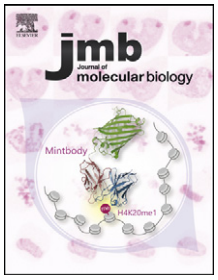

RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly

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, USA*
*2 - **Department of Biochemistry and Chemistry,** Duke University Medical Center, Durham, NC 27710, USA*
*3 - **Department of Biochemistry,** Stanford University, Stanford, CA 94305, USA*

***Correspondence to Rick Russell:** rick_russell@cm.utexas.edu*
http://dx.doi.org/10.1016/j.jmb.2016.07.013
Edited by Philip C Bevilacqua

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, and a tetraloop–receptor interaction, 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 nonnative 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 of an isolated kissing loop similar to P14, and the rate along the induced-fit pathway resembles that of an isolated tetraloop–receptor interaction. Our results indicate substantial modularity in RNA folding and assembly and suggest that these processes can be understood in terms of underlying structural modules.

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 in 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 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, as 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 intermolecular 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. In this pathway, assembly precedes the local rearrangement and is likely nucleated by a tertiary contact that does not require the native P5abc structure. The presence of this pathway allows the highly destabilized P5abc mutants to assemble with the core at rates that are 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 *Tetrahymena* group I intron 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 nonnative base pairs in the helix P5c and one in P5a (Fig. 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 the formation of a bulge, and the formation of additional noncanonical base pairs in P5b (Fig. 1a and 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. 1c) [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 local native structures in the