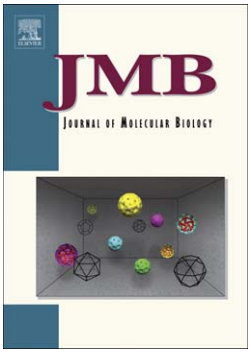


RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly

Brant Gracia, Yi Xue, Namita Bisaria, Daniel Herschlag, Hashim M. Al-Hashimi, Rick Russell

PII: S0022-2836(16)30258-3
DOI: doi: 10.1016/j.jmb.2016.07.013
Reference: YJMBI 65151

To appear in: *Journal of Molecular Biology*

Received date: 24 May 2016
Revised date: 12 July 2016
Accepted date: 14 July 2016

Please cite this article as: Gracia, B., Xue, Y., Bisaria, N., Herschlag, D., Al-Hashimi, H.M. & Russell, R., RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly, *Journal of Molecular Biology* (2016), doi: 10.1016/j.jmb.2016.07.013

RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly

Brant Gracia[1], Yi Xue[2], Namita Bisaria[3], Daniel Herschlag[3], Hashim M. Al-Hashimi[2], and Rick Russell[1,*]

[1]Department of Molecular Biosciences and the Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712

[2]Department of Biochemistry and Chemistry, Duke University Medical Center**,** Durham, NC 27710

[3]Department of Biochemistry, Stanford University, Stanford, CA 94305

*Correspondence and requests for materials should be addressed to R.R. (email: rick_russell@cm.utexas.edu; tel: 512-471-1514).

Abstract

Structured RNAs fold through multiple pathways, but we have little understanding of the molecular features that dictate folding pathways and determine rates along a given pathway. Here we asked whether folding of a complex RNA can be understood from its structural modules. In a two-piece version of the *Tetrahymena* group I ribozyme, the separated P5abc subdomain folds to local native secondary and tertiary structure in a linked transition and assembles with the ribozyme core via three tertiary contacts: a kissing loop (P14), a metal core-receptor interaction (MC/MCR), and a tetraloop-receptor interaction (TL/TLR), the first two of which are expected to depend on native P5abc structure from the local transition. Native gel, NMR, and chemical footprinting experiments showed that mutations that destabilize the native P5abc structure slowed assembly up to 100-fold, indicating that P5abc folds first and then assembles with the core by conformational selection. However, rate decreases beyond 100-fold were not observed because an alternative pathway becomes dominant, with non-native P5abc binding the core and then undergoing an induced fit rearrangement. P14 is formed in the rate-limiting step along the conformational selection pathway but after the rate-limiting step along the induced fit pathway. Strikingly, the assembly rate along the conformational selection pathway resembles that for an isolated kissing loop similar to P14, and the rate along the induced fit pathway resembles that for an isolated TL/TLR. Our results indicate substantial modularity in RNA folding and assembly and suggest that these processes can be understood in terms of underlying structural modules.

Keywords: RNA folding, biomolecular interactions, group I intron, P5abc, three-helix junction

Abbreviations: DMS, dimethyl sulfate; $E^{\Delta P5abc}$, P5abc-deleted *Tetrahymena* group I intron ribozyme; MC/MCR, metal core/metal core receptor tertiary contact; TL/TLR, tetraloop/tetraloop receptor tertiary contact

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 [1], 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 [2, 3]. Further, improper ribosome assembly and incorrect formation of splice junctions are linked to diseases, highlighting the essential nature of accurate and efficient RNA folding [4, 5].

A feature of RNA that has greatly aided understanding is the largely hierarchical nature of its folding [6]. 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 [7, 8]. 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 changes in secondary structure occur in tandem with the formation of tertiary structure, including around the catalytic sites of group I introns [9, 10] and in riboswitches as a

consequence of ligand binding [3, 11-13]. 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 [14]. 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 [15-19], 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 [20, 21]– that is then trapped by intra- or inter-molecular interactions in a process of conformational selection [22-28]. 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 [29-32]. 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 ($E^{\Delta P5abc}$) [33, 34]. 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 (Fig. 1(a)). 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 (Fig. 1(a), (b)) [31, 32, 35]. 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) (Fig. 1(c)) [33, 36, 37]. 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 (Fig. 1(a), (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 [15, 16, 38]. 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, [39].

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 (Fig. 2(a)). 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 (Fig. 1(a) and Fig. S1): 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 [35], 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 (Fig. 2(b)). Each P5abc mutant bound $E^{\Delta P5abc}$ with a pseudo-first-order rate constant that was linearly dependent on $E^{\Delta P5abc}$ concentration (Fig. 2(c), (d) and Fig. S2). The wild-type P5abc assembled with a second-order rate constant of 1.1 (± 0.1) × 10^7 M^{-1} min^{-1} (Fig. 2(d) and Table 1), as measured previously [40, 41]. 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 (Fig. 2(e)).

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 (Fig. 1(a) and Fig. S1). 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+}; [31] 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 (Fig. S3).

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 [31, 32]. 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 [31]. 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 [31]. 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+} [31]. 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 [31].

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 (Fig. 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 [31] (Fig. 3(a)). 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 (Fig. 3(b)). At higher Mg^{2+} concentrations substantial peak broadening and losses of signal were observed, as noted previously [31], 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 (Fig. 3 and Fig. S4). 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 (Fig. 3(b)).

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 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 S1). This increase was accompanied by a modest increase in the relative rate constant of assembly for this mutant (Fig. S5(a)), providing a further link between the equilibrium for the secondary structure transition within P5abc and the rate constant for assembly with the intron core.

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 [42]. 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 [35]. 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 S2). 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 (Fig. S5(b)). Further work will be necessary to test whether this deviation reflects an idiosyncratic effect of the G176A mutation in these backgrounds and more generally to perform quantitative tests of the two-step assembly model.

Mutations can 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 1 and Fig. S1). 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 $\times$ 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 (see Fig. 1(a) and Fig. S1). 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 1), these mutants assembled with rate constants within 2-fold of the values for the single mutants (Fig. 2(e) and Table 1). DMS footprinting results indicated that the second mutation stabilized the alternative secondary structure, as expected (Fig. S6). 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 Supporting Information, *Ribozyme catalytic activity measurements* and Table S3). 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 [43]. 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 (Fig. S7 and Fig. S8), 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, Fig. 4(a)). 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, Fig. 4(a)). 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; Fig. 4(a)). This mutation decreased the assembly rate constant of the wild-type P5abc by 6-fold (Fig. 4(b), 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 (Fig. 4(b) and Table 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 (Fig. 4(a)). 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 Discussion).

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.

Fig. 5(a) depicts the assembly pathway that dominates flux for 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). 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 (Fig. 5(a), 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 [44, 45]. 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.

In contrast, U167C and other P5abc mutants that strongly stabilize the alternative conformation (G174A, U177C and combinations of these mutations) populate a distinct pathway with an 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 Fig. 5(b)). 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 (see Fig. S7 and S8). 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 (see Supporting Information, *Evidence for two pathways*). This new pathway is an induced fit pathway.

In the induced fit pathway, the initial binding step occurs with P5c in the alternative structure and unavailable to form native P14 (see Fig. 4(a)). 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 (as proposed in the cartoon in Fig. 5(b)). 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}) [46-48], consistent with this model (see Supporting Information, *Estimation of the rate constant for TL/TLR formation*). Further work will be necessary to test this hypothesis.

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 (Fig. 5(c)). 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.

At a physical level, the presence of alternative pathways highlights the modularity in RNA folding. While the conformational selection pathway relies on formation of native structure in P5abc and kissing loop P14 formation to nucleate assembly, the induced fit pathway depends on a distinct tertiary module, most likely the TL/TLR contact, which is distant from P5c and not expected to be impacted by the local folding transition of P5abc. 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} [47, 48]. 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 (Fig. 5(d)). This interpretation is supported by recent results indicating that the secondary structure in P5c fluctuates rapidly [49]. Alternatively or in addition, it is possible that the MC/MCR interaction forms while P5c remains in the alternative conformation, which would further highlight the modularity of RNA structure formation.

Implications

Mutations that change the equilibrium for local native structure tune the folding and assembly kinetics, and these observations provide strong evidence for a conformational selection pathway that depends on formation of the native structure. However, when the energetic penalty for native structure formation is sufficiently large, the conformational selection pathway becomes slower than an alternative induced fit pathway, which does not depend on prior formation of the same native structure, so that the induced fit pathway becomes dominant.

RNA folding and assembly reactions have often been described as proceeding through conformational selection or induced fit [22-28, 50-57]. In contrast, we find evidence for *both* pathways, and reveal that the relative flux shifts drastically in response to small sequence changes. Shifts in flux between conformational selection and induced fit pathways have been shown in protein folding, in that case upon changes in concentration of a ligand [58]. Thus, the two ‘classes’ of assembly, rather than being an important or defining characteristic of a given RNA, may be more accurately viewed

as alternative routes whose relative flux can be modulated –and ultimately controlled– in response to small changes to the molecules or even to the experimental conditions.

It is striking that the folding behavior changes drastically with sequence changes as small as single point mutations. This property presumably reflects the strong stability of RNA base pairs and endows structured RNA with an ability to transform folding properties in response to small changes in sequence or conditions. Nature has likely harnessed this ability in generating alternative RNA structures under kinetic control, such as riboswitches that change the dominant folding pathway depending on whether they associate with a regulatory ligand during folding. Intriguingly, the large response to single nucleotide changes is expected to make RNA folding pathways highly evolvable and thus more prone to alteration. The ubiquitous presence of alternative folding and assembly pathways is a feature that will also need to be carefully considered in the design of novel structured and functional RNAs. Most generally, our results suggest that 'deconstruction and reconstitution' is a viable strategy for understanding and ultimately engineering complex RNA structures and conformational transitions.

Materials and Methods

RNA Preparation. RNAs were prepared by *in vitro* transcription essentially as described [59]. 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 [42]. $E^{\Delta P5abc}$ 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.

NMR Sample Preparation. 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.

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 [41]. Bands were quantitated by using a Phosphorimager (GE Healthcare).

NMR measurements. Spectra of the labeled NMR samples were collected by imino SOFAST-HMQC pulse program [60] running in 2D or 1D mode. The spectrum of the unlabeled wild-type tP5abc was collected by a variant of imino SOFAST-HMQC experiment where the ^{1}H magnetization is not transferred to and back from ^{15}N. All of the NMR measurements were conducted on an Agilent DD2 800 spectrometer at 10 °C. Mg^{2+} was added to NMR samples through buffer-exchange using 3-kDa-cutoff centrifugal concentrators (Millipore Corp.). The 2D imino HMQC spectra were processed by NMRPipe [61], and peaks were fitted to Lorentzian lineshape using autofit script included in NMRPipe. The equilibrium constant was calculated according to $K_{eq} = V_{U142*}/V_{U142}$, where V_{U142} and V_{U142*} are the volumes of the U142 peak and the U142* peak, respectively.

DMS footprinting. Reactions were performed using a method developed previously [62] with extensive modifications to increase throughput [63]. 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 [64]. Reactivity values were normalized to the average of the two adenines in the 3' flanking hairpin GAGUA stem-loop.

Acknowledgments

We thank Wipapat Ann Kladwang and Rhiju Das for assistance with high-throughput footprinting, the DNA Sequencing facility in the Institute for Cellular and Molecular Biology at the University of Texas for assistance with capillary electrophoresis, the Duke

Magnetic Resonance Spectroscopy Center for technical support and resources, and Sarah Woodson for the coordinate file of the unfolded P5abc. This work was supported by National Institutes of General Medical Sciences grant P01 GM066275.

References

[1] Miao Z., Adamiak R.W., Blanchet M.F., Boniecki M., Bujnicki J.M., Chen S.J., et al. (2015). RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. RNA 21, 1066-84.

[2] Garst A.D., Batey R.T. (2009). A switch in time: detailing the life of a riboswitch. Biochim Biophys Acta 1789, 584-91.

[3] Savinov A., Perez C.F., Block S.M. (2014). Single-molecule studies of riboswitch folding. Biochim Biophys Acta 1839, 1030-45.

[4] McCann K.L., Baserga S.J. (2013). Genetics. Mysterious ribosomopathies. Science 341, 849-50.

[5] Scotti M.M., Swanson M.S. (2016). RNA mis-splicing in disease. Nat Rev Genet 17, 19-32.

[6] Tinoco I., Jr., Bustamante C. (1999). How RNA folds. J Mol Biol 293, 271-81.

[7] Porschke 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, 361-81.

[8] Porschke D. (1977). Elementary steps of base recognition and helix-coil transitions in nucleic acids. Mol Biol Biochem Biophys 24, 191-218.

[9] Knitt D.S., Narlikar G.J., Herschlag D. (1994). Dissection of the role of the conserved G.U pair in group I RNA self-splicing. Biochemistry 33, 13864-79.

[10] Green R., Szostak J.W. (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, 140-55.

[11] Breaker R.R. (2008). Complex riboswitches. Science 319, 1795-7.

[12] Greenleaf W.J., Frieda K.L., Foster D.A., Woodside M.T., Block S.M. (2008). Direct observation of hierarchical folding in single riboswitch aptamers. Science 319, 630-3.

[13] Neupane K., Yu H., Foster D.A., Wang F., Woodside M.T. (2011). Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. Nucleic Acids Res 39, 7677-87.

[14] Herschlag D., Allred B.E., 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, 125-33.

[15] Cate J.H., Gooding A.R., Podell E., Zhou K., Golden B.L., Kundrot C.E., et al. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. Science 273, 1678-85.

[16] Golden B.L., Gooding A.R., Podell E.R., Cech T.R. (1998). A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. Science 282, 259-64.

[17] Agalarov S.C., Sridhar Prasad G., Funke P.M., Stout C.D., Williamson J.R. (2000). Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain. Science 288, 107-13.

[18] Nikulin A., Serganov A., Ennifar E., Tishchenko S., Nevskaya N., Shepard W., et al. (2000). Crystal structure of the S15-rRNA complex. Nat Struct Biol 7, 273-7.

[19] Wimberly B.T., Brodersen D.E., Clemons W.M., Jr., Morgan-Warren R.J., Carter A.P., Vonrhein C., et al. (2000). Structure of the 30S ribosomal subunit. Nature 407, 327-39.

[20] Sekhar A., Kay L.E. (2013). NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. Proc Natl Acad Sci U S A 110, 12867-74.

[21] Lee J., Dethoff E.A., Al-Hashimi H.M. (2014). Invisible RNA state dynamically couples distant motifs. Proc Natl Acad Sci U S A 111, 9485-90.

[22] Lee S.W., 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, 2943-51.

[23] Duchardt-Ferner E., Weigand J.E., Ohlenschlager O., Schmidtke S.R., Suess B., Wohnert J. (2010). Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. Angew Chem Int Ed Engl 49, 6216-9.

[24] Wunnicke D., Strohbach D., Weigand J.E., Appel B., Feresin E., Suess B., et al. (2011). Ligand-induced conformational capture of a synthetic tetracycline riboswitch revealed by pulse EPR. RNA 17, 182-8.

[25] Haller A., Rieder U., Aigner M., Blanchard S.C., Micura R. (2011). Conformational capture of the SAM-II riboswitch. Nat Chem Biol 7, 393-400.

[26] Wilson R.C., Smith A.M., Fuchs R.T., Kleckner I.R., Henkin T.M., Foster M.P. (2011). Tuning riboswitch regulation through conformational selection. J Mol Biol 405, 926-38.

[27] Xia T., Yuan J., Fang X. (2013). Conformational dynamics of an ATP-binding DNA aptamer: a single-molecule study. J Phys Chem B 117, 14994-5003.

[28] Suddala K.C., Wang J., Hou Q., Walter N.G. (2015). Mg(2+) Shifts Ligand-Mediated Folding of a Riboswitch from Induced-Fit to Conformational Selection. J Am Chem Soc 137, 14075-83.

[29] Sclavi B., Sullivan M., Chance M.R., Brenowitz M., Woodson S.A. (1998). RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. Science 279, 1940-3.

[30] Deras M.L., Brenowitz M., Ralston C.Y., Chance M.R., Woodson S.A. (2000). Folding mechanism of the *Tetrahymena* ribozyme P4-P6 domain. Biochemistry 39, 10975-85.

[31] Wu M., Tinoco I., Jr. (1998). RNA folding causes secondary structure rearrangement. Proc Natl Acad Sci U S A 95, 11555-60.

[32] Zheng M., Wu M., Tinoco I., Jr. (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, 3695-700.

[33] Doherty E.A., Herschlag D., Doudna J.A. (1999). Assembly of an exceptionally stable RNA tertiary interface in a group I ribozyme. Biochemistry 38, 2982-90.

[34] 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, 184-8.

[35] Silverman S.K., Zheng M., Wu M., Tinoco I., Jr., Cech T.R. (1999). Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. RNA 5, 1665-74.

[36] Murphy F.L., Cech T.R. (1993). An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. Biochemistry 32, 5291-300.

[37] 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, 993-1009.

[38] Guo F., Gooding A.R., Cech T.R. (2004). Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. Mol Cell 16, 351-62.

[39] Engelhardt M.A., Doherty E.A., Knitt D.S., Doudna J.A., Herschlag D. (2000). The P5abc peripheral element facilitates preorganization of the *Tetrahymena* group I ribozyme for catalysis. Biochemistry 39, 2639-51.

[40] Russell R., Herschlag D. (1999). New pathways in folding of the *Tetrahymena* group I RNA enzyme. J Mol Biol 291, 1155-67.

[41] Johnson T.H., Tijerina P., Chadee A.B., Herschlag D., Russell R. (2005). Structural specificity conferred by a group I RNA peripheral element. Proc Natl Acad Sci U S A 102, 10176-81.

[42] Kladwang W., Mann T.H., Becka A., Tian S., Kim H., Yoon S., et al. (2014). Standardization of RNA chemical mapping experiments. Biochemistry 53, 3063-5.

[43] Benz-Moy T.L., Herschlag D. (2011). Structure-function analysis from the outside in: long-range tertiary contacts in RNA exhibit distinct catalytic roles. Biochemistry 50, 8733-55.

[44] Duconge F., Di Primo C., Toulme J.J. (2000). Is a closing "GA pair" a rule for stable loop-loop RNA complexes? J Biol Chem 275, 21287-94.

[45] Rist M., Marino J. (2001). Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. Nucleic Acids Res 29, 2401-8.

[46] Qin P.Z., Butcher S.E., Feigon J., Hubbell W.L. (2001). Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: a site-directed spin labeling study. Biochemistry 40, 6929-36.

[47] Hodak J.H., Downey C.D., Fiore J.L., Pardi A., Nesbitt D.J. (2005). Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. Proc Natl Acad Sci U S A 102, 10505-10.

[48] Fiore J.L., Holmstrom E.D., Fiegland L.R., Hodak J.H., Nesbitt D.J. (2012). The role of counterion valence and size in GAAA tetraloop-receptor docking/undocking kinetics. J Mol Biol 423, 198-216.

[49] Xue Y., Gracia B., Herschlag D., Russell R., Al-Hashimi H.M. (2016). Visualizing formation of an RNA folding intermediate through a fast highly modular secondary structure switch. Nat Comm, Jun 13;7 doi: 10.1038/ncomms11768

[50] Webb A.E., Rose M.A., Westhof E., Weeks K.M. (2001). Protein-dependent transition states for ribonucleoprotein assembly. J Mol Biol 309, 1087-100.

[51] Rose M.A., Weeks K.M. (2001). Visualizing induced fit in early assembly of the human signal recognition particle. Nat Struct Biol 8, 515-20.

[52] Gilbert S.D., Stoddard C.D., Wise S.J., Batey R.T. (2006). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. J Mol Biol 359, 754-68.

[53] Ottink O.M., Rampersad S.M., Tessari M., Zaman G.J., Heus H.A., Wijmenga S.S. (2007). Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined induced fit mechanism. RNA 13, 2202-12.

[54] Noeske J., Buck J., Furtig B., Nasiri H.R., Schwalbe H., Wohnert J. (2007). Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. Nucleic Acids Res 35, 572-83.

[55] Adilakshmi T., Bellur D.L., Woodson S.A. (2008). Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. Nature 455, 1268-72.

[56] Lu C., Ding F., Chowdhury A., Pradhan V., Tomsic J., Holmes W.M., et al. (2010). SAM recognition and conformational switching mechanism in the Bacillus subtilis yitJ S box/SAM-I riboswitch. J Mol Biol 404, 803-18.

[57] Heppell B., Blouin S., Dussault A.M., Mulhbacher J., Ennifar E., Penedo J.C., et al. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. Nat Chem Biol 7, 384-92.

[58] Daniels K.G., Tonthat N.K., McClure D.R., Chang Y.C., Liu X., Schumacher M.A., et al. (2014). Ligand concentration regulates the pathways of coupled protein folding and binding. J Am Chem Soc 136, 822-5.

[59] Russell R., Herschlag D. (1999). Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme. RNA 5, 158-66.

[60] Farjon J., Boisbouvier J., Schanda P., Pardi A., Simorre J.P., Brutscher B. (2009). Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. J Am Chem Soc 131, 8571-7.

[61] Delaglio F., Grzesiek S., Vuister G.W., Zhu G., Pfeifer J., Bax A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6, 277-93.

[62] Tijerina P., Mohr S., Russell R. (2007). DMS footprinting of structured RNAs and RNA-protein complexes. Nat Protoc 2, 2608-23.

[63] Cordero P., Kladwang W., VanLang C.C., Das R. (2014). The mutate-and-map protocol for inferring base pairs in structured RNA. Methods Mol Biol 1086, 53-77.

[64] Yoon S., Kim J., Hum J., Kim H., Park S., Kladwang W., et al. (2011). HiTRACE: high-throughput robust analysis for capillary electrophoresis. Bioinformatics 27, 1798-805.

[65] Das R., Travers K.J., Bai Y., Herschlag D. (2005). Determining the Mg^{2+} stoichiometry for folding an RNA metal ion core. J Am Chem Soc 127, 8272-3.

[66] Frederiksen J.K., Li N., Das R., Herschlag D., Piccirilli J.A. (2012). Metal ion rescue revisited: Biochemical detection of site-bound metal ions important for RNA folding. RNA 18, 1123-41.

Figure legends

Fig 1. Folding and assembly of P5abc. **(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 [15, 32]. 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 $E^{\Delta P5abc}$ 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).

Fig 2. P5abc assembly with the $E^{\Delta P5abc}$ ribozyme core. **(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.

Fig 3. Equilibrium measurements of P5abc mutants by NMR. **(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.

Fig. 4. Evidence for an alternative assembly pathway. **(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.

Fig. 5. Two pathways for RNA assembly. **(a)**, The conformational selection pathway (blue), which dominates the flux for the wild-type P5abc and mutants that include the

native-stabilizing G176A mutation. From the alternative conformation ($P5abc_{alt}$), P5abc rearranges to the native conformation (indicated as 'local folding' on the projection below the free energy profiles) and then assembles with the ribozyme core (yellow). This pathway dominates for these P5abc variants because the rate-limiting transition state (blue disc) is lower in free energy than that along the red, induced fit pathway (red disc). **(b)**, The induced fit pathway (red), which dominates for P5abc mutants in which the alternative conformation is strongly stabilized relative to the native conformation. For these mutants, the maximal free energy of the rate-limiting step along the conformational selection pathway (blue disc) is increased sufficiently such that this pathway is disfavored relative to the induced fit pathway. After assembly, these P5abc mutants rearrange to the native conformation (local folding). In panels a and b, the cartoons in brackets depict proposed structural features of the rate-limiting transition states (P14 along the conformational selection pathway and the TL/TLR interaction along the induced fit pathway). **(c)**, 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 model shown in **(a)** and **(b)**. The regime to the right of the vertical black line is occupied by the wild-type P5abc and the native-stabilizing mutant G176A. Here, the equilibrium favors the native P5abc structure relative to the alternative structure ($K_{eq} > 1$) and the observed rate constant for assembly with the ribozyme core (black dashed curve) reflects binding of native P5abc. As native P5abc folding becomes modestly unfavorable in double mutants that include G176A ($0.01 < K_{eq} < 1$, center

regime), the assembly rate constant decreases and the conformational selection pathway remains dominant. Further stabilization of the alternative structure (*e.g.* U167C, left of the vertical red line) results in the conformational selection pathway becoming slower than the induced fit pathway (K_{eq} < 0.01, left regime), such that the induced fit pathway dominates the flux. **(d)**, 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 [42]. See the Discussion for further details.

Fig. 1

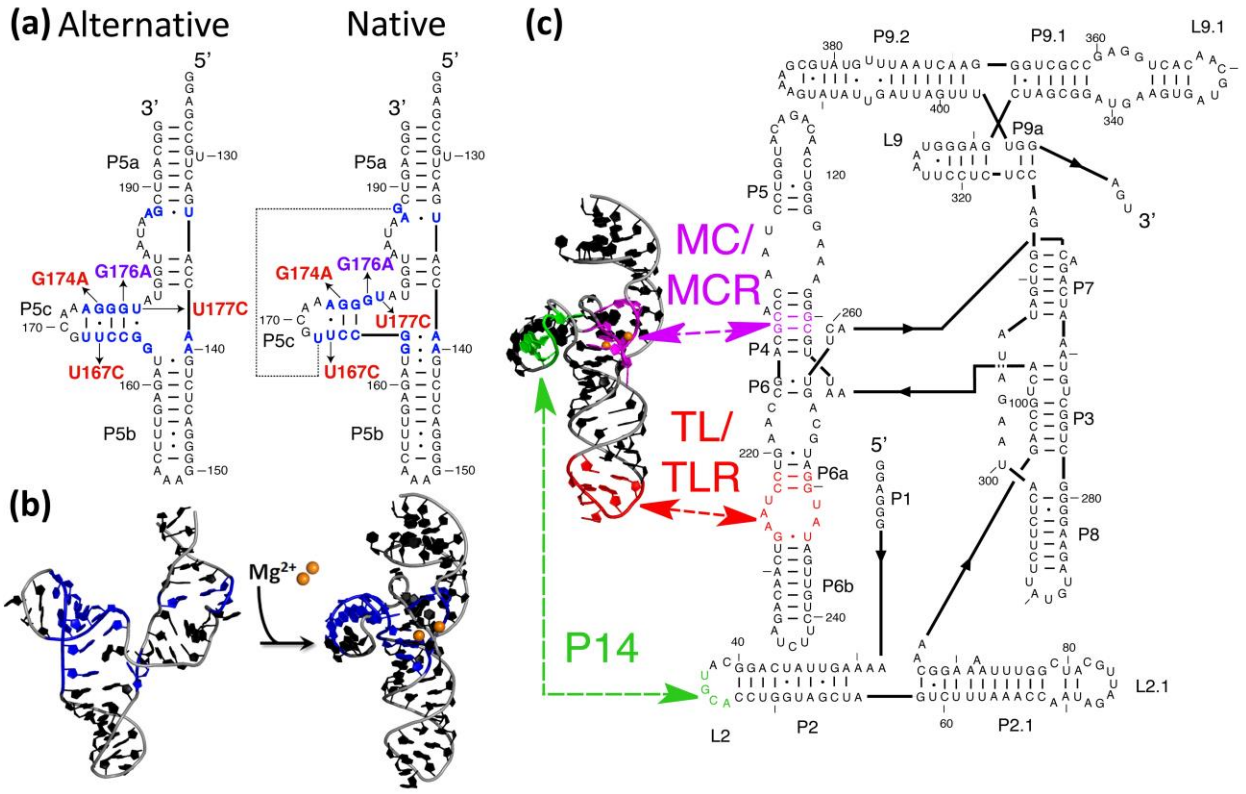

(a) Alternative
Native
5'
3'
P5a
P5b
P5c
G174A
G176A
U177C
U167C
(b)
Mg2+
(c)
MC/
MCR
TL/
TLR
P14
P9.2
P9.1
L9.1
L9
P9a
P5
P4
P6
P6a
P6b
P7
P3
P8
P1
P2
P2.1
L2
L2.1

Fig. 2

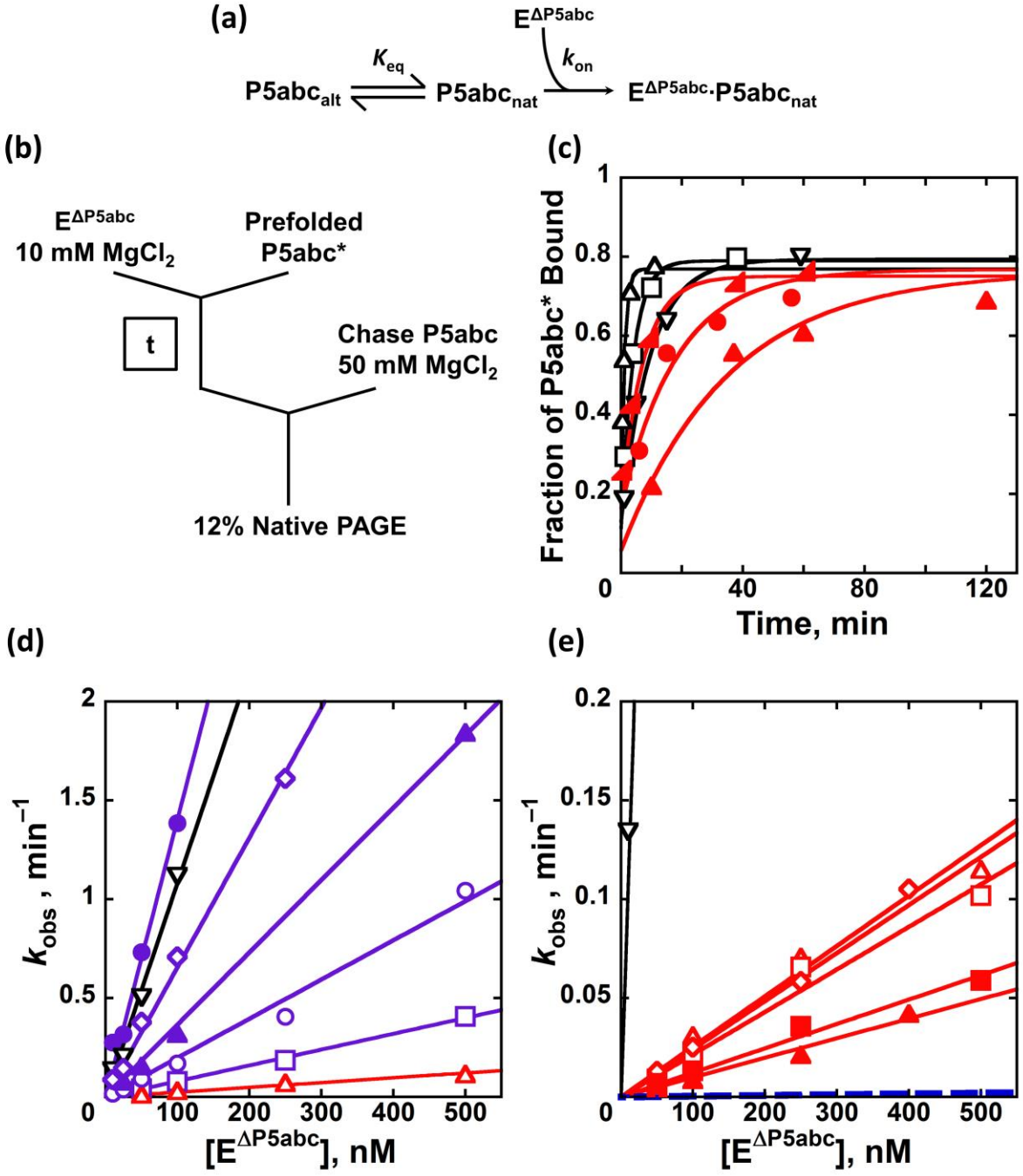

(a)
$E^{\Delta P5abc}$
K_{eq}
k_{on}
$P5abc_{alt}$
$P5abc_{nat}$
$E^{\Delta P5abc}$·$P5abc_{nat}$
(b)
$E^{\Delta P5abc}$
10 mM $MgCl_2$
Prefolded
P5abc*
t
Chase P5abc
50 mM $MgCl_2$
12% Native PAGE
(c)
Fraction of P5abc* Bound
Time, min
(d)
k_{obs}, min^{-1}
$[E^{\Delta P5abc}]$, nM
(e)
k_{obs}, min^{-1}
$[E^{\Delta P5abc}]$, nM

Fig. 3

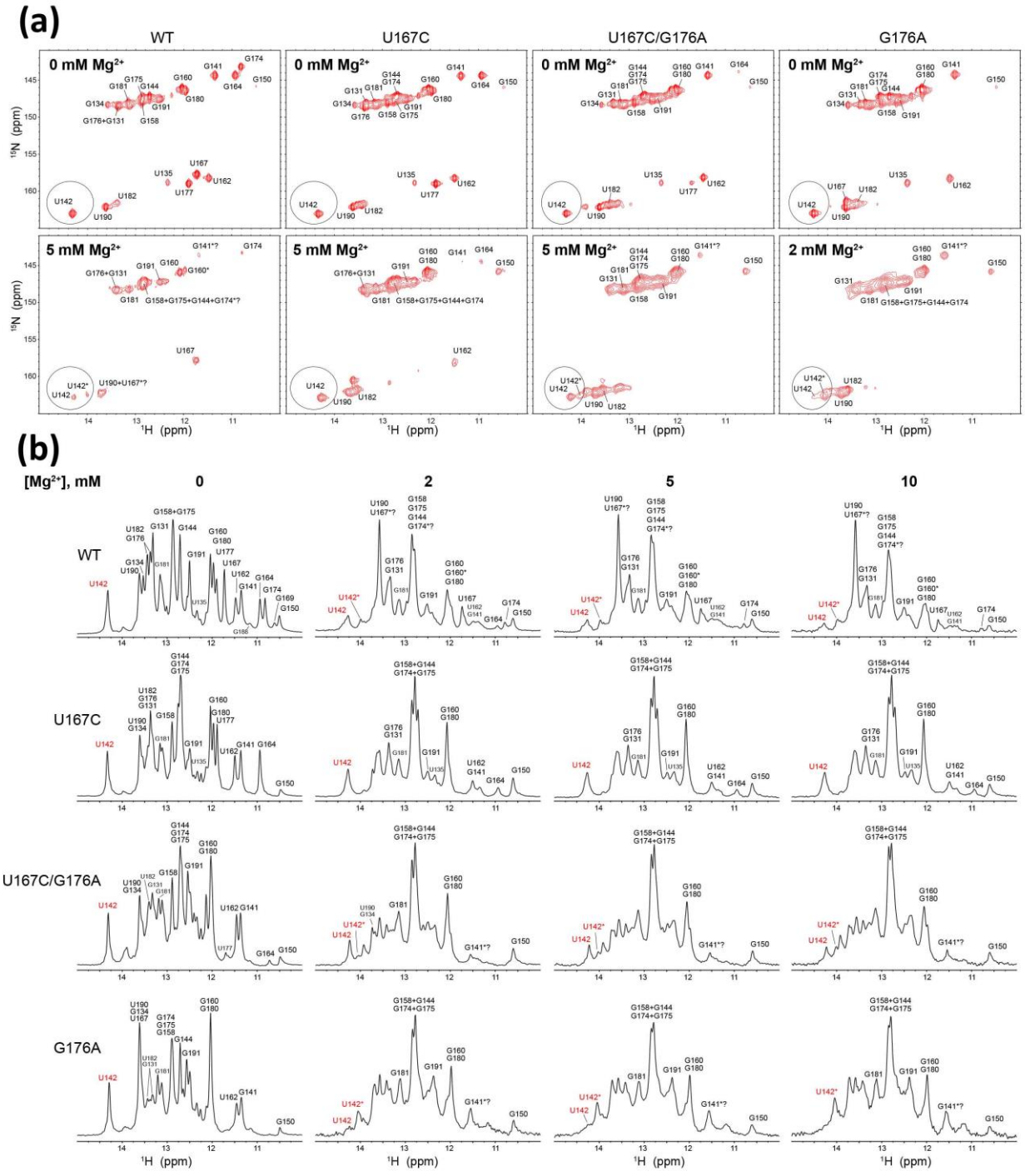

(a)
WT
U167C
U167C/G176A
G176A
0 mM Mg2+
5 mM Mg2+
2 mM Mg2+
15N (ppm)
1H (ppm)
(b)
[Mg2+], mM
0
2
5
10
WT
U167C
U167C/G176A
G176A
1H (ppm)

Fig. 4

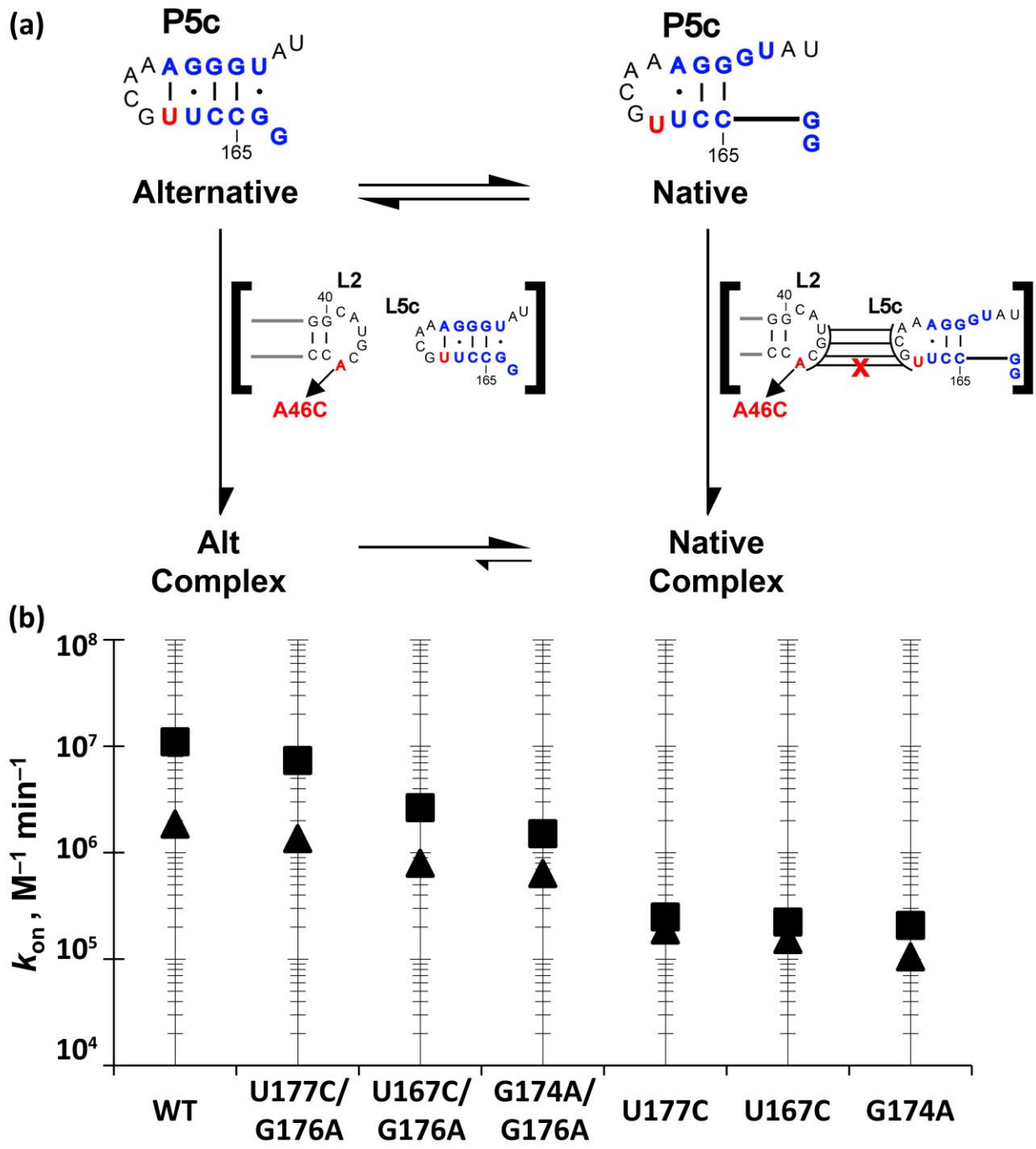

(a)
P5c
Alternative
P5c
Native
L2
L5c
A46C
L2
L5c
A46C
Alt
Complex
Native
Complex
(b)
k_{on}, M^{-1} min^{-1}
10^8
10^7
10^6
10^5
10^4
WT
U177C/ G176A
U167C/ G176A
G174A/ G176A
U177C
U167C
G174A

Fig. 5

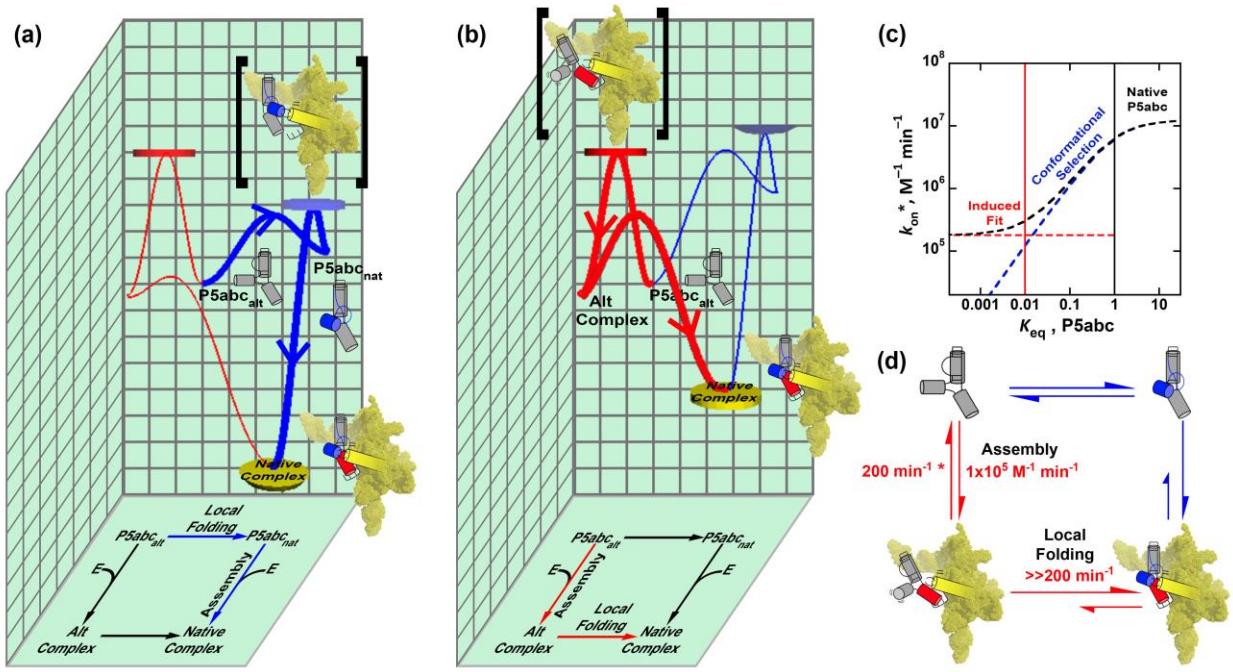

(a)
P5abc$_{nat}$
P5abc$_{alt}$
Native Complex
Local Folding
P5abc$_{alt}$
P5abc$_{nat}$
E
Assembly
E
Alt Complex
Native Complex
(b)
Alt Complex
P5abc$_{alt}$
Native Complex
P5abc$_{alt}$
P5abc$_{nat}$
E
Assembly
Local Folding
E
Alt Complex
Native Complex
(c)
10^8
10^7
10^6
10^5
k_{on} *, M^{-1} min^{-1}
Native P5abc
Conformational Selection
Induced Fit
0.001
0.01
0.1
1
10
K_{eq}, P5abc
(d)
200 min^{-1} *
Assembly
1x10^5 M^{-1} min^{-1}
Local Folding
>>200 min^{-1}

Table 1. Assembly rate constants measured for P5abc variants with $E^{\Delta P5abc}$ or A46C $E^{\Delta P5abc}$ ribozyme cores.*

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.

*Conditions were 50 mM Na-MOPS, pH 7.0, 10 mM $MgCl_2$, 25 °C.

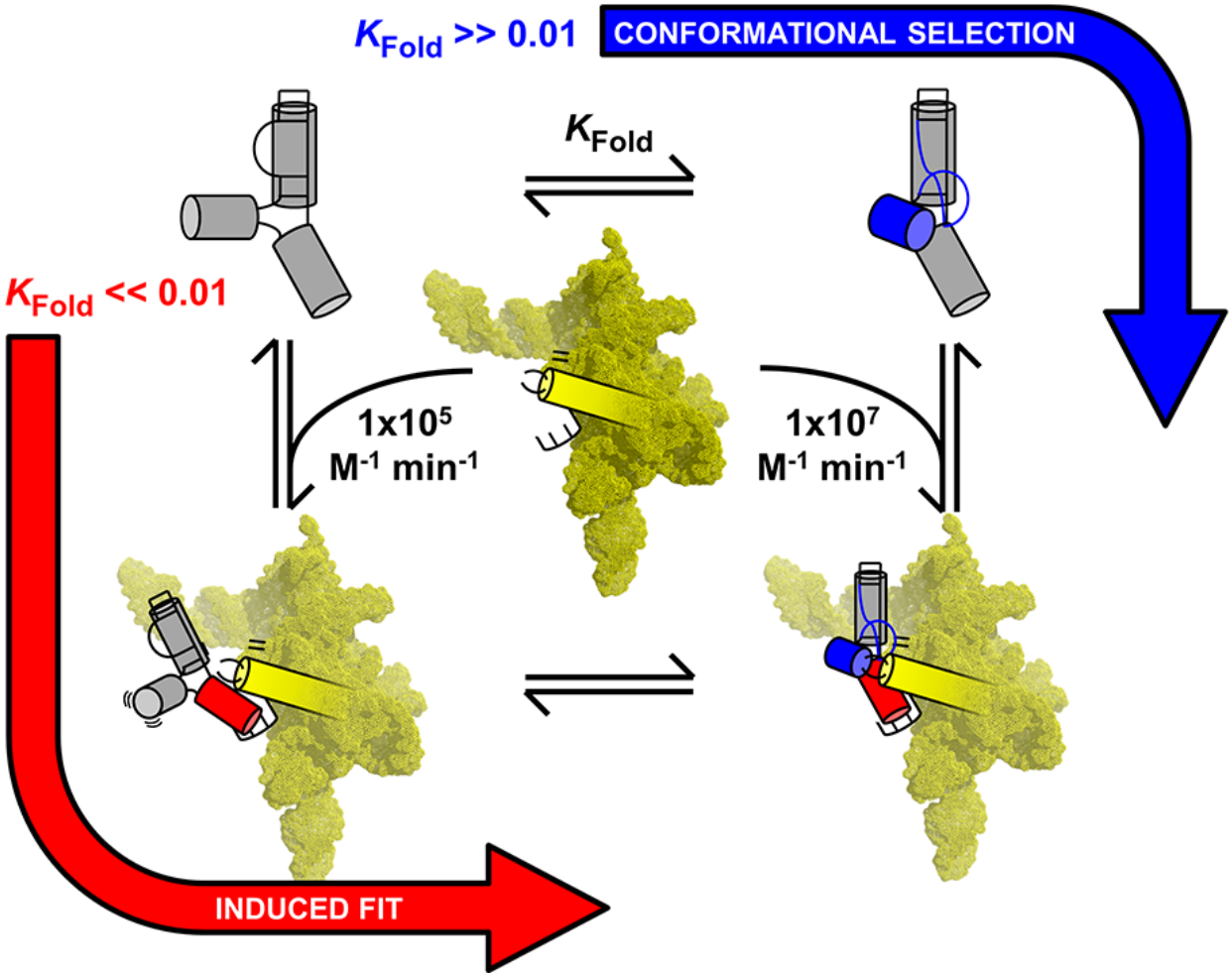
**ΔΔG° values reflect the difference in predicted free energy between the alternative and native secondary structures for each variant relative to the corresponding free energy difference for the wild-type P5abc. Predicted free energies were generated using ViennaFold with solution conditions set to 37 °C and 1 M NaCl.

†k_{rel} P5abc is the assembly rate constant of each P5abc variant divided by the assembly rate constant of wild-type P5abc.

††k_{rel} 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.



Graphical abstract

RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly

Highlights

- Can folding of 2-part group I intron (P5abc & core) be understood from its modules?
- Mutations that destabilize native P5abc decrease the assembly rate up to 100-fold.
- Further P5abc destabilization changes flux to induced fit without further rate change.
- The two pathways use different tertiary contacts in the rate-limiting transition states.
- Results indicate a high degree of modularity in this RNA folding/assembly process.