The Story of RNA Folding, as Told in Epochs

Daniel Herschlag,[1,2,3,4] **Steve Bonilla,**[1,2] **and Namita Bisaria**[5]

[1]Department of Biochemistry, Stanford University, Stanford, California 94305

[2]Department of Chemical Engineering, Stanford University, Stanford, California 94305

[3]Department of Chemistry, Stanford University, Stanford, California 94305

[4]Stanford ChEM-H (Chemistry, Engineering, and Medicine for Human Health), Stanford, California 94305

[5]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142

Correspondence: herschla@stanford.edu

SUMMARY

The past decades have witnessed tremendous developments in our understanding of RNA biology. At the core of these advances have been studies aimed at discerning RNA structure and at understanding the forces that influence the RNA folding process. It is easy to take the present state of understanding for granted, but there is much to be learned by considering the path to our current understanding, which has been tortuous, with the birth and death of models, the adaptation of experimental tools originally developed for characterization of protein structure and catalysis, and the development of novel tools for probing RNA. In this review we tour the stages of RNA folding studies, considering them as "epochs" that can be generalized across scientific disciplines. These epochs span from the discovery of catalytic RNA, through biophysical insights into the putative primordial RNA World, to characterization of structured RNAs, the building and testing of models, and, finally, to the development of models with the potential to yield generalizable predictive and quantitative models for RNA conformational, thermodynamic, and kinetic behavior. We hope that this accounting will aid others as they navigate the many fascinating questions about RNA and its roles in biology, in the past, present, and future.
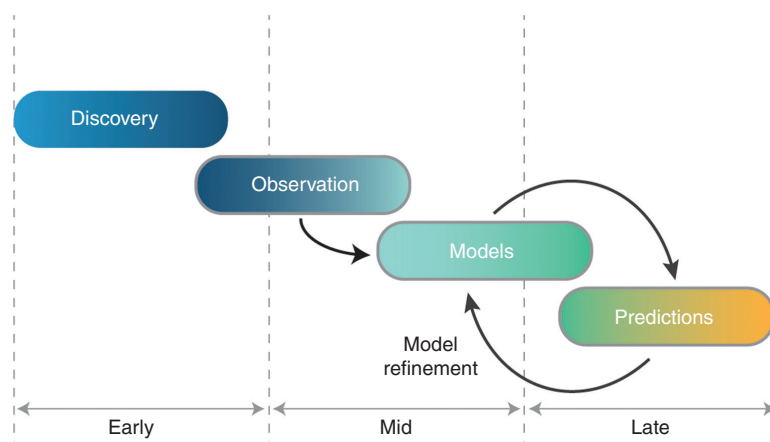
Outline

Figure 1. Schematic of progress in a scientific field in terms of "epochs."

1 INTRODUCTION

The reviews in this RNA World collection highlight new techniques and open questions in the field of RNA biology and emphasize how these technologies can generate new insights and further derive new questions. In this review, we break from this format as we cover a more mature scientific area, RNA folding, and we describe general insights drawn from RNA's molecular behavior and glean lessons from the study of the folding process that are broadly applicable to future studies across RNA biology. Just as geology and evolution can be categorized by epochs, so can the scientific process (Fig. 1). Enumerating the characteristics of each scientific epoch and designating such epochs within a particular scientific field can help inform individuals and communities when there is an opportunity to move toward the next stage.

This review takes the reader on a journey through the epochs in RNA folding research, with the goals of familiarizing the reader with the arc of this field's journey, conveying insights about RNA, technologies, and biology that have been gathered along the way, and laying out important lessons that will help us meet the current colossal challenge of understanding cellular RNAs, their structure and interactions, and their roles in the control of gene expression. As these scientific epochs (Fig. 1) are evidently general to the maturation of any scientific field, be it early astronomy (Crawford 1917) or enzymology (Copeland 2000), the case study provided here may be applicable to other aspects of biology.

Our journey through the epochs begins with observations of RNA's physical properties and the subsequent theories of an RNA World, progresses to a rapid discovery phase of uncovering and characterizing functional RNA structures, and closes with the field's recent entry into the phase in which these observations and principles can be codified into predictive models for designing functional RNAs for use in engineering and medicine.

2 THE RNA FOLDING EPOCHS: EARLY OBSERVATIONS AND THE RNA WORLD

Throughout biology, new traits and new species arise through evolutionary changes rooted in molecules and their interactions. Consequently, understanding these evolutionary solutions builds from an understanding of the properties and limitations of these molecules. The routes taken by evolution, by chance and guided by molecular properties, have led to—and are thus necessary to fully understand—modern-day biology. In this section, we focus on initial observations of the molecular properties of RNA that provided insights into why an RNA World may have been more probable than other options for starting life on Earth.

2.1 Why an RNA World?

The idea of an RNA World that predated life with proteins as the dominant catalysts was popularized by the discovery of catalytic RNA by Cech and by Altman and Pace (Kruger et al. 1982; Guerrier-Takada et al. 1983). But the idea arose more than a decade earlier, when Woese, Crick, and Orgel independently recognized that the "molecular properties" of RNA might endow it with the capabilities needed to support life (Woese 1967; Crick 1968; Orgel 1968). They noted that the ability of one RNA base to recognize another, A with U and G with C, could allow replication with some degree of specificity; the ability of RNA to form stable duplexes could allow templated synthesis of new RNA strands; and the more complex structure of transfer RNA (tRNA) resembled, to some extent, a folded protein and opened the possibility of RNAs as catalysts. Subsequently,

the realization that the ribosome's catalytic core is made of RNA led many to consider this RNA core as a vestige from an RNA World that developed an ability to make proteins and was subsequently superseded by proteins for most functions (Cech 2000; Steitz and Moore 2003; Noller 2012).

A consideration of the molecular properties of RNA provides a foundation for understanding RNA folding and, moreover, supports the idea of an RNA World and the subsequent advancement of proteins. RNA secondary structure is highly stable such that a 6-base-pair helix can be stable for days, in contrast to an α helix of the same length, which unfolds on the microsecond timescale (Narlikar and Herschlag 1997). The ability to form a stable structure with short oligomeric strands would have facilitated development of function under primitive conditions. Consider a folding free-energy profile for formation of a simple three-way junction (Fig. 2A). The formation of each helix is energetically downhill, so that a milieu with short RNA strands would have a reasonable chance to fold intramolecularly or come together intermolecularly. In contrast, proteins fold more cooperatively and generally require longer sequences for stable folding. Of course, most structured molecules do not have function, but RNA can catalyze strand fusion simply via a templated duplex (Joyce et al. 1984; Adamala and Szostak 2013). At a more sophisticated level, catalytic RNAs that act via tertiary structures bring together residues involved in catalysis at junctions like the one in Figure 2A. As any RNA residue can make hydrogen bonding and stacking interactions with any other RNA residue (Tinoco 1993), there is a wealth of possibilities to create a functional site. Indeed, in vitro evolution experi-

ments selected three-way junctions analogous to the one in Figure 2A that catalyze self-cleavage at a remarkably high probability (Salehi-Ashtiani and Szostak 2001). Selection experiments have also yielded a multitude of diverse RNA catalysts, and there has been impressive progress in developing an RNA-based RNA polymerase (Wilson and Szostak 1999; Horning and Joyce 2016).

RNA's stable secondary structure and promiscuous intermolecular interactions that are an asset for the development of life also present limitations, because RNA's stability creates an inherent tendency to also form alternative interactions and to be kinetically trapped in misfolded states (Fig. 2B, black line). RNA misfolding was first observed in the 1960s, by Sueoka and Fresco who independently obtained tRNAs that could not be charged by their respective aminoacyl synthetases until subject to denaturation and renaturation (Gartland and Sueoka 1966; Lindahl et al. 1966). Since then, those who work with RNA have made similar observations—observations that Uhlenbeck memorialized in his entertaining piece "Keeping RNA Happy" (Uhlenbeck 1995).

Because misfolding is an inherent property of RNA, modern biology too has to face this problem, and we reasoned that cells would require "RNA chaperones," proteins that facilitate unfolding and refolding of DNA either through simple binding or through ATP-driven conformational cycles (Fig. 2B, green line) (Herschlag 1995). Indeed, RNA chaperones are now known to be common cellular components, and so-called RNA helicases and dedicated RNA binding proteins serve to promote or mitigate RNA structure in every cellular process involving RNA, from precursor messenger RNA (pre-mRNA) splicing to mRNA
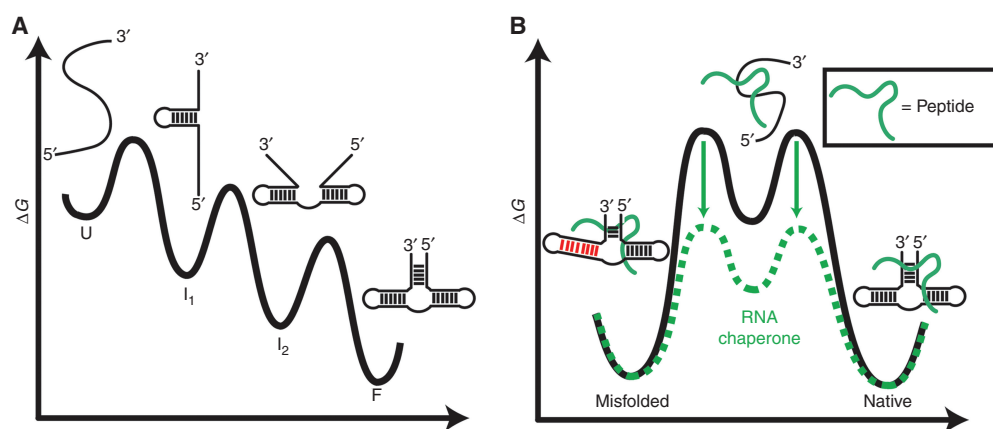
Figure 2. Propensity of RNA to form base-pair interactions and stable secondary structure. (*A*) Short RNAs can form stable structures as a result of high stability of base-pair interactions as shown for the formation of a three-way junction. (*B*) The high stability of secondary structure results in formation of stable misfolded states with alternative secondary structures. Short peptides (green) can act as RNA chaperones and may have aided RNA function in the RNA World and provided a path from the RNA World to the RNA–protein world (Herschlag 1995).

export, localization, translation, and decay (Russell 2008; Jarmoskaite and Russell 2014).

From the perspective of the RNA World, proteins acting as RNA chaperones may have endowed RNA with increased functionality, thereby providing a path out of the RNA World and first steps toward the protein world. The nucleocapsid protein from HIV was shown to increase the rate of conformational steps in hammerhead ribozyme catalysis, despite the fact that these molecules likely never coexist in the same biological context, and certain short peptides with positively charged side chains can also provide this function (Tsuchihashi et al. 1993; Herschlag et al. 1994). More recent work by Holliger and coworkers showed that small positively charged peptides enhance the function of in vitro selected RNA polymerase ribozymes (Tagami et al. 2017). As short peptides would be more readily synthesized by an early ribosome and as nonspecific interactions are more likely than the random occurrence of a peptide with specific recognition and function, it is reasonable that short, nonspecific peptides functioned as early RNA chaperones in an RNA World. The selective advantage from peptides would provide a driving force for evolution to improve the ribosome's accuracy and processivity in peptide synthesis, and with sufficient improvement and time, peptides and short proteins could have emerged that bound specifically to stabilize favorable RNA conformations and that provided one or more side chains that aided molecular recognition or catalysis (Fig. 2B) (Herschlag 1995).

But what would allow proteins to ultimately take over? Several ideas are popular and reasonable, including protein's favorable folding properties, relative to RNA, and the larger number of more catalytically effective side chains of proteins (Narlikar and Herschlag 1997). The diversity of protein side-chains facilitate folding, and the larger number of backbone degrees of freedom of RNA and its uniformly negatively charged backbone are impediments to folding and forming highly preorganized binding pockets and active sites. In addition, from the perspective of the "pathway" of evolution, it may have been easier to diversify the protein than RNA side-chain repertoire given RNA's role in information storage and transmission.

A recently recognized property that may have also contributed to the ultimate takeover by proteins is speed. A survey of association rate constants for ligands binding to RNA revealed that structured RNAs bind ligands orders of magnitude slower than the diffusional limit, whereas many proteins achieve association rates at or near this limit (Gleitsman et al. 2017). Given that slow association rates appear to be universal for RNA, the ability of proteins to more rapidly "grab" and use metabolites may have provided an additional driving force for the ascendancy of proteins. One of several physical mechanisms that may underlie

RNA's slow molecular recognition is its propensity to make stable local interactions that may be difficult to escape from, kinetically and thermodynamically. Considering the roles that RNAs currently carry out in biology, RNA's strong local interactions that allow it to effectively bind and discriminate RNA (and DNA) sequences and its ability to form multiple quasi-stable states may account for its aptitude for and remaining functional roles in processes such as pre-mRNA splicing, telomere maintenance, and RNA interference.

Orgel and Crick emphasized the "principle of continuity" in their evolutionary discussions (Crick 1968; Orgel 1968), meaning that one must be able to draw a stepwise evolutionary path from one biological state to another. Here, we suggest that RNA's stability and local structural promiscuity may have allowed it to be the basis of early life, and later, with these strengths also manifesting as limitations, provided an opportunity for proteins to enter and outgrow RNA's use due to the chemical diversity, faster dynamics, and enhanced molecular recognition of proteins.

3 MIDDLE RNA FOLDING EPOCH: CHARACTERIZATION, MODEL BUILDING, AND REVISION USING CATALYTIC RNAs

Once catalytic RNAs were identified, it was clear that they must contain sophisticated tertiary structures. This realization led to intensive efforts to determine their structures and to dissect their catalytic activities. There existed a synergy between structural and catalytic studies, as catalysis indicated that an active structure had been achieved, guiding crystallization trials and empowering folding studies. In the face of challenges to understand RNA structure and folding, new approaches and concepts were developed and adapted from other fields.

We begin by presenting studies at the start of the middle RNA folding epoch (Fig. 1) that characterized the folding process and that led to initial folding models. These models were tested and further refined, as we describe in the second subsection. These advances, and their limits, clarified the need for an ability to obtain an integrated kinetic, thermodynamic, and structural framework for folding of a model RNA. We describe such a framework for a model RNA in Section 3.3, and how its development helped propel RNA folding research to the final epoch focused on predictive quantitative models.

3.1 Characterization of RNA Folding and Early Models

Early structure mapping of the *Tetrahymena* group I intron by hydroxyl radicals generated by Fe(II)•EDTA in solution

identified protected and nonprotected groups and suggested that this RNA has an inside and outside, akin to a globular protein (Latham and Cech 1989). In addition, the lore that RNA is difficult to crystallize was challenged by several pioneering groups, ultimately leading to critical structural information from overall architecture down to the atomic level (Pley et al. 1994; Cate et al. 1996; Ferré-D'Amaré et al. 1998). Crystallographic studies of ribozymes emphasized active-site crevices and their apparent globular nature, akin to protein enzymes (Ferré-D'Amaré et al. 1998; Herschlag 1998; Kraut et al. 2003). The exception to this rule was the hammerhead ribozyme, but its crystal structure was shown to represent an inactive form that required a large-scale conformational change or an additional long-range tertiary interaction to form or stabilize the active conformation (Pley et al. 1994; Wang et al. 1999; Martick and Scott 2006; Nelson and Uhlenbeck 2006).

Returning to the *Tetrahymena* ribozyme, the development of time-resolved hydroxyl radical cleavage led to the identification of domains that folded at different times, suggesting the existence of multiple stable intermediates (Sclavi et al. 1998; Brenowitz et al. 2002). Evidence for folding intermediates also came from a time-resolved assay that used DNA oligos complementary to stretches of sequence throughout the ribozyme and RNase H to cleave regions that were exposed to form the DNA/RNA hybrids required for cleavage (Zarrinkar and Williamson 1994). Thus, this large ribozyme (~400 nt) did not show two-state folding like many small proteins. Other studies led to the discovery that the P4-P6 region of the ribozyme (Fig. 3A) is an independently folding, stable domain, and then to the P4-P6 X-ray structure (Fig. 3B) and its widespread use as a model in RNA folding studies (see Sec. 3.3).

As studies continued, evidence accumulated supporting models for kinetic traps in folding. In particular, the frequent observation of faster RNA folding on addition of urea and slower folding at higher concentrations of Mg^{2+} suggested the need to disrupt kinetically trapped species to rejoin a productive folding pathway (Pan and Sosnick 1997; Treiber and Williamson 1999; Woodson 2000). For the *Tetrahymena* ribozyme, a long-range helix—that is, one formed by strands distal in primary sequence—referred to as P3 ("paired" region 3; Fig. 3A) could be replaced by a local helix referred to as "altP3," and stabilization of altP3 enhanced formation of the kinetically trapped state or states (Pan and Woodson 1998; Russell et al. 2006). These observations were consistent with the high stability of RNA secondary structure leading to kinetically trapped states, as was emphasized in early tRNA folding studies.

Following the discovery of catalytic RNA, there was much discussion of the role of Mg^{2+} in RNA folding and function. One early model emerged from studies of the P4-P6 domain of the *Tetrahymena* ribozyme invoking nucleation of RNA structure around a core of site-specifically bound Mg^{2+} ions, analogous to a hydrophobic core in protein folding (Cate et al. 1997). This model arose from the structural observation that Mg^{2+} ions bound to a central component of P4-P6 that folds first and folds at lower Mg^{2+} than is needed to complete folding of P4-P6 (Celander and Cech 1991; Cate et al. 1996, 1997; Sclavi et al. 1998).

Emerging single-molecule fluorescence approaches could readily be applied to RNA, including early studies monitoring conformational changes of a four-way junction and monitoring the folding and catalysis by the *Tetrahymena* ribozyme (Zhuang et al. 2000; Tan et al. 2003). Subsequently, single-molecule fluorescence resonance energy transfer (smFRET) experiments were performed with numerous RNAs, including catalytic RNAs, later-discovered riboswitches (Breaker 2011), and RNA/protein complexes (Zhuang 2005). A general finding was that of "molecular individuality," meaning that different RNA molecules within the population displayed different kinetic, thermodynamic, or catalytic behavior (Zhao and Rueda 2013). Molecular individuality meant that RNA molecules did not equilibrate with one another—or swap behaviors—over timescales of minutes or hours. These observations, the preponderance of trapped folding intermediates, and the apparent ease of forming new trapped states led to a dominant model that RNA folding landscapes are "rugged" (Treiber and Williamson 1999; Ditzler et al. 2008).

3.2 Tests of RNA Folding Models in the Middle Epoch

As researchers gained control of model RNA systems and techniques matured there was an emphasis on testing emerging models. Below, we highlight some examples that helped further our understanding of the RNA folding process.

To test the model of nucleation of RNA structure around a Mg^{2+}-core, a mutant *Tetrahymena* ribozyme lacking the Mg^{2+}-binding domain (P5abc domain composed of helices P5a, P5b, and P5c; Fig. 3A) (Engelhardt et al. 2000) was generated and this truncated ribozyme was incubated in Mg^{2+}. Catalytic activity on addition of P5abc in *trans* arose faster than when folding was initiated by the simultaneous addition of Mg^{2+} and P5abc. Thus, the truncated ribozyme formed a folding intermediate that can bind P5abc and progress to its active form, faster than would occur if there were no intermediate or if unfolding of an off-pathway state were required (Russell and Herschlag 1999). Thus, Mg^{2+} core formation is not an obligate early folding step, and the ribozyme can fold along multiple pathways, minimally with P5abc folding early or late.
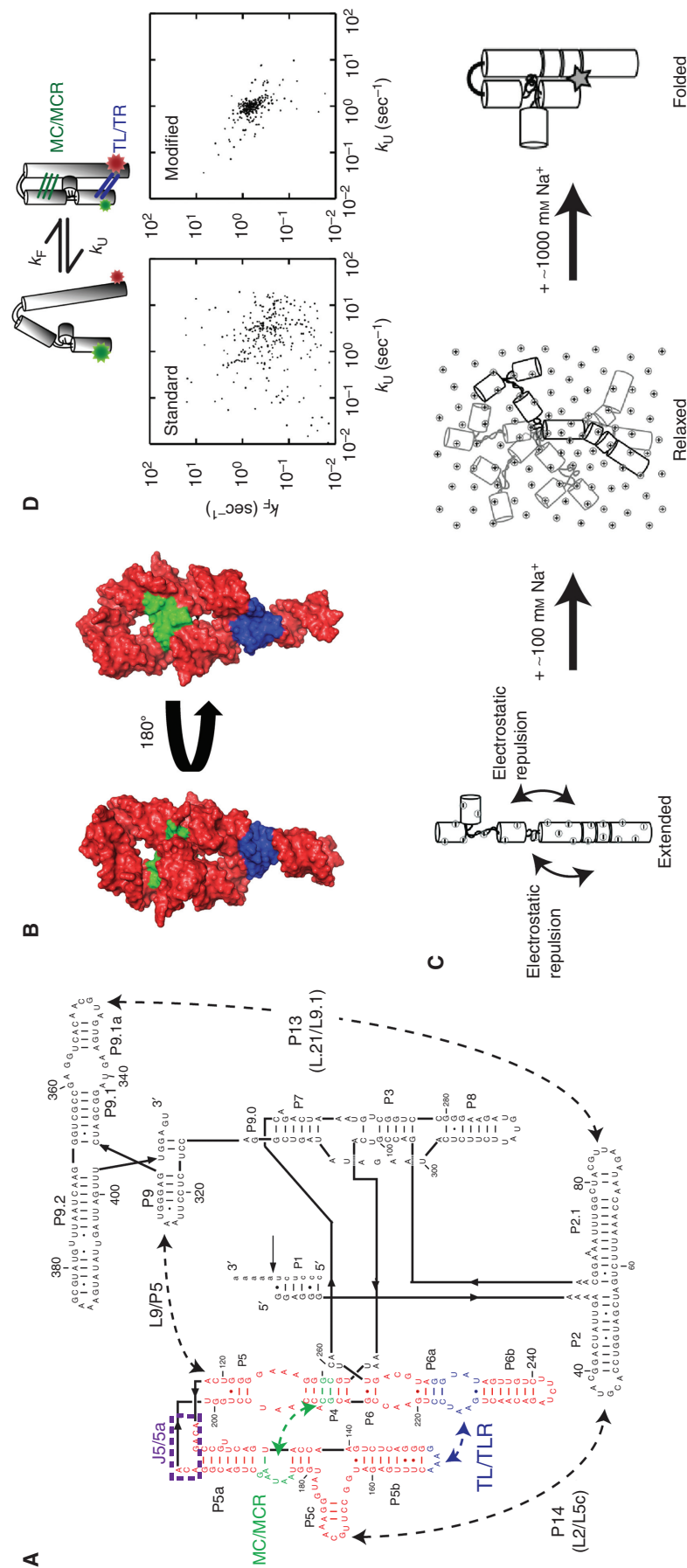
Figure 3. *Tetrahymena* group I intron ribozyme and its P4-P6 domain. (*A*) Secondary structure of the *Tetrahymena* group I intron ribozyme. Long-range contacts are marked with dashed arrows. The P4-P6 domain is highlighted in red, with its metal core/metal core receptor (MC/MCR) and tetraloop/tetraloop receptor (TL/TLR) tertiary contacts colored in green and blue, respectively. Junction J5/5a (purple box) bends and allows coaxial stacks to be roughly parallel to each other. (*B*) X-ray structure of the P4-P6 domain with tertiary contacts colored as in *A* (PDB 1GID; Cate et al. 1996). (*C*) Electrostatic relaxation provided by cations followed by stabilization of a compact state by the formation of tertiary contacts. (*D*) Heterogeneous kinetic behavior of P4-P6 RNA caused by covalent damage during heating and ultraviolet (UV) irradiation. P4-P6 is shown in cartoon form with a pair of donor–acceptor fluorescent dyes (green and red stars) attached for single-molecule fluorescence resonance energy transfer (smFRET) folding studies. Scatter plots show values of the folding and unfolding rate constants (k_F and k_U, respectively) for individual P4-P6 RNA molecules prepared by a standard protocol (*left*) that involves heat/cool renaturation and UV irradiation and by a modified protocol (*right*) in which heating and UV radiation were eliminated during construct preparation and purification. (*C*, from Takamoto et al. 2004; reprinted, with permission, from Elsevier. *D*, from Greenfeld et al. 2011, adapted, with permission, from The American Society for Biochemistry and Molecular Biology.)

The ability to carry out time-resolved small-angle X-ray scattering (SAXS) allowed monitoring of the overall shape of an RNA through early steps in the folding process (Russell et al. 2000). On addition of Mg^{2+}, the *Tetrahymena* ribozyme compacted rapidly (Russell et al. 2002). Compaction occurred well before time-resolved hydroxyl radical probing experiments showed evidence for formation of the first tertiary interactions, and this initial compaction phase was unaffected by ablation of the five long-range tertiary contacts of the *Tetrahymena* ribozyme. Thus, the initial compaction step was distinct from tertiary structure formation, whereas a second compaction step was eliminated for the tertiary-ablated construct (Das et al. 2003). These results revealed that Mg^{2+} is involved in more than one folding step.

To better understand the nature of these folding steps, conformational modeling of SAXS data, in conjunction with hydroxyl radical probing, was performed with simpler model systems (Takamoto et al. 2004). Under low-salt conditions, RNAs tend to adopt extended states, attributable to electrostatic repulsion between the negatively charged phosphates of the RNA backbone (Fig. 3C). Addition of any divalent cation causes compaction, not due to attractive forces but rather from lessening the repulsion that favors extended states and allowing exploration of the larger number of nonextended conformers. This process has been referred to as "electrostatic relaxation" (Das et al. 2003; Takamoto et al. 2004; Lipfert et al. 2014). Monovalent cations also promote electrostatic relaxation, but at far higher concentrations, consistent with expectations from polyelectrolyte theory (Draper et al. 2005; Lipfert et al. 2014).

Specific local metal-ion binding to RNA is of course also observed and is critical for many of RNA's functions. For the P4-P6 domain, its X-ray structure showed five "bound" or localized Mg^{2+} ions (Cate et al. 1996). To determine which of these were directly coupled to folding, Mg^{2+}-promoted folding studies of P4-P6 were performed in the presence of a high concentration of background monovalent cation (Das et al. 2005). The large number of monovalent cations screened nonspecific electrostatic interactions such that the smaller number of available Mg^{2+} would bind only to specific sites. The number of bound Mg^{2+} could be determined by "ion counting," a technique in which associated and free ions are distinguished and quantified by atomic emission spectroscopy or mass spectroscopy. The extent of folding as a function of Mg^{2+} concentration could be followed across the same conditions by hydroxyl radical footprinting. Ion counting revealed the binding of two Mg^{2+} upon folding, and footprinting gave a Hill slope of 2, showing a direct link between the physical association of Mg^{2+} and the thermodynamics of folding. Subsequent experiments using P4-P6 variants with phosphorothioate substi-

tutions to site-specifically alter Mg^{2+} versus Mn^{2+} affinities mapped the identity of the Mg^{2+} sites and further developed thermodynamic models for Mg^{2+} binding (Frederiksen et al. 2012). This approach was borrowed from extensive ribozyme studies that used thio-substitutions to identify and dissect roles for metal ions in catalysis (Piccirilli et al. 1993; Weinstein et al. 1997).

The observation of molecular individuality from smFRET was the subject of much discussion and further tests. Experiments using the hairpin ribozyme showed that this behavior persisted whether molecules were tethered to the surface of a microscope slide or localized in micelles, suggesting that surface effects were not responsible for the observed heterogeneity (Okumus et al. 2004). Persistent molecular behavior over many minutes or even hours and upon denaturation and renaturation could result from deep energetic barriers between the differentially behaving conformations or could result from covalent differences between the molecules caused by synthesis errors or damage. Bulk approaches were used to assess whether covalent heterogeneity was present in populations of P4-P6 RNA molecules that gave persistent heterogeneous behavior in smFRET studies (Fig. 3D) (Greenfeld et al. 2011). Ultraviolet (UV) shadowing used in gel purification and heating used in RNA renaturation were shown to covalently modify the RNA and interfere with folding. A protocol to purify and prepare P4-P6 RNA for smFRET studies without these treatments greatly reduced heterogeneity observed in single-molecule studies close to noise levels (Fig. 3D; compare left and right scatter plots). Evidence for molecular individuality via conversion between states with distinct kinetics and thermodynamics has been shown only for the more complex *Tetrahymena* group I ribozyme, and even for this system interconversion was not complete, suggesting that a portion of the heterogeneous behavior of this molecule may also arise from covalent differences (Solomatin et al. 2010).

Early results, including those described above, identified the presence of and properties of multiple folding intermediates. The next challenge was to determine the order of these species. Russell and colleagues brought to bear multiple approaches to dissect the folding pathways of the *Tetrahymena* ribozyme: smFRET, to provide the time to the native state (N) and the fraction of molecules folding to N versus those folding to a long-lived misfolded state (M); SAXS, to provide information about overall compactness and shape; chemical probing with dimethyl sulfate (DMS), to provide local information about initial and final structural states; and mutations, to stabilize or destabilize known interactions. Depending on the starting monovalent salt conditions, ribozyme molecules preferentially folded along one of three pathways. The pathways are described as "channels" because once a molecule starts down one of the path-

ways it does not switch to a different pathway; in other words, the barriers between pathways are larger than those along the pathways (Russell et al. 2002). Such channels were also observed for RNase P folding (Pan et al. 1999).

With this information, conditions could be found to populate individual states and probe each to better understand the nature of the intermediates. The most mysterious state was M, as it was shown to have nearly the same overall shape, form the same set of long-range tertiary interactions, and have a nearly identical hydroxyl radical protection pattern as N (the active ribozyme form), yet interconversion between M and N took hours (Russell et al. 2006). Structural mapping and kinetic studies led to the model that M represented a topological isomer of N in which the strands in the core are crossed, relative to N. Subsequent experiments provided strong support for this model and evidence that the transition from M to N was slow because it required the peripheral, long-range tertiary interactions to be broken to allow the strands to disentangle and come together in the correct order (Russell et al. 2006; Wan et al. 2010; Mitchell et al. 2013).

3.3 The Value of an Integrated RNA Folding Framework

The results summarized above for the *Tetrahymena* ribozyme revealed complexities and details of its folding landscape and generalizable lessons. Yet, there is more missing even for this advanced model RNA than is known—for example, we do not know the order of formation of the five long-range tertiary contacts nor the quantitative barrier steps along the folding pathways, and structural information is limited to a few of the many states. In an effort to integrate structural, kinetic, and thermodynamic information, and to provide a more comprehensive and quantitative description of a folding process, we turned to the simple P4-P6 domain (Fig. 3).

P4-P6 RNA folds at a junction (J5/5a; purple box in Fig. 3A) that allows two sets of coaxially stacked helices to be placed side-by-side, as occurs in more complex RNAs, and this arrangement is stabilized by two tertiary contacts between these stacks (Fig. 3B). The discrete nature of the two P4-P6 tertiary contacts allowed each to be effectively ablated, without secondary effects. One of the tertiary interactions is a canonical tetraloop/tetraloop receptor (TL/TLR) motif, and the other is a metal core/metal core receptor (MC/MCR) (Fig. 3B, blue and green, respectively).

Its simplicity and the wealth of prior structural information rendered P4-P6 RNA an attractive target for the determination of an integrated folding framework. The availability of an X-ray structure of fully folded P4-P6 in combination with extensive chemical probing data allowed

us to propose a set of structural intermediates on path to the folded state (Fig. 4). Additional intermediates with the P5abc subdomain in an alternative conformation was suggested by nuclear magnetic resonance (NMR), chemical probing, and mutagenesis of an isolated P5abc construct (Wu and Tinoco 1998; Silverman et al. 1999; Xue et al. 2016). Although the overall folding rate was known (Fig. 4, inset), it was not known which step or steps were responsible for the observed folding rate constants. To quantitatively dissect the kinetics and thermodynamics of the individual steps in P4-P6 folding, a set of P4-P6 mutants was generated and their folding properties characterized by smFRET under a variety of solution conditions.

Figure 4 summarizes the folding framework of the P4-P6 domain (Bisaria et al. 2016). Under the conditions studied, the native conformation of the P5abc domain dominates over its alternative conformation, the two critical Mg^{2+} ions bind rapidly and are retained in the thermodynamically favored unfolded state. Following Mg^{2+} binding, the MC/MCR forms first, followed by very rapid formation of the TL/TLR. Superficially, the formation of MC/MCR preceding formation of TL/TLR makes sense as it follows "contact order"—that is, the MC and MCR tertiary contact partners are closer to the J5/5a junction (Fig. 3A) compared with the TL and TLR. However, the scenario is not so simple. P4-P6 folds with ∼90% of the RNA molecules forming MC/MCR first, and ∼10% forming the more remote contact first, so the contact order preference is rather subtle. Further, a mutation in the MC tertiary partner that weakens the MC/MCR tertiary contact and does so in part by slowing the formation of the MC/MCR inverts the preferred folding pathway, such that the TL/TLR tertiary contact forms first ∼90% of the time. Thus, the pathways and choice of pathways are influenced by collision frequencies and also by the properties of the individual tertiary elements. As we describe in the following section, a "reconstitution model" that builds from the properties of RNA structural components can account for, and ultimately predict, these preferences.

3.4 A Rugged Landscape Model for RNA Folding?

As we close this epoch in RNA folding studies, we are positioned to reconsider the model of a rugged folding landscape, which had provided a useful description of early observations. We discuss this model as it has become so prevalent that it has become rare to hear "RNA folding" without reference to its "rugged landscape" and because we can draw lessons from it for understanding the transitions between scientific epochs.

Analogy models, like the rugged landscape, can be beguiling, prompting their acceptance even in the absence of
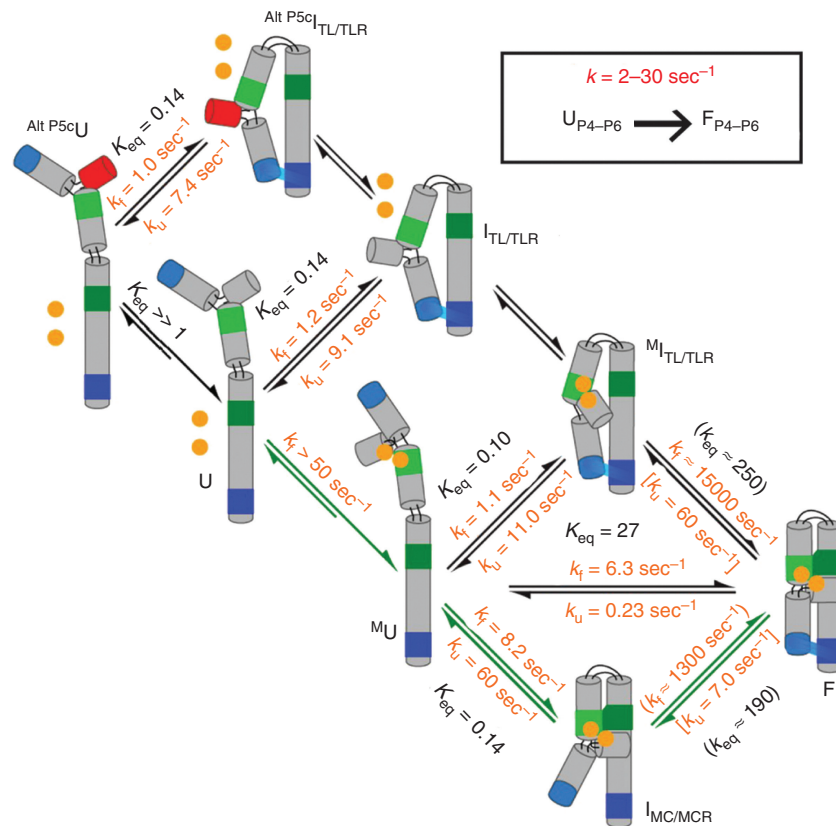
Figure 4. Kinetic and thermodynamic framework of the P4-P6 domain of the *Tetrahymena* group I intron (100 mM KCl, 5 mM MgCl$_2$, pH 8, 23°C). The *inset* shows the overall folding rate constants observed in several studies of overall folding (for references, see Bisaria et al. 2016). (From Bisaria et al. 2016, modified, with permission.)

defined predictions and tests. Indeed, it is not clear what the rugged landscape model predicts. A rugged landscape is consistent with the observation of kinetic traps. But how many, and how deep? We have a concept of rugged hills or mountains, landscape, but this ruggedness is irrelevant to a bird, and what looks smooth to us seems rugged to a smaller mammal or to an insect.

Interestingly, recent results suggest that RNA dynamics may act to smooth the RNA folding landscape. The P5abc subdomain of P4-P6 undergoes a cooperative transition involving a change in secondary structure coupled to tertiary structure formation (Wu and Tinoco 1998; Silverman et al. 1999; Koculi et al. 2012). Previously, it was thought that this transition is "fully" cooperative, such that one rearrangement could not happen without the other. In reality, there must be a limit to this cooperativity, as energetic barriers are finite. Indeed, recent NMR chemical shift exchange experiments have suggested that P5abc occupies a series of local and rare states that are favorable for tertiary structure formation, essentially breaking down a large-scale conformational process into smaller steps that can occur independently and with only modest kinetic and thermo-

dynamic barriers (Xue et al. 2016). This model is supported by recent work that identified folding modules within P5abc that each contribute modestly to an overall cooperativity of ~4 kcal/mol (Gracia et al. 2018). Thus, structural modules within a complex RNA "experiment with" different subconformational states, which smooths the energy landscape and speeds folding transitions.

More broadly, analogies are useful, but only to a point. They help us organize information and think creatively, but analogies cannot substitute for molecular understanding and predictive models.

4 LATE RNA FOLDING EPOCH: TRANSITION TO PREDICTIVE STRUCTURAL AND ENERGETIC MODELS

In science teaching, there is an emphasis on experiments that provide distinct qualitative tests to distinguish between models, such as the molecular nature of genetic material or the semiconservative nature of DNA replication. Although these examples are truly elegant and inspiring, perhaps we underestimate the value of and need for quantitative models

and tests. Quantitative predictive models are readily tested, are needed to describe how biological processes work, allow strongly discriminating tests of models, and can reveal missing features or factors in a process.

Although we would like a quantitative model that predicts RNA three-dimensional structure from sequence, this would not be enough. RNA molecules exist as dynamic ensembles, not as single static single conformations, and their dynamic properties dictate their function in complex and essential multistep processes. Thus, in addition to structural models we need quantitative models of RNA dynamics and energetics that can be used to predict relative stabilities of different states and the transition probabilities between them. It is only through a combination of structural, dynamic, and thermodynamic models that we will be able to predict the wealth of complex interactions and processes involving structured RNAs inside the cell.

4.1 RNA Structure Prediction and Design

We emphasized above lessons from the older and larger field of protein folding that were borrowed and used by the RNA community. For structure prediction, the central lesson was that the ability to make correct predictions cannot be determined without blind trials. Before such trials, solution to the protein folding problem seemed to be right around the corner as several approaches appeared to produce accurate "predictions." However, the limitations of these approaches became apparent through CASP (critical assessment of protein structure prediction) in which modelers were given time to make predictions before release of newly solved structures (Moult et al. 2014). The lesson from this history is that unknown sources of variation in complex problems, unconscious biases, and publication bias provide limitations such that it is insufficient to test models with "retrodictions"; instead, it is only through true predictions, as in CASP, that our understanding of a process can be tested.

Remarkably, structures for many proteins of modest size can be predicted in blind trials using ROSETTA, which uses energetic potentials based largely on empirical information from the large database of protein crystallographic structures (Das and Baker 2007; Kaufmann et al. 2010). Das and colleagues adapted ROSETTA for use in RNA structure prediction, and further integrated information from chemical structure mapping (Cheng et al. 2015). In the RNA-Puzzles blind trials, they have been able to predict overall RNA tertiary structure, in favorable cases to near-atomic resolution (Miao et al. 2017).

Before successful atomic-scale modeling, the presence of recurring RNA motifs was recognized and allowed modeling of RNA structures and engineering of new structures

of modest complexity (Westhof et al. 1996; Jaeger and Leontis 2000). Now, similarly complex structured RNAs can be designed in an automated fashion from a diverse set of helix, junction, and tertiary motif elements (Yesselman et al. 2017).

4.2 Developing a Reconstitution Model for RNA Folding

An early goal in protein folding research was to obtain an additive accounting of energetics. Occasional additive contributions to stability were observed, general principles were established, and the conversion of these rules to approximate energy terms combined with empirical information from the protein data bank (PDB) has led to successes in protein structure prediction (Kaufmann et al. 2010; Huang et al. 2016). However, the larger goal of predicting protein folding energetics has proven elusive, apparently because of the many interacting groups within folded proteins that lead to a vast number of idiosyncratic energetic terms. In contrast, RNA's apparently simpler structural and energetic properties have allowed the development of a "reconstitution model" for RNA folding kinetics and thermodynamics from its component parts (Fig. 5A).

Three simplifying features of folded RNAs are foundational for the "reconstitution model."

1. RNA folding is hierarchical. RNA secondary structure is stable in the absence of tertiary contacts, in contrast to the case with proteins, allowing energy terms for tertiary interactions to be considered distinct from those for helix formation (Fig. 5B). Even in cases with alternative secondary structures (e.g., riboswitches), tertiary folding can be considered in the context of each of the limited number of secondary structure alternatives.

2. Tertiary contacts are sparsely distributed. Folded RNAs appear globular, like proteins; however, closer inspection reveals that most RNA residues interact only with their primary or secondary structure neighbors. Unlike proteins, very few residues are involved in tertiary interactions. This can be seen in the P4-P6 structure, where the side-by-side stacked helices come together only at two tertiary contact interfaces (Fig. 3B).

3. Structured RNAs are built from a limited set of structural motifs. Phylogenetic analysis and structural comparisons revealed the existence of recurring tertiary motifs in RNA (Costa and Michel 1995; Leontis et al. 2006) and Figure 5C shows one example. The TL/TLR in P4-P6 RNA described above is found in several other RNAs. These TL/TLRs, despite being embedded in distinct RNAs, closely superimpose. This structural modularity
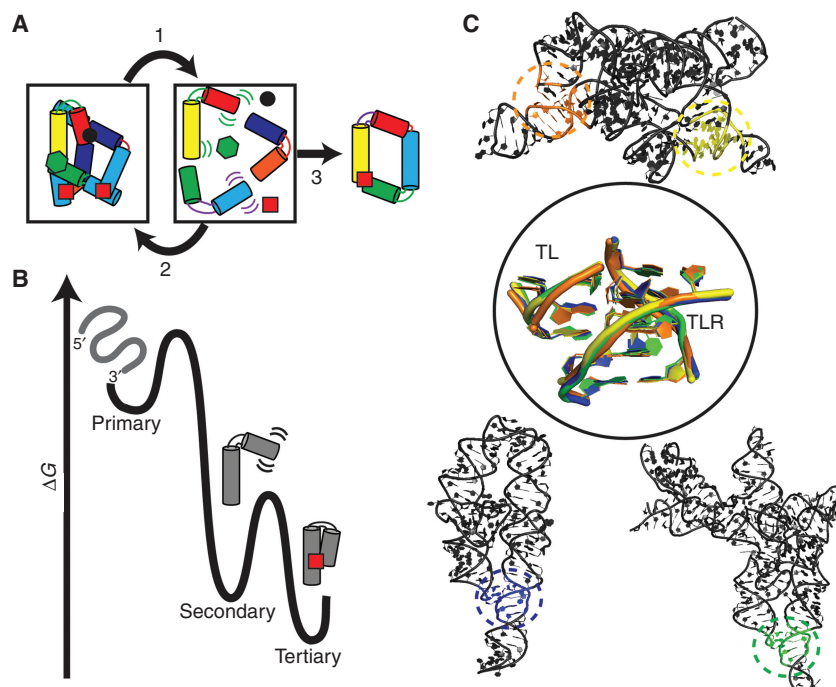
Figure 5. Hierarchical folding, sparsity of tertiary contacts, and structural modularity of RNA lead to a "reconstitution model" for RNA folding. (*A*) Schematic representation of a "reconstitution model" of RNA folding: (*1*) A complex RNA can be broken down into helix-junction-helix elements (HJH; cylinders and strings) and tertiary contact motifs (circles, squares, hexagons), and the energetic and conformational properties of these elements studied in isolation; (*2*) the behavior of the isolated elements can be reconstituted to understand their collective behavior in natural structured RNAs; and (*3*) novel RNAs can be constructed with structural motifs and can be engineered to have tunable energetic and conformational properties. (*B*) Free-energy diagram representing hierarchical folding of RNA. As an example, the folding of a simple RNA stabilized by a single tertiary contact is shown as a cartoon above the free-energy diagram. Helices are represented by cylinders, junctions as strings connecting the helices, and the single tertiary contact as a red square. (*C*) X-ray structures of the GAAA/11ntR tetraloop/tetraloop-receptor (TL/TLR) tertiary contact motif extracted from diverse structured RNAs and superimposed P4-P6 (PDB 1G1D; blue), RNase P (PDB 1NBS; green), and the *Azoarcus* group I intron (PDB 1ZZN; yellow and orange).

suggests an "energetic" modularity that could be foundational for a "reconstitution model."

4.3 The Reconstitution Model

Consider two tertiary contact partners separated by a secondary structure. Figure 6A shows a simple schematic example with two helices connected by a single junction, referred to as a helix-junction-helix (HJH) element, and with each helix containing a tertiary contact partner. The probability of forming the folded structure (F) is a function of two factors: (1) the probability that the two tertiary contact partners will align for formation of the tertiary contact (ΔG_{Align}) and (2) the strength of the tertiary interactions (ΔG_{Tert}; Fig. 6A).

The unfolded RNA (U) exists as an ensemble of conformations and as a narrower ensemble in the folded state (F), and the probability of folding is a function of these conformational spaces (Fig. 6B). For example, a more rigid

junction (e.g., one that prefers stacked states) will be less likely to bend and thus less likely to form the folded state, whereas a junction that has internal interactions that favor a bent state will be more likely to fold, provided that the bent state overlaps with aligned states that allow the tertiary interaction to form. Experimentally, NMR residual dipolar coupling (RDC) and X-ray scattering interference can be used to determine RNA HJH conformational ensembles (Salmon et al. 2014; Shi et al. 2015).

Junctions play an obvious role in shaping conformational ensembles, but the conformational properties of helices also matter. For example, changing helix length rotates and translates the positions of the tertiary contact partners in relation to each other altering the probability of tertiary contact formation. Also, the formation of tertiary contacts between helices connected by floppier junctions will be less sensitive to changes in the conformational properties (e.g., length) of the helices as the flexibility of the junctions can "accommodate" misalignments more easily compared with
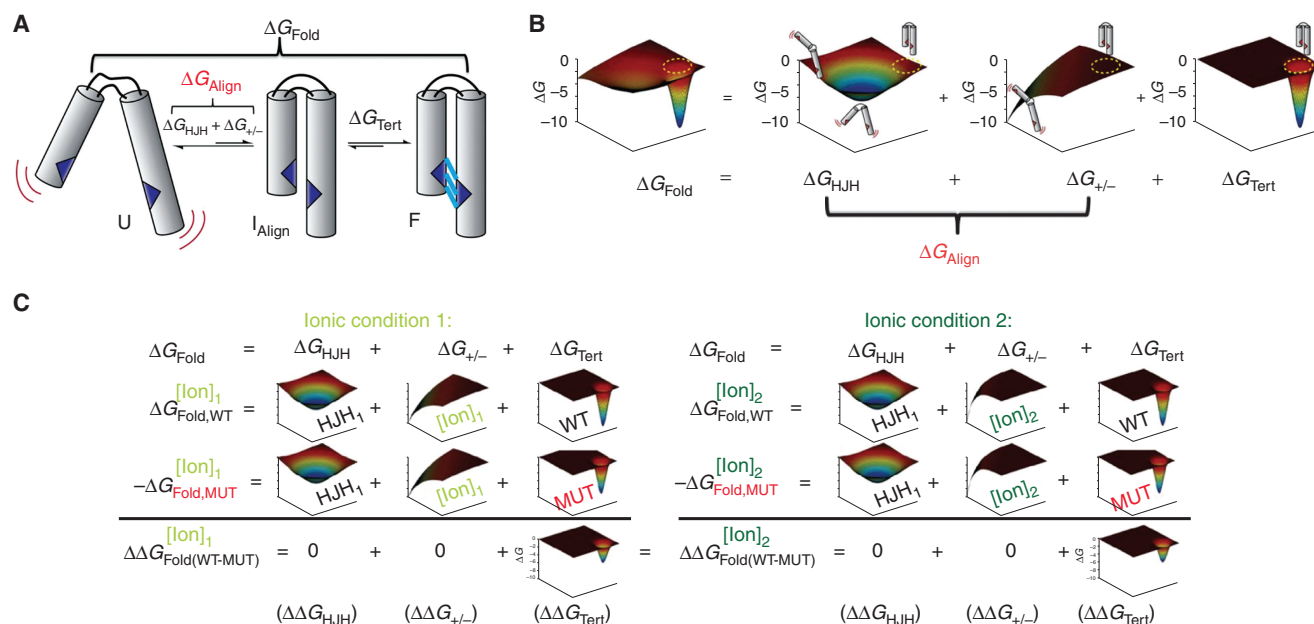
Figure 6. Schematic of a "reconstitution model" of RNA folding energetics. (*A*) Folding of a simple RNA involving the formation of a single tertiary contact (blue triangles) between two helices (cylinders) connected by a junction (strings). The free energy of folding (ΔG_{Fold}) is broken down into two terms: (1) the probability that the tertiary contact partners are found aligned for productive collision (ΔG_{Align}) and (2) the free energy associated with the formation of the tertiary contact (ΔG_{Tert}). (*B*) Each of the energy terms in *A* is represented in 3D free-energy landscapes. ΔG_{Align} can be further broken down into an energy term intrinsic to the conformational preferences of the junction and the helices containing the tertiary contact partners (ΔG_{HJH}) and an energy term reflecting the electrostatic interactions between RNA and surrounding ions and between RNA helices ($\Delta G_{+/-}$). (*C*) Quantitative prediction of the "reconstitution model" of RNA. According to this model, the free energy of formation of a tertiary contact ΔG_{Tert} is separable from the free energies associated with the conformational search of the junction ΔG_{Tert} and the electrostatic interactions with ions $\Delta G_{+/-}$. Hence, this model predicts that the energetic effect of mutating a tertiary contact motif is constant across different ionic conditions. (From Bisaria et al. 2017, reprinted, with permission.)

rigid junctions. These effects have been illustrated using models with more mobile versus more constrained junctions (Chu et al. 2009).

Of course, the tertiary contact motif also matters for folding, most obviously because the folded state, F, is highly unlikely to be populated without stabilization from a tertiary interaction. In the "reconstitution model," the contribution of the tertiary motif to the stabilization of F is represented explicitly by the ΔG_{Tert} term. Thus, the more tertiary hydrogen bonds and stacking interactions between the tertiary contact partners, the more favorable this energy term will be; moreover, different tertiary motifs will also have distinct conformational preferences. Thus, an RNA containing HJH1 may be more stable with tertiary motif A than tertiary motif B, whereas an RNA containing a different HJH (HJH2) may instead be more stable with tertiary motif B than tertiary motif A, because of better matching of the conformational preferences of HJH2 and tertiary motif B.

Considering RNA secondary structure, we can predict duplex stability with reasonable accuracy from a simple

"energetically additive" model that accounts for the identity of each base-pair and its nearest neighbors (Zuker et al. 1999). In contrast, the above considerations indicate that we do not expect this same simple energetic additivity for RNA tertiary structure—we cannot assign universal free-energy values to each structural element and add them to obtain the overall tertiary folding stability. Instead, we need to "add," or convolve, conformational ensembles of the individual elements—that is, the junctions, helices, and tertiary contact motifs present. The "reconstitution model" holds that there is "ensemble additivity" such that a convolution of ensembles determined for isolated RNA elements can describe the conformational ensemble of a complex RNA (after accounting for steric overlaps), and that this ensemble, along with a constant free-energy term indicative of the strength of each particular tertiary contact motif, can be used to "predict" the folding stability for an arbitrary RNA. Correspondingly, the "reconstitution model" readily handles complex scenarios with multiple HJHs and multiple tertiary contacts by combining multiple HJH ensembles

with each tertiary interaction specified as formed or not to model each folding step.

4.4 Testing the Reconstitution Model

Free-energy landscapes are added in the "reconstitution model"; as free energy is directly convertible to probabilities, we can also think of this as combining (multiplying) probability landscapes (Fig. 6B). According to the "reconstitution model," a perturbation in one of the constituent energy landscapes should not alter the others. Thus, a mutation that alters the intrinsic stability of a tertiary contact motif (ΔG_{Tert}) is predicted to have no effect on the energetic consequences from changing the HJH element (ΔG_{HJH}) or from changing the ionic conditions ($\Delta G_{+/-}$). Note that the alignment term discussed above (i.e., $\Delta G_{\mathrm{Align}}$) is broken into two terms (Fig. 6), one representing the conformational preferences of the particular junction sequence (ΔG_{HJH}) and another related to the electrostatic interactions between RNA elements and surrounding ions ($\Delta G_{+/-}$).

These properties lead to predictions that were tested using smFRET (Bisaria et al. 2017). For example, the folding energetics for the P4-P6 domain with the WT TL/TLR tertiary contact motif and a mutant TL/TLR were measured under two different ionic conditions (Fig. 6C). The "reconstitution model" predicts that the overall effect of the mutation on folding ($\Delta\Delta G_{\mathrm{Fold}}$) is the same at each ionic condition, and this predicted result was observed across several ionic conditions and held for several different mutations (Bisaria et al. 2017). Analogously, these mutations to the tertiary motif gave the same energetic effect ($\Delta\Delta G_{\mathrm{Fold}}$) in different HJH contexts—that is, embedded in different RNAs (Bisaria et al. 2017). These results showed the separability of terms foundational to the "reconstitution model."

We have emphasized thermodynamic applications of the "reconstitution model," but it is equally applicable to folding kinetics, as folding kinetics can similarly be described in terms of free-energy landscapes. Indeed, the first evidence suggestive of this model was obtained from kinetic folding data (Bisaria et al. 2016). The folding framework for P4-P6 RNA (Fig. 4) intriguingly revealed the same unfolding rate constant, within error, for all steps in which the TL/TLR tertiary contact motif dissociated. Previously, Nesbitt and colleagues had extensively studied this same tertiary motif in isolation, by embedding the TL and TLR into a simple construct in which the TL and the TLR are connected by a A_7 tether (Hodak et al. 2005); indeed, this construct was used by Bisaria and colleagues in the thermodynamic tests of the "reconstitution model" described above. Remarkably, the unfolding of this simple construct had the same rate constant as those in P4-P6, and this was confirmed across a range of conditions (Bisaria et al. 2016,

2017). These observations suggested that the unfolding rate for a given tertiary motif is an intrinsic property of that motif. In contrast, folding rate constants vary considerably, as expected for the "reconstitution model" and described immediately below.

The kinetic version of the "reconstitution model" can be described as an RNA diffusion-collision model, a model borrowed from protein folding but likely more apt for RNA folding giving the cleaner separation between secondary and tertiary structure formation (Bisaria et al. 2016). As emphasized above, unfolding rate constants are local properties of the tertiary motif involved and thus are transferable across RNAs, for a given set of conditions. In contrast, the rate constants for each folding step are determined by the probability of the tertiary elements encountering one another, which is determined by the free-energy landscapes of the HJH elements and the ionic conditions, and by the probability that, on collision, the tertiary motif will form—that is, the probability of a productive collision.

We noted earlier in this review that RNA association rate constants are considerably slower than diffusion (Gleitsman et al. 2017). There is considerable evidence for conformational changes between free and bound RNA tertiary elements, including the canonical 11ntR TL/TLR highlighted herein (Fig. 7A) (Butcher et al. 1997; Gleitsman et al. 2017). Bonilla and colleagues used smFRET to dissect the folding behavior of this tertiary contact motif along with a set of TL/TLR variants under a wide range of solution conditions. These results in combination with structural characterization of the unbound canonical 11ntR (Butcher et al. 1997) and smFRET studies of the canonical GAAA/11ntR TL/TLR (Fiore et al. 2012) led to a model for the order of conformational changes along the folding pathway and the nature of the transition state for TL/TLR formation (Fig. 7B) (Bonilla et al. 2017). This study also identified an additional free-energy term to be incorporated into the "reconstitution model" ($\Delta G_{\mathrm{Tert},+/-}$), as the formation of some TL/TLR variants involve additional electrostatic and metal ion effects, an observation consistent with the considerable rearrangements observed on binding. Nevertheless, the electrostatic effects specific to a TL/TLR variant were conserved between different HJH contexts and is therefore intrinsic to the TL/TLR variant (Bonilla et al. 2017), consistent with the "reconstitution model."

In summary, there is strong evidence for the basic features of the "reconstitution model," a major step forward because this model is, in principle, predictive. Nevertheless, there are complexities, as noted by the specific ionic effects on certain tertiary contacts and as presented by the large number of RNA structural elements that require quantitative characterization to broadly apply the "reconstitution model." The following section briefly outlines a new
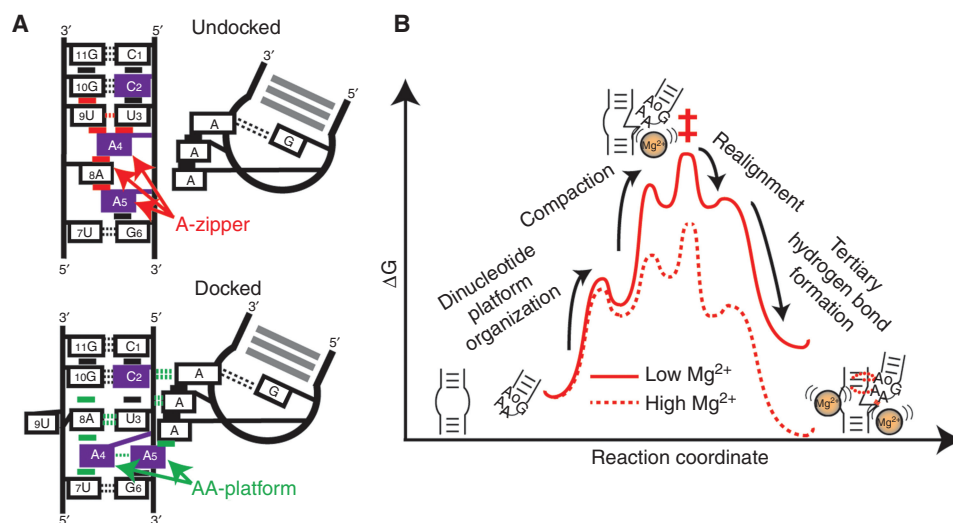
Figure 7. Conformational rearrangements on tetraloop/tetraloop receptor (TL/TLR) formation. (*A*) Comparison of undocked (*top*) and docked (*bottom*) structures of the canonical GAAA/11ntR TL/TLR motif indicate that the 11ntR tetraloop-receptor undergoes a series of conformational changes on binding to the GAAA tetraloop (Cate et al. 1996; Butcher et al. 1997; Davis et al. 2005). The structure of the GAAA tetraloop does not change significantly on docking (Heus and Pardi 1991; Fiore and Nesbitt 2013). (*B*) Single-molecule fluorescence resonance energy transfer (smFRET) dissection of the folding kinetics and thermodynamics of a set of TL/TLR sequence variants across a range of ionic conditions led to a model for the order of conformational changes along the folding pathway of the TL/TLR motif (Bonilla et al. 2017). (From Bonilla et al. 2017, reprinted, with permission, from the American Chemical Society.)

technology that is beginning to provide this needed information.

4.5 High-Throughput Characterization of RNA Elements and Their Ensemble Properties

The broad implementation of the "reconstitution model" rests on the availability of conformational ensembles for many structural elements. A simplifying feature of this model is that atomic-level data about junctions and tertiary contact elements are not required. Instead, what is needed are the preferred orientations of the helices emanating from junctions to determine the spatial probabilities of the tertiary contact elements connected to the helices. Additionally, for tertiary contacts, we need intrinsic stabilities and the orientation of the helices emanating from the formed tertiary contact, which depend on the conformational preferences of the tertiary contact itself. Current approaches to determine RNA conformational ensembles, NMR residual dipolar coupling, and X-ray scattering interference (Salmon et al. 2014; Shi et al. 2015) are powerful tools for obtaining the detailed conformational ensembles of structural elements of interest; however, they are low throughput and thus not capable of providing information for the wide array of RNA structural elements.

Recently, the high-throughput RNA-MaP platform developed by Greenleaf and colleagues (Denny and Greenleaf

2019) has been applied to this problem, yielding insights and models for each class of RNA structural element, helices, junctions, and tertiary contact motifs (Fig. 8) (Denny et al. 2018; Yesselman et al. 2018; S Bonilla, SK Denny, N Bisaria, et al., in prep.) The approach uses an engineered dimer, referred to as "TectoRNA," that assembles via formation of two distinct TL/TLR tertiary contacts (Fig. 8B) (Jaeger and Leontis 2000; Davis et al. 2005; Geary et al. 2008). The TectoRNA system was used as a scaffold for the insertion of a wealth of two-way junctions and components that could be systematically and broadly varied: helix length and content, junction type and sequence, and tertiary motif identity (Fig. 8C). Critically, the assay is direct, provides thermodynamic measurements, and can be performed rapidly, in parallel, and with high precision for tens of thousands of constructs.

4.6 Helices

Comparisons of helices in X-ray structures suggest sequence-specific differences in the conformational preferences of base pairs. Indeed, preliminary studies indicated that changes in the sequence of the TectoRNA helices alter dimerization by nearly 100-fold. Yesselman constructed a computational model that predicts the sequence dependence of the helix ensembles, based on the large numbers of conformations observed for each possible base-pair step in
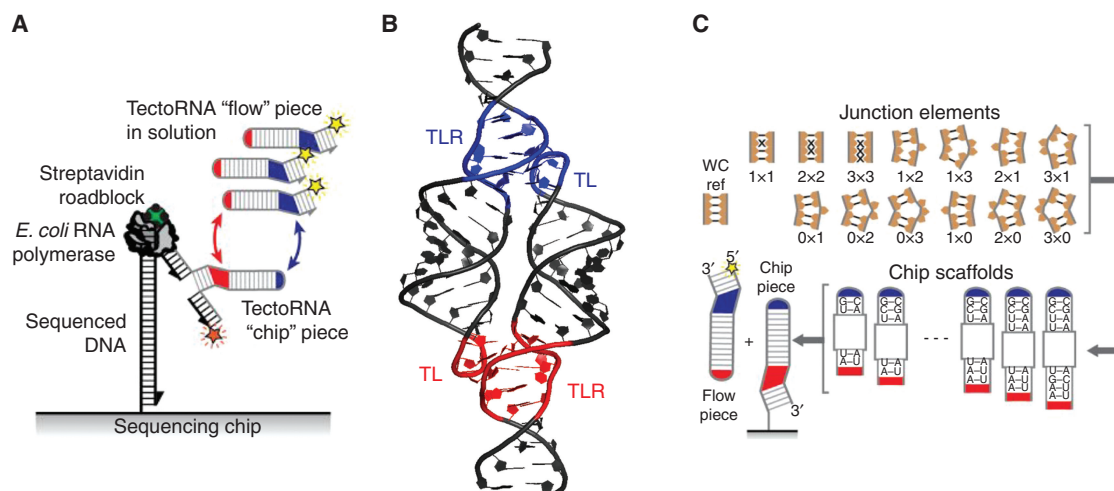
Cold Spring Harbor Perspectives in Biology

Figure 8. RNA-MaP method applied to the study of RNA tertiary elements. (*A*) The RNA-MaP platform with a bound expressed RNA ("chip" piece) and a "flow" piece RNA added in *trans* (Denny et al. 2018). These RNAs combine to form a tertiary "TectoRNA." (*B*) A model for the tertiary structure of the TectoRNA homodimer (PDB: 2ADT; Davis et al. 2005). A heterodimer in which one of the canonical GAAA/11ntR TL/TLRs is replaced by an orthogonal GGAA/R1 (Geary et al. 2008) was used as a "host" for different helix sequences and lengths, different junction elements, and different tertiary motif elements. (*C*) Examples of junction elements used in these studies. (*A,C*, Reproduced, with permission, from Denny et al. 2018.)

X-ray structures from the PDB. This model quantitatively predicted the dimerization affinity for thousands of TectoRNA sequences, with an R^2 of 0.71 and a root-mean-square deviation (RMSD) of 0.34 kcal/mol (Yesselman et al. 2018). Thus, helices can vary significantly in their conformational ensembles and substantially alter tertiary folding energetics, and an initial quantitative model of the ensemble of helix conformations is reasonably predictive.

4.7 Junctions

The TectoRNA was used as a scaffold to embed thousands of junctions. Each junction was embedded in many TectoRNA contexts (Fig. 8C), in which the junction conformational properties could stabilize or destabilize TectoRNA association. The collection of tectoRNA affinities for each junction provided a "thermodynamic fingerprint" for that junction that could be compared with the thermodynamic fingerprints of other junctions. Importantly, these thermodynamic fingerprints reflect the conformational preferences of the junctions; hence, two junctions with nearly identical thermodynamic fingerprints are suggested to have similar conformational ensembles (Denny et al. 2018). Junctions with the same number of unopposed residues tended to have similar thermodynamic fingerprints, and thus likely conformational behavior, as predicted by Al-Hashimi and colleagues (Bailor et al. 2010). Nevertheless, substantial sequence-dependent variation was observed. Further, although there is insufficient X-ray data for any individual junction to build ensembles akin to those developed for helices, Denny and colleagues could group junctions with similar thermodynamic fingerprints and known structures to obtain crude ensemble models for dozens of junctions. This model also had predictive power when combined with the helix model described above, whereas incorporating single X-ray junction structures gave poor predictive power (Denny et al. 2018; Yesselman et al. 2018).

4.8 Tertiary Contact Motifs

A range of tertiary contact motifs can be incorporated in TectoRNA constructs, allowing comparison of their thermodynamic properties. Investigation of more than 1000 TL/TLR variants has revealed different classes of conformational behaviors, most simply those that have conformational preferences similar as that of the canonical GAAA/11ntR, and those with distinct thermodynamic fingerprints (S Bonilla, SK Denny, N Bisaria, et al., in prep.). Whereas the available data suggest a very narrow conformational ensemble for the canonical GAAA/11ntR, the wide tolerance for alignment changes showed by several of the other TL/TLRs suggests broadened conformational landscapes.

The ability to carry out high-throughput and quantitative measurements provides extensive information about conformational ensembles and energetic properties of RNA helices, junctions, and tertiary contact motifs. Although this approach to broadly and deeply interrogate RNA structural elements is in its infancy, it is clear that

we have transitioned to the predictive phase in RNA folding and dynamics. Future work will benefit from additional host systems, beyond the TectoRNA dimer, to more broadly explore conformational space and to profile junctions and tertiary contact motifs with very different conformational preferences. In addition, although the initial ensemble computational models have impressive predictive power, additional tests, including orthogonal methods such as NMR and X-ray scattering interference that give more direct structural and ensemble information, will allow these models to be improved.

5 CLOSING PERSPECTIVE

Scientific endeavors often feel like random walks punctuated by occasional discoveries. Discrete discoveries, such as the observation of new phenomena or the generation of structures of biological complexes are readily recognized as "quantized" steps forward; in contrast, results from characterization and dissection are harder to directly map to progress. Organization of the scientific process by epoch helps clarify the necessity of a breadth of studies, a necessity that is more apparent in retrospect (Fig. 1 and above discussion).

The field of RNA folding was marked by a continued characterization of additional structured RNAs at similar levels of description. The perspective of scientific epochs helped identify what next steps were needed to reach the ultimate goal of predictive models. Given this experience, we encourage those developing and using the powerful genomic-scale technologies and other cutting-edge approaches described in this collection and elsewhere to consider also how these tools may lead to predictive and testable models, to go beyond continued characterization of RNA biological phenomena.

Understanding RNA folding has, in essence by definition, required an atomic and molecular-based perspective. But often, biological questions are framed in terms of sequences, pathways, processes, and cellular features abstracted from their underlying molecular properties. Given this, one might ask how generally applicable are the findings and lessons from RNA folding to areas and questions seemingly more directly connected to biology and medicine. We offer the following perspectives.

We are heartened by the biological insights that have emerged from seemingly specialized and remote studies of RNA folding, including insights into the probability and potential properties of an RNA World, possible steps in the transition to protein-centric life on Earth, the occurrence and roles of RNA chaperones, and reasons for the use of RNA-based machines for several current-day processes involving nucleic acid recognition and complex controlled

reaction cycles. More broadly, these molecules carry out biological functions through physical interactions and chemical reactions, so we should aim for understanding at that level; more practically, given the myriad complexity of biology and biological observations, the ability of molecular models to constrain and steer models, and the ability to alter molecules to incisively distinguish between models, argues for the adaptation of molecular perspectives, especially in this age of genomics and big data.

We see the need for and power of frameworks, here for RNA folding, that break a complex process into individual steps. Cellular processes are even more complex, suggesting that progress in understanding, and ultimately predicting cellular control and function, will benefit greatly from—and likely require—experiments that allow individual steps to be considered and probed. Finally, we reiterate the power of quantitative predictions, because of their ability to finely discriminate between models and the need for quantitation to model, understand, and predict processes and responses in biology.

ACKNOWLEDGMENTS

REFERENCES

*Reference is in this collection.

Adamala K, Szostak JW. 2013. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342:** 1098–1100.

Bailor MH, Sun X, Al-Hashimi HM. 2010. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327:** 202–206.

Bisaria N, Greenfeld M, Limouse C, Pavlichin DS, Mabuchi H, Herschlag D. 2016. Kinetic and thermodynamic framework for P4-P6 RNA reveals tertiary motif modularity and modulation of the folding preferred pathway. *Proc Natl Acad Sci* **113:** E4956–E4965.

Bisaria N, Greenfeld M, Limouse C, Mabuchi H, Herschlag D. 2017. Quantitative tests of a reconstitution model for RNA folding thermodynamics and kinetics. *Proc Natl Acad Sci* **114:** E7688–E7696.

Bonilla S, Limouse C, Bisaria N, Gebala M, Mabuchi H, Herschlag D. 2017. Single-molecule fluorescence reveals commonalities and distinctions among natural and in vitro-selected RNA tertiary motifs in a multistep folding pathway. *J Am Chem Soc* **139:** 18576–18589.

Breaker RR. 2011. Riboswitches and the RNA world. In *RNA worlds: From life's origins to diversity in gene regulation* (ed. Atkins JF, Gesteland RF, Cech TR), pp. 63–77. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Brenowitz M, Chance MR, Dhavan G, Takamoto K. 2002. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting." *Curr Opin Struct Biol* **12:** 648–653.

Butcher SE, Dieckmann T, Feigon J. 1997. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J* **16:** 7490–7499.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273:** 1678–1685.

Cate JH, Hanna RL, Doudna JA. 1997. A magnesium ion core at the heart of a ribozyme domain. *Nat Struct Biol* **4:** 553–558.

Cech TR. 2000. The ribosome is a ribozyme. *Science* **289:** 878–879.

Celander DW, Cech TR. 1991. Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251:** 401–407.

Cheng CY, Chou F-C, Das R. 2015. Modeling complex RNA tertiary folds with Rosetta. *Methods Enzymol* **553:** 35–64.

Chu VB, Lipfert J, Bai Y, Pande VS, Doniach S, Herschlag D. 2009. Do conformational biases of simple helical junctions influence RNA folding stability and specificity? *RNA* **15:** 2195–2205.

Copeland RA. 2000. A brief history of enzymology. In *Enzymes: A practical introduction to structure, mechanism, and data analysis*, pp. 1–10. John Wiley, New York.

Costa M, Michel F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J* **14:** 1276–1285.

Crawford RT. 1917. The important epochs in the development of astronomy. *Publ Astro Soc Pac* **29:** 233–244.

Crick FH. 1968. The origin of the genetic code. *J. Mol Biol* **38:** 367–379.

Das R, Baker D. 2007. Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci* **104:** 14664–14669.

Das R, Kwok LW, Millett IS, Bai Y, Mills TT, Jacob J, Maskel GS, Seifert S, Mochrie SGJ, Thiyagarajan P, et al. 2003. The fastest global events in RNA folding: Electrostatic relaxation and tertiary collapse of the *Tetrahymena* ribozyme. *J Mol Biol* **332:** 311–319.

Das R, Travers KJ, Bai Y, Herschlag D. 2005. Determining the Mg^{2+} stoichiometry for folding an RNA metal ion core. *J Am Chem Soc* **127:** 8272–8273.

Davis JH, Tonelli M, Scott LG, Jaeger L, Williamson JR, Butcher SE. 2005. RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex. *J Mol Biol* **351:** 371–382.

* Denny SK, Greenleaf WJ. 2019. Linking RNA sequence, structure, and function on massively parallel high-throughput sequencers. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a032300.

Denny SK, Bisaria N, Yesselman JD, Das R, Herschlag D, Greenleaf WJ. 2018. High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell* doi: 10.1016/j.cell.2018.05.038.

Ditzler MA, Rueda D, Mo J, Håkansson K, Walter NG. 2008. A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res* **36:** 7088–7099.

Draper DE, Grilley D, Soto AM. 2005. Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **34:** 221–243.

Engelhardt MA, Doherty EA, Knitt DS, Doudna JA, Herschlag D. 2000. The P5abc peripheral element facilitates preorganization of the *Tetrahymena* group I ribozyme for catalysis. *Biochemistry* **39:** 2639–2651.

Ferré-D'Amaré AR, Zhou K, Doudna JA. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395:** 567–574.

Fiore JL, Nesbitt DJ. 2013. An RNA folding motif: GNRA tetraloop-receptor interactions. *Q Rev Biophys* **46:** 223–264.

Fiore JL, Holmstrom ED, Nesbitt DJ. 2012. Entropic origin of Mg^{2+}-facilitated RNA folding. *Proc Natl Acad Sci* **109:** 2902–2907.

Frederiksen JK, Li N-S, Das R, Herschlag D, Piccirilli JA. 2012. Metal-ion rescue revisited: Biochemical detection of site-bound metal ions important for RNA folding. *RNA* **18:** 1123–1141.

Gartland WJ, Sueoka N. 1966. Two interconvertible forms of tryptophanyl sRNA in *E. coli*. *Proc Natl Acad Sci* **55:** 948–956.

Geary C, Baudrey S, Jaeger L. 2008. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res* **36:** 1138–1152.

Gleitsman KR, Sengupta RN, Herschlag D. 2017. Slow molecular recognition by RNA. *RNA* **23:** 1745–1753.

Gracia B, Al-Hashimi HM, Bisaria N, Das R, Herschlag D, Russell R. 2018. Hidden structural modules in a cooperative RNA folding transition. *Cell Rep* **22:** 3240–3250.

Greenfeld M, Solomatin SV, Herschlag D. 2011. Removal of covalent heterogeneity reveals simple folding behavior for P4-P6 RNA. *J Biol Chem* **286:** 19872–19879.

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35:** 849–857.

Herschlag D. 1995. RNA chaperones and the RNA folding problem. *J Biol Chem* **270:** 20871–20874.

Herschlag D. 1998. RNA structure: Ribozyme crevices and catalysis. *Nature* **395:** 548.

Herschlag D, Khosla M, Tsuchihashi Z, Karpel RL. 1994. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J* **13:** 2913–2924.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253:** 191–194.

Hodak JH, Downey CD, Fiore JL, Pardi A, Nesbitt DJ. 2005. Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc Natl Acad Sci* **102:** 10505–10510.

Horning DP, Joyce GF. 2016. Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci* **113:** 9786–9791.

Huang P-S, Boyken SE, Baker D. 2016. The coming of age of de novo protein design. *Nature* **537:** 320–327.

Jaeger L, Leontis NB. 2000. Tecto-RNA: One-dimensional self-assembly through tertiary interactions. *Angew Chem Int Ed* **39:** 2521–2524.

Jarmoskaite I, Russell R. 2014. RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem* **83:** 697–725.

Joyce GF, Inoue T, Orgel LE. 1984. Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, U) templates. *J Mol Biol* **176:** 279–306.

Kaufmann KW, Lemmon GH, Deluca SL, Sheehan JH, Meiler J. 2010. Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49:** 2987–2998.

Koculi E, Cho SS, Desai R, Thirumalai D, Woodson SA. 2012. Folding path of P5abc RNA involves direct coupling of secondary and tertiary structures. *Nucleic Acids Res* **40:** 8011–8020.

Kraut DA, Carroll KS, Herschlag D. 2003. Challenges in enzyme mechanism and energetics. *Annu Rev Biochem* **72:** 517–571.

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31:** 147–157.

Latham JA, Cech TR. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science* **245:** 276–282.

Leontis NB, Lescoute A, Westhof E. 2006. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol* **16:** 279–287.

Lindahl T, Adams A, Fresco JR. 1966. Renaturation of transfer ribonucleic acids through site binding of magnesium. *Proc Natl Acad Sci* **55:** 941–948.

Lipfert J, Doniach S, Das R, Herschlag D. 2014. Understanding nucleic acid-ion interactions. *Annu Rev Biochem* **83:** 813–841.

Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126:** 309–320.

Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, et al. 2017. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23:** 655–672.

Mitchell D III, Jarmoskaite I, Seval N, Seifert S, Russell R. 2013. The long-range P3 helix of the *Tetrahymena* ribozyme is disrupted during folding between the native and misfolded conformations. *J Mol Biol* **425:** 2670–2686.

Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. 2014. Critical assessment of methods of protein structure prediction (CASP)—Round X. *Proteins* **82:** 1–6.

Narlikar GJ, Herschlag D. 1997. Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Annu Rev Biochem* **66:** 19–59.

Nelson JA, Uhlenbeck OC. 2006. When to believe what you see. *Mol Cell* **23:** 447–450.

Noller HF. 2012. Evolution of protein synthesis from an RNA world. In *RNA worlds: From life's origins to diversity in gene regulation* (ed. Atkins JF, Gesteland RF, Cech TR), pp. 141–154. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Okumus B, Wilson TJ, Lilley DMJ, Ha T. 2004. Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J* **87:** 2798–2806.

Orgel LE. 1968. Evolution of the genetic apparatus. *J Mol Biol* **38:** 381–393.

Pan T, Sosnick TR. 1997. 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* **4:** 931–938.

Pan J, Woodson SA. 1998. Folding intermediates of a self-splicing RNA: Mispairing of the catalytic core1. *J Mol Biol* **280:** 597–609.

Pan T, Fang X, Sosnick T. 1999. Pathway modulation, circular permutation and rapid RNA folding under kinetic control. *J Mol Biol* **286:** 721–731.

Piccirilli JA, Vyle JS, Caruthers MH, Cech TR. 1993. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* **361:** 85–88.

Pley HW, Flaherty KM, McKay DB. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372:** 68–74.

Russell R. 2008. RNA misfolding and the action of chaperones. *Front Biosci* **13:** 1–20.

Russell R, Herschlag D. 1999. New pathways in folding of the *Tetrahymena* group I RNA enzyme 1. *J Mol Biol* **291:** 1155–1167.

Russell R, Millett IS, Doniach S, Herschlag D. 2000. Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* **7:** 367–370.

Russell R, Zhuang X, Babcock HP, Millett IS, Doniach S, Chu S, et al. 2002. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci* **99:** 155–160.

Russell R, Das R, Suh H, Travers KJ, Laederach A, Engelhardt MA, et al. 2006. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* **363:** 531–544.

Salehi-Ashtiani K, Szostak JW. 2001. In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature* **414:** 82–84.

Salmon L, Yang S, Al-Hashimi HM. 2014. Advances in the determination of nucleic acid conformational ensembles. *Annu Rev Phys Chem* **65:** 293–316.

Sclavi B, Sullivan M, Chance MR, Brenowitz M, Woodson SA. 1998. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* **279:** 1940–1943.

Shi X, Bonilla S, Herschlag D, Harbury P. 2015. Quantifying nucleic acid ensembles with X-ray scattering interferometry. *Methods Enzymol* **558:** 75–97.

Silverman SK, Zheng M, Wu M, Tinoco I Jr, Cech TR. 1999. Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA* **5:** 1665–1674.

Solomatin SV, Greenfeld M, Chu S, Herschlag D. 2010. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463:** 681–684.

Steitz TA, Moore PB. 2003. RNA, the first macromolecular catalyst: The ribosome is a ribozyme. *Trends Biochem* **28:** 411–418.

Tagami S, Attwater J, Holliger P. 2017. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat Chem* **9:** 325–332.

Takamoto K, Das R, He Q, Doniach S, Brenowitz M, Herschlag D, et al. 2004. Principles of RNA compaction: Insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J Mol Biol* **343:** 1195–1206.

Tan E, Wilson TJ, Nahas MK, Clegg RM, Lilley DMJ, Ha T. 2003. A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc Natl Acad Sci* **100:** 9308–9313.

Tinoco I Jr. 1993. Appendix 1: Structures of base pairs involving at least two hydrogen bonds. In *The RNA world* (ed. Gesteland RF, Atkins JF), pp. 603–607. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Treiber DK, Williamson JR. 1999. Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* **9:** 339–945.

Tsuchihashi Z, Khosla M, Herschlag D. 1993. Protein enhancement of hammerhead ribozyme catalysis. *Science* **262:** 99–102.

Uhlenbeck OC. 1995. Keeping RNA happy. *RNA* **1:** 4–6.

Wan Y, Suh H, Russell R, Herschlag D. 2010. Multiple unfolding events during native folding of the *Tetrahymena* group I ribozyme. *J Mol Biol* **400:** 1067–1077.

Wang S, Karbstein K, Peracchi A, Beigelman L, Herschlag D. 1999. Identification of the hammerhead ribozyme metal ion binding site responsible for rescue of the deleterious effect of a cleavage site phosphorothioate. *Biochemistry* **38:** 14363–14378.

Weinstein LB, Jones BC, Cosstick R, Cech TR. 1997. A second catalytic metal ion in group I ribozyme. *Nature* **388:** 805–808.

Westhof E, Masquida B, Jaeger L. 1996. RNA tectonics: Towards RNA design. *Fold Des* **1:** R78–R88.

Wilson DS, Szostak JW. 1999. In vitro selection of functional nucleic acids. *Annu Rev Biochem* **68:** 611–647.

Woese CR. 1967. The present status of the genetic code. In *Progress in nucleic acid research and molecular biology* (ed. Davidson JN, Cohn WE), pp. 107–172. Academic, New York.

Woodson SA. 2000. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci* **57:** 796–808.

Wu M, Tinoco I Jr. 1998. RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci* **95:** 11555–11560.

Xue Y, Gracia B, Herschlag D, Russell R, Al-Hashimi HM. 2016. Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. *Nat Commun* **7:** 11768.

Yesselman JD, Eiler D, Carlson ED, Ooms AN, Kladwang W, Shi X, Costantino DA, Herschlag D, Jewett MC, Kieft JS, Das R. 2017. Computational design of asymmetric three-dimensional RNA structures and machines. *bioRxiv* doi: 10/1101.223479.

Yesselman JD, Denny SK, Bisaria N, Herschlag D, Greenleaf WJ, Das R. 2018. RNA tertiary structure energetics predicted by an ensemble model of the RNA double helix. *bioRxiv* doi: 10.1101/341107.

Zarrinkar PP, Williamson JR. 1994. Kinetic intermediates in RNA folding. *Science* **265:** 918–924.

Zhao R, Rueda D. 2013. Memory effects in RNA folding dynamics revealed by single-molecule fluorescence. In *Biophysics of RNA folding* (ed. Russell R), pp. 117–133. Springer, New York.

Zhuang X. 2005. Single-molecule RNA science. *Annu Rev Biophys Biomol Struct* **34:** 399–414.

Zhuang X, Bartley LE, Babcock HP, Russell R, Ha T, Herschlag D, Chu S. 2000. A single-molecule study of RNA catalysis and folding. *Science* **288:** 2048–2051.

Zuker M, Mathews DH, Turner DH. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In *RNA biochemistry and biotechnology* (ed. Barciszewski J, Clark BFC), pp. 11–43. Springer, Dordrecht, The Netherlands.

Cold Spring Harbor Perspectives in Biology

The Story of RNA Folding, as Told in Epochs

Daniel Herschlag, Steve Bonilla and Namita Bisaria

Subject Collection   RNA Worlds

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/