Unwinding RNA’s secrets: advances in the biology, physics, and modeling of complex RNAs

Vincent B Chu and Daniel Herschlag

The rapid development of our understanding of the diverse biological roles fulfilled by non-coding RNA has motivated interest in the basic macromolecular behavior, structure, and function of RNA. We focus on two areas in the behavior of complex RNAs. First, we present advances in the understanding of how RNA folding is accomplished in vivo by presenting a mechanism for the action of DEAD-box proteins. Members of this family are intimately associated with almost all cellular processes involving RNA, mediating RNA structural rearrangements and chaperoning their folding. Next, we focus on advances in understanding, and characterizing the basic biophysical forces that govern the folding of complex RNAs. Ultimately we expect that a confluence and synergy between these approaches will lead to profound understanding of RNA and its biology.

Study of catalytic RNAs, which allows detection of correct folding via activity measurements, has established basic thermodynamic and kinetic properties of RNA folding. Large catalytic RNAs such as the group I intron from *Tetrahymena thermophila* and the RNase P RNA from *Bacillus subtilis* fold at vastly different rates under different conditions upon addition of Mg$^{2+}$ and have been shown to fold via multiple parallel pathways, in which different molecules follow different routes with different rates and intermediates, to a common final folded structure [4,5]. These observations support the common view that RNAs traverse rugged energy landscapes as they fold [6,7]. However, little is known about the true shape of these landscapes, and our understanding of the underlying physical behavior of RNA is severely limited.

RNA’s propensity to misfold has been attributed to its intrinsic ‘character flaws’: the low information content of its four side chains (relative to the 20 for proteins), the sequestration of these side chains in helices (relative to the outward-facing side chains in peptide beta sheets and alpha helices), the promiscuous ability of any base to form multiple hydrogen bonds with any other base, and the high stability of stacking interactions of rigidly oriented groups upon the cyclic base (relative to the more flexible protein side chains) [12,13].

It has been suggested that the problems encountered by RNA provided an inroad for peptides and proteins in the presumed early RNA world [13–15]. RNA chaperones, proteins that can reorient RNA structures, presumably evolved to address this misfolding, and considerable evidence for the existence and roles of RNA chaperones has accumulated over the past decade [16–19].

In this review, we first focus on recent advances in the field of chaperone-assisted RNA folding—one of the more complex processes in RNA biology. We then turn to recent conceptual and experimental advances in understanding the basic physical forces that underlie complex RNAs and RNA/protein complexes. We end with a brief discussion of recent technical advances that will be crucial for developing a deep energetic, dynamic, and
structural understanding of RNA and, ultimately, for drawing profound connections between these physical properties and behaviors and the biology of RNAs and RNA/protein complexes.

By necessity, more has been omitted than included in this short review. Fortunately, there are insightful reviews in these omitted areas (e.g. [20–28]). It is a remarkable testament to those in the rather small world of RNA folding and function that so much progress has been made in the relatively short time of modern studies of RNA folding and dynamics.

How ATP-dependent RNA chaperones assist folding

The predisposition of RNA for misfolding or slow folding under physiological conditions poses serious challenges for RNA in vivo [7,13]. Paradoxically, some RNA structural elements seem to exacerbate misfolding; studies of the group I intron from *Tetrahymena* have shown that its P5abc domain stabilizes misfolded intermediates and dramatically slows refolding of these intermediates into the native state [11,29,30]. P5abc’s detrimental effect on folding kinetics is balanced by its ability to confer thermodynamic stability to the native state and enhance its catalytic activity [31,32]. RNA chaperones that rescue misfolded structures allow an escape from these opposing demands for catalytic and folding efficiency by relieving selective pressure against slow folding or misfolding RNAs.

Members of the DExD/H-box protein family, principally from the major subfamily DEAD-box (the eponymous DEAD motif is one-letter code for the amino acid sequence Asp-Glu-Ala-Asp), are involved in virtually every cellular process involving RNA, ranging from splicing, transport, localization, translation, and decay [23,33]. Although DEAD-box proteins have been casually referred to as ‘RNA helicases’ on the basis of their structural similarity to DNA helicases, they are unlikely to processively unwind helices, as RNA rarely contains the long helical stretches found in genomic DNA. Furthermore, functional RNAs often require conformational transitions other than simple helical rearrangements. Since DEAD-box proteins may function differently from canonical DNA helicases, they have been less judgmentally referred to as RNA-dependent ATPases or RNA ‘unwindases’ [34,35,36**].

The discovery of DEAD-box proteins that promote splicing in the well-characterized group I and II introns opened the door for recent mechanistic breakthroughs [18,37,38**,39]. Interestingly, these DEAD-box proteins were shown to be interchangeable to a substantial extent, demonstrating that they function rather promiscuously. Deletion of Mss116p, a mitochondrial DEAD-box protein from *Saccharomyces cerevisiae*, adversely affected the splicing of group I and II introns in vivo; these deleterious effects could be largely rescued by the expression of a related DEAD-box protein, CYT-19 from *Neurospora crassa* [38**]. Further, Mss116p and CYT-19 were shown to promote group I and group II intron splicing in vitro, as was the cytosolic DEAD-box protein, Ded1p [37,38**,39]. Each of these proteins was also shown to unwind RNA helices non-specifically in an ATP-dependent manner [36**,37,39–41].

The ability of these DEAD-box proteins to promote proper folding of structurally diverse RNAs and their common structural features suggest that mechanistic details obtained for one particular DEAD-box protein and RNA substrate may yield general insights for a wide variety of DEAD-box proteins. Recently, the mechanism of CYT-19 has been probed in the folding and function of a ribozyme derived from the *Tetrahymena* group I intron. CYT-19 was shown to act non-specifically, disrupting both misfolded and native forms of the ribozyme. However, the efficiency of unfolding activity was dependent on the stability of the RNA species, with CYT-19 unfolding the misfolded RNA at least 50-fold faster than the native ribozyme [42*,43].

Unexpectedly, CYT-19 was also able to unwind the base paired P1 duplex, formed between the ribozyme and its oligonucleotide substrate, much more efficiently than the same duplex free in solution, demonstrating that local unwinding activity was enhanced by the presence of the larger intron structure. Further, unwinding activity was abolished if tertiary contacts were formed between P1 and the intron. This curious result immediately suggested a physical model for the chaperone activity of CYT-19 and other DEAD-box proteins. Chaperones function by first binding non-specifically to structured RNA, then rearranging RNA structure by unwinding double-stranded regions and perhaps other accessible structures [36**]. This general model was also supported by a clever experiment in which the unwinding activity of Ded1p was shown to be enhanced by an ssRNA tethered through a streptavidin linker to a model RNA duplex [44].

These and related results lead to a general model for RNA chaperone function (Figure 1) [40,45**]. The proteins do not recognize any specific structural features of their target RNAs; unfolding efficiency is dependent on the accessibility of the target. If a structural element is loosely associated with the rest of the RNA, it is easy to ‘grab’ and unwind. This model elegantly explains the enhanced activity of CYT-19 on misfolded relative to native RNA, as properly folded structure is more compact and will have fewer conformational excursions of its structural elements.

A consequence of non-specific unfolding activity is that DExD/H-box proteins can drive folding away from
thermodynamic equilibrium, and such effects have recently been demonstrated for Ded1p and CYT-19 [42,43,46]. These results suggest that, subsequent to each ATP-driven unfolding event, the RNA repartitions between folding forms based on the kinetics of this partitioning rather than the stability of the ultimate folded states. If the stability and accessibility differences between the alternative folded states of the RNA are not too large, the chaperone can unfold both species and allow accumulation of a less stable but kinetically preferred state.

The ability of DExD/H-box proteins to perturb RNAs away from thermodynamic equilibrium may also have profound biological implications beyond its implications for the chaperone mechanism. Structured RNAs are involved in complex, multi-step processes, such as pre-mRNA splicing and translation. Although the details of these processes remain elusive, it is plausible that DExD/H-box proteins capture the energy of ATP hydrolysis to drive the transient formation of thermodynamically unstable states that are part of the normal reaction cycle. Although non-specific DEAD-box chaperones could be important for these processes, presumably most of the required structural changes are mediated by dedicated DExD/H-box proteins, which use specific RNA or protein interactions to position themselves within complexes for rearrangement of the desired structural element [47–49].

A future challenge in this area is dissecting the individual reaction steps in complex RNA-mediated processes and unraveling the physical roles played by DExD/H-box proteins. The recently attained ability to follow pre-mRNA splicing by FRET at the single molecule level represents a technical breakthrough toward this important goal [50].

Simple forces underlying the complex behavior of RNA

As noted in the Introduction, one of the primary lessons from prior work is that RNA folding is highly complex. However, work from several groups, aimed at understanding the underlying physical properties of RNA and the behavior of RNA components, suggests the feasibility and potential of a ‘bottom up’ approach to RNA folding.

Figure 2 presents a simple hierarchy of forces and features that underlie the folding of all RNAs. Understanding the underlying behavior of each component, from theory or empirical observation, would provide the foundation for describing and understanding of the behavior of complex RNAs. In this section, we explain this hierarchy and emphasize some recent foundational advances in our understanding.

The formation of secondary structure, encoded in RNA sequence, is typically an early step in folding. The stability of RNA secondary structure and the fact that formation of secondary and tertiary structure is often temporally separated simplifies the consideration of the folding process [51] (see also [28]). However, RNA secondary structure topology is more complex than the topology of proteins, and understanding this complexity...
and the folding constraints introduced by secondary structure topology will require advanced modeling and experimental approaches [52,53].

In the absence of other forces, the predominant force on secondary structure helices (depicted in Figure 2 as cylinders) is the enormous inter-helical electrostatic repulsion due to the polyelectrolyte nature of the RNA. Draper has written extensively on the importance of the ‘ion atmosphere’ that surrounds any polyelectrolyte and has noted that the ions in this atmosphere are likely to provide the dominant electrostatic contribution to the folding energetics, far greater than any ions bound to specific sites within the RNA [54,55].

Nevertheless, the ion atmosphere has been a difficult concept for experimentalists to grasp, both because of limited familiarity with the underlying theory and because of our inability to directly visualize this dynamic sheath of ions via X-ray crystallography or other direct approaches. Thus, the recent ability to decipher the shape of this atmosphere by anomalous small angle X-ray scattering provided an unusually direct experimental window into properties of the atmosphere [56,57]. Further, the contents of the atmosphere have recently been quantitatively dissected, by essentially counting the ions associated with a nucleic acid under a variety of ionic conditions. This technique allows RNA systems to be characterized stoichiometrically in terms of the number of each cation and anion species present in the ion atmosphere [58].

The next challenge is the calculation of electrostatic energies, as these energies determine the energetic penalty of assembling helices in densely packed arrangements. Most simply, helices at low ionic concentrations repel; at higher ionic concentration, this repulsion is reduced owing to the enhanced screening, allowing helices to approach one another (Figure 2a). However, some have proposed the existence of attractive forces in polyelectrolyte systems under certain ionic conditions that may be responsible for the coalescence of helices in the folded state [59,60].

While simulation experiments and the aggregation of DNA and RNA in ethanol during standard precipitation procedures and other conditions demonstrate the existence of such attractive forces under some conditions, their exact magnitude and significance for folding in vitro was unknown [61,62]. The magnitude of these attractive forces was tested with a simple model: two DNA helices tethered together by a neutral polyethylene glycol tether. Small angle X-ray scattering (SAXS) revealed no significant attractive forces under standard in vitro folding conditions [63]. Thus, the observed ‘electrostatic collapse’ of large RNAs upon addition of ions appears to arise from a screening of repulsive force rather than the action of an attractive one (e.g. [64–66]).

There have also been recent advances in understanding and quantitating electrostatic forces. The predominant electrostatic theory used by biochemists has been Poisson-Boltzmann (PB) theory, a computationally convenient method for calculating approximate electrostatic forces [67]. PB theory neglects potentially important excluded volume effects from ion-size and positional correlation effects from ion–ion electrostatic interactions. In particular, correlation effects are predicted to be more important for multivalent ions.

Simulation has suggested that PB theory is inadequate for describing electrostatic energies for nucleic acids in the presence of divalent ions such as Mg$^{2+}$, a conclusion supported by experiments assessing the effects of ions on DNA duplex stability ([68,69] and references therein). However, the opposite conclusion has been drawn from fitting RNA folding transitions to PB predictions [70–74]. Ion-counting and SAXS experiments have provided a simple test for these effects and have revealed effects of ion size and valence that are not accounted for by PB theory, with deviations up to an order of magnitude in the systems investigated [58,75]. Advances in theory and simulation are required to account for these discrepancies.
and obtain accurate electrostatic energies for use in dissecting RNA folding and RNA/protein energetics [76,77].

The junctions in secondary structure bias the conformations of the helices they join (Figure 2b). Ground-breaking work by Lilley on Holliday junctions – four-way DNA junctions involved in recombination – and junctions found in structured RNAs has shown that these junctions position their attached helices into preferred conformations [78–80]. Such preferred geometries are integral to the adoption of stable, active structure by functional RNAs. Ha and Lilley have used single molecule FRET to probe the underlying kinetics and thermodynamics of hairpin ribozyme folding, showing that the conformational dynamics of its junction accelerate folding and promote formation of its active site [81**]. An important challenge for understanding and predicting RNA structure and dynamics remains to interrogate the behaviors and structures of isolated RNA junctions to reveal their intrinsic behaviors and then how these behaviors determine and are modulated by the rest of the RNA molecule in which they are found. The development of a database of RNA junctions extracted from structured RNAs will undoubtedly help advance this goal [82*].

RNAs often utilize motifs for ‘gluing’ RNA secondary structure together to fold to precise tertiary structures (Figure 2c). Long-range interactions in RNA are often found by phylogenetic covariation, beginning with tRNA whose tertiary structure was first modeled in 1969 [83]. The tetraloop/tetraloop receptor motif is perhaps the most famous RNA motif. It was first identified by phylogeny in group I and II introns by Michel, Westhof, and colleagues and observed for the first time in the seminal crystal structure of the P4–P6 domain of the Tetrahymena group I intron, in which this motif reinforces the side-by-side packing architecture of RNA helices [84–86]. This motif and others are ubiquitous in the ribosome and other structured RNAs [87*-91]. We wish to emphasize that information on motif partners and secondary structure – determined from phylogeny – provides a powerful tool in modeling RNA structure. Westhof has developed and exploited this tool to create impressive first generation three-dimensional models for many structured RNAs [92].

Early studies emphasized special roles for Mg$^{2+}$ in RNA folding and function. However, some RNAs have been shown to fold without Mg$^{2+}$ [93–95]. Some, such as the hairpin and hammerhead ribozymes, can even function in its absence [96,97]. A recent dissection of folding the P4–P6 domain from the Tetrahymena group I intron has helped clarify the roles of metal ions in RNA folding and, more generally, the underlying energetics of folding [94]. This RNA can adopt a compact shape with its tetraloop motif formed in the presence of monovalent ions alone. Instead of a special role for Mg$^{2+}$, folding of P4–P6 and of RNAs in general can be accounted for by a simple balance of forces and interactions outlined in Figure 2. Unfavorable electrostatic repulsion and loss of conformational entropy in the folded state resist folding, whereas junction preferences and tertiary motif formation favor it. Electrostatic forces are modulated by ion concentration and valence; thus, even monovalent cations can mitigate electrostatic repulsion to allow folding to be favorable if present at sufficient concentrations. Fewer divalent ions are needed since they are much more effective at charge screening (localizing a single divalent ion around RNA is entropically less costly than two monovalents). Specific divalent ion binding sites – if present in the folded RNA – contribute additional stability. The contributions from specifically bound divalent cations have been isolated by saturating the ion atmosphere with monovalent cations and following the binding of two specific divalent metal ions to the core of the P4–P6 RNA [98,99] (see also [100]).

**RNA: present and future**

We have highlighted some of the conceptual and experimental advances over the past few years and some of the future challenges. In this section, we briefly note some additional methodological advances in the context of challenges that remain in unraveling the structure, energetics, and dynamics of RNA.

Single molecule force experiments have far exceeded simple proof-of-principle experiments and have been used to elucidate the folding thermodynamics and kinetics for several RNAs [101–104,105*,106**]. This approach has answered one of the most fundamental questions in nucleic acid structure formation, providing direct evidence for the long-held nucleation model for duplex involving formation of two to three base pairs before rapid formation of the remainder of the duplex [107]. Further, it has been shown that force unfolding data can be deconstructed into a complete energy landscape [108]. It will be fascinating to see if this approach can be applied to increasingly complex RNAs.

There have been important advances in studying the kinetics of the folding of medium and large RNAs. Time-resolved hydroxyl radical footprinting experiments at millisecond resolution are now routinely conducted using standard rapid-quench devices, without the need for synchrotron radiation to generate radicals [109*]. Thus, a large body of kinetic data on RNA folding at the residue level is readily obtainable. This large body of data can be analyzed in minutes instead of hours, using freely available semi-automated software (https://simtk.org/home/safa) [110]. Still higher throughput data acquisition is now possible using capillary electrophoresis and a software analysis package called CAFA [111]. Parallel advances have been made with a new and powerful footprinting approach called ‘SHAPE’, which shows...
protection data for residues in secondary and tertiary structures [112–114]. New kinetic modeling approaches have greatly simplified and systematized the development of kinetic models from large sets of footprinting data; such an approach has been recently applied to characterize the folding pathways of wild type and mutant Tetrahymena group I intron [115•,116].

Early single molecule FRET studies on RNA folding kinetics revealed a wealth of information inaccessible to traditional pre-steady state kinetic methods, demonstrating the utility of these studies [117–119]. Since then, there has been an explosion of single molecule fluorescence studies on RNA [120]. One of the most puzzling observations in these studies is of molecular ‘memory’—the fact that different molecules tend to display similar kinetic behavior over long time scales that differ from that of other molecules of the same sequence [81**,121]. Future work will undoubtedly address the basis for such quixotic behavior by single molecules.

RNA structures remain difficult to obtain relative to protein structures despite the solution of the ribosome structure other RNA structures. Piccirilli and co-workers, following on observations that proteins may help RNA crystallization, have pioneered a potentially high-throughput approach for crystallization of RNA/antibody complexes [122•]. Nevertheless, many important RNA structures may be partially or intrinsically unstructured and thus resistant to crystallization, necessitating the development of new approaches to RNA structure. Medium resolution structural information is now obtainable from a method that allows pairwise distance constraints to be read from a single polyacrylamide gel. This method has been used to obtain structural information about a molten globule-like state of the P4–P6 RNA [123]. Also, additional modeling approaches are needed for RNA, and a first pass fragment assembly method, following the highly successful approach of David Baker for protein structure prediction, has been pioneered for RNA [124].

The intrinsically dynamic behavior of RNA that complicates structure determination has also been highly resistant to study. However, recent advances in NMR techniques over the past two years are now providing an unprecedented view into the structural dynamics of RNA [125–127]. Particularly powerful is the ability to determine the amplitude and relative directionality of the internal motion of helical domains through residual dipolar coupling measurements [128,129].

In this review we have emphasized advances in mechanistic and physical understanding using model RNAs of varying complexity. In parallel, the biology community is increasingly recognizing the diverse roles of RNA in basic gene expression and regulation. We suspect that biological and physical approaches to understanding RNA will converge in the coming decade as biological systems are reconstituted and characterized and the tools for understanding RNA are developed and perfected. While there is much work ahead, we eagerly anticipate that union.

**Conflicts of interest**

The authors declare no conflicts of interest.

**Acknowledgements**

The authors thank and apologize to the many investigators in the RNA world whose work has contributed greatly to our overall understanding of RNA but could not be cited owing to space limitations. It is truly a testament to this vigorous community that so much has been accomplished of note in the recent years. We would like to thank Yu Bai, Max Greenfeld, Jan Lapfert, Jungwon Park, Adelene Sim, and Sergey Solomatin for critical discussion and insightful comments and all of our collaborators. Our work on RNA folding is supported by NIH grant PO1 GM066275. VBC is supported by a Pfizer Bio-X Graduate fellowship. We have adopted an agnostic view in highlighting references to our own work to avoid the moral dilemma of assigning ‘significance stars’ to our own work.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- • of outstanding interest

The authors demonstrate the promiscuity of DEAD-box proteins CYT-19 from Neurospora crassa and Mss116p from Saccharomyces cerevisiae by showing that splicing reactions in Mss116p-deletion mutants can be rescued by expression of CYT-19 in vivo.

The authors investigate the activity of Ded1p, a DEAD-box protein from Neurospora crassa, to unwind the native form of the Tetrahymena ribozyme, albeit 50-fold less efficiently than the misfolded form. Furthermore, the ATP-driven unwind function activity actively redistributes the conformational ensemble of structured RNAs, populating intermediates that otherwise would be rare at thermodynamic equilibrium.

The authors demonstrate the ability of CYT-19, a DEAD-box protein from Neurospora crassa, to unwind the native form of the Tetrahymena ribozyme, albeit 50-fold less efficiently than the misfolded form. Furthermore, the ATP-driven unwind function activity actively redistributes the conformational ensemble of structured RNAs, populating intermediates that otherwise would be rare at thermodynamic equilibrium.

The authors demonstrate the ability of CYT-19, a DEAD-box protein from Neurospora crassa, to unwind the native form of the Tetrahymena ribozyme, albeit 50-fold less efficiently than the misfolded form. Furthermore, the ATP-driven unwind function activity actively redistributes the conformational ensemble of structured RNAs, populating intermediates that otherwise would be rare at thermodynamic equilibrium.

The authors demonstrate the ability of CYT-19, a DEAD-box protein from Neurospora crassa, to unwind the native form of the Tetrahymena ribozyme, albeit 50-fold less efficiently than the misfolded form. Furthermore, the ATP-driven unwind function activity actively redistributes the conformational ensemble of structured RNAs, populating intermediates that otherwise would be rare at thermodynamic equilibrium.
83. The authors present a database of RNA junctions extracted from structural RNA data, offering researchers a centralized collection of data to perform structural analyses and structure design.
The authors demonstrate the feasibility of rapid acquisition of structural

10. Li PT, Budiaman C, Tinoco I Jr: Real-time control of the
energy landscape by force directs the folding of RNA molecules.

11. Li PT, Budiaman C, Tinoco I Jr: Unusual mechanical stability
of a minimal RNA kissing complex. Proc Natl Acad Sci U S A
2006, 103:15847-15852.

Using optical tweezers, the authors probe the energetics of the folding
and unfolding of a kissing-loop motif. They find that the stability of the
kissing-loop interaction is unusually strong.

12. Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM:
Direct observation of hierarchical folding in single riboswitch

Using optical tweezers, the authors follow the entire folding trajectory of a
simple adenine riboswitch aptamer. With extremely high accuracy, the authors observe the formation of secondary and tertiary structure, as well as the ligand-induced stabilization of the folded aptamer state, resolving the hierarchical folding of this important regulatory ncRNA.

13. Woodside MT, Behnke-Parks WM, Larizadeh K, Travers K,
Herslag D, Block SM: Nanomechanical measurements of the
sequence-dependent folding landscapes of single nucleic acid hairpins.

14. Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K,
Herslag D, Block SM: Direct measurement of the full,
sequence-dependent folding landscape of a nucleic acid.

footprinting: a laboratory-based method for the time-resolved

The authors demonstrate the feasibility of rapid acquisition of structural
footprinting data at the millisecond time scale using hydroxyl radicals
generated using a chemical source and standard quench-flow mixers.
Such a technique makes hydroxyl radical footprinting techniques acces-
sible to the wide scientific community without the use of a synchrotron source.

