Published in final edited form as:

Curr Opin Chem Biol. 2008 December ; 12(6): 655–666. doi:10.1016/j.cbpa.2008.09.017.

Energy barriers, pathways and dynamics during folding of large, multi-domain RNAs

Inna Shcherbakova,

Department of Biochemistry, Albert Einstein College of Medicine 1300 Morris Park Avenue, Bronx, NY 10461

Somdeb Mitra,

Department of Biochemistry, Albert Einstein College of Medicine 1300 Morris Park Avenue, Bronx, NY 10461

Alain Laederach, and

Developmental Genetics and Bioinformatics Wadsworth Center, Albany NY 12208

Michael Brenowitz

Department of Biochemistry, Albert Einstein College of Medicine 1300 Morris Park Avenue, Bronx, NY 10461

Summary

Large, multi-domain RNA molecules are generally thought to fold following multiple pathways down rugged landscapes populated with intermediates and traps. A challenge to understanding RNA folding reactions are the complex relationships that exist between the structure of the RNA and its folding landscape. The identification of intermediate species that populate folding landscapes and characterization of elements of their structures are key components to solving the RNA folding problem. This review explores recent studies that characterize the dominant pathways by which RNA folds, structural and dynamic features of intermediates that populate the folding landscape and the energy barriers that separate the distinct steps of the folding process.

Introduction

Initially believed to be humble carriers of genetic information, RNA molecules acquire intricate structures and rely on these structures when playing important roles in translation of the genetic code into protein sequence, in regulation of gene expression and in biological catalysis. Unlike DNA, RNA is designed and synthesized as a self complementary single-stranded molecule. The short stretches of helical duplex, often formed by adjacent regions, provide the basic architecture of the molecule. These regions are separated by internal loops that confer flexibility and therefore allow significant conformational dynamics of the base paired domains. The conformational dynamics in turn enable regions of the backbone to be brought together that are distant in primary sequence and whose juxtaposition is essential for biological function. Long range tertiary contacts between these elements of secondary structure confine the conformational space of the molecule by locking the ensemble of compacted structures into a predominantly single low energy conformation (Figure 1, left panel).

Correspondence should be addressed to M.B. (brenowit@aecom.yu.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The polyanionic nature of an RNA molecule contributes to the initial hurdle that must be overcome to acquire a compact native structure. At low cation concentrations, RNA molecules are highly expanded and potentially rigid due to electrostatic repulsion between the backbone's negatively charged phosphate moieties [1] that outweighs the propensity to form stable tertiary interactions between enthalpically favorable preformed elements of secondary structure [2]. Most compaction arises from the non-specific neutralization of the charge. However, many crystal structures of folded RNAs reveal specifically bound, partially dehydrated divalent [3–8] (and sometimes monovalent [6,9]) cations that neutralize pockets of high negative potential. This final touch of local charge neutralization is critical to some folding reactions; some molecules cannot acquire their native structure in the absence of the divalent cations. Therefore most *in vitro* folding kinetics studies aim to elucidate the time evolution of structural changes initiated by addition of the divalent cations to the RNA molecules whose initial state varies in the degrees of backbone charge neutralization, compaction and preformed tertiary structure. Furthermore, *in vitro* folding experiments generally study the formation of tertiary structure formation, as secondary structure formation occurs on extremely short (sub-microsecond) timescales and is relatively independent of counterion concentrations.

From the thermodynamic point of view, the entropy loss upon folding associated with confining a molecule's conformational space is offset by the enthalpy gain arising from tertiary contact formation and solvent release. When viewed in coordinates of conformational space and energy, a folding landscape is visualized as a funnel (Figure 1, right panel). The initial ensemble of molecules is widely distributed in conformational space. The ensemble progressively approaches conformational homogeneity as it descends toward a unique native structure. It has been shown in numerous *in vitro* studies that real folding landscapes are rarely represented by a smoothly descending funnel. Folding landscapes for large RNA molecules are typically described as 'rough' and 'rugged' [10] populated with intermediates located in local energy minima that slow down the folding time. In many cases, more than a single pathway leading from the initial to the final state is appreciably populated. That an unbiased search for the structure with minimal energy through all possible conformations would take an astronomical amount of time (the Levinthal paradox) suggests that the folding is kinetically guided. Folding kinetics studies can illuminate the paths each RNA molecule follows to reach its unique native structure.

Among the approaches designed to probe the kinetics of RNA folding are methods that can provide more than a single measure reporting on the folding, namely folding kinetics of local reporters distributed throughout the molecule. In this review we will discuss how the information about formation of local tertiary contacts can be placed into context of global molecular compaction observed during RNA folding. Another focus of our discussion is the distillation of the kinetic information about the formation of the individual tertiary contacts acquired by time-resolved millisecond hydroxyl radical footprinting [11] into a kinetic folding model [12]. We show how these ensemble experiments provide glimpses of the intermediate structures, population of distinct folding pathways and rates of the individual reaction steps. Systematic variation of a solution variable (such as cation type and/or concentration) and careful evaluation of the effects of the changes allow us to discriminate distinct roles of variables in determining features of the folding landscapes of the RNA molecules.

We center our discussion around one of the best studied of large RNA molecules, the group I intron ribozyme derived from the ciliate *Tetrahymena thermophila*. Kinetic and thermodynamic information derived from this molecule can be generalized to most large multi-domain RNA molecules because of its structural complexity, diversity of tertiary interactions and the common topological features that it shares with most other large RNA molecules such as parallel packing of coaxially stacked helices [13].

Seeing the forest *and* the trees: global and local perspectives

Small-angle x-ray scattering (SAXS) and analytical ultracentrifugation report on the overall conformation of a biological molecule (reviewed in [14] and [15], respectively). While both techniques provide insight into the thermodynamic folding landscapes under various ionic conditions, only SAXS has been applied to the study of folding kinetics [16–19]. For both thermodynamic and kinetic studies, information about global compaction of the macromolecule is complemented by site-specific information. Recent studies of the thermodynamic folding landscape of RNAs using SAXS and site-specific techniques revealed the sequence of events during cation mediated folding [1], the character of the thermodynamic transitions [20] and the structure of the equilibrium intermediates [21,22]. In particular, comparison of the global compaction and tertiary contact formation as a function of increasing concentration of the monovalent cations revealed two distinct phases in Na⁺-induced folding of the P4-P6 domain of the *Tetrahymena* ribozyme [1]: the first phase is rationalized by electrostatic relaxation of the molecule, the second one is tertiary structure-driven compaction. A thermodynamic study of the Mg²⁺-induced folding of a group I intron from *Azoarcus sp.* revealed sharp cooperative transition between the unfolded and folded states [20]. Mutational analysis of this ribozyme revealed the tertiary contacts that differentially influence global compaction of the folded state and the character of the folding transition [23].

The comparison of the global compaction with tertiary structure formation of a glycine-dependent riboswitch by increasing the concentration of Mg²⁺ and glycine revealed a three-state thermodynamic model for compaction [22]: addition of Mg²⁺ in the absence of glycine resulted in the formation of the partially folded riboswitch, with some tertiary structure formed, whereas final tertiary structure formation required addition of glycine and some additional uptake of the Mg²⁺ cations. Low-resolution models for the unfolded and intermediate structure were proposed based on the SAXS profiles. A nucleotide-level model of the structure of the thermodynamic intermediate arising upon folding of another model RNA, the S-domain of the RNaseP from *Bacillus subtilis*, has been presented [21]. This model utilized the radius of gyration (R_g) as a constraint and was validated by variety of global and site-specific experiments. The studies revealed that this major intermediate in the thermodynamic folding landscape possesses extensive tertiary structure.

Kinetic folding studies of the larger group II introns ribozymes, which are thought to be the precursors of eukaryotic spliceosomes, demonstrate the importance of comparative analysis of phylogenetically unrelated or distantly related and functionally divergent RNAs. As opposed to the fast collapse followed by slow organization of the catalytic core of group I introns ribozymes, the rate limiting step in the folding pathway of the D135 ribozyme derived from a group II intron is a slow Mg²⁺-dependent collapse to an unstable intermediate [24]. Appearance of this intermediate, characterized by formation of the κ - ζ region within domain D1, during the early stages of folding, is followed by a fast and direct assembly of the remaining domains to the native state without detectable kinetic traps [24–28].

The kinetic folding landscape of the *Tetrahymena* ribozyme was interrogated by time-resolved SAXS [16,17]. Comparison of the kinetics of global compaction and the kinetics of tertiary contact formation collected under the same experimental condition allows non-specific electrostatic collapse and the compaction driven by the formation of the tertiary contacts to be distinguished [18]. In studies of cation mediated folding of the *Tetrahymena* ribozyme, non-specific compaction unaccompanied by tertiary contact formation occurs within the first milliseconds after initiation of the reaction. Thus, the conformational search for the partners of the tertiary contacts occurs within a limited conformational space. Formation of the tertiary structure is reflected by a second phase of global molecular compaction that occurs from 10 to 100 msec followed by a small additional compaction occurring at 100 seconds that reflects

the reorganization of the elements of the tertiary structure within the compact macromolecular ensemble.

Folding of the P4-P6 domain is essential to the second phase of global molecular compaction, since this phase is only present for the ribozyme with native contacts present that lock together its characteristic hairpin-like structure. Recent studies of the time-dependent folding of the P4-P6 domain in isolation [19] demonstrated that its formation is a two phase process and the rate of its formation depends on the ionic strength. Folding is limited by hinge stiffness at low ionic strength and formation of tertiary contacts at high ionic strength. In contrast to the important role that tertiary contacts play within the P4-P6 domain, destabilization of the tertiary contacts that link the ring of the peripheral helices of the full length ribozyme only minimally affects the kinetics of global compaction [18]. Thus, ordering of the periphery locks the preformed RNA structure together rather than driving compaction. Next we consider tertiary contact mediated folding transitions that occur subsequent to the initial non-specific compaction of the RNA.

Early energy barriers (Up then down)

Many of the early studies of the Mg^{2+} -mediated folding kinetics of the *Tetrahymena* ribozyme were conducted in a very low ionic strength background (~ 10 mM Na^+ or K^+) [29,30]. Time-resolved hydroxyl radical footprinting studies carried out under these solution conditions revealed mono-exponential time progress curves whose rates ranged from a moderate ~ 2 s $^{-1}$ to a very slow 0.01 s $^{-1}$. The unique feature of the Mg^{2+} -mediated folding landscape observed in the very low ionic strength background is the presence of early barriers preventing the ribozyme from instantaneous initiation of tertiary contact formation. The presence of this early barrier was confirmed by the two orders of magnitude facilitation of the initial folding events by sub-denaturing concentrations of urea [31], presumably accelerating folding by destabilizing the kinetic intermediates [32]. Several studies suggest that the initial barriers are connected with a non-native conformation of the P4-P6 domain arising under the initial low salt condition. Indeed, destabilization of the P4-P6 domain by mutating its tetraloop also accelerates the overall folding of the ribozyme [33]. The initial folding steps for the mutated ribozyme are accelerated ≥ 50 fold; the temporal sequence of folding events is also drastically altered [31].

The *Azoarcus* group I ribozyme whose P4-P6 domain consists of a single coaxially stacked set of helices [34], due to the absence of the P5abc sub-domain folded alongside of the longer arm of the P4-P6 domain in the *Tetrahymena* ribozyme, folds rapidly under these experimental conditions [23]. In contrast, the *Tetrahymena* ribozyme lacking the P5abc extension folds only slightly faster (5-fold) compared to the wild type [35]. The P4-P6 domain in isolation and within the ribozyme folds at comparable rates, with only the three-helix junction within P5abc structuring significantly faster in isolation [36]. Clearly, the rate-limiting step to P4-P6 folding does not involve interdomain interactions under these solution conditions.

The above described early barrier to folding by the *Tetrahymena* ribozyme is clearly dependent upon the very low ionic strength background from which the Mg^{2+} -mediated folding was initiated. Mg^{2+} -mediated folding is much faster when the reaction is initiated from a higher ionic strength background [37,38]. (The folding landscape observed at higher ionic strength is discussed later in this review.) However, low ionic strength is necessary but insufficient to raise the initial folding barriers. Folding onset is instantaneous (> 50 s $^{-1}$) when initiated by high concentrations of monovalent cations from this very low ionic strength initial condition [39]. Comparison of monovalent- and Mg^{2+} -induced folding landscape suggests that Mg^{2+} intimately interacts with the ribozyme. These interactions stabilize fast forming non-native structures, generating early barriers to folding. Mg^{2+} -ribozyme interactions clearly play an

important role in the determination of the intermediate structures and their stabilization [40, 41]. The increased intermediate stability creates kinetic traps that molecules must overcome to proceed down the folding funnel. The next section explores recent progress in the identification and characterization of RNA folding intermediates.

Folding intermediates and kinetic traps

Kinetic intermediates were suggested to populate the *Tetrahymena* ribozyme folding pathways from the earliest kinetics studies that utilized oligonucleotide hybridization to follow folding [29]. Since the P4-P6 domain of the *Tetrahymena* ribozyme folds faster than its catalytic core, the presence of a kinetic intermediate containing the folded P4-P6 was proposed [29]. The millisecond temporal resolution of time-resolved hydroxyl radical footprinting resolved a folding hierarchy for the Mg^{2+} -induced folding of the *Tetrahymena* ribozyme. In a very low ionic strength background the hierarchy was originally interpreted to be sequential with $P5c > P4-P6 \gg$ Periphery $\gg$ Catalytic Core [30]. Subsequent analysis of these and additional time progress curves acquired under the same experimental conditions, utilizing new methods of data acquisition [42], clustering and ‘structural – kinetic’ (KinFold) modeling of the resultant time-progress curves for individual sites distributed throughout the molecule, yielded a robust kinetic folding model configuration where three parallel pathways partition from the very onset of folding [12]. Below we summarize the KinFold algorithm and discuss the kinetic models discerned for the Mg^{2+} -induced folding of the *Tetrahymena* ribozyme.

KinFold clusters similar time progress curves; the number of clusters determines the minimum number of intermediates required to describe the folding reaction (one less than the number of clusters). Exhaustive testing of all the possible kinetic model configurations yields the model that best describes the folding reaction along with information about the structured portions of each intermediate [12]. The structural information inherent in a time resolved footprinting experiment stems from the association of time-progress curve clusters with intermediates; KinFold samples hydroxyl radical reactivity at intermediate folding times, identifies regions in the RNA that exhibit similar behavior and thus defines the kinetic subdomain architecture of the molecule. In our intermediate models, we make the assumption that the observed hydroxyl radical reactivity is a consequence of native structure and thus constrain these regions. Only the secondary structure is constrained for the rest of the molecule. The resulting models therefore represent likely conformations comprised of tertiary contacts that are appreciably formed. The identification of non-native structures that constitute kinetically trapped intermediates is an important challenge for modeling intermediate structures. It is likely that the inclusion of additional experimental measures which complement the solvent accessible surface reported by hydroxyl radical footprinting will be required for the robust characterization of non-native intermediate structures.

To illustrate the information gleaned from a KinFold analysis of hydroxyl radical time progress curves, we walk through the model resolved for the Mg^{2+} -induced folding of the *Tetrahymena* ribozyme in a very low ionic strength background introduced above. Folding is dominated by a small number of pathways that progress through three structurally distinct intermediates before converging to the native conformation (Figure 2; [12]). The most populated pathway proceeds through I2 that has P4-P6 domain folded; I2 then slowly structures the remainder of the ribozyme to form the native molecule. This pathway reflects the general view that P4-P6 is a folding scaffold. The improved experimental data obtained with contemporary methods [42] allowed an alternate pathway to be statistically distinguished in which folding of the P5abc subdomain (I1) precedes structuring of the remainder of P4-P6 (I2). Structuring of P5abc also precedes population of I3 where both P4-P6 and the periphery are organized; these pathways are the second and third most populated on the folding landscape. It is interesting that I1, closest to the unfolded ribozyme in energy, is predominantly a precursor

to formation of I2 or I3. The direct transition between I1 and the final state is virtually unpopulated. The fourth most populated pathway is the direct formation of I3, which has all the ribozyme except its catalytic core folded. Other detectably populated pathways include those where the ribozyme wanders among intermediates before descending to the final state suggesting that for this condition the barriers separating the intermediates are lower than those separating the intermediates from the final state.

Catalytic core folding is the slowest step of the ribozyme's conversion into its native form for all of the discerned intermediates because of the second line of energy barriers separating them from the folded structure. That tertiary contacts must be broken and reformed to fold the catalytic core from these intermediates is shown by the acceleration of the catalytic core formation by a sub-denaturing concentration of urea [31]. Another interesting feature of this folding landscape is that the direct transition $U \rightarrow F$ is minimally populated under these folding conditions. Presumably, its initial barrier is significantly higher than those of the alternative initial steps. It is interesting that the *Tetrahymena* ribozyme folding under this condition of high temperature by a single type of cation is described by a very rugged folding landscape (Figure 2).

The influence of Mg^{2+} in carving this rugged landscape can be gleaned by comparison with monovalent cation-mediated folding initiated from the same initial state [39]. High concentrations of monovalent cations yield a structure that is close to the active, Mg^{2+} -folded, molecule [43]. Monovalent cation-mediated folding is characterized by a very smooth folding funnel (Figure 3; [12]). The two intermediates populated this landscape have structures that are very different from those observed during Mg^{2+} -mediated folding. In this case, the sites known to specifically bind divalent cations [4] do not fold first. Rather folding proceeds from the outside to the inside of the ribozyme. The P9 peripheral extension is structured in I1. P5abc, P6 and P2 peripheral helix join the P9 extension to form I2. The ribozyme proceeds rapidly down the folding funnel either through one of these intermediates or directly to the final state. Intermediate interconversion is minimal suggesting that the barriers between them are higher than those blocking folding to the final state. This reaction is notable not only for its faster onset (due to the absence of initial barriers) but also for the speed of the latter folding steps. These observations suggest much lower barriers along the monovalent cation mediated folding pathways due to the reduced stability of the intermediates.

When Mg^{2+} -mediated folding is conducted in a background of higher salt, the kinetic model configuration is comparable to that resolved in the low ionic strength background although there are significant differences in reaction rates and intermediate structures (Figure 4; [38]). As noted above, the initial barrier impeding folding in the low salt background is not present at higher salt concentrations. Within a cohort of six high ionic strength folding conditions, striking difference in the initial flux partitioning are observed. The differences in folding flux partitioning suggest that once the initial barrier of electrostatic repulsion is overcome, the intermediate structures are minimally affected by differences in compaction and extent of tertiary contact formation in the unfolded ensemble.

The consistency of the kinetic model configuration as a function of background salt concentration contrasts with the completely different set of folding intermediates that was revealed upon destabilization of the tetraloop-tetraloop receptor interaction that is the principal stabilizing contact of P4-P6 domain [38]. While this mutant ribozyme acquires essentially the same catalytically active final folded state as the wild type, the structural features of its folding intermediates are dictated by the overall stability and contact order of tertiary interactions in addition to the topology of its native state.

KinFold modeling of the folding of the *Tetrahymena* ribozyme under several folding conditions discussed above suggests that RNA folding is parallel from its very onset, proceeding down more or less defined folding ‘channels’ [58], with the ‘sideways’ barriers between the channels higher or lower depending on the folding condition and stability of the intermediates. For example, comparison of the folding landscapes of the *Tetrahymena* ribozyme in the background of high concentration of monovalent cations revealed a phenomenon worthy of further exploration: folding proceeds down the channels in K^+ containing solutions while significant exchange among the folding intermediates occurs in Na^+ containing solution [38]. It is tempting to speculate that one or more specific K^+ binding events stabilize intermediate structures and thus restrict their exchange. Earlier works, done both on the *Tetrahymena* ribozyme [58] and circularly permuted RNase P [44], indicated the presence of several parallel pathways separated by energy barriers.

For the full-length RNase P RNA the kinetic trap that limits its folding is associated with the resolution of non-native inter-domain interactions [47]. It was demonstrated that folding of the C-domain of the same RNA proceeds through fast formation of an intermediate [32,48] that possesses native secondary and some extent of tertiary structure [49,50]. Subsequent formation of the native tertiary structure occurs along a single robust pathway, comprising of two on-pathway intermediates [48,51]. The S-domain was found to follow the same $U \rightarrow I \rightarrow N$ scheme. Through a detailed investigation of the pathway for detection of kinetic traps [52] it was proposed that the slow consolidation of tertiary structure around a metal-ion binding site, similar to the D135 group II intron ribozyme, constitutes the rate-limiting step in folding. Recent single molecule papers [45,46] revealed the presence of high energy barriers between two distinct regions of the free energy landscape, separating two populations of the RNA molecules of significantly different dynamics. Last but not the least, information about the kinetics of folding and assembly of the grandest of all RNA structures, the ribosomal RNA, has started flowing steadily into the RNA folding literature [53]. An article about this topic is included in this volume [Editors: Please reference Woodson article in volume].

The exploration of structural features and interconversion of kinetic intermediates during folding of the *Tetrahymena* ribozyme may be expanded into a more general scenario that could potentially elucidate how RNA molecules with minimal sequence homology often fold into conserved three dimensional structures. Our studies indicate that arrangement of simple base paired helices into complex yet conserved three dimensional topologies may be an evolutionary strategy adopted by divergent RNA molecules. Energetically favorable patterns of tertiary interactions might preferentially guide the ensemble of folding molecules through native like intermediates that can serve as stable building blocks for higher order structures under widely different folding conditions.

The playing field and who determines the rules

What does it take to change a folding landscape? Obviously, the folding landscapes observed for monovalent and divalent cation-mediated folding of the *Tetrahymena* ribozyme are dramatically different. Destabilizing a key tertiary contact also changes the landscape. However, perhaps more informative to understanding the barriers that define landscapes are subtle perturbations that affect some features of the landscape but leave its overall topology intact. Our first foray into subtle perturbation of the Mg^{2+} -induced folding of the *Tetrahymena* ribozyme sought to determine the stability of a folding landscape as a function of the electrostatic environment [38]. We performed Mg^{2+} -mediated folding as a function of background monovalent cation type (Na^+ vs. K^+) and concentration (200 vs. 600 mM). ‘Salt jump’ experiments were also performed in which the monovalent cation concentration was changed concurrently with Mg^{2+} for initiating folding in order to differentiate between the effects of monovalent salt concentration on the initial conformation and on the folding reaction.

As noted above, a kinetic model configuration composed of three intermediates is common to all of these solution conditions for the wild type ribozyme (Figure 4). The structures of the intermediates are minimally affected by changes in the folding conditions. Four pathways dominate folding, the direct pathway and three pathways populated with intermediates. All the intermediates are separated from the folded state by significant energy barriers. Of particular importance is the fact that changing the folding conditions does not alter the overall configuration of the kinetic model; it simply results in a rearrangement of the molecular flux through the different pathways. This observation argues for conserved RNA folding mechanisms that are dictated by structure and not by the conditions under which the molecules fold. Thus, altered electrostatic condition mainly modulates the height and subtle geometry of energy barriers although dismantling and erecting new ones is observed as well. For example, the barrier between initial state and I3 is present for the folding reaction that occurs in a background of Na^+ but not K^+ (Figure 4).

That differences in the initial conformational ensemble modulate the initial folding rates suggests that it is not uniform but is confined to several niches separated by low energy barriers. The addition of Mg^{2+} opens the downhill channels of the landscape, the molecules descend faster than they exchange between the distinct initial states. Once folding commences, the solution environment modulates the relative heights of the 'sideways' and 'downhill' barriers. The observed structures of the intermediates are minimally affected by changes in the initial state and the concentration of background cations. This correspondence between structure and folding is not surprising as is the opposite, destabilization of a tertiary contact results in gross alteration of intermediate structures [38] accompanied by a radical change in the kinetic model configuration.

Concluding remarks

Focusing our studies on the folding of the *Tetrahymena* ribozyme and developing tools that provide us improved spatial and temporal resolution, we have made progress towards understanding the dynamics of folding intermediates and energy barriers along its folding landscape. The individual-site time progress curves generated by the ensemble method of time-resolved hydroxyl radical footprinting analyzed by KinFold [12] yields kinetic model configurations for folding reactions populated with intermediates and preferred pathways. This approach combines millisecond time-resolution and the ability to meld the kinetic behavior of multiple reporters of folding into testable models replete with the structural character of the intermediates. The result is the distillation of a cornucopia of information into global kinetic models that can be interpreted in terms of folding landscapes.

We observe that the folding of the *Tetrahymena* ribozyme is parallel since its earliest steps and under most solution conditions, involves partitioning of the molecular flux between the direct folding pathway and several pathways populated by intermediates. The structures of the folding intermediates are dictated by the type of cations used to fold the molecule as well as by the macromolecular stability and order of tertiary contact formation. The initial conformational ensemble determines the initial partitioning of the molecular flux through different pathways whereas the solution environment modulates the rates of individual reactions affecting the height of the barriers separating the folding pathways and local energy minima. Whether these behaviors are general to other RNA molecules is being actively investigated.

If these behaviors are general, the principles emerging from these studies could provide the framework for developing knowledge based structure prediction tools. By doing so, not only shall we spearhead our way into the vast and largely unexplored realm of unsolved biologically important RNA structures, but also effectively design RNA sequences that would fold into novel structures. For example, comparative phylogenetic analysis has proved to be an effective

tool for predicting RNA secondary structures [54,55]; a similar approach could be adopted to test whether specific energetically favorable sets of intra-molecular interactions or ‘folding pathways’ have been conserved in structurally homologous RNA molecules derived from evolutionarily divergent biological species. To date we have very few examples of studies that explore the folding landscapes of closely related RNAs (and even fewer at comparable folding conditions) to allow us to make general conclusions about conserved themes in RNA folding.

A recurring theme among kinetic folding of large RNA molecules that have been studied is that folding pathways are rarely fast and direct; instead, they are impeded either by misfolded intermediates and kinetic traps that need to be resolved by partial unwinding of the non-native structure or by slow nucleation of the folding domains around specific cation binding sites. With the current information, it is hard to predict the nature of the rate-limiting step in the folding of a particular RNA and horizontal expansion into the phylogenetic space of diverse model RNAs could significantly contribute to the understanding of the general rules guiding RNA folding.

While this review has exclusively discussed experiments that measure the average behavior of a large ensemble of molecules, recent advances in single molecule techniques have started providing unprecedented details about the behavior of individual molecules within an ensemble. Such studies have begun to detect transient intermediates and their conformational dynamics that are largely precluded in ensemble measurements. The time is ripe to set up complementary studies employing both single molecule and bulk approaches to increase the depth as well as the breadth of our understanding of the folding landscape of complex RNA structures.

Acknowledgments

The writing of this article was supported by the Institute of General Medical Sciences of the National Institutes of Health grants PO1-GM066275 and K99/R00-GM079953.

References

1. Takamoto K, Das R, He Q, Doniach S, Brenowitz M, Herschlag D, Chance MR. Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J Mol Biol* 2004;343:1195–1206. [PubMed: 15491606]
2. Manning GS. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* 1978;11:179–246. [PubMed: 353876]
3. Cate JH, Doudna JA. Metal-binding sites in the major groove of a large ribozyme domain. *Structure* 1996;4:1221–1229. [PubMed: 8939748]
4. Cate JH, Hanna RL, Doudna JA. A magnesium ion core at the heart of a ribozyme domain. *Nat Struct Biol* 1997;4:553–558. [PubMed: 9228948]
5. Stahley MR, Strobel SA. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 2005;309:1587–1590. [PubMed: 16141079]
6. Stahley MR, Adams PL, Wang J, Strobel SA. Structural metals in the group I intron: a ribozyme with a multiple metal ion core. *J Mol Biol* 2007;372:89–102. [PubMed: 17612557]
7. Jovine L, Djordjevic S, Rhodes D. The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg(2+) in 15-year old crystals. *J Mol Biol* 2000;301:401–414. [PubMed: 10926517]
8. Golden BL, Kim H, Chase E. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 2005;12:82–89. [PubMed: 15580277]
9. Basu S, Rambo RP, Strauss-Soukup J, Cate JH, Ferre-D'Amare AR, Strobel SA, Doudna JA. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat Struct Biol* 1998;5:986–992. [PubMed: 9808044]

10. Rook MS, Treiber DK, Williamson JR. Fast folding mutants of the Tetrahymena group I ribozyme reveal a rugged folding energy landscape. *J Mol Biol* 1998;281:609–620. [PubMed: 9710534]
11. Sclavi B, Woodson S, Sullivan M, Chance MR, Brenowitz M. Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J Mol Biol* 1997;266:144–159. [PubMed: 9054977]
12. Laederach A, Shcherbakova I, Liang MP, Brenowitz M, Altman RB. Local kinetic measures of macromolecular structure reveal partitioning among multiple parallel pathways from the earliest steps in the folding of a large RNA molecule. *J Mol Biol* 2006;358:1179–1190. [PubMed: 16574145]
13. Laederach A, Chan JM, Schwartzman A, Willgohe E, Altman RB. Coplanar and coaxial orientations of RNA bases and helices. *Rna* 2007;13:643–650. [PubMed: 17339576]
14. Lipfert J, Doniach S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu Rev Biophys Biomol Struct* 2007;36:307–327. [PubMed: 17284163]
15. Howlett GJ, Minton AP, Rivas G. Analytical ultracentrifugation for the study of protein association and assembly. *Curr Opin Chem Biol* 2006;10:430–436. [PubMed: 16935549]
16. Russell R, Millett IS, Tate MW, Kwok LW, Nakatani B, Gruner SM, Mochrie SG, Pande V, Doniach S, Herschlag D, et al. Rapid compaction during RNA folding. *Proc Natl Acad Sci U S A* 2002;99:4266–4271. [PubMed: 11929997]
17. Russell R, Millett IS, Doniach S, Herschlag D. Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* 2000;7:367–370. [PubMed: 10802731]
18. Kwok LW, Shcherbakova I, Lamb JS, Park HY, Andresen K, Smith H, Brenowitz M, Pollack L. Concordant exploration of the kinetics of RNA folding from global and local perspectives. *J Mol Biol* 2006;355:282–293. [PubMed: 16303138]
19. Schlatterer JC, Kwok LW, Lamb JS, Park HY, Andresen K, Brenowitz M, Pollack L. Hinge stiffness is a barrier to RNA folding. *J Mol Biol* 2008;379:859–870. [PubMed: 18471829]
20. Chauhan S, Caliskan G, Briber RM, Perez-Salas U, Rangan P, Thirumalai D, Woodson SA. RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. *J Mol Biol* 2005;353:1199–1209. [PubMed: 16214167]
21. Baird NJ, Westhof E, Qin H, Pan T, Sosnick TR. Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. *J Mol Biol* 2005;352:712–722. [PubMed: 16115647]
22. Lipfert J, Das R, Chu VB, Kudaravalli M, Boyd N, Herschlag D, Doniach S. Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae*. *J Mol Biol* 2007;365:1393–1406. [PubMed: 17118400]
23. Chauhan S, Woodson SA. Tertiary interactions determine the accuracy of RNA folding. *J Am Chem Soc* 2008;130:1296–1303. [PubMed: 18179212]
24. Waldsich C, Pyle AM. A kinetic intermediate that regulates proper folding of a group II intron RNA. *J Mol Biol* 2008;375:572–580. [PubMed: 18022197]
25. Waldsich C, Pyle AM. A folding control element for tertiary collapse of a group II intron ribozyme. *Nat Struct Mol Biol* 2007;14:37–44. [PubMed: 17143279]
26. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science* 2008;320:77–82. [PubMed: 18388288]
27. Su LJ, Waldsich C, Pyle AM. An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res* 2005;33:6674–6687. [PubMed: 16314300]
28. Su LJ, Brenowitz M, Pyle AM. An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *J Mol Biol* 2003;334:639–652. [PubMed: 14636593]
29. Zarrinkar PP, Williamson JR. Kinetic intermediates in RNA folding. *Science* 1994;265:918–924. [PubMed: 8052848]
30. Sclavi B, Sullivan M, Chance MR, Brenowitz M, Woodson SA. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 1998;279:1940–1943. [PubMed: 9506944]
31. Shcherbakova I, Brenowitz M. Perturbation of the hierarchical folding of a large RNA by the destabilization of its Scaffold's tertiary structure. *J Mol Biol* 2005;354:483–496. [PubMed: 16242711]

32. Pan T, Sosnick TR. Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat Struct Biol* 1997;4:931–938. [PubMed: 9360610]
33. Treiber DK, Williamson JR. Concerted kinetic folding of a multidomain ribozyme with a disrupted loop-receptor interaction. *J Mol Biol* 2001;305:11–21. [PubMed: 11114243]
34. Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. Crystal structure of a self-splicing group I intron with both exons. *Nature* 2004;430:45–50. [PubMed: 15175762]
35. Russell R, Tijerina P, Chadee AB, Bhaskaran H. Deletion of the P5abc peripheral element accelerates early and late folding steps of the Tetrahymena group I ribozyme. *Biochemistry* 2007;46:4951–4961. [PubMed: 17419589]
36. Deras ML, Brenowitz M, Ralston CY, Chance MR, Woodson SA. Folding mechanism of the Tetrahymena ribozyme P4-P6 domain. *Biochemistry* 2000;39:10975–10985. [PubMed: 10998234]
37. Uchida T, Takamoto K, He Q, Chance MR, Brenowitz M. Multiple monovalent ion-dependent pathways for the folding of the L-21 Tetrahymena thermophila ribozyme. *J Mol Biol* 2003;328:463–478. [PubMed: 12691754]
38. Laederach A, Shcherbakova I, Jonikas MA, Altman RB, Brenowitz M. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc Natl Acad Sci U S A* 2007;104:7045–7050. [PubMed: 17438287]
39. Shcherbakova I, Gupta S, Chance MR, Brenowitz M. Monovalent ion-mediated folding of the Tetrahymena thermophila ribozyme. *J Mol Biol* 2004;342:1431–1442. [PubMed: 15364572]
40. Koculi E, Hyeon C, Thirumalai D, Woodson SA. Charge density of divalent metal cations determines RNA stability. *J Am Chem Soc* 2007;129:2676–2682. [PubMed: 17295487]
41. Koculi E, Thirumalai D, Woodson SA. Counterion charge density determines the position and plasticity of RNA folding transition states. *J Mol Biol* 2006;359:446–454. [PubMed: 16626736]
42. Shcherbakova I, Mitra S, Beer RH, Brenowitz M. Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins. *Nucleic Acids Res* 2006;34:e48. [PubMed: 16582097]
43. Takamoto K, He Q, Morris S, Chance MR, Brenowitz M. Monovalent cations mediate formation of native tertiary structure of the Tetrahymena thermophila ribozyme. *Nat Struct Biol* 2002;9:928–933. [PubMed: 12434149]
44. Pan T, Fang X, Sosnick T. Pathway modulation, circular permutation and rapid RNA folding under kinetic control. *J Mol Biol* 1999;286:721–731. [PubMed: 10024446]
45. Qu X, Smith GJ, Lee KT, Sosnick TR, Pan T, Scherer NF. Single-molecule nonequilibrium periodic Mg^{2+} -concentration jump experiments reveal details of the early folding pathways of a large RNA. *Proc Natl Acad Sci U S A* 2008;105:6602–6607. [PubMed: 18448679]
46. Smith GJ, Lee KT, Qu X, Xie Z, Pesic J, Sosnick TR, Pan T, Scherer NF. A large collapsed-state RNA can exhibit simple exponential single-molecule dynamics. *J Mol Biol* 2008;378:941–951.
47. Kent O, Chaulk SG, MacMillan AM. Kinetic analysis of the M1 RNA folding pathway. *J Mol Biol* 2000;304:699–705. [PubMed: 11124019]
48. Fang XW, Thiagarajan P, Sosnick TR, Pan T. The rate-limiting step in the folding of a large ribozyme without kinetic traps. *Proc Natl Acad Sci U S A* 2002;99:8518–8523. [PubMed: 12084911]
49. Fang X, Pan T, Sosnick TR. A thermodynamic framework and cooperativity in the tertiary folding of a Mg^{2+} -dependent ribozyme. *Biochemistry* 1999;38:16840–16846. [PubMed: 10606517]
50. Fang X, Littrell K, Yang XJ, Henderson SJ, Siefert S, Thiagarajan P, Pan T, Sosnick TR. Mg^{2+} -dependent compaction and folding of yeast tRNA^{Phe} and the catalytic domain of the B. subtilis RNase P RNA determined by small-angle X-ray scattering. *Biochemistry* 2000;39:11107–11113. [PubMed: 10998249]
51. Fang XW, Pan T, Sosnick TR. Mg^{2+} -dependent folding of a large ribozyme without kinetic traps. *Nat Struct Biol* 1999;6:1091–1095. [PubMed: 10581546]
52. Baird NJ, Fang XW, Srividya N, Pan T, Sosnick TR. Folding of a universal ribozyme: the ribonuclease P RNA. *Q Rev Biophys* 2007;40:113–161. [PubMed: 17931443]
53. Adilakshmi T, Ramaswamy P, Woodson SA. Protein-independent folding pathway of the 16S rRNA 5' domain. *J Mol Biol* 2005;351:508–519. [PubMed: 16023137]

54. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 2002;3:2. [PubMed: 11869452]
55. Gutell RR, Lee JC, Cannone JJ. The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol* 2002;12:301–310. [PubMed: 12127448]
56. Russell R, Herschlag D. Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway. *J Mol Biol* 2001;308:839–851. [PubMed: 11352576]
57. Russell R, Das R, Suh H, Travers KJ, Laederach A, Engelhardt MA, Herschlag D. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* 2006;363:531–544. [PubMed: 16963081]
58. Russell R, Zhuang X, Babcock HP, Millett IS, Doniach S, Chu S, Herschlag D. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci U S A* 2002;99:155–160. [PubMed: 11756689]

Figure 1. Left panel: Folding of the *Tetrahymena* ribozyme *in vitro* is initiated by addition of Mg^{2+} . The initial ensemble of RNA molecules possesses native secondary structure and varying extents of backbone charge neutralization and compaction depending on the solution condition. Right panel: Upon addition of Mg^{2+} , the RNA molecules descend down a folding funnel such that the loss in entropy associated with reduced conformational freedom of the molecule is compensated by the negative enthalpy change arising from formation of the tertiary contacts. The color coding of the secondary and tertiary structures reflects regions of the RNA that fold at the same rate as shown by clustering of local time progress curves as described later in the text.

Figure 2.

Top left panel: Folding of the *Tetrahymena* ribozyme initiated by addition of 10 mM Mg^{2+} in a background of ~ 8 mM Na^+ ('low ionic strength condition'). Middle left panel: The kinetic model configuration resolved for the Mg^{2+} -induced folding of the *Tetrahymena* ribozyme at this solution condition. The structured regions of each intermediate are modeled from the hydroxyl radical reactivity patterns and rendered as bold ribbons, the unstructured regions can potentially acquire alternative conformations. The colored arrows illustrate the major folding pathways and their colors correspond to that of the histograms depicting the relative flux through the each of the parallel folding pathways (Lower left panel). The left hand panels are reproduced from [12] with permission from Elsevier. Right panel: Schematic representation of the Mg^{2+} -induced folding landscape of the *Tetrahymena* ribozyme at the low ionic strength condition, as inferred by the resolved kinetic model. The horizontal axes represent conformational space, and the vertical axis is energy. Three structurally distinct intermediates are observed - I1, I2 and I3 (models representing their respective structural features are shown below the funnel). Early barriers separating the unfolded states (U) from intermediates (I) are specific to folding under low salt conditions and are absent in studies conducted in the presence of higher background concentrations of monovalent cations. The direct pathway, $U4 \rightarrow F$, is minimally populated under this condition because of either a small initial population of U4 or a significantly higher barrier separating U4 and F (the pathway is shown by a dashed line). The

tertiary structure of a long-lived misfolded intermediate ('M') [56] is nearly identical to that of the native ribozyme by hydroxyl radical footprinting [57] and is therefore not included in our scheme.

Figure 3.

Folding of the *Tetrahymena* ribozyme initiated by addition of 1.5 M Na⁺ in a background of ~8 mM Na⁺ ('low ionic strength condition') organized as shown in Figure 2 for the Mg²⁺-mediated folding reaction. Note that the direct pathway U → F is appreciably populated in this reaction, the folding landscape is much smoother and the structures of the intermediates quite different from the Mg²⁺-mediated folding reaction. The left hand panels are reproduced from [12] with permission from Elsevier.

Figure 4.

Top left panel: Summary of the reaction schemes for Mg²⁺-mediated folding reactions conducted in the background of higher concentrations of the monovalent ions. Reactions A and B are folding experiments in which Mg²⁺ is added over a constant concentration of Na⁺ or K⁺. In Reaction C, Mg²⁺ and Na⁺ or K⁺ are added concomitantly to initiate folding. Bottom left panel: The kinetic model configuration resolved for the Mg²⁺-induced folding of the *Tetrahymena* ribozyme. The intermediate structures are modeled from the hydroxyl radical reactivity patterns as described in the legend to Figure 2. Each arrow represents a particular structural arrangement during folding and the thickness of each arrow is proportional to the resolved microscopic rate constant for that particular transformation. The arrows for the reverse reactions I1 → U and F → I1 reactions are omitted for clarity. Black arrows indicate that there is no significant difference in that particular rate measured at a different salt concentration. Gray-black arrows indicate differences in the resolved rates between the two salt concentrations indicated (grey and black represent, 200 and 600 mM Na⁺, respectively). The rates which are dramatically different for folding in the presence of K⁺ rather than Na⁺ are shown by black dashed arrows. The numbers accompanying the arrows indicate the resolved microscopic rate constants (s⁻¹) for the folding reaction conducted in the presence of 200 mM Na⁺. The left hand panels are reproduced from [38] Copyright 2007 National Academy of Sciences, U.S.A. Right panel: Schematic representation of the Mg²⁺-induced folding landscape of the *Tetrahymena* ribozyme conducted in the background of the 200 mM Na⁺ as inferred from the kinetic model. The axes are as described in the preceding figures. Depending on the initial state, the basins of the unfolded molecules can be located either higher or lower along the energy axis. Three structurally distinct intermediates were observed - I1, I2 and I3 (models

representing their respective structural features are shown below the funnel). The barrier separating U3 from I3 is characteristic for the folding in the presence of Na^+ and is diminished or absent for folding in the presence of K^+ . Adjustments in the initial conditions affect the relative distribution of molecular flux throughout the folding landscape by changing the relative abundance of different species in a heterogeneous mixture of RNA conformations.