubated for 15 min at 25 °C or 60 min at 10

°C. Reactions were quenched by adding 2.4 M 2-mercaptoethanol (Fisher), 500 mM NaCl, and 6.7 nM of an oligonucleotide used to isolate RNA and to prime reverse transcription (5′-FAM-A20-Tail2 primer, IDT). This oligonucleotide hybridizes to the 3′ end of the RNA and has a flanking poly-A tail at the 5′ end, allowing the RNA to be isolated by association with oligo dT magnetic beads (Poly-A Purist Kit, Life Technologies). RNA was incubated with beads for 10 min and then washed twice with 5 volumes of 70% ethanol. Samples were left to dry at room temperature for ~20 min and then resuspended in 5 μL of superscript III reverse transcriptase mix as recommended by the manufacturer (Life Technologies). Reverse transcription was performed for 30 min at 55 °C. RNA was degraded by incubation with 0.2 M NaOH at 90 °C for 3 min. The solution was neutralized with 0.3 M HCl and then 0.4 M Na-acetate, pH 4.5, with 700 mM NaCl. cDNA was again isolated by magnetic beads, washed twice, as described above, and eluted from the oligo dT beads with Hi-Di Formamide. cDNA products were resolved by capillary electrophoresis (ABI 3730) and aligned and quantified using HiTRACE (117). Reactivity values were normalized to the average of the two adenines in the 3' flanking hairpin GAGUA stem-loop.

Estimation of rate constant for TL/TLR formation

The value of 5×10^5 M^{-1} min^{-1} for this association was estimated from measurements of the equilibrium constant for binding in a bi-molecular system (0.40 mM) and the dissociation (undocking) rate of the same contact in a model, unimolecular system (3.3 s^{-1}). The two measurements were performed under somewhat different conditions from each other and from the current work so the value obtained is an estimate. The equilibrium measurement was performed at ambient temperature, 50 mM Na-phosphate, pH 6.0, 125 mM $MgCl_2$, 500 mM NaCl, 500 mM KCl (86). The kinetics

measurement was performed at 21 °C, 50 mM Na-HEPES, pH 7.5, 100 μM EDTA, 10 mM Mg^{2+}, 100 mM NaCl (26). The important differences are likely to be the Mg^{2+} concentration, which was much higher for the equilibrium measurement, and the monovalent ion concentration, which is also much higher for the equilibrium measurement. The kinetics measurements were performed under conditions quite similar to ours, whereas the equilibrium experiments were under conditions much more different from our conditions. Additional equilibrium binding experiments in Qin et al. suggested that omitting the added monovalent ions did not have a large effect on the affinity in the presence of 125 mM Mg^{2+}, but decreasing the Mg^{2+} concentration to 10 mM in the background of 100 mM NaCl gave weakened binding, which was not quantitated but appeared to an effect of 5-10-fold. If the binding is indeed weaker at lower Mg^{2+} concentrations, which more closely approximate our experimental conditions, the rate constant for binding would be somewhat less than 5×10^5 M^{-1} min^{-1}. A 5-10-fold decrease would result in a rate constant of 0.5-1 $\times 10^5$ M^{-1} min^{-1}, the same as our observed rate constant for the alternative structure-stabilizing mutants of P5abc.

Ribozyme catalytic activity measurements

To probe whether mutant versions of P5abc rearrange to the native state when they are in complex with the $E^{\Delta P5abc}$ ribozyme core, catalytic activity was measured for the complexes using the standard, all-ribose substrate (rSA_5) and a substrate with an atomic-level substitution (–3m,rSA_5, $CCCmUCUA_5$) (Table S3). This latter substrate contains a 2′-methoxy substitution of the uridine at the –3 position, which destabilizes docking of the substrate-containing helix into tertiary contacts with the core (152). Thus, if the wild-type and mutant P5abc both rearrange to the native state in the complex, the decrease in activity from the –3m substitution is expected to be the same for both wild-

type and mutant complexes. Additionally, the absence of stable docking eliminates the possibility that cooperative tertiary structure stabilization from substrate docking induces native state formation of P5abc and thereby produces a result not reflective of the conditions under which P5abc binding experiments were performed. A subsaturating guanosine concentration (55 μM) was used in the catalytic activity measurements for the same reason. Wild-type or mutant P5abc was 10 μM, sufficient to saturate binding of $E^{\Delta P5abc}$ (200 nM) based on equilibrium values calculated from rate constants for binding and dissociation (Table 1). Reactions were performed as described previously with trace 5′-^{32}P-labeled substrate (59, 75). Solution conditions were 50 mM Na-MOPS, pH 7.0, 10 mM Mg^{2+} at 25 °C.

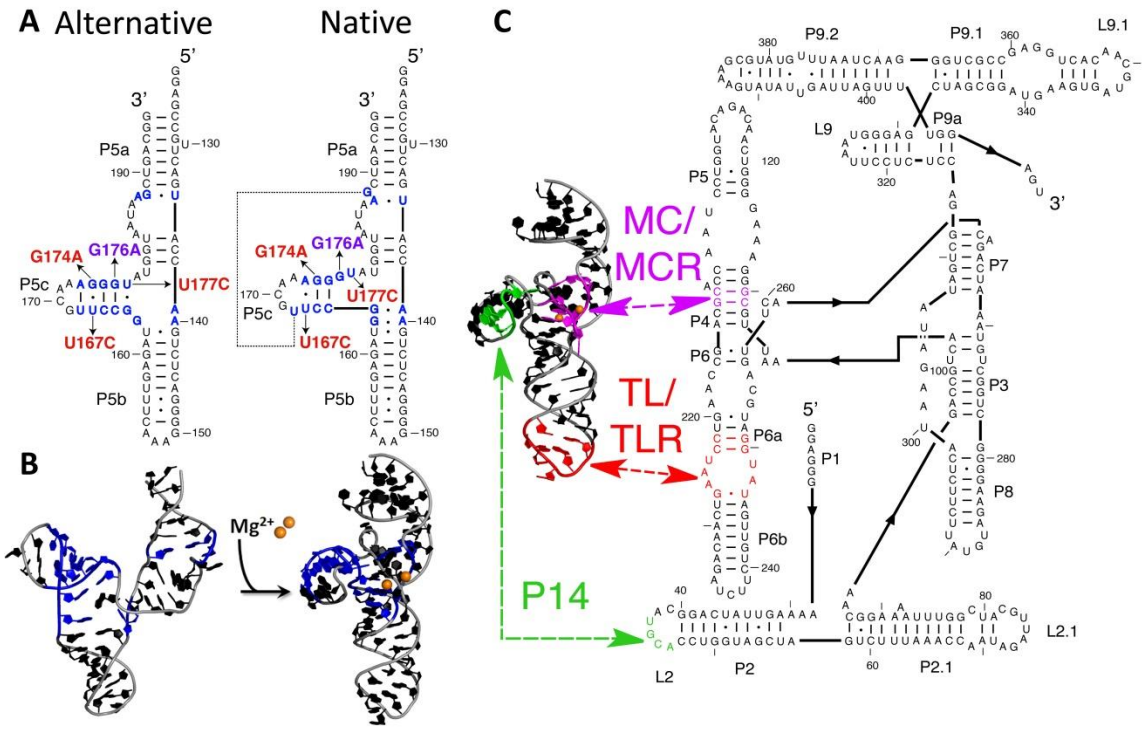

Figure 3.1: Mutations in P5c to probe P5abc assembly.

Figure 3.1: continued.,

(*A*) Secondary structure changes in P5abc induced by Mg^{2+} binding (blue). Point mutations designed to shift the equilibrium between these two structures are indicated with arrows. Mutations indicated in red stabilize the non-native secondary structure and the mutation in purple (G176A) stabilizes the native secondary structure. (*B*) The overall Mg^{2+}-induced folding transition depicted with structures of the alternative and native P5abc structures based on NMR and X-ray crystallography, respectively (66, 71). Nucleotides that change base pairing are colored blue as above, and site-bound Mg^{2+} ions are depicted as orange spheres [15, 65, 66]. (*C*) P5abc assembly with the EΔP5abc ribozyme core. The complex includes three long-range tertiary contacts: a tetraloop-tetraloop receptor interaction (TL/TLR, red), a metal core-metal core receptor interaction (MC/MCR, magenta), and a kissing loop formed by base pairing between L5c and L2 (P14, green).

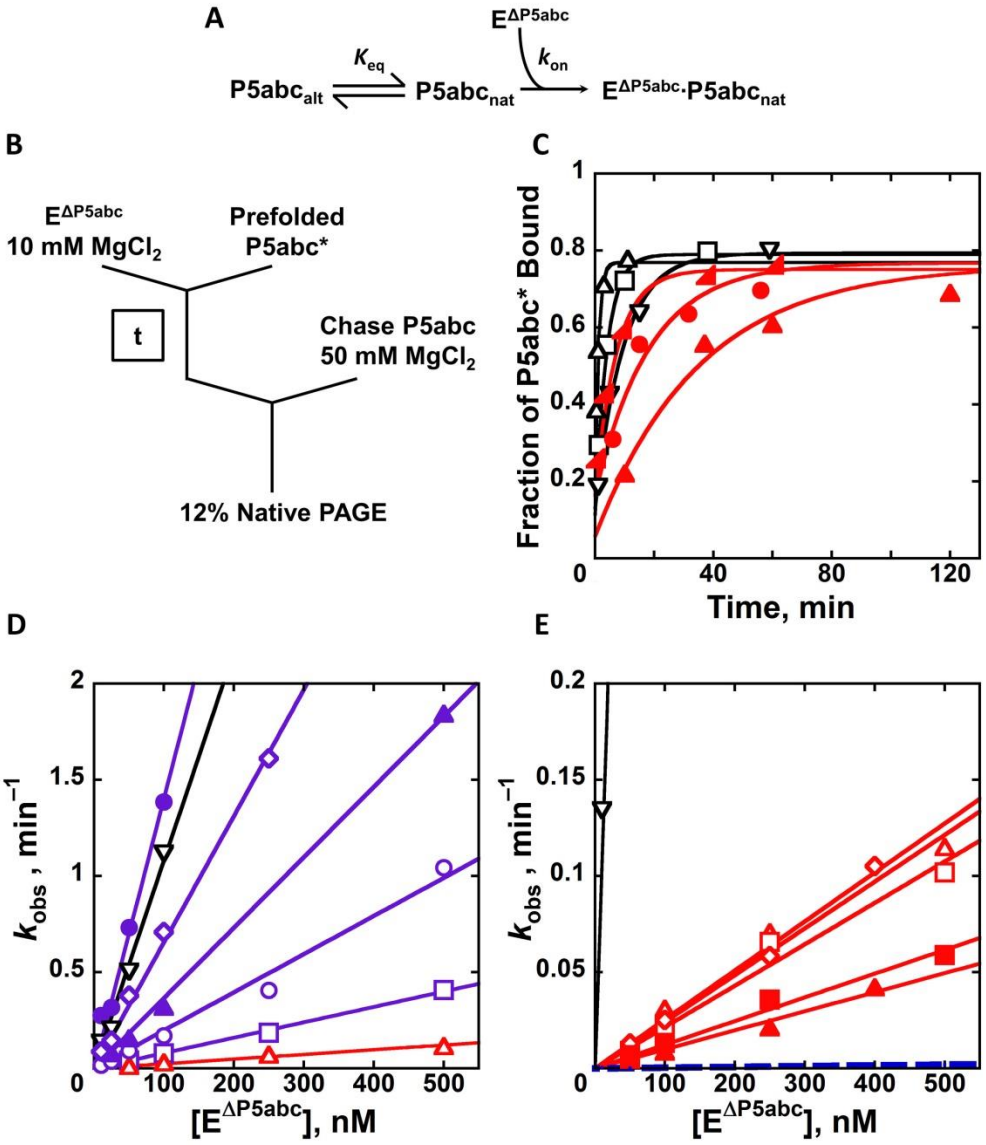

Figure 3.2: The kinetics of P5abc assembly with $E^{\Delta P5abc}$ can be tuned.

Figure 3.2: continued.,

(*A*) Minimal model for coupled folding and assembly of P5abc with $E^{\Delta P5abc}$ by conformational selection. (*B*) Pulse-chase assay to measure assembly kinetics. $E^{\Delta P5abc}$ and ^{32}P-labeled P5abc (P5abc*) were separately incubated in Mg^{2+} solution to permit folding, and then they were mixed together to initiate assembly. After various times t, further association of P5abc* was blocked by adding excess unlabeled P5abc. The concentration of Mg^{2+} was also increased to 50 mM to prevent dissociation of bound P5abc*. (*C*) Progress curves for assembly of wild-type P5abc (black) and U167C P5abc (red) at 25 °C, 10 mM Mg^{2+}. The $E^{\Delta P5abc}$ concentration was 10 nM (▽), 25 nM (□), and 100 nM (△) for reactions with wild-type P5abc, and 100 nM (▲), 250 nM (●), and 1 uM (◢) for reactions with U167C P5abc. (*D*) Dependence of the observed rate constant on $E^{\Delta P5abc}$ concentration. P5abc variants were: wild type (▽), G176A (●), U167C (△), U167C/G176A (▲), G174A/G176A (○), U177C/G176A (◇), and U167C/U177C/G176A (□). All P5abc mutants that include the G176A mutation are shown in purple. (*E*) The slowly assembling mutants shown on a scale that allows visualization of the concentration dependences. Wild-type P5abc is shown in black (▽) and mutants U167C (△, reproduced from panel d for comparison), G174A (□), U177C (◇), U167C/U177C (▲), and G174A/U177C (■) are red. The blue dashed line shows the behavior of U167C/U177C predicted from the conformational selection model in panel (*A*) with additive effects of the two mutations.

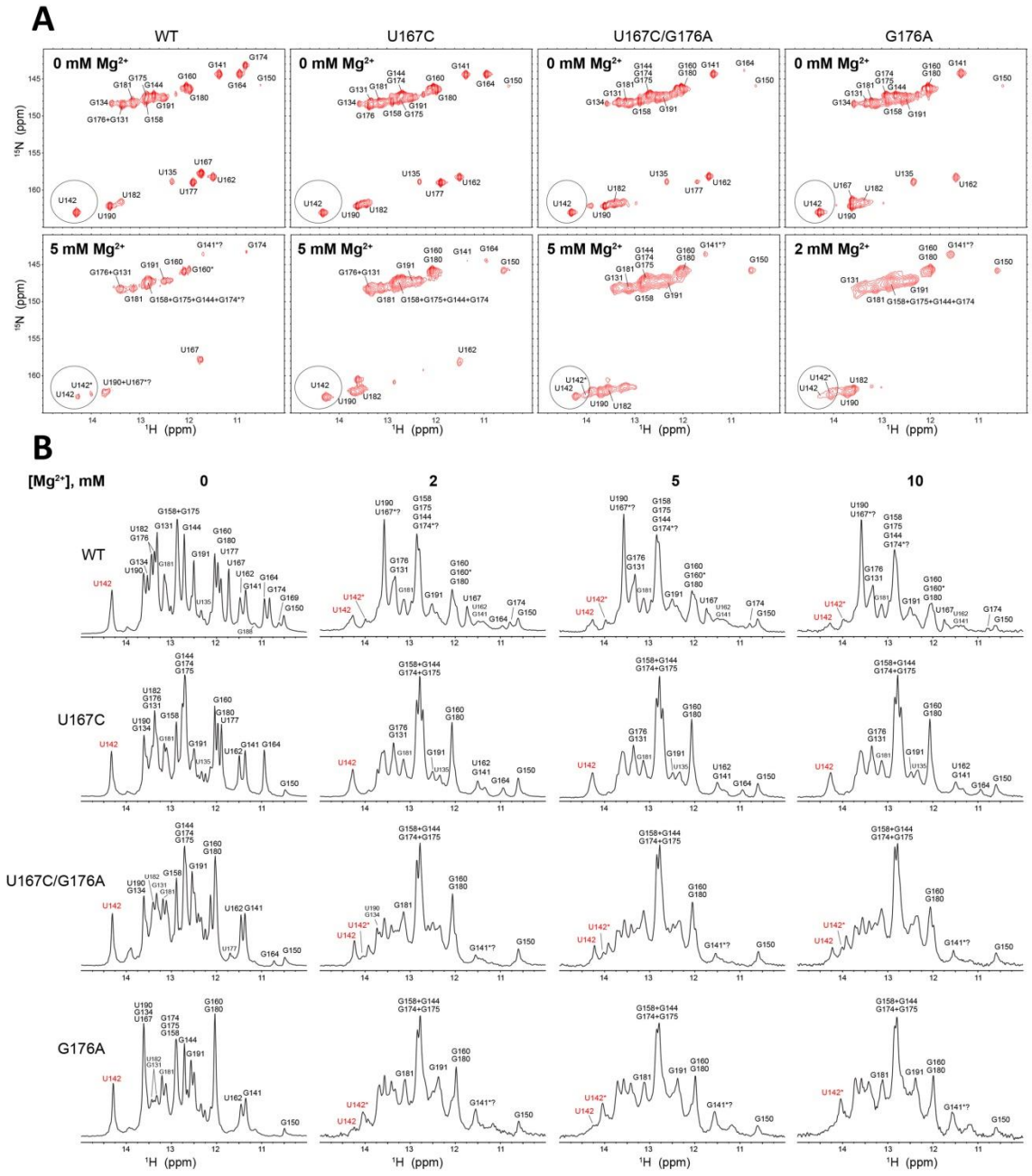

Figure 3.3: Equilibrium measurements of P5abc mutants by NMR.

Figure 3.3: continued.,

(*A*) 2D ^{1}H-^{15}N SOFAST-HMQC spectra measured in the absence of Mg^{2+} (upper panels) and in the presence of 5 mM Mg^{2+} (lower panels; 2 mM Mg^{2+} was used for G176A to avoid severe signal deterioration). The U142 and U142* peaks are highlighted by black circles. (*B*) 1D imino SOFAST-HMQC spectra measured under the specified Mg^{2+} concentrations. The U142 and U142* peaks are highlighted by red labels. For all NMR experiments, RNA concentrations were 0.1 mM and spectra were collected at 10 °C.

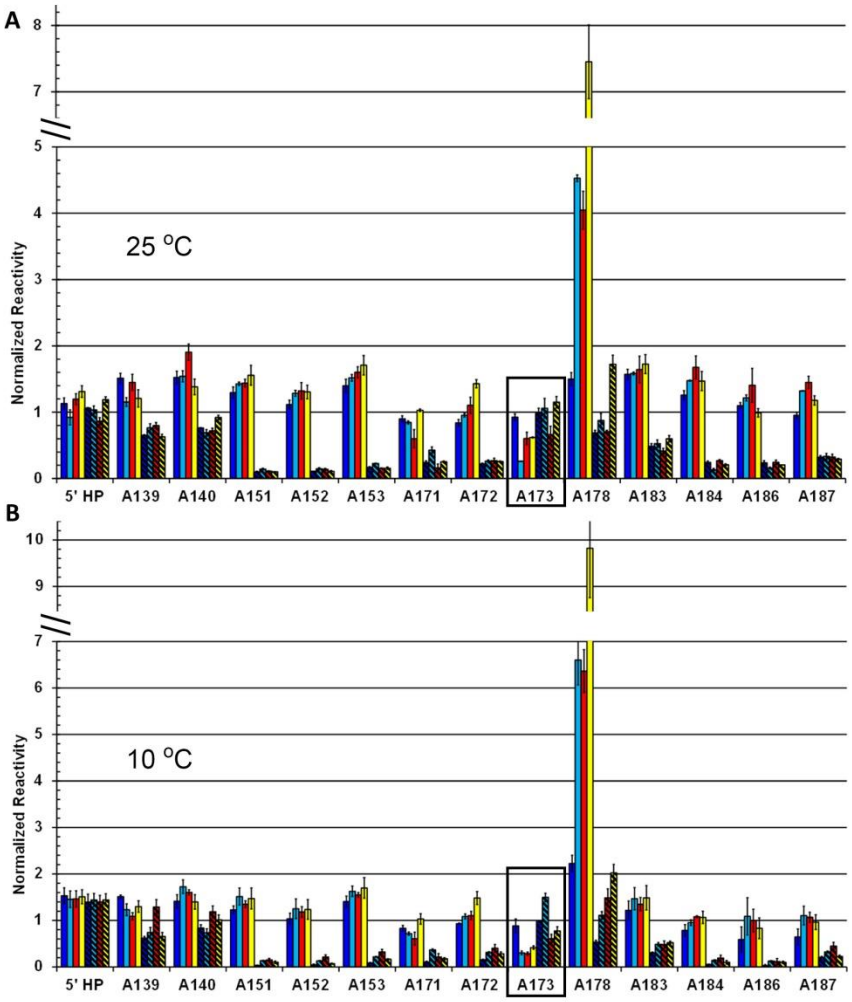

Figure 3.4: DMS footprinting of P5abc variants in complex with $E^{\Delta P5abc}$.

Figure 3.4: continued.,

DMS footprinting of P5abc within the assembled ribozyme complex. Measurements were performed at 25 °C (*A*) or 10 °C (*B*) for wild-type P5abc (blue), U167C (cyan), G174A (red), and U177C (yellow). Solid bars show data in the absence of $E^{\Delta P5abc}$ and striped bars show data for the complex with 1 μM $E^{\Delta P5abc}$. A173 (enhanced in native) is boxed. 5′ HP (Hairpin) represents the average reactivity of the adenines in the loop of this reference hairpin (loop sequence GAGUA). For 25 °C, error bars represent the standard error from two measurements, and for 10 °C error bars represent the standard error from three measurements.

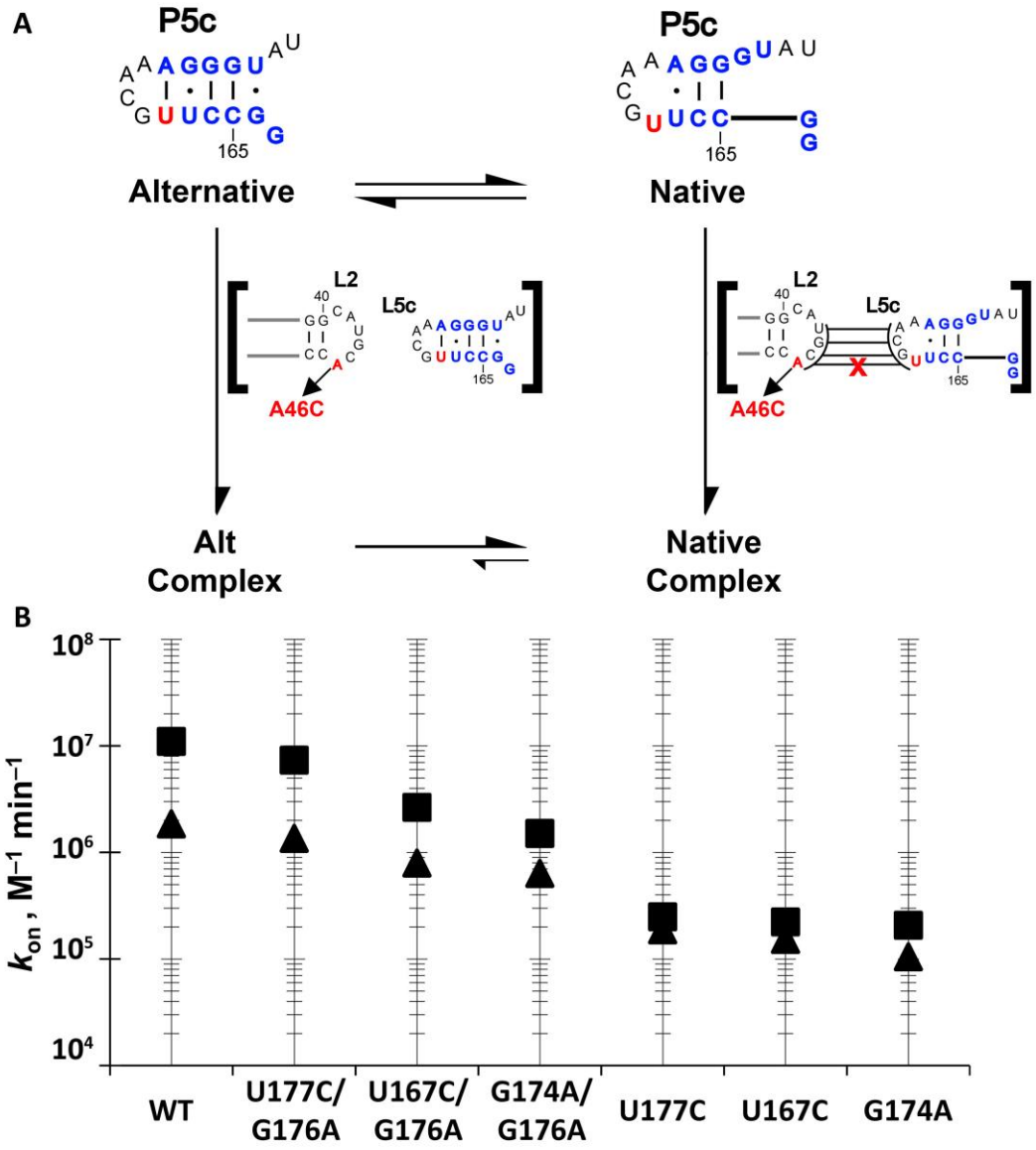

Figure 3.5: Assembly kinetics with A46C $E^{\Delta P5abc}$ provides evidence for an alternative assembly pathway.

Figure 3.5: continued.,

(*A*) Model for core assembly by the alternative (left) and native (right) P5abc structures. The L2 mutation A46C is expected to disrupt a base pair in the transition state for the native P5abc and therefore slow assembly (right), whereas this mutation is not expected to slow assembly with alternative P5abc. Nucleotides A46 and U168, which form this P14 base pair in the native conformation, are red, and P5c nucleotides that change secondary structure between the alternative and native structures are blue. (*B*) Assembly rate constants for wild-type P5abc and the indicated P5abc mutants with the wild-type $E^{\Delta P5abc}$ core (squares) or the $E^{\Delta P5abc}$ mutant A46C (triangles) at 10 mM $MgCl_2$ and 25 °C. Error bars representing the standard error from at least two determinations are present but not visible because they are smaller than the markers.

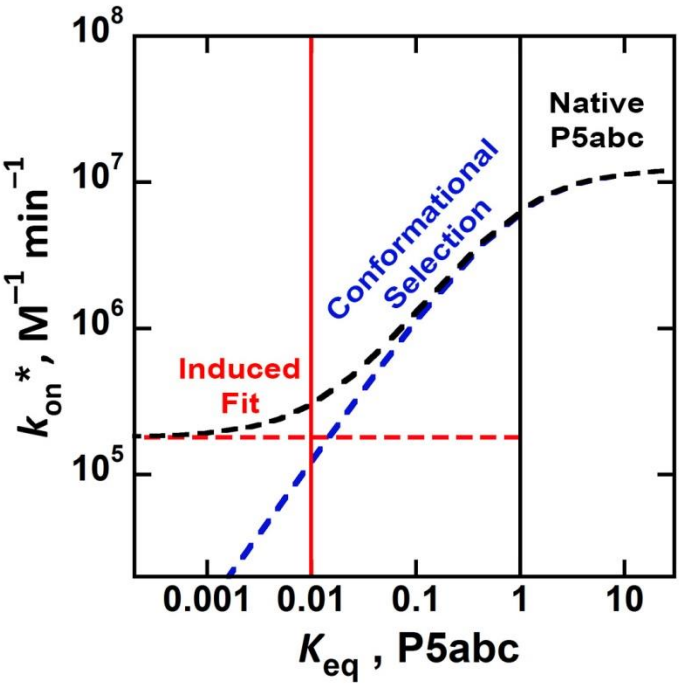

Figure 3.6: The local folding transition controls rate and assembly pathway.

Dependence of the observed assembly rate constant and dominant pathway on the equilibrium for native P5abc formation. The dependences of rate constant on K_{eq} for each pathway are shown as dashed lines with colors corresponding to the labels and the black dashed line shows the observed rate constant predicted by the observed rates.

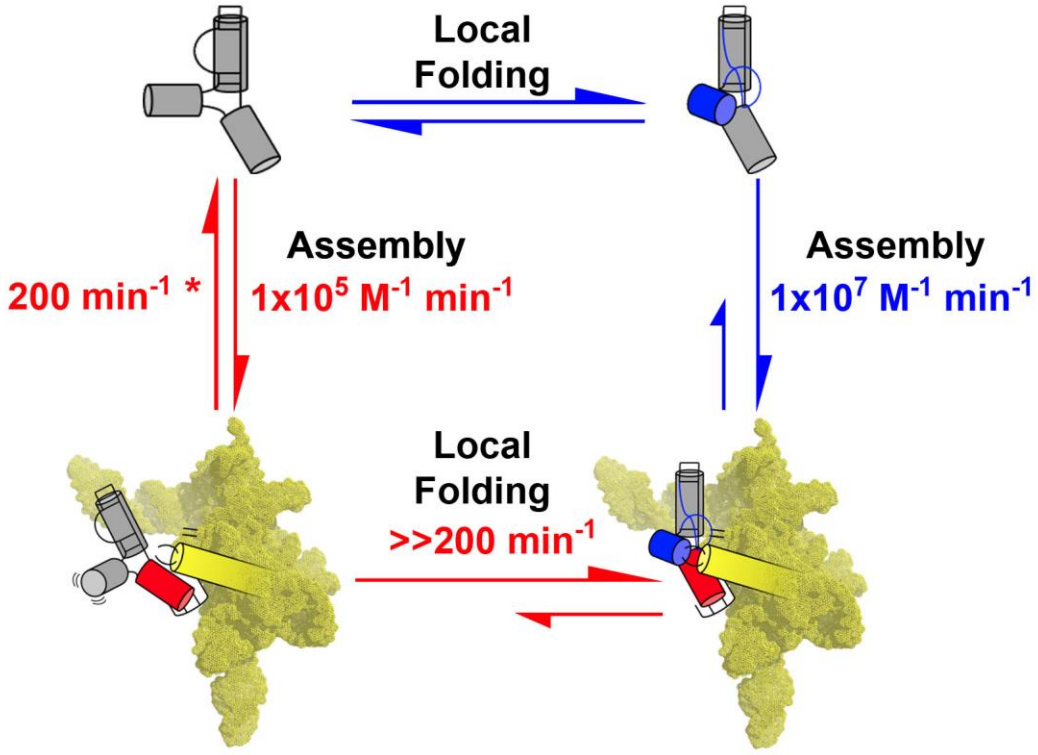

Figure 3.7: Two pathways for RNA assembly.

Rapid rearrangement of P5abc follows initial assembly along the induced fit pathway. The asterisk indicates that the value reflects the measured rate of dissociation for an isolated TL/TLR contact (26) (see Methods section: *Estimation of rate constant for TL/TLR formation*).

P5abc Variant	ΔΔG°, kcal/ mol**	k_{on} P5abc, M^{-1} min^{-1}	k_{rel} P5abc†	k_{on} A46C, M^{-1} min^{-1}	k_{rel} A46C††
wild-type	0	$1.1 \pm 0.1 \times 10^7$	(1)	$1.9 \pm 0.3 \times 10^6$	0.17 ± 0.05
G176A	-6.7	$1.4 \pm 0.3 \times 10^7$	1.2 ± 0.3	N.D.	N.D.
U177C/G176A	-4.5	$7.3 \pm 1.0 \times 10^6$	0.65 ± 0.10	$1.4 \pm 0.2 \times 10^6$	0.18 ± 0.04
U167C/G176A	-2.3	$2.6 \pm 0.2 \times 10^6$	0.23 ± 0.03	$7.9 \pm 0.1 \times 10^5$	0.30 ± 0.06
G174A/G176A	0.1	$1.5 \pm 0.2 \times 10^6$	0.13 ± 0.02	$6.3 \pm 0.3 \times 10^5$	0.42 ± 0.07
U167C/U177C/ G176A	-0.1	$7.8 \pm 0.3 \times 10^5$	0.069 ± 0.007	N.D.	N.D.
U177C	2.4	$2.5 \pm 0.1 \times 10^5$	0.022 ± 0.002	$1.9 \pm 0.2 \times 10^5$	0.75 ± 0.10
U167C	4.4	$2.2 \pm 0.3 \times 10^5$	0.020 ± 0.003	$1.5 \pm 0.2 \times 10^5$	0.68 ± 0.13
G174A	6.8	$2.1 \pm 0.3 \times 10^5$	0.018 ± 0.003	$1.1 \pm 0.1 \times 10^5$	0.51 ± 0.10
U167C/U177C	6.8	$1.2 \pm 0.2 \times 10^5$	0.010 ± 0.002	N.D.	N.D.
G174A/U177C	9.2	1.2×10^5	0.010	N.D.	N.D.

Table 3.1: Assembly rate constants measured for P5abc variants with $E^{\Delta P5abc}$ or A46C $E^{\Delta P5abc}$ ribozyme cores. *Conditions were 50 mM Na-MOPS, pH 7.0, 10 mM $MgCl_2$, 25 °C. **ΔΔG° values reflect the difference in predicted free energy between the alternative and native secondary structures for each variant relative to the corresponding free energy difference for the wild-type P5abc. Predicted free energies were generated using ViennaFold with solution conditions set to 37 °C and 1 M NaCl. †krel P5abc is the assembly rate constant of each P5abc variant divided by the assembly rate constant of wild-type P5abc. ††krel A46C is the assembly rate constant of each P5abc variant with the A46C $E^{\Delta P5abc}$ divided by the corresponding rate constant for assembly with the wild-type $E^{\Delta P5abc}$. Uncertainty values reflect the standard error of at least two measurements. The reported value for the G174A/U177C variant does not include an uncertainty estimate because assembly for this variant was measured only once under these conditions. N.D., not determined.

	K_{eq}		
P5abc variant	2 mM Mg^{2+}	5 mM Mg^{2+}	10 mM Mg^{2+}
wild-type	*N.Q.*	1.0	*N.Q.*
G176A	2.2	*N.Q.*	*N.Q.*
U167C/G176A	0.31	0.71	0.79
U167C	<0.032†	<0.032†	<0.032†

Table 3.2: P5abc folding equilibrium measured by NMR. Equilibrium values were calculated as the ratio of the peak volumes observed at U142 and U142*. †An upper limit on Keq was established from the absence of a detectable U142* peak. N.Q., not quantified.

P5abc	Fraction Native P5abc*
wild-type	0.60
G176A	0.81
U167C/G176A	0.35
U167C	0.02

Table 3.3: P5abc folding measured by quantitative DMS footprinting. Data were collected at 5 mM $MgCl_2$, 50 mM Na-MOPS, pH 7.0, 10 °C. *The fraction of native P5abc was determined by normalizing the reactivity of A173 by the maximal reactivity of the G176A variant determined by Mg^{2+} titration.

P5abc•$E^{\Delta P5abc}$	***k*, rSA_5 (min^{-1})**	***k*, -3m,rSA_5 (min^{-1})**
wild-type	2.6	0.012
U177C	4.1	0.0062
G174A/U177C	3.9	0.017

Table 3.4: Catalytic activity measurements of P5abc variants assembled with the ribozyme core. Data were collected at 200 nM $E^{\Delta P5abc}$, 10 μM P5abc, 10 mM $MgCl_2$, 55 μM guanosine, 50 mM Na-MOPS, pH 7.0, 25 °C using trace ^{32}P-labeled substrate.

Chapter 4: RNA assembly pathways utilize different tertiary contacts in the assembly transition states

INTRODUCTION

Structured RNAs are ubiquitous in cellular metabolism. The exponential growth of high resolution RNA structures has led to a clear view of the cell where RNA-protein machines regulate higher order processes. Indeed the ribosome, spliceosome, and telomerase are all machines that include structured RNAs, and the dynamic movements of the structured RNA components are directly linked to catalytic function (10, 91, 93, 153-156). It might not be naïve to think that RNPs are made up from a 'skeleton' of RNA structure required for function, while proteins are the decorative 'additions' that function in signaling and improve efficiency.

Transient RNA structures are widespread in biology (29). At the start of biogenesis, RNAs fold *in vivo* during transcription. The lifetime of RNA secondary structure relative to mRNA degradation rate (157-160) suggests that the first secondary structure to form during transcription dictates the fate of the RNA. For example, during transcription a riboswitch RNA folds along a given folding pathway that leads to a particular structure and functional outcome (30, 32, 161). Under certain cellular conditions, the folding pathway can be switched leading to a different structure and outcome (162). In the case of a riboswitch, a metabolite ligand changes the structural state by physically binding to the RNA, stabilizing a secondary structure that is incompatible with alternative structural states. To fully understand the factors that determine RNA folding pathway followed requires a fundamental understanding of how tertiary structure and long range interactions impact RNA folding pathways.

To more closely understand the determinants of RNA folding pathway followed during folding and assembly of an RNA secondary structure switch, we dissected the

assembly pathways of P5abc with the P5abc-deleted ribozyme core ($E^{\Delta P5abc}$). The P5abc subdomain can fold in isolation from an extended conformation with an alternative set of base pairs to the native structure with a switched base pairing register (69-71) (see Figure 1.5, Chapter 1). The folded P5abc forms three distinct long range tertiary interactions with the ribozyme core ($E^{\Delta P5abc}$) to form a long lived complex with femtomolar affinity and a lifetime on the order of years (74, 75). We previously demonstrated that assembly of P5abc with $E^{\Delta P5abc}$ occurs through two pathways: assembly of the high affinity folded form of P5abc (conformational selection) or assembly of the slower binding of the alternative extended conformation (induced fit) (101). Because the P5c folding transition fundamentally changes the P5c helix that forms the P14 contact with the ribozyme core, we hypothesized that the failure to follow the faster conformational selection pathway was due to an inability to form P14 in the alternative P5abc conformation. We proposed that the rate of folding along each pathway might reflect the rate of tertiary contact formation for an individual tertiary contact nucleating in the assembly transition state, but no further dissection was carried out to address this hypothesis (101).

Here, we use mutagenesis and pulse-chase assays to measure the contribution of each tertiary contact (P14, TL/TLR, or MC/MCR) to each folding pathway (conformational selection or induced fit). The results suggest that P14 forms early in the conformational selection pathway and the MC/MCR forms late in folding, after the rate-limiting step. The slower assembling induced fit pathway utilizes the TL/TLR tertiary contact because P14 is unable to form with the alternative P5abc, followed by formation of P14 and MC/MCR. P14 deletion mutants switch the preference for assembly pathway from conformational selection to induced fit because the TL/TLR contact is now the rate limiting assembly step. These results suggest that the rate or tertiary contact formation

dictates the order of tertiary contact assembly and the folding pathway that receives the dominant flux.

Results

To measure the contribution of each tertiary contact to the transition state for assembly, we generated a two-piece version of the *Tetrahymena thermophila* ribozyme (Figure 4.1A). The rate constants for association and dissociation can be directly measured by a pulse-chase assay using radiolabeled P5abc (Figure 4.1B and C). Point mutations in the ribozyme core at each of the three tertiary contacts were previously shown to have an effect on folding or assembly (28, 63, 163). First, we test the contribution of each tertiary contact in the assembly transition state of the conformational selection pathway (pre-folded modules). Then, we carry out similar measurements for the assembly transition state of the induced fit pathway with P5abc mutants that lock P5c into the alternative conformation.

Probing tertiary contacts in the transition state of the pre-folded modules

Under the conditions of assembly, wild-type P5abc is fully folded (Figure 4.2A). We measured the rate constant for binding trace P5abc at various concentrations of $E^{\Delta P5abc}$ and determined the second order rate constant for binding (Figure 4.2B and C). Mutations in the TLR contact slowed the assembly rate by 5-fold, and mutations in P14 slowed the assembly rate by 4-6 fold compared to wild-type. These results suggest that P14 and the TLR form before the assembly transition state, because mutations that destabilize these contacts decrease the rate constant for assembly. At high concentrations of $E^{\Delta P5abc}$ (>10 μM) there is observable precipitation consistent with previous reports (previous observations, R.R. unpublished). When we made mutations in the MCR, we observed no rate dependence on the concentration of $E^{\Delta P5abc}$ up to the concentrations

feasible for measurements of assembly, potentially owing to the large effect on k_{off} for MCR mutants.

To compare dissociation rate constants, we had to find solution conditions that would allow for measurement of wild-type and mutant dissociation kinetics on a feasible time scale. Previous measurements showed that the half-life of the wild-type P5abc complex is on the order of years under standard conditions (10 mM $MgCl_2$ and 25 °C) (75), so we adopted solution conditions for association and dissociation measurements that accelerate the dissociation rate to a measurable time-scale (10 mM $MgCl_2$, 100 mM KCl, 37 °C). We found that mutations in P14 had modest effects on the dissociation rate (up to 10-fold faster than wild-type P5abc), suggesting that P14 forms early in the folding pathway. Mutations in the TLR had larger effects (up to 50-100-fold faster than wild-type) suggesting that the TL/TLR contact forms after the P14 contact. Consistent with our previous observations in k_{on} measurements, mutations in the MCR contact substantially increased the rate constant for dissociation to close to 0.4 min^{-1}, meaning that association rates can only be measured at concentrations of $E^{\Delta P5abc}$ where k_{on} is much faster than 0.4 min^{-1} (approaching the limit of reliable rate constants that can be measured by hand). Taken together, the assembly rate constant is equally affected by mutations in P14, TLR, and MCR with near equal contributions of each contact to the assembly rate (Figure 4.2F). The dissociation rate constants show a preference for particular tertiary contacts, with P14 mutations contributing the smallest effect and MCR mutations contributing the largest effect (Figure 4.2F).

Phi-value analysis (Φ-value) can be used to determine the contribution of the mutated contact to the transition state for assembly (164-166). By measuring the association and dissociation rate constants for wild-type and mutant complexes, the contribution of the effect from a given mutation to the transition state can be assessed by

comparing the total effect on the equilibrium for complex formation. Phi-values typically are expressed from '0' to '1' where lower values indicate that contact is not formed in the transition state and higher values suggest that the contact is fully formed in the transition state. This analysis showed that the P14 tertiary contact is more formed in the assembly transition state than the MCR contact, though the phi-values for the P14 contact are modest (about 0.4). The partial phi-value observed for the TLR contact suggests that this contact may also partially form before the transition state of the conformational selection pathway.

Probing tertiary contacts in the transition state during induced fit assembly

We utilized a similar approach to probe the assembly transition of the induced fit pathway, making mutations in each of the three tertiary contacts and measuring the second order rate constant for binding. Similar to what we observed when probing the conformational selection pathway, mutations in the MCR contact weakened binding to an immeasurable extent. Still, we observed that mutations in P14 had almost no effect on the assembly rate of alternative P5abc, while the TLR mutants slowed the assembly rate constant between 5-20 fold compared to assembly of alternative P5abc with wild-type $E^{\Delta P5abc}$ (Figure 4.3A). These results strongly suggest that assembly through the induced fit pathway nucleates the TL/TLR contact before the transition state, and forms P14 late in the folding pathway after assembly, in agreement with previous predictions (101).

To further test this model, we conducted dissociation rate measurements with P5abc mutants that follow the induced fit assembly pathway. First principles of microscopic reversibility demand that the dominant pathway is the same in both directions. Our previous results showed that the induced fit pathway binds alternative P5abc and switches P5abc to the native conformation in the assembled ground state

complex after the assembly transition state (101). Thus, the measured dissociation rate through the induced fit pathway is expected to strongly depend on the stability of native P5c relative to the alternative conformation, because P5c must switch to the alternative conformation before reaching the transition state for disassembly (Figure 4.3B). Indeed the dissociation rate of mutants that stabilize alternative P5c secondary structure over a 9 kcal/mol range varied by 50 fold compared to each other (Figure 4.3C). We also observed an initial fast phase for most P5abc mutants, and further work demonstrated that this fast phase reflects dissociation from misfolded $E^{\Delta P5abc}$ or damaged RNA that rapidly dissociate. These results further confirm our previous model: the induced fit pathway assembled with P5abc in the alternative conformation and P5c switching and P14 formation form after the transition state.

We then probed the contribution of each tertiary contact to the transition state by utilizing the $E^{\Delta P5abc}$ variants shown above. Shown are progress curves for dissociation of P5abc mutant U167C/U177C. We observed that point mutations in the TLR increase the dissociation rate by no more than 3-fold, a much smaller effect than for mutants in P14, which gave 8-10-fold increases in dissociation rate. Overall, the TL/TLR contact makes a larger contribution than P14 for assembly through the induced fit pathway but a smaller contribution for dissociation (Figure 4.4A and B). We calculated phi-values for each of the mutants that follow the induced fit pathway, and indeed, for nearly all P5abc mutants, the TL/TLR contact contributes more to the transition state than P14 (for U167C/U177C: Φ-value of 0.7 for TL/TLR and about 0.2 for P14). Mutations in the MCR had a large effect on k_{off} through the induced fit pathway (about 300-fold faster dissociation rate) which weakens the affinity so that k_{on} was once again immeasurable. Thus, the contribution of the effect on k_{off} relative to the total effect on the equilibrium cannot be determined for MCR mutants, precluding determination of a phi-value. Still, given the

threshold of observed effects on k_{on} for mutants in the TLR contact and P14 (at most a 10-fold effect), it is reasonable to assume that most of the observed effect on k_{off} for MCR mutants (300-fold) is not due to weakening of k_{on} to the same extent (10-fold for TLR).

We conclude that in the induced fit assembly pathway the TL/TLR is formed in the transition state for assembly, in agreement with calculations of the assembly rate constant for an isolated TL/TLR (see Chapter 3 Methods section: *Estimation of rate constant for TL/TLR formation*). After the assembly transition state, P5c switches to the native conformation and P14 forms. We previously demonstrated that there is cooperativity between P5c switching and metal core (MC) folding in isolated P5abc (see Chapter 2), suggesting that after P14 forms and P5c switches, the MC forms. Thus, the last step in folding of induced fit is presumably formation of the MC/MCR contact. Full determination of the phi-value for the MC/MCR contact is required to confidently make this conclusion.

Controlling the rate and pathway with rational mutagenesis of the structural modules

With a model in hand for how P5abc assembles and the knowledge of the tertiary contacts formed in the assembly transition state along each pathway, we designed constructs that could specifically test predictions made by our model. If the faster conformational selection pathway is faster because kissing loop contacts (P14) form 100-fold faster than TL/TLR contacts, then we predicted that in the absence of P14, the wild-type P5abc would now assemble through the slower induced fit pathway with a rate constant in agreement with nucleation of a TL/TLR contact. We generated P14 knock out mutants by mutating loop L2 to UUCG in the ribozyme core or mutating the P5abc loop L5c to UUCG (Figure 4.5A). Unexpectedly, our initial P14 KO mutants had a 10,000-fold effect on the binding rate constant of wild-type P5abc, 100 fold slower than what is

expected for assembly through the induced fit pathway (Figure 4.5B, black square vs. black diamond and circle). This result perplexed us because if a faster forming TL/TLR contact (k_{on} near 1×10^5 M^{-1} min^{-1}) is available for contact formation, then that pathway should receive the greatest flux when P14 is deleted. This led us to propose the possibility that when P14 is unable to form, the P2 helix itself within the ribozyme core somehow occludes the TLR binding site, resulting in a pre-equilibrium for moving the P2 helix out of the way and a penalty for the rate of association. To test this model, we generated a construct where the P14 contact is knocked out by mutating L2 to UUCG and substantially truncating the P2 helix to three base pairs (Figure 4.5A, ΔP2). Truncation of P2 accelerated the rate constant of binding for wild-type P5abc 100-fold relative to L2UUCG and L5cUUCG to the level predicted for nucleation of the TL/TLR contact in the assembly transition state (Figure 4.5B, red diamond).

We also tested double mutants containing the G176A mutation because these mutants were previously established to use the conformational selection pathway when assembling with the wild-type $E^{\Delta P5abc}$. When we measured the assembly rate constants with ΔP2 ribozyme core, all three G176A double mutants assembled with a rate constant almost equal to that observed for the wild-type P5abc with ΔP2 (Figure 4.5B, red dashed line). We observed that ΔP2 had a 10-fold slowing effect on the rate constant for binding mutants that stabilize alternative P5abc, possibly because these mutants do not have a properly formed MC within P5abc, resulting in a penalty for forming this contact and a slower association rate than that observed for TL/TLR. In agreement with this model, the G176A double mutant series and wild-type P5abc's all properly form the MC in P5abc under the conditions of the association measurements and assemble with a rate constant equal to the TL/TLR contact and induced fit pathway.

Lastly, we hypothesized that the assembly rate observed for the induced fit pathway in the context of wild-type $E^{\Delta P5abc}$ could be rescued if the faster forming P14 contact were re-established. To test this prediction, we measured the rate constant for association of mutant G174A with $E^{\Delta P5abc}$ and then generated a mutant that restores the A46-U168 base pair in P14 by inserting an additional 'U' nucleotide into loop L5c (Figure 4.5A, G174A+Uins). Assembly of the Uins mutant was accelerated 5-10-fold relative to G174A assembling with the $E^{\Delta P5abc}$ ribozyme core. Our results suggest that the faster forming P14 kissing loop is a kinetic highway for structure formation, and when the full P14 is unable to form, an alternative pathway is available through the slower forming TL/TLR contact. Following assembly through the TL/TLR contact, P14 is able to form and P5c switches to the native state and the MC forms in the assembled ground state complex (101).

Discussion

At a physical level, the presence of alternative pathways highlights the modularity in RNA folding. The conformational selection pathway relies on formation of native structure in P5abc and kissing loop P14 formation to nucleate assembly (Figure 4.6A, blue pathway). The induced fit pathway is an alternative assembly pathway that comes to dominate when P14 is unable to form (Figure 4.6B, red pathway). Instead, the TL/TLR contact, which is distant from P5c and not expected to be impacted by the local folding transition of P5abc, is formed in the transition state for assembly, and this assembly occurs with the same rate constant as that of an isolated TL/TLR (26). The commonality between rate constants suggests that the rate of RNA assembly even for multi tertiary contact forming RNA assembly may be determined simply by the probability of

productive collisions for a given tertiary contact structure module. This principle could prove powerful for engineering of precisely timed novel structured RNAs.

The pre-folded modules assemble through multiple pathways

Phi-value analysis of the assembly transition state in the conformational selection pathway suggests that no contact is fully formed in the transition state (highest Φ-value 0.45). One model to account for this result is initial and partial nucleation of a particular tertiary contact followed by rapid downhill formation of all other contacts. Taking the phi-value analysis literally suggests that P14 is partially formed in the transition state, and after the transition state there is rapid and downhill formation of the TL/TLR and MC/MCR contacts. These results are in agreement with measurements of the affinity of the native complex (75), one of the tightest binders currently measured in biochemistry. The small phi-value for forming the MC/MCR contact suggests that it is the last contact formed during conformational selection and induced fit assembly, potentially owing to the instability of the a-minor interaction in isolation (28) (see Chapter 1 Table 1.1).

Our measurements of k_{on} suggest an alternate model where the TL/TLR contact is partially formed in the transition state (Figure 4.2C). We propose a second possible model where there is partial P14 formation, because it forms faster (22, 26), followed by partial formation of the TL/TLR (Figure 4.6A). This model requires that the TL/TLR form 100-fold faster in the context of P14 because the maximal rate constant for assembly in conformation selection is 100-fold faster than that of an isolated TL/TLR. In addition, both models could be correct, with multiple assembly pathways possible through a conformational selection process. A third possible model would be rapid formation of the TL/TLR, followed by rapid formation of P14, leading to the transition

state, but this is unlikely because the initial formation rate of TL/TLR is 100-fold slower than the maximal rate of the conformational selection pathway and the rate of P14.

A TL/TLR alternative assembly pathway switches P5c and forms P14 late during folding

When P5abc folding to the native conformation is perturbed with mutation, an alternative induced fit assembly and folding pathway is followed (101). The local folding transition within the P5abc structural modules controls the rate and pathway of RNA folding and assembly. Phi-value analysis of the assembly transition state in the induced fit pathway suggests that the TL/TLR contact is near fully formed in the transition state (highest Φ-value 0.85). After the assembly transition state, there is an intramolecular rearrangement within the complex where P14 forms and P5abc folds to its native conformation. Thus, substantial cooperativity arises from the tertiary contacts that form with the ribozyme core because the penalty for switching the destabilized P5abc mutants that assemble through induced fit is overcome in the ground state complex, a requirement of induced fit. We observe an upward trend in our measured phi-values for the TL/TLR with decreasing stability for P5abc folding (Figure 4.4C, A225U and A226U). One possibility is that the flux is not entirely dominated by the induced fit pathway under these conditions (10 mM $MgCl_2$, 100 mM KCl, 37 °C) for mutants that only modestly destabilize P5c switching (Figure 4.4C, U177C in purple).

In P5abc folding and assembly, individual tertiary contacts dictate the folding pathway followed. When P5abc is folded, the P14 contact is available to form resulting in a kinetic highway to the ground state complex. If P5abc is in the alternative conformation and P14 is unable to form, then the folding flux is shifted to a pathway that forms the TL/TLR contact in the transition state. Thus, the local folding transition switches the assembly pathway by changing the order of tertiary contact formation.

Methods

RNA preparation

The same protocol was used as described previously in Methods of Chapter 3.

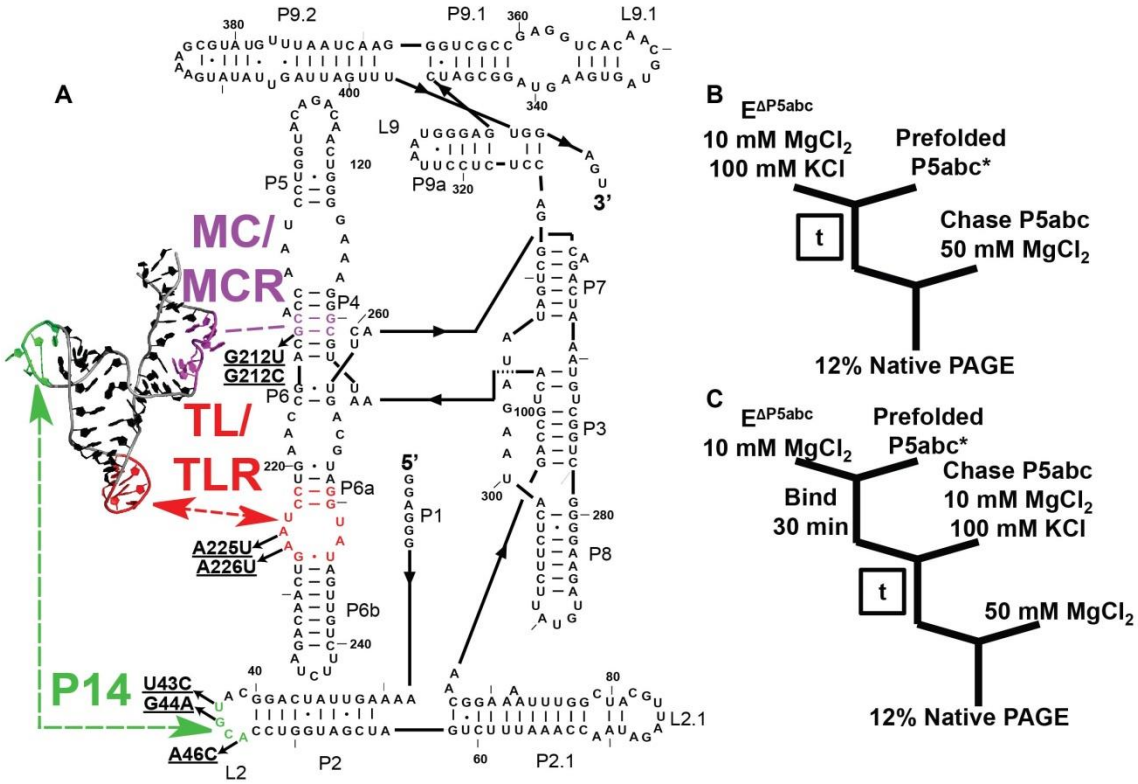

Figure 4.1: Mutations in $E^{\Delta P5abc}$ to probe the assembly transition state.

(*A*) P5abc forms three long range tertiary contacts with $E^{\Delta P5abc}$ and mutations in each of these contacts are used to probe the particular tertiary contact (P14 green, TL/TLR red, MC/MCR purple). Pulse-chase assay scheme used to measure the kinetics of association (*B*) and dissociation (*C*) are shown.

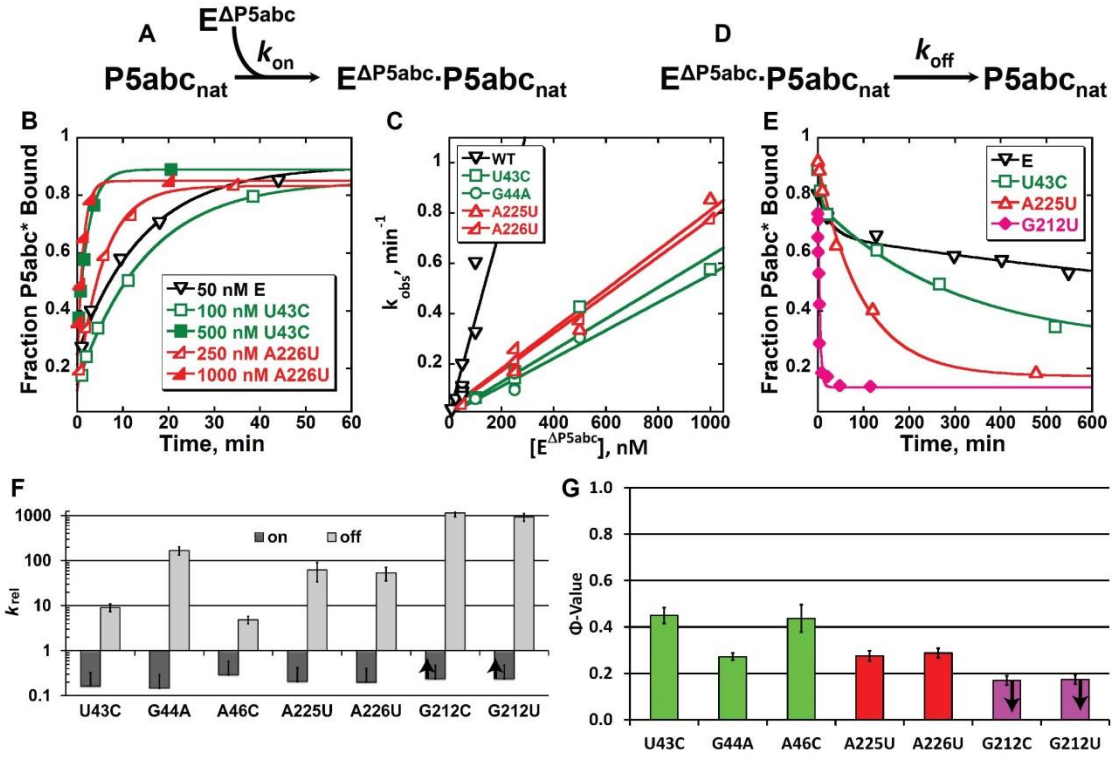

Figure 4.2: Probing assembly and dissociation of the pre-folded modules.

(*A-C*) Association and dissociation (*D, E*) kinetics measured for WT P5abc with $E^{\Delta P5abc}$ and mutants that destabilize P14 (green), the TLR (red), and MCR (magenta). (*F*) Rate constants are shown relative to the WT complex. k_{rel} is calculated by dividing the rate of the mutant by the rate of WT. (*G*) Phi-value analysis for the WT complex. Φ is calculated from $\ln(k_{on,mut} / k_{on,WT}) / \ln((k_{on,mut} \times k_{off,WT}) / (k_{on,WT} \times k_{off,mut}))$. A limit on the phi-value was determined for the MC/MCR contact because mutants of this contact result in binding that is too weak to measure k_{on}.

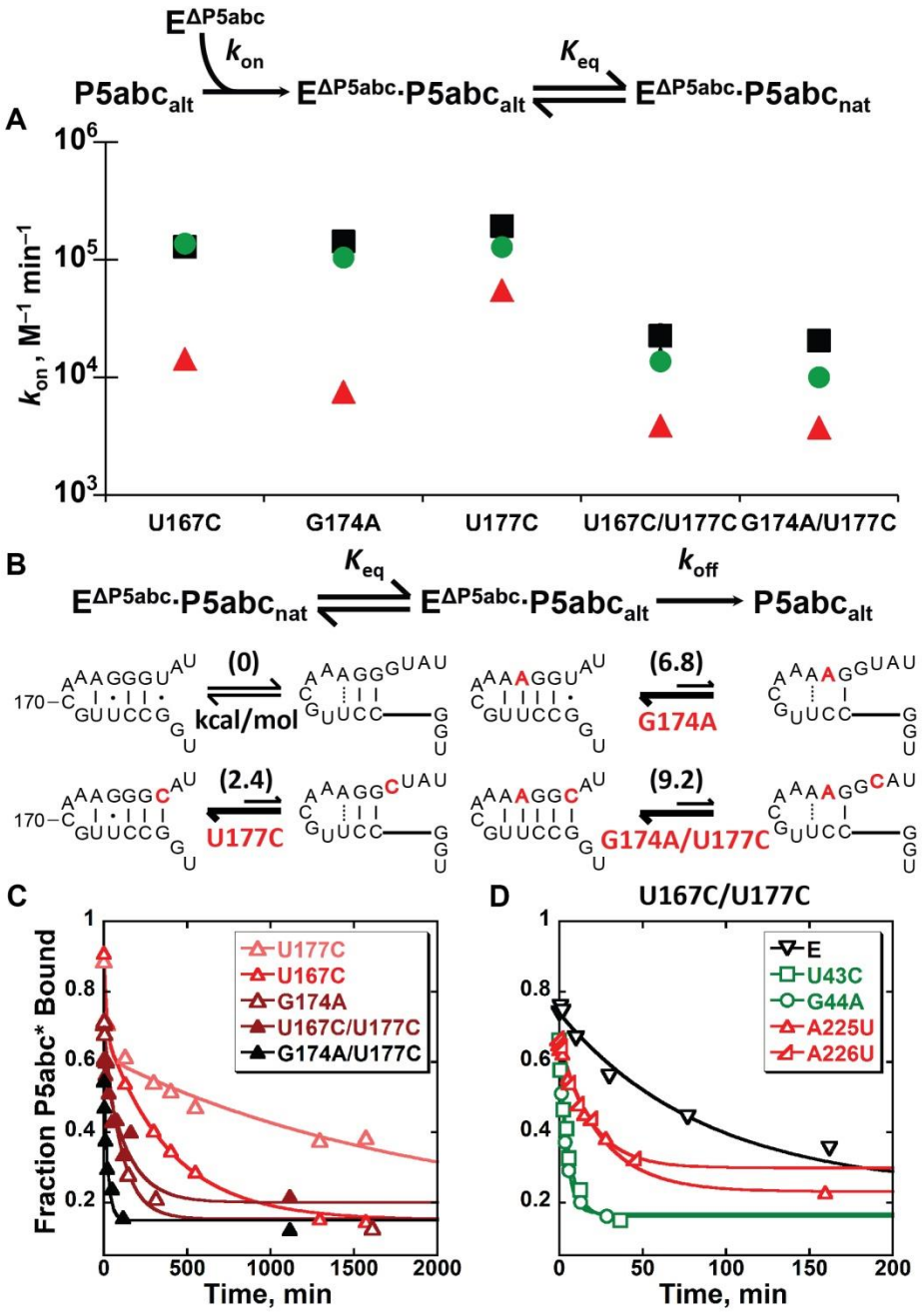

Figure 4.3: Probing assembly and dissociation of the induced fit pathway.

Figure 4.3: continued.,

(*A*) Association kinetics measured for mutants that follow the induced fit pathway. Assembly with WT $E^{\Delta P5abc}$ is shown in black squares. U43C (green cirlces) targets P14 and A225U (red triangles) targets the TLR. (*B*) Dissociation kinetics measured for mutants that stabilize the alternative P5c secondary structure. For space limitations, U167C and U167C/U177C are not shown. (*C*) Dissociation kinetics from WT $E^{\Delta P5abc}$. (*D*) Dissociation kinetics for mutant U167C/U177C with $E^{\Delta P5abc}$ mutants.

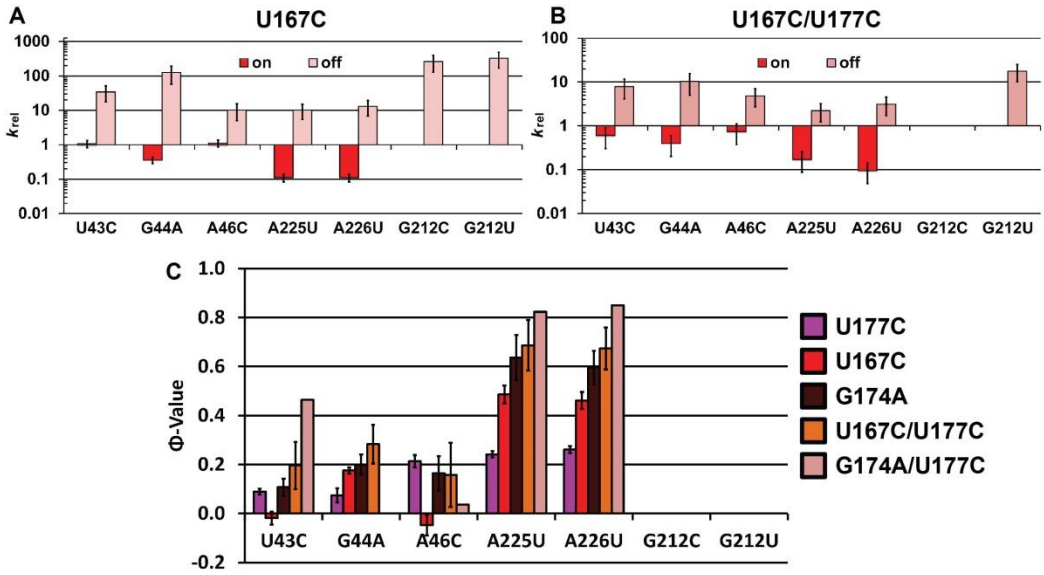

Figure 4.4: Phi-value analysis of the induced fit assembly pathway.

Rate constants for U167C (*A*) and U167C/U177C (*B*) are shown relative to the WT complex. k_{rel} is calculated by dividing the rate of the mutant by the rate of WT. (*C*) Phi-value analysis for the induced fit pathway. Φ is calculated from $\ln(k_{on,mut} / k_{on,WT}) / \ln((k_{on,mut} \times k_{off,WT}) / (k_{on,WT} \times k_{off,mut}))$.

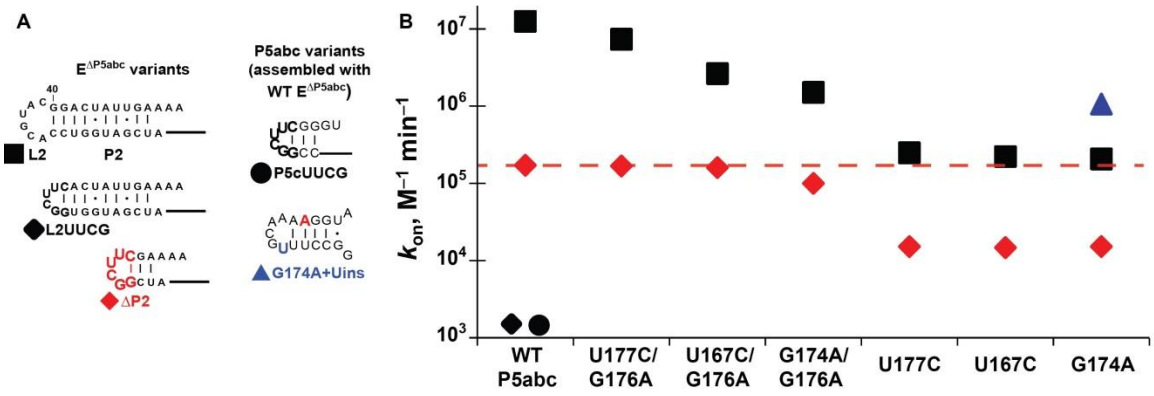

Figure 4.5: Mutating the ribozyme core to control the assembly pathway.

(*A*) Shown are mutants that target the P14 contact. For simplicity, only the relevant region is shown for $E^{\Delta P5abc}$ and P5abc variants. (*B*) Rate constants for binding of $E^{\Delta P5abc}$ variants with P5abc mutants. The dashed red line shows the association rate constant calculated for a TL/TLR tertiary contact from isolated measurements (see Chapter 3 Methods). Error bars from at least two measurements are not seen because they are smaller than the size of the marker.

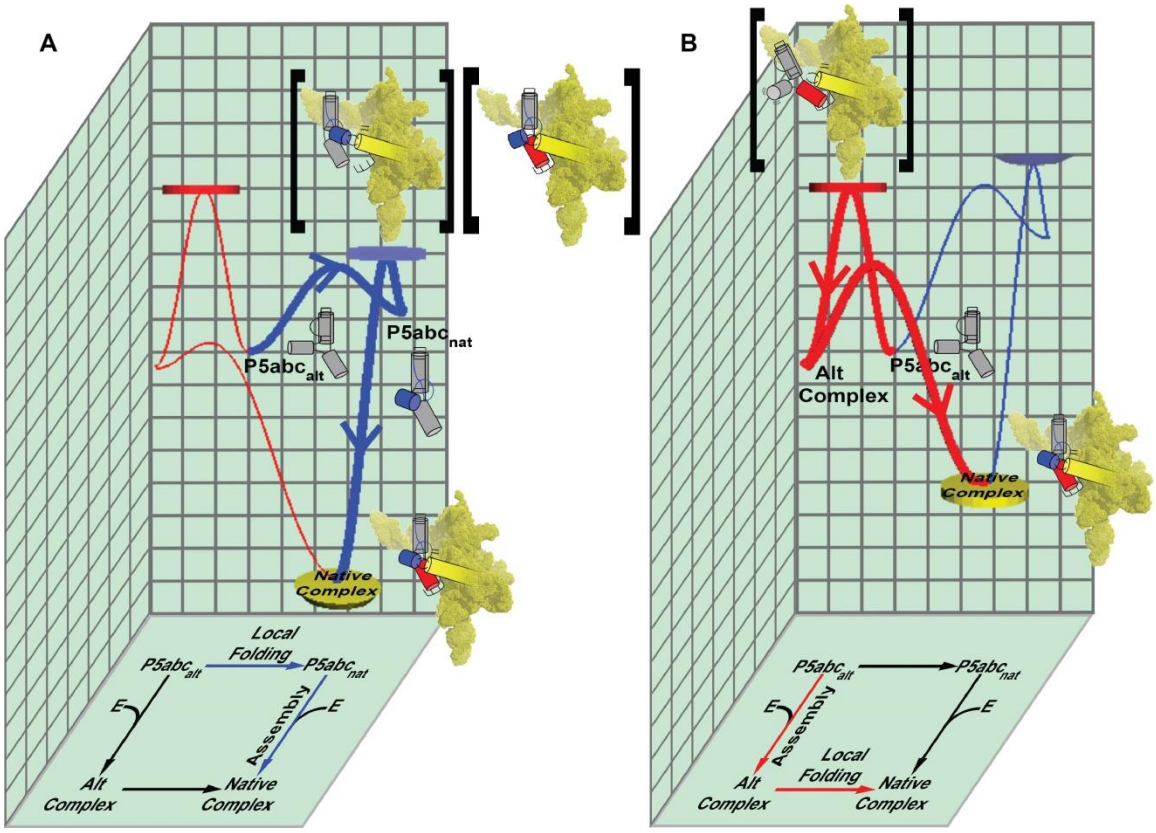

Figure 4.6: The local folding transition controls the assembly pathway by switching the order of tertiary contact formation.

Conformational selection (*A*) and induced fit (*B*) of P5abc with $E^{\Delta P5abc}$. The P2 helix is shown as a yellow cylinder, P5c and P14 formation are blue, and P5b and TL/TLR is shown in red.

$E^{\Delta P5abc}$ Variant **P5abc Mutant**	$E^{\Delta P5abc}$	U43C	G44A	A46C	A225U	A226U	G212C	G212U
wild-type	(1)	0.16	0.15	0.29	0.21	0.20	0.24	0.24
G176A	(1)	0.38	0.17	0.39	0.26	0.24	N.D.	N.D.
U167C	(1)	1.07	0.36	1.11	0.11	0.11	N.D.	N.D.
G174A	(1)	0.73	0.55	0.76	0.05	0.11	N.D.	N.D.
U177C	(1)	0.66	0.69	0.64	0.28	0.26	N.D.	N.D.
U167C/U177C	(1)	0.60	0.40	0.74	0.17	0.10	N.D.	N.D.
G174A/U177C	(1)	0.49	N.D.	0.96	0.18	0.11	N.D.	N.D.

Table 4.1: Relative association rate constants measured for P5abc mutants and $E^{\Delta P5abc}$ variants. N.D. is not determined.

$E^{\Delta P5abc}$ Variant **P5abc Mutant**	$E^{\Delta P5abc}$	U43C	G44A	A46C	A225U	A226U	G212C	G212U
wild-type	(1)	9.2	170	4.9	63	54	1200	930
G176A	(1)	12	67	6.2	22	36	N.D.	N.D.
U167C	(1)	34	130	10	10	13	260	330
G174A	(1)	13	11	4.2	5.4	4.5	N.D.	N.D.
U177C	(1)	67	100	5.3	53	44	N.D.	N.D.
U167C/U177C	(1)	7.9	10	4.9	2.2	3.1	N.D.	18
G174A/U177C	(1)	2.3	3.6	3.1	1.4	1.5	N.D.	N.D.

Table 4.2: Relative dissociation rate constants measured for P5abc mutants and $E^{\Delta P5abc}$ variants. N.D. is not determined.

Appendix A

STRUCTURAL CONNECTIONS IN THE 3WJ OF P5ABC

We tested a point mutation, G176A, which stabilizes the P5c native secondary structure to further affirm the connection between P5c switching and the MC (100, 101) (Figure A1.A). Surprisingly, we found that G176A does not stabilize formation of the MC (Figure A1.B). Further experiments in the background of the Nat+3 mutant showed that G176A destabilizes MC formation. These results indicated to us that the G176A substitution weakens tertiary structure formation in the 3WJ, coincidentally offsetting the G176A stabilization of native P5c. Two points arise from these results. First, there is an energetic connection between the 3WJ and native secondary structure in P5c. Second, care must be taken in designing and testing mutations to ensure that they do not have effects beyond the desired ones.

In regards to the energetic connectivity, there is also a structural basis for the connection between the 3WJ and P5c switch. G164, which forms a G-A sheared pair within the folded 3WJ, is liberated in the transition to native P5c, as are G176 and U177 (Figure 2.1A). We attempted to rescue structure formation by restoring the nucleotides present in the folded 3WJ (Figure A1.C). Insertion of only the G (a proxy for G164) or the GU (a proxy for G176 and U177) individually did not rescue folding relative to U167C (Figure A1.D). We found that both segments were required for proper structure formation in the 3WJ (Figure 2.10A, B, and Figure 2.9B). Taken together, these results suggest that nucleotides G164, G176, and U177 form tertiary contacts in the folded 3WJ. Investigation of the crystal structure shows non-covalent interactions between these nucleotides in the 3WJ and MC when folded (Figure A1.E). If these junction nucleotides were the only source of cooperativity, then the 3WJ restoration mutant would have formed structure in the MC and 3WJ with the same magnesium dependence as the Nat+3

mutants, which require less magnesium than wild-type P5abc (Figure 2.10B). The identity of the base pair proximal to the 3WJ may play a role, as this is a CG pair in the wild-type P5abc but a GC pair in the mutant used here. The absence of a protection at position U177 in the 3WJ restoration mutant compared to Nat+3 could explain the result if the G164-U177 junction interaction does not form for the 3WJ restoration mutant but does form in Nat+3 (Figure 2.9).

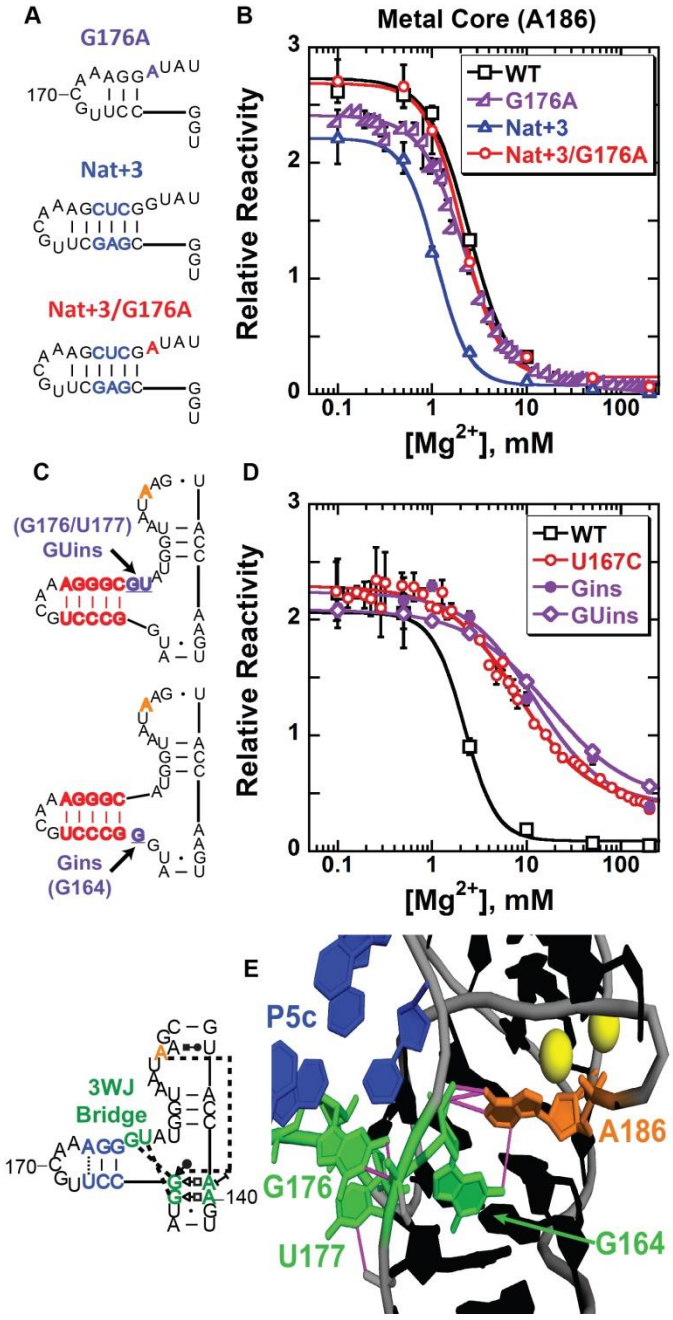

Figure A1: Three nucleotides in the junction are critical for P5abc folding.

Figure A1: continued.,

(*A* and *B*) Mutant G176A destabilizes tertiary structure. (*C* and *D*) GUins or Gins is not sufficient for proper MC and 3WJ formation. (*E*) Structural connectivity between the P5c (blue), 3WJ (green) and MC (orange). Hydrogen bonds are shown in magenta. Two site bound Mg^{2+} ions are shown in yellow (65).

Appendix B

MOLECULAR CROWDERS SLOW THE FOLDING OF A LARGE MISFOLDED RNA

Fundamental *in vitro* studies are typically carried out in dilute solutions mimicking physiological conditions. Recent studies have proposed that the cellular milieu is quite crowded, with estimates of up to 30% wt/vol cosolutes and macromolecules (167). How do biological macromolecules respond to a solution that is crowded? For RNA specifically, *in vitro* studies in the presence of molecular crowders (e.g. polyethylene glycol (PEG) or dextrans) demonstrated enhanced catalysis (168, 169). Additionally, molecular crowders enhance the stability of small structured RNAs and tertiary contacts, such as tRNA and TL/TLR, respectively (170, 171). Single molecule FRET measurements in presence of crowders were applied to folding of the isolated TL/TLR to determine the effects on the kinetics of folding. Molecular crowders predominantly enhance the folding rate without substantially effecting the unfolding rate (171). These results are initially attractive because one can propose that biochemistry works better in the crowded cellular environment, but for larger more complex structured RNAs within the cell we propose an alternative model.

For a large structured RNA that accumulates stable intermediates or forms kinetic traps, molecular crowders are expected to enhance the stability of these intermediate structures to the same extent as the native structure (172-174). If the transition state for folding between structures is unchanged, then RNA folding kinetics are predicted to be substantially slower in the crowded cellular environment. To test this idea, we conducted refolding measurements on the near-native misfolded structure of the *Tetrahymena thermophila* group I intron ribozyme (67) (Figure B1A). Misfolded ribozymes were generated by briefly incubating RNA at 10 mM Mg^{2+} for 10 min at 25 °C. To accelerate refolding to a measurable time-scale in the presence of crowders, the concentration of

Mg^{2+} was lowered to 5 mM, and the temperature was raised to 50 °C (Figure B1B). At various times, a fraction of the solution was diluted to 50 mM $MgCl_2$ to slow the refolding rate, essentially blocking any refolding of misfolded or native ribozymes (175, 176). Guanosine and labeled substrate (rSA_5*) were added to measure the fraction of native ribozyme by catalytic activity (175, 176).

We observed a burst of product formed in the first turnover of cleavage that steadily increased from about 0.1 to 0.8 with a half-life of 2 minutes (Figure B1C). Strikingly, when we supplemented the solution with dextran70, the refolding rate was decreased with concentration (Figure B1D). At 21% wt/vol dextran70, the refolding rate was slowed by 30-fold in agreement with our prediction that molecular crowders might slow refolding of near native kinetic traps. We also tested PEG200 and PEG8000 crowders and observed a similar slowing effect on the refolding rate from the misfolded state (Figure B2A).

Our results suggest that misfolded structures are stabilized by molecular crowders relative to the transition state for unfolding. Thus, molecular crowders can deepen the valleys of the RNA folding landscape (Figure B2B). As mentioned before, single molecular crowders increased the rate constant for folding, but had little effect on the rate constant for unfolding (171). Refolding of the misfolded ribozyme is thought to involve breaking of all tertiary contacts and global unfolding because mutations that knock out any one tertiary contact substantially accelerate the rate of refolding (67). We propose that the misfolded ribozyme structure in molecular crowders reforms tertiary contacts rapidly after spontaneous dissociation of contacts required for unfolding. This would increase the stability of the misfolded and native structures relative to the unfolded transition state, generating a substantial penalty for refolding misfolded structures in the crowded cellular environment. Molecular crowders may serve to increase the ruggedness

of the RNA folding landscape, necessitating the requirement of general RNA chaperone proteins *in vivo* to non-specifically unfold RNA structures (177-179).

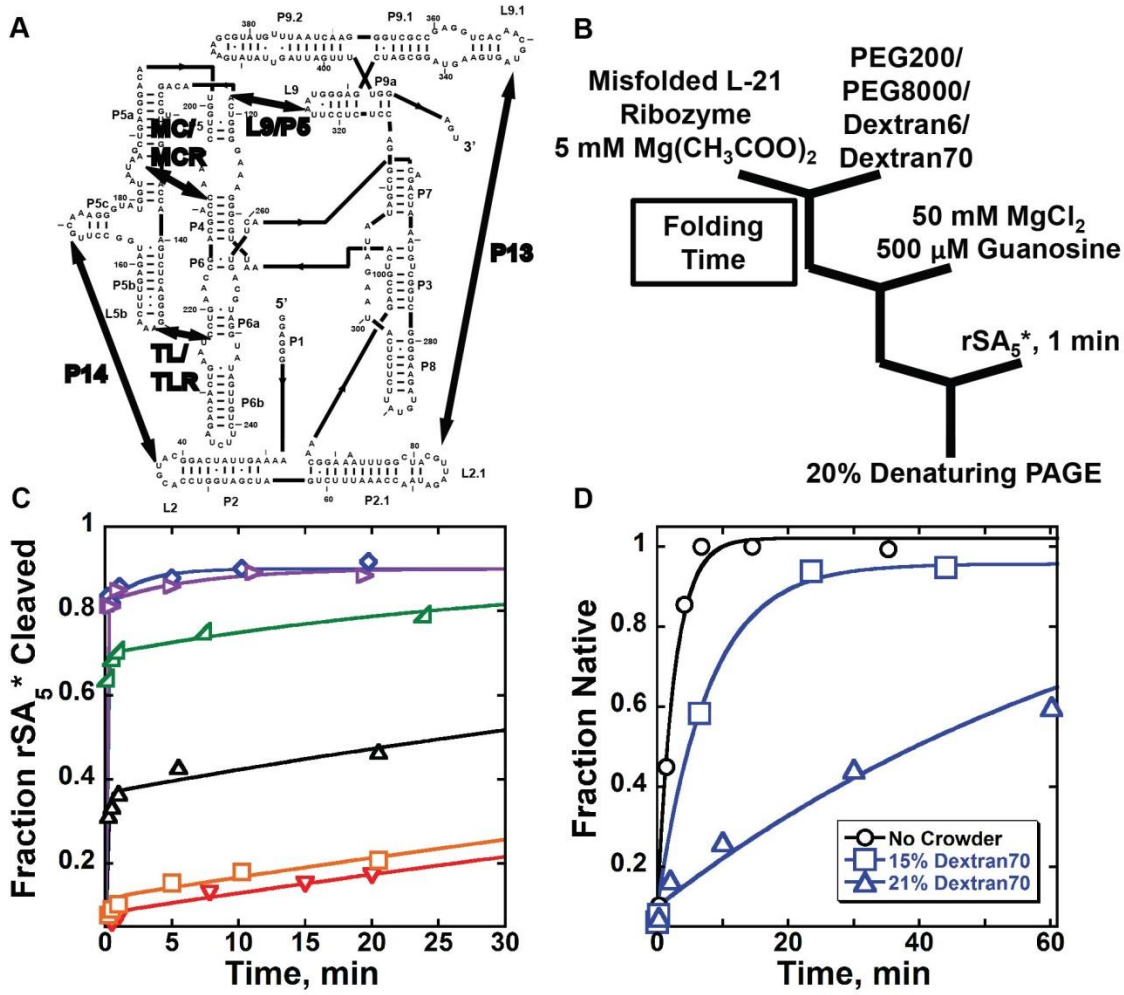

Figure B1: Molecular crowders slow the re-folding rate of the misfolded ribozyme.

(*A*) Secondary structure of the L-21 ribozyme from *Tetrahymena thermophila* used for folding studies herein. Long range tertiary contacts are indicated with arrows and labeled. (*B*) Catalytic assay for measuring the refolding rate from the misfolded ribozyme structure. (*C*) Cleavage time courses measured at different folding times (*B*), the burst of product formation observed reflects the fraction of ribozyme that has folded to the native state. (*D*) Folding time courses at 50 °C in the presence of the molecular crowder Dextran70 (167, 170).

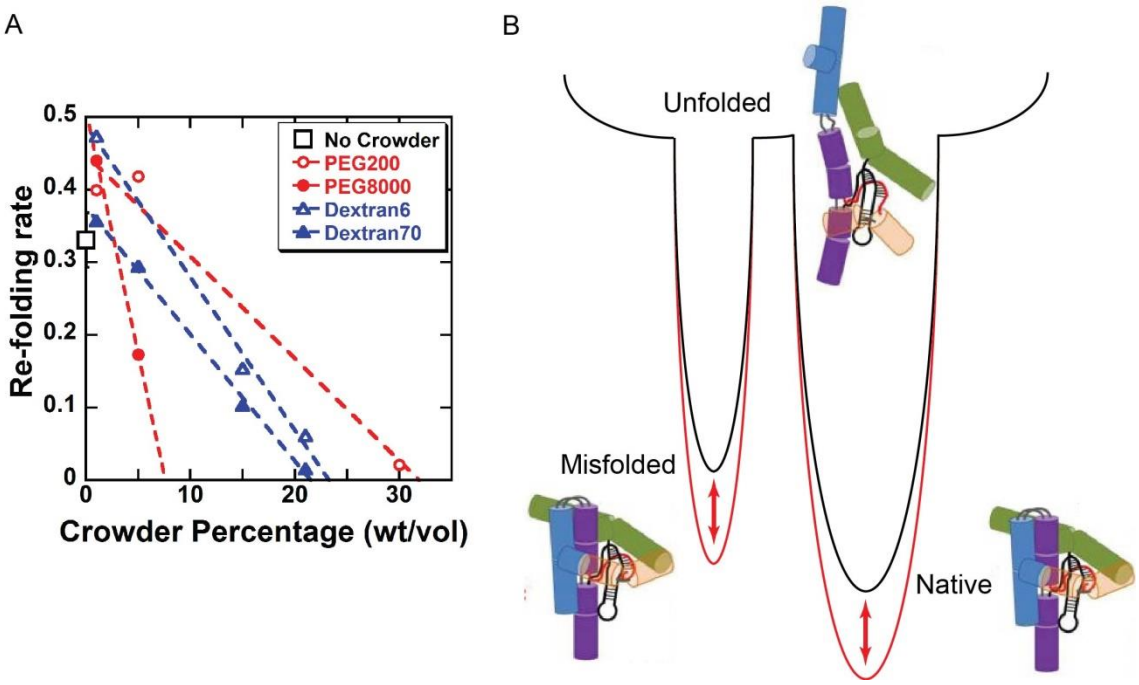

Figure B2: Molecular crowders deepen valleys of the RNA folding landscape exaggerating folding transition state barriers.

(*A*) The dependence on re-folding rate at different concentrations of crowding agents. (*B*) Free energy diagram illustrating the effect crowders have (red arrow) on the stability of the folded structures relative to the unfolded transition state ensemble.

References

1. Kruger K*, et al.* (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31(1):147-157.
2. Sigler PB (1975) An analysis of the structure of tRNA. *Annu Rev Biophys Bioeng* 4(00):477-527.
3. Brion P & Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113-137.
4. DEVOE H & TINOCO I (1962) The stability of helical polynucleotides: base contributions. *J Mol Biol* 4:500-517.
5. Turner DH (1996) Thermodynamics of base pairing. *Curr Opin Struct Biol* 6(3):299-304.
6. Tinoco I, Uhlenbeck OC, & Levine MD (1971) Estimation of secondary structure in ribonucleic acids. *Nature* 230(5293):362-367.
7. Freier SM*, et al.* (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci U S A* 83(24):9373-9377.
8. Turner DH & Mathews DH (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res* 38(Database issue):D280-282.
9. Wimberly BT*, et al.* (2000) Structure of the 30S ribosomal subunit. *Nature* 407(6802):327-339.
10. Ban N, Nissen P, Hansen J, Moore PB, & Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* 289(5481):905-920.
11. Gutell RR, Lee JC, & Cannone JJ (2002) The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol* 12(3):301-310.
12. Noller HF*, et al.* (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* 9(22):6167-6189.
13. Woese CR*, et al.* (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 8(10):2275-2293.
14. Wyatt JR, Puglisi JD, & Tinoco I (1989) RNA folding: pseudoknots, loops and bulges. *Bioessays* 11(4):100-106.
15. Beales LP, Rowlands DJ, & Holzenburg A (2001) The internal ribosome entry site (IRES) of hepatitis C virus visualized by electron microscopy. *RNA* 7(5):661-670.
16. Yamamoto H*, et al.* (2015) Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. *EMBO J* 34(24):3042-3058.
17. Dibrov SM*, et al.* (2014) Hepatitis C virus translation inhibitors targeting the internal ribosomal entry site. *J Med Chem* 57(5):1694-1707.
18. Lilley DM (2014) The K-turn motif in riboswitches and other RNA species. *Biochim Biophys Acta* 1839(10):995-1004.

19. Huang L & Lilley DM (2016) The Kink Turn, a Key Architectural Element in RNA Structure. *J Mol Biol* 428(5 Pt A):790-801.
20. Lilley DM (2012) The structure and folding of kink turns in RNA. *Wiley Interdiscip Rev RNA* 3(6):797-805.
21. Amano M (2003) The new loop-loop interactions between the peripheral domains and three-dimensional model of Tetrahymena group I intron. *Nucleic Acids Res Suppl* (3):173-174.
22. Salim N*, et al.* (2012) Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended duplex. *Biophys J* 102(5):1097-1107.
23. Jucker FM, Heus HA, Yip PF, Moors EH, & Pardi A (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* 264(5):968-980.
24. Pley HW, Flaherty KM, & McKay DB (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* 372(6501):68-74.
25. Fiore JL & Nesbitt DJ (2013) An RNA folding motif: GNRA tetraloop-receptor interactions. *Q Rev Biophys* 46(3):223-264.
26. Hodak JH, Downey CD, Fiore JL, Pardi A, & Nesbitt DJ (2005) Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc Natl Acad Sci U S A* 102(30):10505-10510.
27. Lescoute A & Westhof E (2006) The A-minor motifs in the decoding recognition process. *Biochimie* 88(8):993-999.
28. Bisaria N*, et al.* (2016) Kinetic and thermodynamic framework for P4-P6 RNA reveals tertiary motif modularity and modulation of the folding preferred pathway. *Proc Natl Acad Sci U S A* 113(34):E4956-4965.
29. Zhu JY & Meyer IM (2015) Four RNA families with functional transient structures. *RNA Biol* 12(1):5-20.
30. Serganov A & Nudler E (2013) A decade of riboswitches. *Cell* 152(1-2):17-24.
31. Winkler WC, Nahvi A, Sudarsan N, Barrick JE, & Breaker RR (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* 10(9):701-707.
32. Eschbach SH, St-Pierre P, Penedo JC, & Lafontaine DA (2012) Folding of the SAM-I riboswitch: a tale with a twist. *RNA Biol* 9(5):535-541.
33. Haller A, Rieder U, Aigner M, Blanchard SC, & Micura R (2011) Conformational capture of the SAM-II riboswitch. *Nat Chem Biol* 7(6):393-400.
34. Heppell B*, et al.* (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat Chem Biol* 7(6):384-392.
35. Wilson RC*, et al.* (2011) Tuning riboswitch regulation through conformational selection. *J Mol Biol* 405(4):926-938.
36. Nou X & Kadner RJ (1998) Coupled changes in translation and transcription during cobalamin-dependent regulation of btuB expression in Escherichia coli. *J Bacteriol* 180(24):6719-6728.
37. Nou X & Kadner RJ (2000) Adenosylcobalamin inhibits ribosome binding to btuB RNA. *Proc Natl Acad Sci U S A* 97(13):7190-7195.

38. Cheah MT, Wachter A, Sudarsan N, & Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447(7143):497-500.
39. Croft MT, Moulin M, Webb ME, & Smith AG (2007) Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci U S A* 104(52):20770-20775.
40. Tuerk C & Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249(4968):505-510.
41. Ellington AD & Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346(6287):818-822.
42. Stoltenburg R, Reinemann C, & Strehlitz B (2007) SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng* 24(4):381-403.
43. Famulok M, Mayer G, & Blind M (2000) Nucleic acid aptamers-from selection in vitro to applications in vivo. *Acc Chem Res* 33(9):591-599.
44. Paige JS, Wu KY, & Jaffrey SR (2011) RNA mimics of green fluorescent protein. *Science* 333(6042):642-646.
45. Huang H*, et al.* (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat Chem Biol* 10(8):686-691.
46. Warner KD*, et al.* (2014) Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat Struct Mol Biol* 21(8):658-663.
47. You M & Jaffrey SR (2015) Structure and Mechanism of RNA Mimics of Green Fluorescent Protein. *Annu Rev Biophys* 44:187-206.
48. Germer K, Leonard M, & Zhang X (2013) RNA aptamers and their therapeutic and diagnostic applications. *Int J Biochem Mol Biol* 4(1):27-40.
49. Burnett JC & Rossi JJ (2012) RNA-based therapeutics: current progress and future prospects. *Chem Biol* 19(1):60-71.
50. Kole R, Krainer AR, & Altman S (2012) RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat Rev Drug Discov* 11(2):125-140.
51. Sankaran N (2016) The RNA World at Thirty: A Look Back with its Author. *J Mol Evol* 83(5-6):169-175.
52. Fedorov A & Fedorova L (2004) Introns: mighty elements from the RNA world. *J Mol Evol* 59(5):718-721.
53. Saldanha R, Mohr G, Belfort M, & Lambowitz AM (1993) Group I and group II introns. *FASEB J* 7(1):15-24.
54. Vicens Q & Cech TR (2006) Atomic level architecture of group I introns revealed. *Trends Biochem Sci* 31(1):41-51.
55. Golden BL, Gooding AR, Podell ER, & Cech TR (1998) A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* 282(5387):259-264.
56. Paukstelis PJ, Chen JH, Chase E, Lambowitz AM, & Golden BL (2008) Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* 451(7174):94-97.

57. Golden BL, Kim H, & Chase E (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12(1):82-89.
58. van der Horst G, Christian A, & Inoue T (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc Natl Acad Sci U S A* 88(1):184-188.
59. Engelhardt MA, Doherty EA, Knitt DS, Doudna JA, & Herschlag D (2000) The P5abc peripheral element facilitates preorganization of the tetrahymena group I ribozyme for catalysis. *Biochemistry* 39(10):2639-2651.
60. Doudna JA & Cech TR (1995) Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* 1(1):36-45.
61. Shi X*, et al.* (2014) Roles of Long-Range Tertiary Interactions in Limiting Dynamics of the Tetrahymena Group I Ribozyme. *J Am Chem Soc*.
62. Benz-Moy TL & Herschlag D (2011) Structure-function analysis from the outside in: long-range tertiary contacts in RNA exhibit distinct catalytic roles. *Biochemistry* 50(40):8733-8755.
63. Lehnert V, Jaeger L, Michel F, & Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. *Chem Biol* 3(12):993-1009.
64. Mitra S, Laederach A, Golden BL, Altman RB, & Brenowitz M (2011) RNA molecules with conserved catalytic cores but variable peripheries fold along unique energetically optimized pathways. *RNA* 17(8):1589-1603.
65. Cate JH, Hanna RL, & Doudna JA (1997) A magnesium ion core at the heart of a ribozyme domain. *Nat Struct Biol* 4(7):553-558.
66. Cate JH*, et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273(5282):1678-1685.
67. Russell R*, et al.* (2006) The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* 363(2):531-544.
68. Lescoute A & Westhof E (2006) Topology of three-way junctions in folded RNAs. *RNA* 12(1):83-93.
69. Murphy FL & Cech TR (1993) An independently folding domain of RNA tertiary structure within the Tetrahymena ribozyme. *Biochemistry* 32(20):5291-5300.
70. Wu M & Tinoco I (1998) RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci U S A* 95(20):11555-11560.
71. Zheng M, Wu M, & Tinoco I (2001) Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the Tetrahymena group I ribozyme. *Proc Natl Acad Sci U S A* 98(7):3695-3700.
72. Silverman SK, Zheng M, Wu M, Tinoco I, & Cech TR (1999) Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA* 5(12):1665-1674.
73. Koculi E, Cho SS, Desai R, Thirumalai D, & Woodson SA (2012) Folding path of P5abc RNA involves direct coupling of secondary and tertiary structures. *Nucleic Acids Res* 40(16):8011-8020.

74. Doherty EA, Herschlag D, & Doudna JA (1999) Assembly of an exceptionally stable RNA tertiary interface in a group I ribozyme. *Biochemistry* 38(10):2982-2990.
75. Johnson TH, Tijerina P, Chadee AB, Herschlag D, & Russell R (2005) Structural specificity conferred by a group I RNA peripheral element. *Proc Natl Acad Sci U S A* 102(29):10176-10181.
76. Hampel KJ & Burke JM (2001) Time-resolved hydroxyl-radical footprinting of RNA using Fe(II)-EDTA. *Methods* 23(3):233-239.
77. Schlatterer JC & Brenowitz M (2009) Complementing global measures of RNA folding with local reports of backbone solvent accessibility by time resolved hydroxyl radical footprinting. *Methods* 49(2):142-147.
78. Russell R & Herschlag D (1999) New pathways in folding of the Tetrahymena group I RNA enzyme. *J Mol Biol* 291(5):1155-1167.
79. Stephenson JD, Kenyon JC, Symmons MF, & Lever AM (2016) Characterizing 3D RNA structure by single molecule FRET. *Methods* 103:57-67.
80. Sattin BD, Zhao W, Travers K, Chu S, & Herschlag D (2008) Direct measurement of tertiary contact cooperativity in RNA folding. *J Am Chem Soc* 130(19):6085-6087.
81. Greenfeld M & Herschlag D (2010) Measuring the energetic coupling of tertiary contacts in RNA folding using single molecule fluorescence resonance energy transfer. *Methods Enzymol* 472:205-220.
82. Biesiada M, Pachulska-Wieczorek K, Adamiak RW, & Purzycka KJ (2016) RNAComposer and RNA 3D structure prediction for nanotechnology. *Methods* 103:120-127.
83. Batey RT (2011) Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdiscip Rev RNA* 2(2):299-311.
84. Adams PL*, et al.* (2004) Crystal structure of a group I intron splicing intermediate. *RNA* 10(12):1867-1887.
85. Gruber AR, Bernhart SH, & Lorenz R (2015) The ViennaRNA web services. *Methods Mol Biol* 1269:307-326.
86. Qin PZ, Butcher SE, Feigon J, & Hubbell WL (2001) Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: a site-directed spin labeling study. *Biochemistry* 40(23):6929-6936.
87. Yanofsky C (1981) Attenuation in the control of expression of bacterial operons. *Nature* 289(5800):751-758.
88. Bhattacharyya D, Diamond P, & Basu S (2015) An Independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit. *Biochemistry* 54(10):1879-1885.
89. Hille F & Charpentier E (2016) CRISPR-Cas: biology, mechanisms and relevance. *Philos Trans R Soc Lond B Biol Sci* 371(1707).
90. Ng EW*, et al.* (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat Rev Drug Discov* 5(2):123-132.

91. Galej WP, Nguyen TH, Newman AJ, & Nagai K (2014) Structural studies of the spliceosome: zooming into the heart of the machine. *Curr Opin Struct Biol* 25:57-66.
92. Yusupova G & Yusupov M (2014) High-resolution structure of the eukaryotic 80S ribosome. *Annu Rev Biochem* 83:467-486.
93. Chen W & Moore MJ (2014) The spliceosome: disorder and dynamics defined. *Curr Opin Struct Biol* 24:141-149.
94. Parks JW & Stone MD (2017) Single-Molecule Studies of Telomeres and Telomerase. *Annu Rev Biophys*.
95. Petrov A, Chen J, O'Leary S, Tsai A, & Puglisi JD (2012) Single-molecule analysis of translational dynamics. *Cold Spring Harb Perspect Biol* 4(9):a011551.
96. Chen J, Petrov A, Tsai A, O'Leary SE, & Puglisi JD (2013) Coordinated conformational and compositional dynamics drive ribosome translocation. *Nat Struct Mol Biol* 20(6):718-727.
97. Winkler W, Nahvi A, & Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419(6910):952-956.
98. Mironov AS*, et al.* (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111(5):747-756.
99. Winkler WC, Cohen-Chalamish S, & Breaker RR (2002) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A* 99(25):15908-15913.
100. Xue Y, Gracia B, Herschlag D, Russell R, & Al-Hashimi HM (2016) Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. *Nat Commun* 7:ncomms11768.
101. Gracia B*, et al.* (2016) RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly. *J Mol Biol*.
102. Frederiksen JK, Li NS, Das R, Herschlag D, & Piccirilli JA (2012) Metal-ion rescue revisited: biochemical detection of site-bound metal ions important for RNA folding. *RNA* 18(6):1123-1141.
103. Pabit SA, Sutton JL, Chen H, & Pollack L (2013) Role of ion valence in the submillisecond collapse and folding of a small RNA domain. *Biochemistry* 52(9):1539-1546.
104. Correll CC, Beneken J, Plantinga MJ, Lubbers M, & Chan YL (2003) The common and the distinctive features of the bulged-G motif based on a 1.04 A resolution RNA structure. *Nucleic Acids Res* 31(23):6806-6818.
105. Travers KJ, Boyd N, & Herschlag D (2007) Low specificity of metal ion binding in the metal ion core of a folded RNA. *RNA* 13(8):1205-1213.
106. Das R, Travers KJ, Bai Y, & Herschlag D (2005) Determining the Mg2+ stoichiometry for folding an RNA metal ion core. *J Am Chem Soc* 127(23):8272-8273.
107. Russell R, Millett IS, Doniach S, & Herschlag D (2000) Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* 7(5):367-370.

108. Takamoto K, *et al.* (2004) Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J Mol Biol* 343(5):1195-1206.
109. Murphy FL & Cech TR (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J Mol Biol* 236(1):49-63.
110. Jolley EA & Znosko BM (2016) The loss of a hydrogen bond: Thermodynamic contributions of a non-standard nucleotide. *Nucleic Acids Res*.
111. Creighton TE (1995) Protein folding. An unfolding story. *Curr Biol* 5(4):353-356.
112. Creighton TE, Darby NJ, & Kemmink J (1996) The roles of partly folded intermediates in protein folding. *FASEB J* 10(1):110-118.
113. Davis JH, *et al.* (2016) Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell* 167(6):1610-1622.e1615.
114. Russell R & Herschlag D (1999) Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme. *RNA* 5(2):158-166.
115. Kladwang W, *et al.* (2014) Standardization of RNA chemical mapping experiments. *Biochemistry* 53(19):3063-3065.
116. Lee J, Vogt CE, McBrairty M, & Al-Hashimi HM (2013) Influence of dimethylsulfoxide on RNA structure and ligand binding. *Anal Chem* 85(20):9692-9698.
117. Yoon S, *et al.* (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* 27(13):1798-1805.
118. Ha SH & Ferrell JE (2016) Thresholds and ultrasensitivity from negative cooperativity. *Science* 352(6288):990-993.
119. Leontis NB & Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7(4):499-512.
120. Wang J, Daldrop P, Huang L, & Lilley DM (2014) The k-junction motif in RNA structure. *Nucleic Acids Res* 42(8):5322-5331.
121. Miao Z, *et al.* (2015) RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* 21(6):1066-1084.
122. Garst AD & Batey RT (2009) A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta* 1789(9-10):584-591.
123. Savinov A, Perez CF, & Block SM (2014) Single-molecule studies of riboswitch folding. *Biochim Biophys Acta* 1839(10):1030-1045.
124. McCann KL & Baserga SJ (2013) Genetics. Mysterious ribosomopathies. *Science* 341(6148):849-850.
125. Scotti MM & Swanson MS (2016) RNA mis-splicing in disease. *Nat Rev Genet* 17(1):19-32.
126. Tinoco I & Bustamante C (1999) How RNA folds. *J Mol Biol* 293(2):271-281.
127. Pörschke D & Eigen M (1971) Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic--oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J Mol Biol* 62(2):361-381.

128. Pörschke D (1977) Elementary steps of base recognition and helix-coil transitions in nucleic acids. *Mol Biol Biochem Biophys* 24:191-218.
129. Knitt DS, Narlikar GJ, & Herschlag D (1994) Dissection of the role of the conserved G.U pair in group I RNA self-splicing. *Biochemistry* 33(46):13864-13879.
130. Green R & Szostak JW (1994) In vitro genetic analysis of the hinge region between helical elements P5-P4-P6 and P7-P3-P8 in the sunY group I self-splicing intron. *J Mol Biol* 235(1):140-155.
131. Breaker RR (2008) Complex riboswitches. *Science* 319(5871):1795-1797.
132. Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, & Block SM (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science* 319(5863):630-633.
133. Neupane K, Yu H, Foster DA, Wang F, & Woodside MT (2011) Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res* 39(17):7677-7687.
134. Herschlag D, Allred BE, & Gowrishankar S (2015) From static to dynamic: the need for structural ensembles and a predictive model of RNA folding and function. *Curr Opin Struct Biol* 30:125-133.
135. Agalarov SC, Sridhar Prasad G, Funke PM, Stout CD, & Williamson JR (2000) Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* 288(5463):107-113.
136. Nikulin A*, et al.* (2000) Crystal structure of the S15-rRNA complex. *Nat Struct Biol* 7(4):273-277.
137. Sekhar A & Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci U S A* 110(32):12867-12874.
138. Lee J, Dethoff EA, & Al-Hashimi HM (2014) Invisible RNA state dynamically couples distant motifs. *Proc Natl Acad Sci U S A* 111(26):9485-9490.
139. Lee SW, Zhao L, Pardi A, & Xia T (2010) Ultrafast dynamics show that the theophylline and 3-methylxanthine aptamers employ a conformational capture mechanism for binding their ligands. *Biochemistry* 49(13):2943-2951.
140. Duchardt-Ferner E*, et al.* (2010) Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angew Chem Int Ed Engl* 49(35):6216-6219.
141. Wunnicke D*, et al.* (2011) Ligand-induced conformational capture of a synthetic tetracycline riboswitch revealed by pulse EPR. *RNA* 17(1):182-188.
142. Xia T, Yuan J, & Fang X (2013) Conformational dynamics of an ATP-binding DNA aptamer: a single-molecule study. *J Phys Chem B* 117(48):14994-15003.
143. Suddala KC, Wang J, Hou Q, & Walter NG (2015) Mg(2+) shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. *J Am Chem Soc* 137(44):14075-14083.

144. Sclavi B, Sullivan M, Chance MR, Brenowitz M, & Woodson SA (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279(5358):1940-1943.
145. Deras ML, Brenowitz M, Ralston CY, Chance MR, & Woodson SA (2000) Folding mechanism of the Tetrahymena ribozyme P4-P6 domain. *Biochemistry* 39(36):10975-10985.
146. Guo F, Gooding AR, & Cech TR (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16(3):351-362.
147. Ducongé F, Di Primo C, & Toulme JJ (2000) Is a closing "GA pair" a rule for stable loop-loop RNA complexes? *J Biol Chem* 275(28):21287-21294.
148. Rist M & Marino J (2001) Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. *Nucleic Acids Res* 29(11):2401-2408.
149. Fiore JL, Holmstrom ED, Fiegland LR, Hodak JH, & Nesbitt DJ (2012) The role of counterion valence and size in GAAA tetraloop-receptor docking/undocking kinetics. *J Mol Biol* 423(2):198-216.
150. Tijerina P, Mohr S, & Russell R (2007) DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc* 2(10):2608-2623.
151. Cordero P, Kladwang W, VanLang CC, & Das R (2014) The mutate-and-map protocol for inferring base pairs in structured RNA. *Methods Mol Biol* 1086:53-77.
152. Narlikar GJ & Herschlag D (1996) Isolation of a local tertiary folding transition in the context of a globally folded RNA. *Nat Struct Biol* 3(8):701-710.
153. Huang J*, et al.* (2014) Structural basis for protein-RNA recognition in telomerase. *Nat Struct Mol Biol* 21(6):507-512.
154. Montemayor EJ*, et al.* (2014) Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. *Nat Struct Mol Biol* 21(6):544-551.
155. Cash DD & Feigon J (2017) Structure and folding of the Tetrahymena telomerase RNA pseudoknot. *Nucleic Acids Res* 45(1):482-495.
156. Chan H, Wang Y, & Feigon J (2017) Progress in Human and Tetrahymena Telomerase Structure. *Annu Rev Biophys*.
157. Ma H*, et al.* (2006) Exploring the energy landscape of a small RNA hairpin. *J Am Chem Soc* 128(5):1523-1530.
158. Chen SJ (2008) RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* 37:197-214.
159. Zhang W & Chen SJ (2002) RNA hairpin-folding kinetics. *Proc Natl Acad Sci U S A* 99(4):1931-1936.
160. Radhakrishnan A & Green R (2016) Connections Underlying Translation and mRNA Stability. *J Mol Biol* 428(18):3558-3564.
161. Nelson JW*, et al.* (2013) Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat Chem Biol* 9(12):834-839.
162. Lutz B, Faber M, Verma A, Klumpp S, & Schug A (2014) Differences between cotranscriptional and free riboswitch folding. *Nucleic Acids Res* 42(4):2687-2696.

163. Bisaria N & Herschlag D (2015) Probing the kinetic and thermodynamic consequences of the tetraloop/tetraloop receptor monovalent ion-binding site in P4-P6 RNA by smFRET. *Biochem Soc Trans* 43(2):172-178.
164. Raleigh DP & Plaxco KW (2005) The protein folding transition state: what are Phi-values really telling us? *Protein Pept Lett* 12(2):117-122.
165. Weikl TR & Dill KA (2007) Transition-states in protein folding kinetics: the structural interpretation of Phi values. *J Mol Biol* 365(5):1578-1586.
166. Kim J & Shin JS (2010) Probing the transition state for nucleic acid hybridization using phi-value analysis. *Biochemistry* 49(16):3420-3426.
167. Foffi G, Pastore A, Piazza F, & Temussi PA (2013) Macromolecular crowding: chemistry and physics meet biology (Ascona, Switzerland, 10-14 June 2012). *Phys Biol* 10(4):040301.
168. Strulson CA, Yennawar NH, Rambo RP, & Bevilacqua PC (2013) Molecular Crowding Favors Reactivity of a Human Ribozyme Under Physiological Ionic Conditions. *Biochemistry*.
169. Desai R, Kilburn D, Lee HT, & Woodson SA (2014) Increased ribozyme activity in crowded solutions. *J Biol Chem* 289(5):2972-2977.
170. Strulson CA, Boyer JA, Whitman EE, & Bevilacqua PC (2014) Molecular crowders and cosolutes promote folding cooperativity of RNA under physiological ionic conditions. *RNA* 20(3):331-347.
171. Dupuis NF, Holmstrom ED, & Nesbitt DJ (2014) Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics. *Proc Natl Acad Sci U S A* 111(23):8464-8469.
172. Treiber DK & Williamson JR (2001) Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* 11(3):309-314.
173. Treiber DK & Williamson JR (1999) Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* 9(3):339-345.
174. Russell R (2008) RNA misfolding and the action of chaperones. *Front Biosci* 13:1-20.
175. Wan Y, Mitchell D, & Russell R (2009) Catalytic activity as a probe of native RNA folding. *Methods Enzymol* 468:195-218.
176. Gracia B & Russell R (2014) RNA catalytic activity as a probe of chaperone-mediated RNA folding. *Methods Mol Biol* 1086:225-237.
177. Bhaskaran H & Russell R (2007) Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* 449(7165):1014-1018.
178. Jarmoskaite I & Russell R (2014) RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem* 83:697-725.
179. Akabayov SR, Akabayov B, Richardson CC, & Wagner G (2013) Molecular crowding enhanced ATPase activity of the RNA helicase eIF4A correlates with compaction of its quaternary structure and association with eIF4G. *J Am Chem Soc* 135(27):10040-10047.