The Dissertation Committee for Inga Jarmoskaite Certifies that this is the approved version of the following dissertation:

Mechanistic Studies of the RNA Chaperone Activities of the DEAD-box RNA Helicase CYT-19

Committee:

Rick Russell, Supervisor

Karen S. Browning

Kevin Dalby

Alan M. Lambowitz

Andreas Matouschek

Mechanistic Studies of the RNA Chaperone Activities of the DEAD-box RNA Helicase CYT-19

by

Inga Jarmoskaite, Diplom.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2014

Dedication

To my Teachers

Acknowledgements

The years leading to this dissertation have been an exciting journey and I would like to use this opportunity to thank all the people who guided me along the way. First and foremost, I want to thank my advisor Rick Russell for his exceptional mentorship and for providing such an intellectually stimulating environment in which to grow as a scientist. Science, however inspiring, would be so much harder on the long term without human role models, and I am extremely fortunate to have had one as my graduate mentor. I am also grateful to my committee members Karen Browning, Kevin Dalby, Alan Lambowitz and Andreas Matouschek for providing valuable feedback and ideas along the way and for being great scientists and educators that I look up to. I especially thank Dr. Lambowitz and his group, with whom it has been an honor and an incredible learning experience to collaborate.

Members of the Russell lab, past and present, have contributed to this project with their time, knowledge, and reagents, and I want to thank all of them. I am grateful to former lab members Hari Bhaskaran, Amanda Chadee, David Mitchell and Yaqi Wan, who so warmly welcomed me to the lab, patiently taught me lab techniques, and, with their hard work and contagious enthusiasm for science, motivated me to be a good scientist and graduate student as well. Brian Cannon, Woongsoon Choi, Alexis Davis, Brant Gracia, Cynthia Pan, Jeff Potratz, Pilar Tijerina and Luke Ward have all been amazing colleagues and friends. The trips to the synchrotron were arguably one of the most character-building experiences in graduate school, and I want to especially thank Pilar Tijerina and Soenke Seifert (Advanced Photon Source) for being such an incredible team to work with. Woongsoon, Brant, Jeff also generously contributed their time to help with these experiments, and Woongsoon, Alexis and Pilar purified CYT-19. I am also

very grateful to Anna Mallam (Lambowitz lab) for skillfully leading the collaborative project and for teaching me new methods and analysis tools.

On a more personal note, I want to thank my fiancé and the friends I met in Texas for making me feel at home in Austin during all these years. Finally, these acknowledgements would be incomplete without thanking my family, who were my first and most influential teachers and who have been my biggest supporters during the work on this dissertation and all my other pursuits.

Mechanistic Studies of the RNA Chaperone Activities of the DEAD-box RNA Helicase CYT-19

Inga Jarmoskaite, Ph.D.

The University of Texas at Austin, 2014

Supervisor: Rick Russell

Structured RNAs are pervasive in biology, spanning a functional repertoire that includes messengers, regulators of gene expression and catalysts of translation and splicing. From the relatively simple tRNAs and riboswitches to the highly structured ribosomal RNAs, the ability of RNAs to function is dependent on well-defined secondary and tertiary structures. However, studies of RNA folding in vitro have revealed an extreme propensity to form alternative structures, which can be long-lived and interfere with function. In the cell, a diverse array of RNA binding proteins and RNA chaperones guide RNAs towards the correct structure and disrupt misfolded intermediates. Among these proteins, DEAD-box protein family stands out as one of the largest groups, with its members ubiquitously involved in RNA metabolism across all domains of life. DEAD-box proteins can function as both specific and general RNA chaperones by disrupting RNA structures in an ATP-dependent manner. Here I describe my work studying the general RNA chaperone mechanism of the *Neurospora crassa* protein CYT-19, a model DEAD-box protein and a biological RNA chaperone that is required for efficient folding of self-splicing group I intron RNAs in vivo. After an introduction to DEAD-box proteins and their mechanisms as RNA remodelers (Chapter 1), I will first describe studies of group I intron unfolding by CYT-19, focusing on the effects of RNA tertiary structure

stability on CYT-19 activity and targeting to RNA substrates (Chapter 2). I will then describe the characterization of ATP-dependent mechanisms during CYT-19-mediated refolding of the misfolded group I intron (Chapter 3). In Chapter 4, I will present small-angle X-ray scattering (SAXS) studies of structural features of DEAD-box proteins that allow them to efficiently interact with large structured RNA substrates. Finally, I will turn to studies of DEAD-box protein involvement during early steps of RNA compaction and folding, using SAXS and activity-based approaches (Chapter 5). I will conclude with a general discussion of superfamily 2 RNA helicases, which include DEAD-box and related proteins, and their functions and mechanisms as remodelers of structured RNAs and RNPs.

Table of Contents

List of Tables

List of Figures

Chapter 1 has been adapted from and contains figures reproduced from the following reviews, which were co-written with Rick Russell and Alan Lambowitz:

Jarmoskaite, I. and Russell, R. (2011) DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdisciplinary Reviews RNA* 2(1): 135-52.

Russell, R., Jarmoskaite, I., and Lambowitz, A. M. (2013) Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biology* 10(1): 44-55.

Chapter 1: DEAD-box proteins as RNA helicases and chaperones

INTRODUCTION

DEAD-box proteins are the largest family within the superfamily 2 (SF2) helicases and play important roles in RNA metabolism in all three domains of life. They perform a diverse set of functions in an even more diverse set of cellular processes, including assembly and function of the ribosome and spliceosome, translation initiation, nuclear export of mRNA, folding of self-splicing RNA introns, and others (1, 2) (Figure 1.1 and see Chapter 6). DEAD-box proteins share with other SF2 proteins a helicase core that is composed of two RecA-like domains and contain a common set of sequence motifs, including the eponymous D-E-A-D sequence in motif II (Figure 1.2) (3). Since their discovery 25 years ago(4), the central defining biochemical activity of DEAD-box proteins has been their ability to bind and hydrolyze ATP in a cycle that is stimulated by binding to single-stranded or double-stranded RNA (ssRNA or dsRNA). Subsequent work has shown that the energetic communication implied by the stimulation of ATPase activity by RNA binding forms a common set of properties that is shared by DEAD-box proteins despite their diverse activities. Progression through the ATPase cycle leads to changes in RNA affinity, allowing DEAD-box proteins to bind and release RNA in a regulated manner. This regulated binding can then be used by different DEAD-box proteins to separate the strands of short duplex regions of RNA, to promote rearrangements of complex, structured RNAs, to remodel RNA-protein (RNP)

complexes, or even to form a highly stable complex on ssRNA. Some DEAD-box proteins are tuned to operate in the context of specific RNA or RNP substrates and even to promote specific rearrangements, whereas others function more broadly, interacting productively with a range of substrates. This chapter first outlines the basic properties of DEAD-box proteins as RNA strand separators or helicases and describes how the fundamental mechanistic feature, ATP-dependent cycling of RNA affinity, allows the proteins to perform this function. Then, it reviews the work that probes how DEAD-box proteins use their RNA helicase activity to promote RNA structural rearrangements and folding events, focusing on self-splicing introns as a model.

THE DISTINCTIVE RNA HELICASE ACTIVITIES OF DEAD-BOX PROTEINS

Despite extensive structure and sequence homology, DEAD-box proteins show important mechanistic differences from conventional helicases. They act efficiently on short RNA helices or RNA-DNA heteroduplexes, and this basic activity is certain to underlie many of their functions. The properties of unwinding differ in fundamental ways from other helicases, both in terms of the features required of the substrates and the mechanisms of unwinding.

Declining efficiency with increasing helical length and stability

The general mechanism that conventional helicases use to separate the strands of a duplex is to bind at the end of the helix and use energy from cycles of ATP binding and hydrolysis to move, or translocate, directionally along one of the strands (5, 6). Most SF1 and SF2 helicases translocate in the 3′ to 5′ direction. By gripping tightly to one of the strands as they translocate, helicases cause the other strand to be displaced. With each cycle, there is a finite probability that the helicase will dissociate from the strand prematurely, before completely unwinding the helix. The probability of continuing

onward rather than dissociating defines the processivity, with a more highly processive helicase being more likely to continue translocating. Therefore, a common means of evaluating the processivity of a helicase is to examine how the fraction of successful unwinding events decreases as the helix length increases.

When evaluated in this way, the DEAD-box proteins that have been tested are poor helicases. Helicase processivity is commonly measured in the hundreds or thousands of base pairs, but the DEAD-box proteins become ineffective when the helix length rises above 10-15 base pairs, and above 20-25 base pairs unwinding is often not detected at all (7-10). Further, unlike most helicases, the efficiency of strand separation for DEAD-box proteins is strongly dependent on the stability of the helix, not just its length. This behavior was first demonstrated for mammalian eIF4A, a small DEAD-box protein involved in translation initiation. It was observed that eIF4A can readily unwind certain short duplexes but is much less efficient for duplexes that are rich in G-C pairs and calculated to be more stable (11, 12). An analogous correlation has been suggested for at least one other DEAD-box protein (13).

While the apparent lack of processivity makes DEAD-box proteins rather inefficient in comparison to other helicases, it is important to note that the helical elements within structured RNAs are rarely longer than 10 base pairs. Thus, processivity may not be required in proteins that function by rearranging structured RNAs. In fact, duplexes that are unwound efficiently are so short that a single protein monomer could bind along most of the length of the duplex, raising the possibility that strand separation by DEAD-box proteins does not involve any translocation at all. This would indeed represent a substantially different mechanism from conventional helicases (11, 13-17). Notably, ancillary domains that are known to be involved in processive translocation in

other SF2 and SF1 helicases are absent from DEAD-box proteins, providing structural insight into the apparent lack of translocation in DEAD-box proteins (18).

Enhancement of unwinding by RNA extensions or appendages

Besides processivity, a second general feature of conventional helicases is that most have a strong preference for helices that include a single-stranded extension (6). The role of the extension is to provide an initial site of interaction, from which the helicase can begin to translocate into the helix. Because most helicases translocate from 3′ to 5′, they require that the extension include a free 3′ end (Figure 1.3*A*).

In this regard as well, DEAD-box proteins behave quite differently from most other SF1 and SF2 helicases. First, some DEAD-box proteins are as efficient at separating the strands of a blunt-ended helix as the same helix with an extension, as first demonstrated for eIF4A (12). The equivalent activity on a blunt-ended duplex suggests that eIF4A can initiate unwinding by contacting a double-stranded region, most simply suggesting that unwinding could begin internally, within a duplex. Such a mechanism would be very unlikely to succeed for long duplexes, even for a processive helicase, because initiating unwinding from an internal region of the duplex would leave intact the region of the duplex that was not in the preferred direction of translocation. Nevertheless, this could be a viable strategy for short duplexes. The protein could generate local strand separation within the helix, and any base pairs that remained intact at one or both ends of the helix could dissociate spontaneously (12). Indeed, it was demonstrated directly that the *Saccharomyces cerevisiae* proteins Ded1 and Mss116 can initiate an unwinding reaction by interacting with an internal region of an RNA duplex (19).

Second, many DEAD-box proteins are enhanced for unwinding by extensions, but unlike the behavior of conventional helicases, the extension can be beneficial for

unwinding regardless of its polarity (Figure 1.3*B*) (8-10, 14, 19-21). This non-polar activation argues strongly against a mechanism that requires directional translocation. Although a mechanism involving random, bi-directional translocation would not be ruled out, further data suggest that the extension is not used as a starting point for translocation at all, but rather as an additional handle to maintain an interaction with the duplex. Thus, the *Neurospora crassa* protein CYT-19, which functions in folding of mitochondrial group I and group II introns (see *RNA chaperone functions of DEAD-box proteins* below), is activated by double-stranded extensions at least as well as single-stranded ones, and extensions composed of DNA are as activating as RNA extensions (Figure 1.3*C*) (22). DNA extensions also activate Ded1, and this activation is preserved even if the extension is not covalently linked to the duplex but is instead held in proximity through linkages to streptavidin (Figure 1.3*D*) (23). Finally, activation of CYT-19 is even greater when the helix is attached to a highly structured group I intron RNA (22). Together, these observations lead to a model in which the role of the extension, whether it be single-stranded, double-stranded, or higher-order RNA structure, is to provide a site of interaction that can localize the DEAD-box protein in position to disrupt nearby RNA structure.

Targeting DEAD-box proteins to certain RNAs or RNPs

Many DEAD-box proteins have specialized functions that require them to act on specific RNAs or RNPs. In analogy with the DEAD-box proteins described above that bind to helix extensions with low or no specificity, DEAD-box proteins that function specifically are thought to be directed by forming additional interactions with their targets. These interactions can be formed in two general ways, either through specific

surfaces within the helicase core or by ancillary domains adjacent to the helicase core (Figure 1.4).

One of the best-understood examples of DEAD-box protein targeting to a specific RNA is the *E. coli* protein DbpA (YxiN in *Bacillus subtilis*), which functions in biogenesis of the large ribosomal subunit and uses a C-terminal ancillary domain to recognize a specific hairpin within the 23S rRNA (Figure 1.4*A*) (24-27). The small C-terminal domain adopts the fold of an RNA recognition motif (RRM) (28). The connections between the domains are flexible, suggesting that a specific interaction of the C-terminal domain with its target helix would allow the helicase core to be localized and yet still be sufficiently flexible to interact with multiple target sites nearby on the assembling ribosomal subunit (26, 29). The *Thermus thermophilus* protein Hera also binds structured RNAs through an RRM-like domain, but in a less specific way (30-32). Analogously, CYT-19, Mss116 and other DEAD-box proteins such as Ded1 and SrmB possess a basic 'tail' region at the C-terminus that interacts with structured RNAs with no apparent specificity (16, 33, 34) (see Chapter 5) (Figure 1.4*B*). Another means of non-specific targeting may be multimerization, which could allow additional interactions to be formed by the helicase core of an interacting subunit (Figure 1.4*C*) (19). A dimerization domain is present in Hera and may contribute to binding of several different structured RNAs or interactions with large RNA substrates (30, 35-37).

Several DEAD-box proteins are localized to their substrates through specific protein-protein interactions (Figure 1.4*D*). eIF4A, which is thought to disrupt 5′ UTR structures during translation initiation, is targeted to the 5′ UTR through interactions with eIF4G, a central component of the mRNA cap-binding complex (38, 39). eIF4G forms extensive contacts with the helicase domains of eIF4A and in addition to a role in localization can modulate the ATPase and unwinding activities (38-41) (see *ATPase*

cycle kinetics below). The C-terminal RecA-like domain (domain 2) of eIF4A can further bind the translation initiation factors eIF4H and eIF4B (39). eIF4H was shown by single-molecule fluorescence to tether eIF4A to hairpin structures, allowing the DEAD-box protein to perform repeated cycles of unwinding without dissociating (42). Another well-characterized example of protein-mediated localization of a DEAD-box protein is Dbp5 (DDX19 in humans), which functions in nuclear export of mRNAs and is localized to the nuclear pore through interactions with the nucleoporin Nup159 (NUP214 in humans) (43) and also forms interactions with Gle1 (44). Both of these proteins also regulate the ATPase cycle of Dbp5 (see below). Protein-protein interactions are also important for the localization and function of the exon-junction complex (EJC) component eIF4AIII, which forms extensive contacts with other EJC proteins and can also interact with the nonsense-mediated decay protein UPF3b and the splicing factor CWC22 (45-48). Interestingly, the same region in domain 2 of the DEAD-box core is contacted by eIF4G, UPF3b, CWC22, Gle1 and at least one other protein (49). Overall, these results suggest that specific interactions between the helicase core and protein components of RNPs are a common mechanism for localization of DEAD-box proteins.

Tethered strand separation without translocation

In combination, the biochemical properties described above suggest a model for RNA unwinding by DEAD-box proteins that is quite different from the mechanism of conventional helicases. A tethering interaction holds the DEAD-box helicase core in proximity to targeted RNA duplexes, allowing the core to generate local separation of the strands. Different DEAD-box proteins could use this basic mechanism to promote RNA remodeling, protein displacement, or to facilitate disruption of RNA tertiary contacts, as the critical event is simply the tight, regulated binding of RNA. A critical distinction of

this mechanism relative to conventional helicases, which may also be localized via interactions with other proteins, is that for the non-processive actions of DEAD-box proteins, the tethering interaction could remain intact through the entire process. Thus, a DEAD-box protein could remain positioned to disrupt the same structure repeatedly or to facilitate additional structure disruptions of intermediates that form after the initial disruption of structure.

COUPLING ATP BINDING AND HYDROLYSIS TO RNA STRAND SEPARATION

The helicase core of DEAD-box proteins consists of two RecA-like domains that include a series of conserved motifs, lining the protein surfaces that bind RNA and an adenosine nucleotide. Crystallography, fluorescence and SAXS studies have shown that the two domains are very flexible relative to each other in the absence of bound ligands, adopting highly extended, or 'open', conformations in several DEAD-box proteins (34, 50-53). In striking contrast, the domains become packed together, or 'closed', in structures with bound RNA and ATP analog, as shown in Figure 1.2 (34, 44, 45, 47, 48, 52, 54-57). This movement is also supported by solution studies in which conformational changes were detected by sensitivity to protease (58). The domain closure is thought to underlie the cooperative binding of ssRNA and ATP (see below), and intimate energetic communication between the bound RNA and nucleotide leads to changes from cooperativity to anti-cooperativity during the ATPase cycle. The sections below describe the molecular basis for RNA and nucleotide binding by DEAD-box proteins and describe the current understanding of how energy from ATP binding and hydrolysis is transduced to the RNA binding site to promote strand separation.

Molecular basis for RNA and ATP recognition by DEAD-box proteins

Important insights into the mechanism of RNA strand separation by DEAD-box proteins came from a series of crystal structures of the helicase core domains bound to one or both substrates. Early structures of DEAD-box proteins bound to ssRNA and ATP analogs revealed that the conserved motifs of the helicase core line the interface between the two domains, shaping the binding sites for the RNA and adenosine nucleotide (Figure 1.2) (44, 45, 47, 48, 54-56, 59). ATP is contacted primarily by conserved amino acids in domain 1 and can be bound by isolated domain 1 (60-64). The specificity for adenine is provided by Q motif, which contacts the adenine base through a conserved glutamine residue, while the side chain of a phenylalanine stacks with the base (65, 66). Motif I contains a P-loop, which is found in many nucleotide-binding proteins and interacts with the triphosphate moiety. Amino acids from the D-E-A-D sequence within motif II also form contacts with the phosphates. Upon domain closure, motif VI within domain 2 interacts with the triphosphate, while motif Va contacts the ribose moiety. Mutations in motifs I, II, and VI weaken or abolish nucleotide binding and hydrolysis (67-75). Motif III, located between the RNA and ATP binding sites, participates in positioning of the γ-phosphate and has recently been implicated in coordinating the other conserved motifs for formation of a high-affinity RNA-binding site in the ATP-bound state and for hydrolysis of ATP (76).

Structural basis for dsRNA recognition was recently established by the Lambowitz lab (64). Domain 2 of Mss116 can bind dsRNA with the same affinity as the full-length helicase core, suggesting that it mediates initial dsRNA binding (64). Crystal structures of domain 2 bound to double-stranded nucleic acid substrates revealed a pocket for A-form helices, with conserved amino acids from motifs IV, IVa, V and Va forming extensive contacts with the phosphate groups of one of the RNA strands. The second

strand is only contacted by the C-terminal extension (CTE), which is present in a subset of fungal DEAD-box proteins, including Mss116 and CYT-19. The specificity for A-form helices provides an explanation for why DEAD-box proteins can unwind dsRNA and RNA/DNA hybrid helices, but not DNA helices, while lack of contacts with the RNA bases is consistent with non-sequence specific binding (64).

Crystal structures of ternary complexes with bound ssRNA and nucleotide, representing the post-unwinding state, have provided important insights into the mechanism of RNA strand separation by DEAD-box proteins. The ssRNA lies along a basic crevice that extends across both domains and contains contributions from both domain 2 motifs IV, IVa, V and Vb (forming essentially identical interactions as those observed with isolated domain 2 above) and domain 1 motifs Ia, Ib and Ic (44, 45, 47, 48, 54-56, 59, 64). These contacts are exclusively to the sugar-phosphate backbone of the RNA, including several interactions with 2´OH groups. In all available structures that include bound ssRNA, a domain 1 helix that contains motif Ic imposes a sharp bend in the RNA backbone, and in Mss116 structures a second bend is introduced by the C-terminal extension of domain 2 (55). In addition to deformation of the backbone of the bound strand, the second strand is sterically excluded from the RNA binding cleft. Together, these observations suggest that during ATP- and dsRNA-dependent domain closure, domain 1 collides with the RNA helix, resulting in displacement of one strand and tightened binding of the other strand. This structure-based model is consistent with local, non-processive unwinding and predicts that binding of a single ATP molecule should be sufficient for unwinding of a short helix — a prediction that has indeed been experimentally validated (see below).

Cooperative binding of RNA and adenosine nucleotide

At the heart of the ATP-dependent activities of DEAD-box proteins is a cycle of RNA affinity changes that is tightly linked to the ATPase cycle. The effect of these affinity changes is to generate strong, yet regulated binding to a strand of RNA, displacing this segment from a partner strand and temporarily preventing it from forming intramolecular or intermolecular contacts with other RNA segments or proteins components. Indeed, cooperativity of binding to ssRNA and adenosine nucleotides has been observed for many DEAD-box proteins. For eIF4A, cooperativity was demonstrated by steady-state ATPase activity (77) and for *E. coli* DbpA and other DEAD-box proteins — by binding assays using nucleotide analogs such as AMP-PNP (52, 71, 78-80). A critical feature of the ATPase cycle is that the level of cooperativity changes dramatically at different stages, regulating the affinity for RNA. In general, positive cooperativity is associated with the presence of the γ-phosphoryl group of the adenosine nucleotide (Figure 1.5). Thus, in the presence of ATP or an ATP analog binding of RNA is typically tightened. However, a limitation of ATP analogs is that they do not mimic all of the binding properties of ATP, as most do not stimulate RNA unwinding (13, 17, 81), so it is unclear whether the level of cooperativity is the same as with ATP. In contrast, ADP binding is typically observed to give anti-cooperativity (77, 79) or energetically independent binding of RNA (82).

Detailed, quantitative characterization of the energetic coupling between nucleotide and RNA binding at each stage of the ATPase cycle was recently performed for Mss116 (Figure 1.6) (83). Fluorescence approaches with a truncated protein construct that includes the tandem RecA-like core domains and the CTE but lacks the C-tail and a short N-terminal extension (NTE) showed that ATP binding increases the affinity for a small double-stranded, hairpin RNA (2–7-fold). In contrast, RNA binding is negatively

coupled with ADP binding (3-fold). Further experiments showed that RNA accelerates hydrolysis of ATP and suppresses its resynthesis on the enzyme, such that there is a dramatic shift of the equilibrium toward products. This knowledge allows calculation of the affinity for RNA when the protein is bound to ADP and P_i (denoted the ADP-P_i state), which indicates very tight RNA binding (140–500-fold tighter than with ATP). The RNA affinity cycle that emerges has important parallels with that derived from the same approaches for DbpA (82). For both proteins, the tightest binding of RNA occurs in the ADP-P_i state and is substantially weaker in the ATP and ADP-bound states. Unlike Mss116, cooperativity with ATP and anti-cooperativity with ADP were not detected for DbpA.

All current models link positive cooperativity to the closing of the two core domains (36, 54). In the closed conformation, formed upon binding of RNA and ATP or an ATP analog, both ligands form contacts with both domains, and thus the domain closure is highly likely to be the main source of cooperativity (54). The increase in cooperativity in the transition from the ATP to the ADP-P_i state most likely reflects rearrangements that occur upon ATP hydrolysis during or after core closure and affect the affinity for the bound RNA. The key predictions from this model are that first, ligand combinations that do not give cooperative binding should not give stable formation of the correct closed conformation. Consistent with this prediction, ADP and RNA binding to DbpA do not give a stable, closed conformation as viewed by FRET (52). Second, the cooperativity of binding should be particularly sensitive to changes at the domain interface. Indeed, mutations in motifs III and Va, both of which line the domain interface, have been observed to reduce the cooperativity without substantially weakening binding of the individual ligands (75, 76).

ATPase cycle kinetics

A general picture of the ATPase kinetics of DEAD-box proteins has emerged from extensive data sets for DbpA and Mss116 (Figure 1.6) (82, 83), along with previous work on eIF4A and the yeast protein Ded1, which functions in the nucleus and cytosol in processes including RNA splicing and translation (84). While the absolute rate constants will differ for different proteins and may depend on the specific RNA substrate and on solution conditions, there are likely to be general themes for the relative rate constants that are useful as a starting point. ATP and RNA can bind in either order (77). The affinity for ATP is typically between 50 and 500 μM in the absence of RNA (77, 82, 83, 85, 86) and can vary across this range for different proteins and even for the same protein under different solution conditions (77). In the absence of RNA, ATP hydrolysis is slow (< 0.3 s^{-1}), such that the equilibrium constant for binding is expected to be equal to the K_M measured by steady state ATPase activity, and thus the affinity estimates above include K_M values. Upon binding RNA, ATP hydrolysis is stimulated to 5–10 s^{-1}, generating an ADP-P_i state that remains bound to RNA. As noted above, RNA affinity is maximized in this state. With bound RNA, ATP hydrolysis is largely irreversible because P_i release is faster than ATP re-synthesis, but release of P_i is slow enough to be at least partially rate-limiting for the overall ATPase cycle (2–6 s^{-1}). Therefore, under conditions of saturating ATP and RNA the protein would be expected to spend a significant fraction of the time in the tight-binding, ADP-P_i state (82, 83). Upon release of P_i, the affinities for RNA and ADP are dramatically reduced. ADP is released rapidly (30–100 s^{-1}), and RNA presumably is as well. At this point, the protein can re-bind weakly to RNA and can again bind ATP to begin another cycle. It may also re-bind ADP, which for some DEAD-box proteins binds substantially tighter than ATP (77, 82). In vivo, at cellular

concentrations of ADP, the net exchange of ADP for ATP may be partially rate-limiting for the overall cycle for some DEAD-box proteins (82).

For many DEAD-box proteins, another level of regulation in the ATPase kinetics is conferred by interactions with partner proteins. One of the most extreme examples is regulation of eIF4AIII within the EJC complex. When bound to mRNA and the other protein components of the EJC, Magoh and Y14, eIF4A-III is arrested in the ADP-P_i state, resulting in a very stable complex (87). This allows EJC to remain stably bound to the spliced RNA, which is required for its function in nonsense-mediated decay (88). Interactions of eIF4A with other components of the translation initiation machinery accelerate the ATPase cycle (39, 89-92). Co-crystal structures indicate that the binding of eIF4G promotes a conformation of the two core domains of eIF4A that is neither fully open nor fully closed (41, 89), presumably increasing the rates of nucleotide exchange and P_i release (89). Interestingly, Ded1 also forms a complex with eIF4G, but the complex is arrested rather than accelerated through the ATPase cycle (93). In another example, interaction between Dbp5 and Nup159 at the nuclear pore accelerates ADP release by Dbp5, which is suggested to be inherently slow (94). Gle1 interacts strongly with Dbp5 in the presence of the cofactor IP_6 and in the crystal structure of the complex with Gle1-IP_6 the two core domains of Dbp5 are separated, disrupting the RNA binding site (44). Thus, Gle1-IP_6 binding is strongly anti-cooperative with RNA binding, which may result in rapid release of Gle1-IP_6 upon binding of RNA (94) and/or acceleration of RNA release to allow recycling of Dbp5 (44). On the other hand, Gle1-IP_6 binding is cooperative with nucleotide binding, perhaps contributing to the requirement for Nup159 to accelerate ADP release (94).

ATP utilization during RNA unwinding

An important goal for a mechanistic understanding of DEAD-box proteins has been to understand how the ATPase cycle is coupled RNA unwinding. Considerable progress has been made in the last few years, demonstrating for a series of DEAD-box proteins that unwinding of short duplexes can be completed with just binding of ATP, while ATP hydrolysis is required primarily to recycle the enzymes (13, 17).

Studies using Ded1, eIF4A, Mss116 (17) and later – YxiN (81) found that the non-hydrolyzable ATP analog ADP-BeF_x, although not other analogs, is sufficient for unwinding of a short helix in a single-turnover reaction. However, the extremely tight binding of RNA in the presence of ADP-BeF_x prevents efficient unwinding under multiple-turnover conditions, leading to the suggestion that ATP hydrolysis is required to generate the weaker binding states that allow release of the bound RNA strand and enzyme recycling (17). In a separate study, ATP hydrolysis and RNA unwinding were measured in parallel to determine the number of ATPs consumed during unwinding (13). A stoichiometry of one ATP per duplex unwound was observed for short duplexes under a range of conditions, consistent with the above finding. Furthermore, at a low Mg^{2+} concentration (2 mM), which destabilizes the helix, only half of unwinding events were accompanied by ATP hydrolysis, providing further support for the conclusion that unwinding precedes ATP hydrolysis and requires only ATP binding. Together, these results provide strong evidence for a simple coupling between the ATPase cycle and the RNA binding cycle (19, 23). A model in which unwinding occurs in a single cycle of ATP binding and hydrolysis is consistent with the known ability of most DEAD-box proteins to unwind short RNA helices but not helices longer than 10–15 bp (7, 8, 12). More ATP is hydrolyzed during unwinding of longer or more stable helices, indicating that ATP hydrolysis does not always lead to complete unwinding of such duplexes and

that some ATPase cycles are futile (13). Other studies have since supported hydrolysis of a single ATP during unwinding of short RNA helices by DEAD-box proteins (83, 95).

Overall, the model that emerges is that binding of ATP and dsRNA can lead to a conformational change that generates local strand separation. For a short, relatively unstable duplex, this conformational change and resulting strand separation can be sufficient to generate complete unwinding, and the strand that is not tightly bound by the DEAD-box protein dissociates. The conformational change and strand separation would follow ATP and RNA binding and be slower than these processes, but it would be faster than ATP hydrolysis. The rate and extent of this conformational change could also depend on the stability of the RNA duplex, because this would influence the energy barrier that must be overcome. Multi-phase kinetics have indeed been observed for the binding of nucleotides (82, 83), supporting models in which strand separation is linked to a conformational change that follows ATP binding.

OTHER MECHANISMS OF DEAD-BOX PROTEINS DURING RNA REARRANGEMENTS

While it is clear that ATP-dependent unwinding is important for the physiological functions of many DEAD-box proteins (see Chapter 6), DEAD-box proteins can also use other mechanisms. Several DEAD-box proteins have been shown to possess annealing activity, which is the reverse of the unwinding reaction and is ATP-independent (7, 21, 33, 96, 97). A prominent example of a DEAD-box protein that functions as an annealer is Rok1, which is essential for ribosome biogenesis in yeast and can transiently stabilize a helix that is required for correct processing of small-subunit rRNA (98, 99). However, Rok1 has also been implicated in processes that require unwinding (100)

The RNA substrates of many DEAD-box proteins form complexes with RNA-binding proteins in vivo, and the function of some DEAD-box proteins may be to

remodel ribonucleoprotein complexes by dissociating proteins from the RNA. The ability to dissociate proteins bound to RNA has indeed been demonstrated in vitro for Ded1 (101, 102) and several other proteins, including Dbp5 (103) and Sub2 (104) are thought to function by dissociating proteins from their physiological RNA substrates. Together, these examples illustrate the versatility of DEAD-box proteins as RNA and RNP remodelers.

RNA CHAPERONE MECHANISMS OF DEAD-BOX PROTEINS

In vivo, many DEAD-box proteins function as chaperones in folding and assembly of structured RNAs and RNPs, such as the spliceosome, the ribosome and self-splicing introns. Thus, there has been considerable interest in understanding how DEAD-box proteins integrate their basic RNA binding and unwinding activities during rearrangements of complex RNA structures, which may include tertiary structure and bound proteins. This section will focus on studies of the DEAD-box proteins CYT-19 and Mss116, which have served as model systems for delineating the RNA chaperone mechanisms of DEAD-box proteins. For a broader discussion of chaperone activities of DEAD-box and related helicases, see Chapter 6.

RNA misfolding and chaperone functions of CYT-19 and Mss116

RNA molecules fold through rugged free energy landscapes and can populate long-lived alternative structures in addition to their functional structures (105-107). Starting with pioneering studies of tRNA and 5S rRNA (108, 109), misfolding has been recognized as a dominant theme in RNA folding, and misfolded intermediates have been demonstrated for several group I introns (110-116), the bacterial RNase P RNA (117) and the hepatitis delta virus ribozyme (118), among others. While also observed in proteins, the problem of misfolding is particularly pronounced in RNA and is primarily a result of

the extremely high stability of RNA secondary structure. Locally base-paired regions can form and persist independently from the global tertiary structure, leading to kinetic traps that may take hours to unfold spontaneously. Therefore, in order to fold on biologically meaningful time scales, RNA molecules often require RNA chaperones, many of which function by destabilizing RNA structure, accelerating transitions between conformations (106, 119).

DEAD-box proteins constitute a major family among the diverse group of proteins that function as RNA chaperones (119). Mechanistically, DEAD-box proteins stand out among other RNA chaperones for their ability to undergo regulated, ATP-dependent cycles of RNA binding and unwinding, conceptually resembling ATP-dependent chaperones that function in protein folding, such as Hsp70 (120). The *Neurospora crassa* DEAD-box protein CYT-19 was the first biological RNA chaperone to be discovered and has since become a model system for studying DEAD-box protein-mediated rearrangements of structured RNAs (85). CYT-19 is required for efficient splicing of three mitochondrial group I introns, and its yeast homolog Mss116 promotes splicing of all nine mitochondrial group I introns and all four group II introns (121). Group I and group II introns are highly structured, autocatalytic RNAs that function by splicing themselves out of precursor RNAs (122, 123). Representatives of both groups have been shown to misfold substantially, suggesting that CYT-19 and Mss116 may function by resolving their misfolded conformations (124, 125). The diversity of their physiological substrates indicates that CYT-19 and Mss116 are general RNA chaperones and this is further supported by the results that CYT-19 can suppress the defects from loss of functional Mss116 in yeast and that it can promote folding of a series of non-physiological group I and group II RNAs in vitro (116, 121, 126).

Group I intron structure and folding

The well-characterized group I intron from *Tetrahymena thermophila* has provided a valuable model system for probing the chaperone mechanisms of DEAD-box proteins (16, 22, 85, 127-129). The intron is comprised of a conserved catalytic core and a series of peripheral elements and structurally resembles the physiological group I intron substrates of CYT-19 and Mss116 (130-132) (Figure 1.7). The group I intron core consists of two extended helical domains (the P4-P6 and P3-P9 domains) (130, 133), and these domains interact to form the active site, which catalyzes a two-step splicing reaction (131, 134). The core is encircled and stabilized by a series of peripheral elements, which form long-range tertiary interactions with each other and with the core RNA (132) (Figure 1.7).

Most folding studies have used a ribozyme construct of the *Tetrahymena* group I intron, in which the exons are removed, leaving an RNA that can bind and cleave a substrate oligonucleotide in trans, in a reaction that mimics the first step of splicing (135). The advantage of such constructs is that the fraction of native, catalytically active ribozyme can be determined in a robust way by measuring the fraction of substrate that is cleaved rapidly upon addition at various folding times and conditions (136, 137). Chemical footprinting and catalytic activity approaches showed that the *Tetrahymena* ribozyme folds through a series of intermediates (138, 139), and most of the population becomes trapped in a long-lived, extensively structured misfolded intermediate (113, 140-143), which refolds to the native state on the time scale of hours under typical in vitro conditions (144, 145).

The misfolded intermediate is remarkably similar to the native ribozyme, forming a complete native network of secondary and tertiary interactions and a globally compact fold, with only local differences in solvent accessibility observed within the core (144,

145). Despite these similarities, refolding to the native state was shown to require extensive unfolding, including disruption of all five peripheral tertiary contacts and the long-range core helix P3 (145, 146). Thus, a topological error was proposed, in which two single-stranded segments between the core domains are crossed incorrectly in the misfolded ribozyme (145, 146). Correction of such a topology error would involve large-scale motions between the core domains and thus would require transient global disruption of ribozyme tertiary structure and the P3 helix, consistent with the experimental results (145, 146).

Mechanisms of RNA chaperone activity of CYT-19

CYT-19 accelerates the refolding of the misfolded ribozyme in an ATP-dependent manner and is thought to do so by disrupting the ribozyme structure (127, 147). Consistent with its non-specific nature, CYT-19 can also unfold the native ribozyme (127). However, the unfolding activity is much less efficient for the native ribozyme apparently because this structure is more stable, as indicated by the enhancement of CYT-19-mediated unfolding by mutations in the ribozyme that weaken tertiary structure (127) (see Chapter 2). Thus, CYT-19 efficiently disrupts the misfolded ribozyme, while allowing the native ribozyme to accumulate, leading to an overall effect of acceleration of misfolded ribozyme refolding (Figure 1.8).

SUMMARY AND OBJECTIVES

The research over the past two decades has led to an in depth understanding of the ATP-dependent helix unwinding activity of DEAD-box proteins. However, mechanistic understanding of how DEAD-box proteins rearrange RNAs that contain tertiary structure has remained incomplete. This is an important limitation, as the RNA molecules that DEAD-box proteins typically encounter in vivo are extensively structured, presenting

DEAD-box proteins with unique challenges such as tertiary packing of RNA helices and cooperativity between covalently linked parts of a larger RNA. The overarching goal of this dissertation has been to establish the mechanistic basis for DEAD-box protein-mediated rearrangements of a highly structured RNA. Using CYT-19 and the *Tetrahymena* group I intron ribozyme as a model system, I first focused on the roles of global tertiary stability in CYT-19-mediated unfolding and targeting to unstable and incorrectly folded RNAs. I then investigated the mechanism of misfolded ribozyme refolding by CYT-19 and, building on the earlier studies of ATP utilization for helix unwinding, measured ATP utilization for refolding of the misfolded group I intron. To understand the structural basis for the ability of CYT-19 to promote global RNA rearrangements, I studied, in collaboration with the Lambowitz lab, the role of a basic C-terminal tail in tethering CYT-19 to its RNA substrates. Finally, I probed the effects of CYT-19 binding and remodeling activities on the initial folding of the group I intron, starting with globally unfolded RNA. Together, this work has advanced the mechanistic understanding of CYT-19-mediated RNA rearrangements, encompassing several stages of group I intron folding, and many of the features described here will likely apply to other members of the DEAD-box protein family.

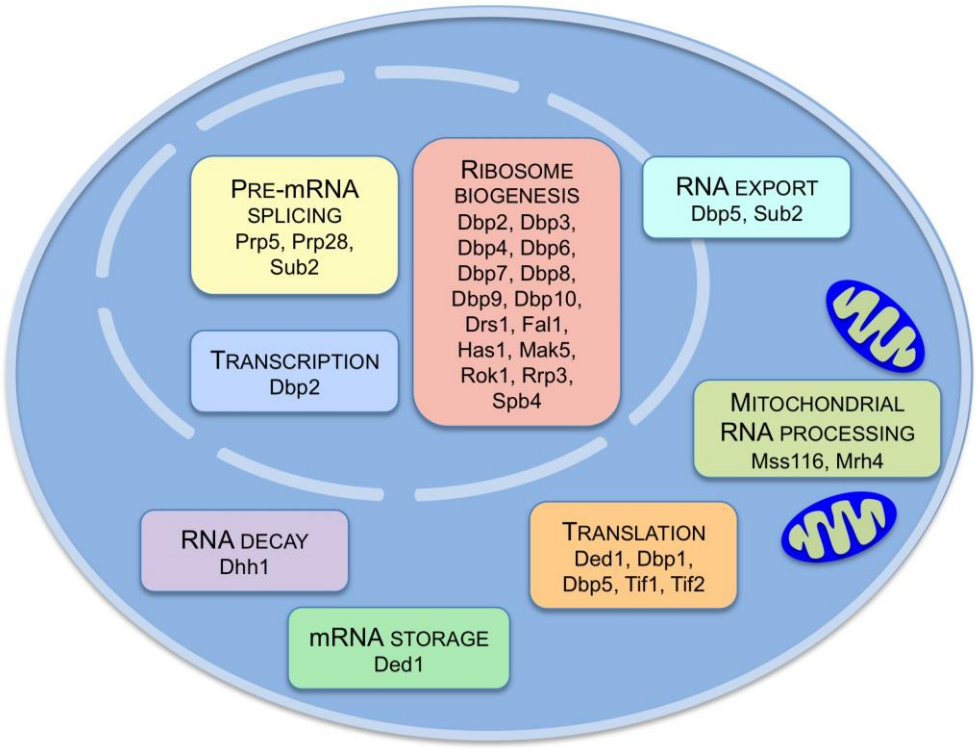


Figure 1.1: Functions of DEAD-box proteins in *Saccharomyces cerevisiae*.

The functions of *S. cerevisiae* DEAD-box proteins are shown as described in Ref. (1) and in more recent primary literature (93, 148). A detailed discussion of the physiological roles of DEAD-box proteins and related RNA helicases can be found in Chapter 6.

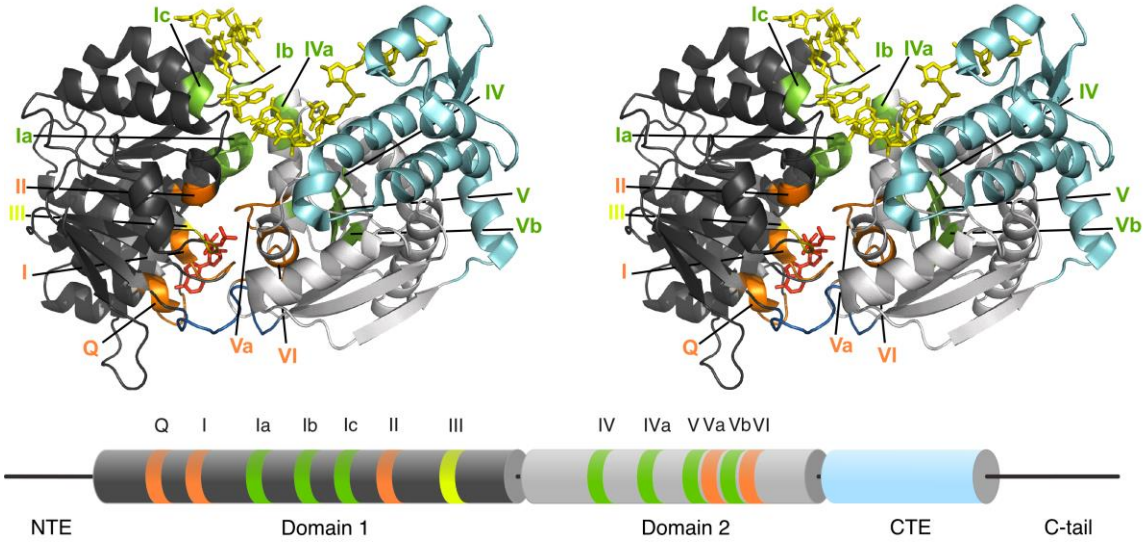


Figure 1.2: Domain structure of DEAD-box proteins.

The top panel shows a stereo view of Mss116 bound to AMP-PNP and ssRNA (U10) (55). Motifs that interact with ATP are orange and motifs that interact with ssRNA are green. Motif III is olive. Domain 1 is dark gray, domain 2 is silver, the linker between them is blue, and the CTE, present in Mss116 and CYT-19, is cyan. Below the structure is a cylinder representation showing the domains and motifs in the same colors. Also shown are the N-terminal extension (NTE) and the C-tail of Mss116, which were not present or not resolved in the crystal structure.

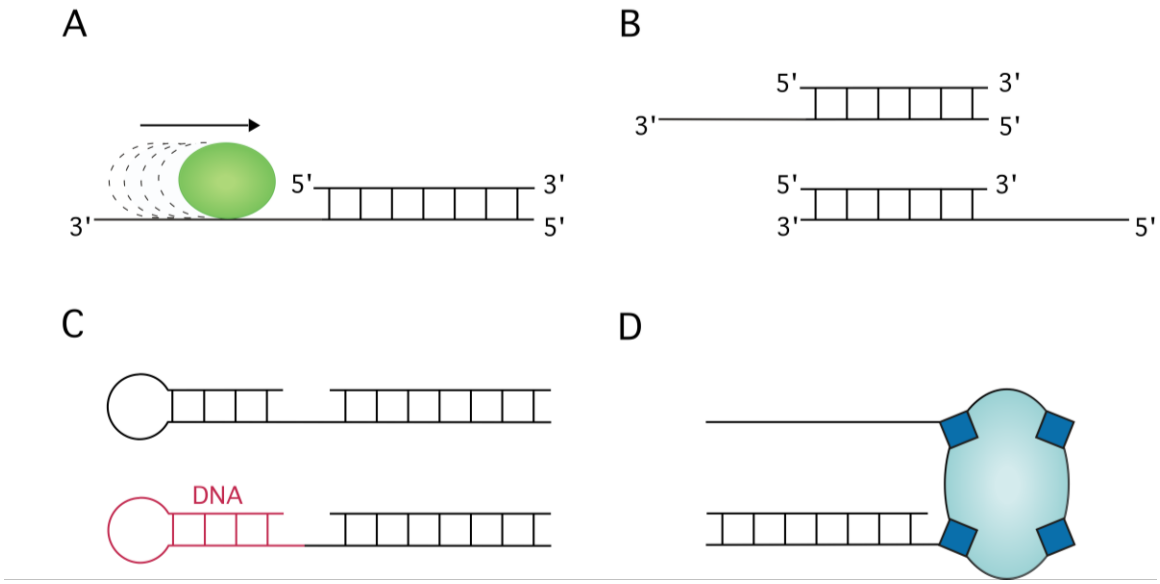


Figure 1.3: Roles of single-stranded and double-stranded extensions in duplex unwinding.

(*A*) Strand separation by a processive, 3′-5′ helicase begins when the helicase loads on a single-stranded extension with a dangling 3′-end. The helicase then translocates in the 3′ to 5′ direction into the adjacent double-stranded region. (*B*) Single-stranded overhangs enhance strand separation by some DEAD-box proteins, but are not required to have a defined polarity with respect to the adjacent duplex. (*C*) Double-stranded RNA or DNA extensions can also enhance unwinding for some DEAD-box proteins. (*D*) A single-stranded RNA or DNA segment can activate RNA unwinding by DEAD-box proteins even when physically separated from the neighboring duplex, as demonstrated when biotinylated RNA constructs were linked to streptavidin (23).

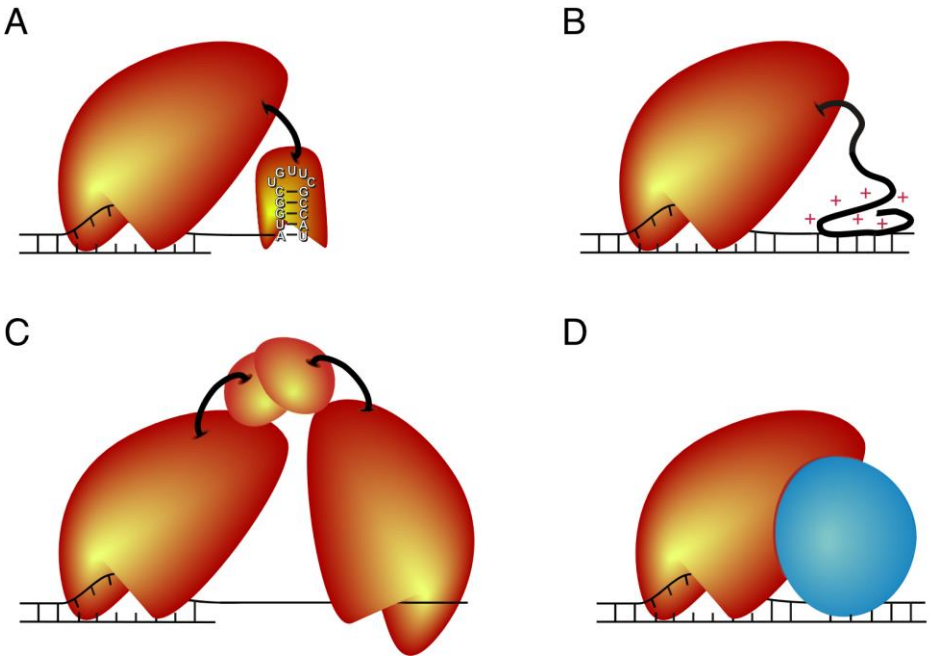


Figure 1.4: Targeting DEAD-box proteins to RNA substrates through specific and non-specific interactions.

(*A*) Recognition of a specific RNA motif via an ancillary domain. Hairpin 92 from the 23S rRNA, shown by nucleotide letters, is recognized with high specificity by the bacterial proteins DbpA and YxiN. (*B*) Electrostatically-driven RNA binding through an ancillary domain or unstructured extension. (*C*) Helicase core loading on RNA substrates via dimerization, with a tethering interaction being formed by the helicase core from one monomer and the unwinding being carried out by the other monomer. (*D*) Targeting through specific interactions between the DEAD-box core and a protein component of the targeted RNP complex.

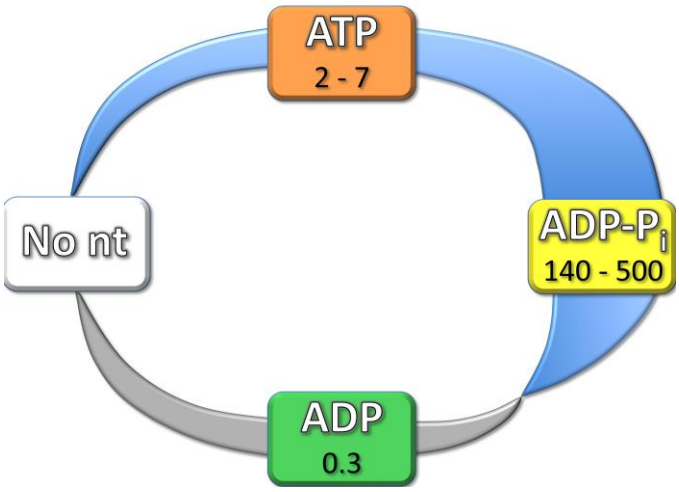


Figure 1.5: RNA affinity changes resulting from changes in cooperativity during the ATPase cycle.

In the absence of nucleotide ("No nt"), cooperativity is absent by definition. Cooperativity is typically observed in the ATP-bound state and is maximal in the ADP-Pi state, as shown by the thickening blue stripe. Anti-cooperativity in the ADP-bound state is shown by the gray stripe. Specific values of cooperativity for Mss116 are indicated for each state (83). The colors for the nucleotide states match those in Figure 1.6.

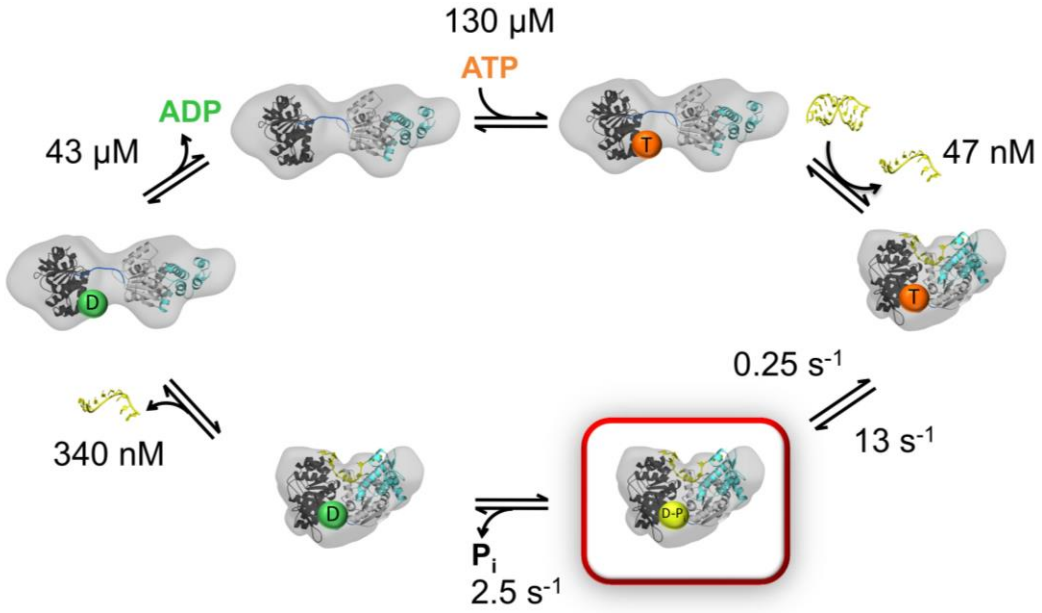


Figure 1.6: ATPase kinetics of DEAD box proteins.

Specific rate and equilibrium constants are shown for Mss116 in the presence of saturating RNA (data from Ref. (83)) The domain structures shown are based on SAXS studies of the open and closed complexes (34). Nucleotide states are abbreviated by T (ATP), D-P_i (ADP and P_i), and D (ADP). The ADP-P_i state is highlighted in red to indicate that it is populated to the greatest extent in the steady state.

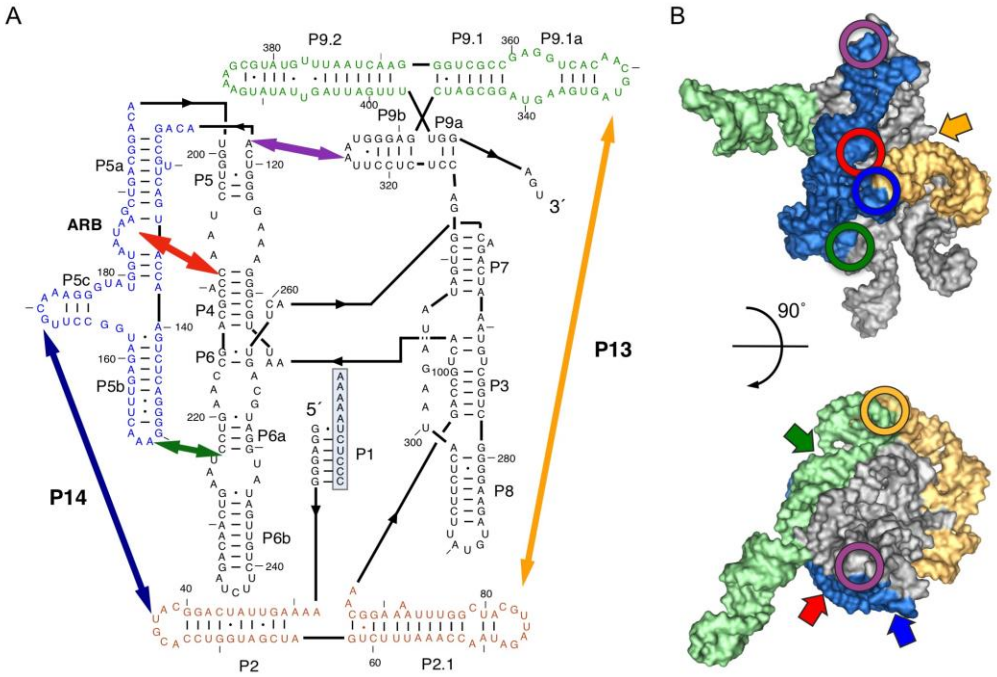


Figure 1.7: The *Tetrahymena thermophila* group I intron ribozyme.

(*A*) Secondary structure and tertiary contacts. The core nucleotides are shown in black, peripheral elements are colored, and bold arrows mark the long-range peripheral tertiary interactions. The boxed nucleotides in the P1 helix indicate the oligonucleotide mimic of the 5´ exon, which is used to measure the catalytic activity of the ribozyme. (*B*) Tertiary structure model of the ribozyme (132). Peripheral elements (colored surface) and the locations of the long-trange peripheral tertiary contacts (circles) are highlighted using the same color scheme as in *A*. The ribozyme core is shown in silver. The block arrows indicate the approximate positions of tertiary contacts not visible in each respective view of the ribozyme. The figures were prepared using PyMOL.

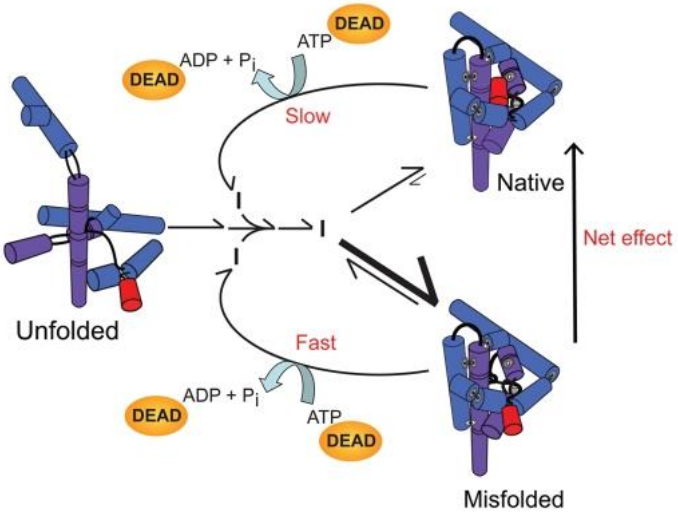


Figure 1.8: Model for refolding of the misfolded *Tetrahymena* group I intron ribozyme by CYT-19.

During initial folding of the *Tetrahymena* ribozyme, misfolding occurs faster than folding to the native state, as indicated by differences in the arrow thickness. This kinetic preference allows the misfolded ribozyme to accumulate despite its lower thermodynamic stability. The DEAD-box protein disrupts the misfolded structure in an ATP-dependent manner, allowing the ribozyme to re-enter the folding pathway from a largely unfolded state. A fraction of unfolded ribozymes enter the native state and because the much more stable native ribozyme is unfolded by the DEAD-box protein inefficiently, the native state accumulates, leading to a net effect of accelerated refolding from the misfolded to the native state. This model was originally published in Ref. (127) and the figure was reproduced from Ref. (149) with the authors' permission.

Chapter 2: The DEAD-box protein CYT-19 is activated by exposed helices in a group I intron RNA

INTRODUCTION

DEAD-box proteins function in all stages of RNA metabolism and have been implicated in assembly and conformational rearrangements of many large structured RNAs and RNPs (3). Thus, it is important to establish how these proteins use their basic mechanisms of RNA binding and helix unwinding to interact with and remodel higher-order RNA structures. The interactions of DEAD-box proteins with structured RNAs have been extensively studied using the *Neurospora crassa* protein CYT-19 from and Mss116 from *Saccharomyces cerevisiae*. In vivo, CYT-19 promotes splicing of several mitochondrial group I introns and can promote splicing of group I and group II introns in yeast mutants that lack functional Mss116 (85, 121). Both proteins have been shown to act as general RNA chaperones during group I and group II intron folding *in vitro* and are thought to act primarily by reversing misfolding of the intron RNAs, although additional mechanisms may be used for some substrates (85, 121, 150-154). Importantly, the chaperone activities of these and other DEAD-box proteins correlate with their ATP-dependent helix unwinding activities, suggesting that DEAD-box proteins function by lowering the energy barriers for transitions between alternative structures that involve disruption of base pairs (128, 129).

In vitro studies using the group I intron ribozyme from *Tetrahymena thermophila* have been instrumental in probing the RNA chaperone mechanism of CYT-19 (22, 85, 127, 155). This ~400 nt RNA folds into a compact, globular structure that consists of a conserved catalytic core and a series of peripheral elements that encircle the core by forming long-range tertiary contacts (Figure 1.7) (131, 132, 156). Upon addition of Mg^{2+}

ions, the majority of the ribozyme population becomes trapped in a long-lived misfolded intermediate, which contains all native secondary and tertiary structure elements, but appears to differ in topology within the core (145, 157). Spontaneous refolding of the misfolded ribozyme is slow, taking hours under most conditions, as it requires transient disruption of all five peripheral tertiary contacts and a core helix (see Chapter 1) (145, 146). Given the structural similarity between the native and misfolded ribozyme, it is interesting that CYT-19 can accelerate refolding of the misfolded intermediate by an order of magnitude, while not detectably unfolding the native ribozyme (147). Insights into this apparent preference for the misfolded ribozyme came from studies of two ribozyme mutants in which the tertiary structure was destabilized, making the stability of the native ribozyme comparable to that of the misfolded intermediate (127). CYT-19 unfolded the native and misfolded conformers of these mutants with comparable efficiencies, suggesting that the efficiency of chaperone-mediated unfolding depended on the stability of ribozyme tertiary structure. However, the mutations studied were concentrated in one region of the ribozyme, leaving open the possibility that CYT-19 recognizes local disruptions rather than global stability.

Here we investigate the roles of RNA stability in CYT-19-mediated unfolding of the *Tetrahymena* ribozyme by using a series of ribozyme mutants with disruptions of each of the five peripheral tertiary contacts. We observe a strong correlation between CYT-19 activity and global stability of ribozyme tertiary structure. Further, we find that the RNA-dependent ATPase activity of CYT-19 depends on the accessibility of secondary structure in the ribozyme. Our results lead to a general model for recognition and remodeling of unstable or incorrectly folded RNAs by DEAD-box proteins.

RESULTS

Efficiency of CYT-19-mediated unfolding increases with lower ribozyme stability

To systematically address the roles of RNA stability in chaperone-mediated unfolding, we used a set of previously characterized *Tetrahymena* ribozyme variants in which the five long-range peripheral tertiary contacts are individually mutated (Figure 2.1) (145, 158, 159). The mutations include substitutions of the A-rich bulge (ARB) of the metal ion core/metal ion core receptor interaction; the tetraloops L9 and L5b, which form two tetraloop/tetraloop receptor (TL/TLR) interactions; the L2 loop of the "kissing loop" interaction P14, and the L2.1 and L9.1 loops of the P13 kissing loop (Figure 2.1). The ARB mutant was one of the ribozymes used in the previous study of CYT-19-mediated ribozyme unfolding (127). Footprinting approaches have shown increased Mg^{2+}-requirements for tertiary folding by each of these and related tertiary contact mutants, indicating decreased global stability (159-161). Here we focused on the native, catalytically active state of each ribozyme, allowing us to accurately quantify the stability and chaperone-mediated unfolding by measuring catalytic activity (127, 136). Although rearrangements of native group I intron structure are presumably not a physiological function of CYT-19, the principles of CYT-19-mediated unfolding are likely to be largely the same for resolving misfolded intermediates, which can strongly resemble their native counterparts.

To determine the stabilities of the mutant ribozymes in the native state, we measured the Mg^{2+}-concentration dependences for equilibrium folding using a two-stage catalytic activity assay (136). The ribozymes were folded to equilibrium at 0–10 mM Mg^{2+} and then transferred to catalysis conditions, which included 50 mM Mg^{2+}, guanosine and an oligonucleotide substrate. The high Mg^{2+} concentration blocks folding transitions from the native state and traps most of the non-native ribozyme in the long-

lived misfolded conformation. Because both the native and misfolded structures can bind the oligonucleotide substrate, but only the native ribozyme can cleave it, the fraction of substrate that is cleaved in a single turnover reflects the fraction of native ribozyme (136).

All of the mutants displayed greater Mg^{2+} requirements than the wild-type ribozyme, consistent with previous determinations by hydroxyl radical footprinting and oligonucleotide hybridization (159, 161-163) (Figure 2.2*A*, Figure A1 in Appendix A). We calculated the native ribozyme stabilities based on the equilibrium values at 2 mM Mg^{2+}, where all of the mutants populated quantifiable ratios of native to non-native ribozyme, avoiding extrapolation-based approaches (see *Materials and Methods*) (164). The native ribozyme stabilities at 2 mM Mg^{2+} ranged across ~3 kcal/mol from the L9 mutant (most stable) to the ARB mutant (least stable). Note that the native wild-type ribozyme is too stable under these conditions for accurate determination of the folding equilibrium.

To measure how stability affects CYT-19-mediated unfolding, we formed native ribozymes at 10 mM Mg^{2+} and initiated unfolding by adding CYT-19 and ATP while diluting Mg^{2+} to 2 mM. Loss of the native ribozyme was measured as above by using the discontinuous ribozyme activity assay (127, 128). We found that CYT-19 accelerated unfolding of all mutant ribozymes, with rates that increased with ATP concentration and with little or no acceleration in the absence of ATP (Figure 2.2*B*, Figures A2-A4) (127). Importantly, the apparent efficiencies of CYT-19-mediated unfolding, derived from CYT-19 concentration dependences of the unfolding rates, tracked with ribozyme stability (Figure 2.2*C*, Figures A2*B*, A3). The differences in unfolding efficiency between the most stable (L9) and the least stable (ARB) mutant were ~70-fold and ~20-fold at subsaturating (100 μM) and near-saturating (2 mM) ATP concentrations. Interestingly,

the L5b tetraloop mutant showed one of the lowest CYT-19-mediated unfolding efficiencies despite an intermediate stability. This mutant also had one of the lowest spontaneous unfolding rates, suggesting that the intrinsic ribozyme propensity to unfold may play a role in CYT-19-mediated disruption (Figure 2.2*C*, Figure A5). Overall, our results indicate that CYT-19 efficiency is tightly linked to the global stability of the ribozyme and suggest a role for spontaneous unfolding events in promoting CYT-19-mediated ribozyme unfolding.

ATPase activity increases with lower tertiary stability

To understand the mechanistic origins of the relationship between CYT-19-mediated unfolding and RNA tertiary stability, we tested two simple models. In the first model, less stable ribozyme mutants would provide more opportunities for productive CYT-19 interactions, for example because their structures are more dynamic and on average less compact at 2 mM Mg^{2+}. In the second model, CYT-19 would perform the same initial rearrangements in all mutants, with less stable ribozymes being more likely to unfold following these rearrangements. We tested these two possibilities by measuring the RNA-dependent ATPase activity of CYT-19 under the conditions of the unfolding assays, with the prediction that the initial ATPase rates should increase with lower ribozyme stability (model 1) or remain constant regardless of stability (model 2).

The ATPase stimulation by each ribozyme was measured immediately following transfer of the pre-folded native ribozyme to 2 mM Mg^{2+} under conditions identical to those of unfolding assays. We also measured ATPase stimulation by the native wild-type ribozyme, which is stably folded under these conditions (see Figure 2.2*A*) and a quintuple mutant, which contains mutations of all five peripheral tertiary contacts and has been shown by SAXS to lack a stable compact structure (165). The ATPase rates showed a

clear increase with decreasing native ribozyme stability, with larger differences observed in the presence of subsaturating ATP (Figure 2.3, Figure A6). The strongest ATPase stimulation was measured with the quintuple mutant, indicating that the lack of compact tertiary structure stimulated CYT-19 activity. Thus, our results support the first model, in which CYT-19 has more opportunities to interact with the less stable ribozymes.

CYT-19 activity tracks with Mg^{2+}-dependent ribozyme folding

As an orthogonal approach to modulating ribozyme stability by mutations, we tested the effects of Mg^{2+} ion concentration on the ATPase stimulation. We hypothesized that the ATPase rates should mirror the Mg^{2+}-dependent folding isotherms of each individual ribozyme, with the ATPase rate decreasing at increasing Mg^{2+} concentrations as more native ribozyme is present at equilibrium.

A subset of ribozyme variants (L2, L5b and ARB mutants) were folded for ~24 h at various Mg^{2+} concentrations and ATPase measurements were initiated by adding CYT-19 and 100 μM ATP. As expected, the ATPase rates decreased with increasing Mg^{2+} concentration around the respective folding midpoints of each mutant (Figure A7). While the ATPase rates increased with Mg^{2+} concentration at <2 mM Mg^{2+}, this increase was independent of tertiary structure formation, as it was also observed with the quintuple mutant and oligonucleotide helices (Figure A7) (13). The weak ATPase activity at low Mg^{2+} concentrations may result from slow release of RNA, limiting nucleotide turnover (166), or from Mg^{2+}-dependent secondary structure stabilization (13). In order to isolate the effects of tertiary folding, we normalized the ATPase rates for each mutant by the corresponding rate for the quintuple mutant ribozyme at the same Mg^{2+} concentration (Figure 2.4). The normalization uncovered a clear correspondence between the Mg^{2+}-dependent fraction of native ribozyme and the ATPase stimulation across the whole

tested range of Mg^{2+} concentrations. The continued decrease in ATPase stimulation at high Mg^{2+} concentrations could be caused by decreasing dynamics of native structures. Overall, the results support the role of global tertiary structure stability in modulating CYT-19 activity. Further, the observation that the largest changes in ATPase stimulation are centered on the folding midpoints, where the ratios of native to unfolded ribozyme are changing the most strongly, suggest that CYT-19 acts preferentially upon unfolded, non-compact ribozyme molecules within a population.

To directly address the role of ribozyme compactness in the stability dependence of CYT-19 activity, we compared the ribozyme mutants by SAXS at 2 mM and 10 mM Mg^{2+}. As expected, at 2 mM Mg^{2+} the radius of gyration (R_g) values of equilibrated ribozymes increased with lower native ribozyme stability, reflecting increased populations of non-compact forms (Figure 2.5*A*, Figure A8). Differences in compactness were also reflected in the shapes of the Kratky plots, where a pronounced parabolic peak indicates compact, globular particles, e.g., for the native wild-type ribozyme, while flat or ascending profiles indicate non-compact structures, e.g., for the quintuple mutant (167) (Figure 2.5*B*). The peaks of the single tertiary contact mutants fall between those of the wild-type and the quintuple mutant ribozymes in an order corresponding to their stabilities (Figure 2.5*B*). Further, at 10 mM Mg^{2+}, where all of the ribozymes except the quintuple mutant populate predominantly the native conformation, both the R_g values and the ATPase rates were clustered around the corresponding values for the wild-type ribozyme (Figure 2.5*C, D*). As expected, the quintuple mutant maintained an extended structure and gave strong stimulation of ATPase activity. In addition, the ARB mutant ribozyme gave both a slightly larger R_g value and slightly greater ATPase stimulation than the wild-type ribozyme (Figure 2.5*C, D*). Together, the SAXS and ATPase

measurements indicate that CYT-19 activity scales with average ribozyme compactness, with non-compact ribozymes providing more productive interaction opportunities.

CYT-19 interacts with exposed ribozyme helices

Interestingly, the range of ATPase rates with the ribozyme mutants was much narrower than that of the unfolding efficiencies, as substantial ATPase stimulation was observed even with the most stable native ribozymes (~4-fold range from the wild-type ribozyme to the quintuple mutant) (Figure 2.3). Considering CYT-19 activation by non-compact structures, we hypothesized that CYT-19 may interact with helices that are exposed on the surface of the folded ribozyme. The most likely candidates are P6b, P8 and P9.2, which extend prominently away from the ribozyme body and are not constrained by tertiary interactions (Figure 2.6*A*). Therefore, we generated a set of mutants with these helices truncated (see Figure 2.1). The mutants resembled the wild-type ribozyme in their stabilities and substrate cleavage rates, indicating that the structures were not perturbed beyond the truncated helices (Figure A9). ATPase measurements revealed small but reproducible decreases in all individual truncation mutants, and an additive, 30% decrease with a P9.2 and P6b double truncation mutant (Figure 2.6*B*). Counterintuitively, a triple truncation mutant showed no detectable decrease in ATPase stimulation relative to the wild-type ribozyme at 2 mM Mg^{2+}. However, at 10 mM Mg^{2+} the ATPase stimulation decreased to 20% below wild-type, suggesting that decreased local stability of the mutant may oppose the expected decrease in ATPase rate (Figure 2.6*C*). Adding the differences from the individual truncation mutants yields an estimated 2-fold reduction in ATPase stimulation relative to the wild-type ribozyme, suggesting that interactions with these helices underlie at least half of the ATPase stimulation by the folded ribozyme. This estimate is a lower limit because some

residual ATPase activity for the mutants may arise from CYT-19 interactions with the 'stumps' of the truncated helices, and for both the wild-type and mutant ribozymes small amounts of damaged, incompletely folded ribozyme may contribute to ATPase activity.

To further address the roles of helix accessibility in stimulating CYT-19 activity, we compared ATPase stimulation by equimolar concentrations of the wild-type ribozyme and two isolated oligonucleotide helices corresponding to parts of the ribozyme (Figure 2.6*D*). Interestingly, the oligonucleotides stimulated the ATPase activity to nearly the same level as the wild-type ribozyme (within 20%) (Figure 2.6*D*). The result that a single helix in solution can provide nearly the same ATPase stimulation as the 388 nt ribozyme is consistent with the notion that CYT-19 preferentially interacts with accessible secondary structure elements and is unable to efficiently interact with the helices that are stably packed in tertiary structure. Together, the results in this section reveal three CYT-19 interaction sites within the fully folded *Tetrahymena* ribozyme and provide further support for a model in which CYT-19 interacts preferentially with exposed or protruding RNA helices on the surfaces of structured RNAs.

DISCUSSION

Here, we investigated the ATP-dependent interactions of the DEAD-box protein CYT-19 with a group I intron ribozyme, a relative of the physiological group I intron substrates of CYT-19. We found that the efficiency of CYT-19-mediated unfolding tracks with the global tertiary structure stability of the ribozyme, and that the ATPase activity of CYT-19 increases with decreasing compactness of the ribozyme structure. Based on these and previous results, we describe a model for general RNA chaperone activity in which CYT-19, and likely other DEAD-box proteins, remodel structured RNAs and RNPs primarily by disrupting accessible secondary structure. Importantly, the preference for

non-compact RNA structures can bias DEAD-box proteins to remodel misfolded RNAs and RNPs, in a manner conceptually similar to unrelated RNA and protein chaperones.

Model for DEAD-box protein-mediated RNA rearrangements

The model that emerges for DEAD-box protein-mediated global rearrangements of group I intron structure is shown in Figure 2.7. The basic features of this model are likely to apply generally to DEAD-box protein-mediated remodeling of misfolded RNAs as well as native RNAs that undergo rearrangements during their functions. Based on our finding that compact tertiary packing of the ribozyme limits productive interactions with CYT-19, the first step is a spontaneous disruption of the compact RNA structure. This disruption, shown as a dynamic fluctuation at a peripheral tertiary contact, exposes one or more helices that can then be bound by the protein. Helices that are exposed constitutively in the folded ribozyme can also interact and be unwound locally. Indeed, the ATPase activation by the folded ribozyme and the decreased ATPase activity upon truncation of extended helices support this hypothesis. However, the strong dependence of unfolding efficiency on ribozyme stability suggests that unfolding of the *Tetrahymena* ribozyme is driven primarily by CYT-19 interactions with transiently exposed helices.

The binding of an exposed helix and its subsequent unwinding by CYT-19 are both likely to contribute to larger-scale unfolding. By capturing a helix, CYT-19 can block re-formation of tertiary contacts between the helix and the rest of the ribozyme, which may also destabilize structure that forms cooperatively, promoting further unfolding steps. In support of this hypothesis, CYT-19 was observed to bind a transiently undocked ribozyme helix for tens of seconds and to block redocking (C. Pan, J. P. Potratz, B. Cannon, Z. B. Simpson, J. L. Ziehr, P. Tijerina, R. Russell, submitted). ATP-dependent unwinding of the captured helix then furthers the overall unfolding, both by

itself and because the distortions of local structure may propagate to adjacent or nearby helical segments. Both from helix binding and unwinding, propagated disruptions are likely to be largest for transiently exposed helices because the tertiary interactions and cooperativity that localize them to the RNA structure in the ground state are lost upon interaction with the DEAD-box protein.

As additional helices are exposed by initial unfolding events, they are also expected to become productive sites for DEAD-box protein binding and unwinding. Multiple protein monomers may participate in structure disruption, and the upward curvature in the CYT-19 concentration dependence of both the unfolding and ATPase activities for some ribozymes is consistent with this hypothesis (Figures A2*B*, A3, A6*B*) (129). As unfolding progresses, spontaneous formation of new contacts and reformation of disrupted structure compete with the unfolding steps, and the ribozyme can ultimately fold to the native state or the long-lived misfolded conformation.

A key feature of this model is that CYT-19 does not directly disrupt tertiary contacts, but instead relies on spontaneous losses of tertiary structure. While it remains possible that CYT-19 can actively disrupt some types of tertiary interactions, the dependence of ATPase activity on overall compaction of the ribozyme indicates that the dominant activity of CYT-19 is to interact with accessible secondary structure and promote ATP-dependent unwinding. Further, the dependence of unfolding efficiency on ribozyme stability indicates that binding and unwinding of these helical segments leads to global unfolding of the RNA. Thus, it appears that even in the context of RNAs with extensive tertiary structure, the ATP-dependent activity of CYT-19 is primarily as a helicase, with ATP being used to unwind helical segments during RNA structural rearrangements.

Implications for misfolded RNAs and RNPs

Analogous to the tertiary contact mutants studied here, misfolded RNAs will typically have less compact structures than the intact native RNA, either dynamically, due to lower stability and thus increased frequency of fluctuations, or constitutively, due to incorrect packing of helices that may leave some helices unprotected and accessible to DEAD-box proteins. The increased accessibility of the helical segments is expected to lead to preferential disruption of misfolded structures by DEAD-box proteins, providing the RNA with new opportunities to refold to the native state. Ultimately, the native conformer can accumulate if it is more stable and therefore protects its helical segments more effectively against fortuitous interactions with DEAD-box proteins.

In addition to signaling incorrect tertiary folding, exposed helices could also indicate failure to bind proteins during RNP assembly, analogous to recognition of exposed single-stranded termini by Ro and La chaperones (168-170). The preferential disruption of exposed secondary structure by DEAD-box proteins could play a role, for example, in proofreading mechanisms and structural transitions of the spliceosome or during ribosome maturation, which require multiple helicases (171, 172).

Comparisons with other chaperones

Our results are reminiscent of previous observations for other general RNA chaperones, which often act by recognizing structural elements that are protected by stable folding and RNP formation but can be exposed in defective RNAs. The retroviral nucleocapsid (NC) peptide and the unwinding domain of the human hnRNP A1 protein bind single-stranded RNA and contain motifs for exposed guanosine bases (106, 119). These proteins are proposed to chaperone rearrangements during dimerization of retroviral RNA by interacting with accessible guanosines in the monomer structure (173). The ATP-independent bacterial protein StpA shows increased refolding activity with

destabilized mutants of a viral group I intron and may rely on enhanced exposure of potential interaction sites within the structurally unstable introns (174). A key difference is the use of ATP by DEAD-box proteins, which allows them to efficiently disrupt structure while also being able to switch via ATP hydrolysis and P_i release to a state that dissociates rapidly from the single-stranded RNA after unwinding (13, 17).

Despite important differences between structural properties of RNA and proteins, the preference of CYT-19 for exposed RNA helices also bears a strong conceptual resemblance to the substrate recognition modes by several protein chaperones. For instance, Hsp70 chaperones bind to hydrophobic regions, which are buried in native proteins and, when exposed, are common markers for protein misfolding (175). Hsp90 chaperones preferentially bind unstable, flexible protein structures, and the heat shock cognate protein 70 (Hsc70) initiates clathrin coat disassembly by capturing conformational fluctuations at the C-termini of clathrin units (176-178). Hence, stimulation by open, flexible structures appears to be a general theme in recognition of folding errors by molecular chaperones.

MATERIALS AND METHODS

Materials

Ribozyme mutants were prepared using the QuikChange protocol (Agilent Technologies) with primers from Integrated DNA Technologies. All mutations were confirmed by sequencing. The ribozymes were transcribed by T7 RNA polymerase and purified using Qiagen RNeasy columns as described (113). Concentrations were determined by absorbance at 260 nm using an extinction coefficient of 3.9×10^6 M^{-1} cm^{-1} for the wild-type ribozyme and tertiary contact mutants (113). Extinction coefficients for

the truncation mutants were calculated based on sequence changes relative to the wild-type ribozyme and were 2.5–12.4 % lower.

The RNA oligonucleotides (Dharmacon) were deprotected following the manufacturer's instructions and purified by ethanol precipitation. Oligonucleotide concentrations were determined using extinction coefficients supplied by the manufacturer (P2: 371,400 M^{-1} cm^{-1}; P9.2: 421,500 M^{-1} cm^{-1}). The ribozyme substrate rSA_5 ($CCCUCUA_5$) was 5′-end-labeled with [γ-^{32}P] ATP (Perkin Elmer) using T4 polynucleotide kinase (179), purified by non-denaturing PAGE (20%; 7 M urea, 1x TBE buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.3), and stored at -20 ˚C in TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA).

CYT-19 was expressed and purified as described (127) and stored in single-use aliquots at -80 ˚C in storage buffer ("CYT-19 buffer") containing 20 mM Tris-Cl pH 7.5, 500 mM KCl, 1 mM EDTA, 0.2 mM DTT and 50% (v/v) glycerol. The protein concentration was determined via Bradford assay using bovine serum albumin as a standard.

Stability determination of *Tetrahymena* ribozyme mutants

The Mg^{2+} dependence for native ribozyme folding was measured using catalytic activity essentially as described (136). Briefly, the ribozymes (200 nM) were incubated in the presence of 0–10 mM $Mg(OAc)_2$, 50 mM Na-MOPS (pH 7.0), 10% CYT-19 buffer (final 2 mM Tris-Cl pH 7.5, 50 mM KCl, 0.1 mM EDTA, 0.02 mM DTT, 5% glycerol (v/v)), and 2 mM ATP•Mg^{2+} at 25 ˚C. ATP and CYT-19 buffer were included in the equilibrium incubations to match the conditions used to probe CYT-19-dependent unfolding. Here and below, ATP was added in an equimolar mix with $Mg(OAc)_2$. After ~24 h, aliquots of the folding reactions were transferred to a 2-fold excess of folding

quench, containing final concentrations of 50 mM $MgCl_2$, 500 μM guanosine, and 50 mM Na-MOPS (pH 7.0). After ≥ 2 min at 25 ˚C to ensure that the RNA was trapped in the stable native or misfolded conformations, a trace amount of 5′-[^{32}P]-labeled substrate was added (1/7th total volume) and the cleavage reaction was stopped after 1 min by adding two volumes of 72% formamide (v/v), 100 mM EDTA and 0.4 mg/ml each of xylene cyanol and bromophenol blue. The fraction of product generated in a burst in a single-turnover reaction was previously shown to provide a good measure of the fraction of native ribozyme, because the native and misfolded conformers bind the substrate oligonucleotide with comparable rates but only the native ribozyme cleaves the substrate (140). Control reactions showed that a 1 min time point was representative of the burst amplitude for all mutants. The 5′ cleavage product was separated from the uncleaved substrate by 20% denaturing PAGE (7 M urea, 1x TBE) and quantified using a phosphorimager (GE Healthcare). Data were quantified using ImageQuant 5.2 (GE Healthcare) and analyzed with Kaleidagraph (Synergy Software).

To calculate the folding equilibria, we normalized the equilibrium fraction of native ribozyme at 2 mM Mg^{2+} ($f_{N,obs}$) by the maximal value ($f_{N,max}$; formed at 10 mM Mg^{2+}), which ranged from 0.8 to 0.9 and varied between ribozyme preparations, presumably due to a small fraction of damaged RNA. The calculations also took into account the small fraction of native folding that occurs in the 50 mM Mg^{2+} folding quench (~10% of the unfolded ribozyme evades the long-lived misfolded conformation and folds to the native state) (158). This "partitioning" value (f_{part}) was derived for each mutant from the average fraction of native ribozyme measured in the bottom plateau of the Mg^{2+}-dependence curve, where essentially all of the ribozyme is unfolded. Overall, the fraction of native ribozyme present at equilibrium at 2 mM Mg^{2+} (f_N) can be derived from the observed fraction as follows:

$$f_{N,obs} = f_N + f_{part}(f_{N,\,max} - f_N) \quad (1)$$

$$f_N = (f_{N,obs} - f_{part} f_{N,max})/(1 - f_{part}) \quad (2)$$

The final equation used to calculate native ribozyme stabilities at 2 mM Mg^{2+} was:

$$\Delta G = -RT\ln K_{eq} = -RT\ln\,(f_{N,norm}/(1 - f_{N,norm})) \quad (3)$$

where

$$f_{N,norm} = f_N / f_{N,max} \quad (4)$$

The correction for partitioning in eq. 1 and 2 assumes that all non-native ribozyme is unfolded, which is supported by the good agreement between Mg^{2+} dependences measured by ribozyme activity or hydroxyl radical footprinting (159, 180).

After establishing that 2 mM Mg^{2+} was appropriate for stability determination of all mutants, we obtained the equilibrium values from the endpoints of ribozyme unfolding time courses in the absence of CYT-19 (see next section) and performed the same calculations as above (eq. 1–4).

We found that varying the ATP•Mg^{2+} concentration in the folding reactions had a small but reproducible effect on the folding equilibrium, with greater ATP concentrations leading to lower stability of the native ribozyme. This effect is most likely caused by residual free ATP in the added equimolar mixtures of ATP and Mg^{2+}. We therefore report the ribozyme stabilities separately for experiments performed in the presence of 100 μM and 2 mM ATP.

Monitoring native ribozyme unfolding by cleavage activity

Unfolding time courses were measured essentially as described (127). The ribozymes (2 μM) were prefolded to the native state by incubating them in submerged

tubes at 50 ˚C for 45 min–1 h with 10 mM $Mg(OAc)_2$ and 50 mM Na-MOPS (pH 7.0) and stored on ice until use. Unfolding was initiated by diluting the ribozyme 10-fold into the reaction mix, bringing the concentrations to 200 nM ribozyme, 100–500 nM CYT-19, 2 mM $Mg(OAc)_2$, 50 mM Na-MOPS (pH 7.0), 0–2 mM ATP•Mg^{2+}, and 10% (v/v) CYT-19 buffer. CYT-19 was kept on ice at 10x final concentration and added to the reaction mix immediately prior to ribozyme addition. Unfolding was monitored at 25 ˚C and the fraction of native ribozyme over time was determined as above (see *Stability determination*), except that the quench also included 1 mg/ml Proteinase K. The zero time points for the unfolding reactions were obtained by adding prefolded native ribozyme directly to the quench solution.

The unfolding data were fit by a single-exponential equation, with the exception of the L5b mutant ribozyme, which unfolded in two phases, presumably due to loss of CYT-19 activity during the slow unfolding of this mutant. The slow phase had the same rate constant as spontaneous unfolding of the L5b ribozyme (Figure A2*B*, A3). Loss of CYT-19 activity was also observed with other mutants when folding was monitored over longer times (> 20 min), as indicated by a slow return to the intrinsic equilibrium value. To account for the variable endpoints in the unfolding reactions, the rate constants obtained from single exponential fits (or each phase of the double exponential fit in the case of L5b unfolding) were normalized by the ratio of the maximal possible unfolding amplitude ($A_{max} = f_{N,prefold} - f_{N,min}$) and the observed amplitude ($A_{obs} = f_{N,prefold} - f_{N,end}$), where $f_{N,prefold}$ represents the initial fraction of native ribozyme, formed by folding at 10 mM Mg^{2+}, 50 ˚C; $f_{N,min}$ represents the fraction of native ribozyme formed in the folding quench, which limits the observed amplitude of unfolding (see *Stability determination*); and $f_{N,end}$ is the observed endpoint. This approach is analogous to obtaining initial rates from linear fits of the earliest data points and normalizing by the observable range

($f_{N,prefold} - f_{N,min}$); however, exponential fits are more accurate as they are constrained by a greater number of available time points.

The rate constants for spontaneous unfolding were calculated from the observed rate constants as follows. The observed rate constant, k_{obs}, is defined as the sum of the unfolding rate constant, k_{unf}, and the rate constant for folding, k_{fldg}:

$$k_{obs} = k_{unf} + k_{fldg} .$$

The endpoint of the unfolding reaction yields the equilibrium $K_{eq} = k_{fldg}/k_{unf}$, allowing substitution of k_{fldg} with $K_{eq} \bullet k_{unf}$. Thus, the unfolding rate constant was calculated as:

$$k_{unf} = k_{obs} /(1+ K_{eq}).$$

The unfolding efficiencies were derived from linear fits of the CYT-19 concentration dependences of unfolding rates (Figures A2*B*, A3). Upper and lower limits of the slopes were determined where systematic deviations from linearity occurred.

ATPase assays

ATP hydrolysis was monitored under conditions identical to those of the folding reactions but in the presence of trace [γ-^{32}P]-ATP (Perkin Elmer) in addition to unlabeled ATP. Reactions were initiated by adding prefolded native ribozyme and time points were collected during the initial linear phase, not exceeding ~20% of the total ATP concentration, with a quench of 100 mM (final) EDTA. RNA-independent ATPase rates were measured by substituting TE buffer for the RNA. Reactions were analyzed by thin-layer chromatography, using a mobile phase of 1 M formic acid and 0.5 M LiCl, and quantified using a phosphorimager and Kaleidagraph software (181).

To measure ATPase stimulation by ribozymes pre-equilibrated at various $Mg(OAc)_2$ concentrations, the RNAs (250 nM) were folded for ~24 h at 25 ˚C in the

presence of 0–10 mM $Mg(OAc)_2$, 50 mM Na-MOPS (pH 7.0) and 10% CYT-19 buffer (v/v). ATPase reactions were initiated by adding CYT-19 and ATP (with trace [γ-^{32}P]-ATP) in the desired Mg^{2+} concentration, bringing final concentrations to 200 nM RNA, 500 nM CYT-19, 100 μM ATP•Mg^{2+}, and 10% CYT-19 buffer (v/v).

The oligonucleotide helices (2 μM) were incubated for 5 min at 25˚C with 10 mM $Mg(OAc)_2$ and 50 mM Na-MOPS (pH 7.0) and kept on ice until use. The helices were diluted 10-fold into the ATPase reaction mix containing final 1–10 mM $Mg(OAc)_2$, 100 μM ATP•Mg^{2+}, 50 mM Na-MOPS (pH 7.0), with 500 nM CYT-19 (final) added immediately before the RNA.

SAXS experiments

Data were collected at the Advanced Photon Source (APS) beamline 12-ID-C, using a sample-detector distance of 2 m. To probe the compactness of ribozyme variants at 2 mM Mg^{2+}, the ribozymes (600 nM) were equilibrated at 25 ˚C in the presence of 2 mM $Mg(OAc)_2$, 50 mM Na-MOPS (pH 7.0) and 10% CYT-19 buffer (v/v). The samples were concentrated using Amicon Ultra filters (EMD Millipore; 30 kD or 50 kD molecular weight cut-off), followed by multiple cycles of washing with the folding buffer to ensure that the free Mg^{2+} concentration remained at 2 mM even with the micromolar ribozyme concentrations required for SAXS measurements (182, 183). The samples were flash-frozen in liquid nitrogen and shipped to the APS on dry ice. To measure scattering from ribozymes at 10 mM Mg^{2+}, the RNAs were folded on site by incubation at 45–50 ˚C for 1 h (2 μM ribozyme, 50 mM Na-MOPS pH 7.0, 10 mM $Mg(OAc)_2$) and cooled on ice. All SAXS measurements were performed at 25 ˚C. Scattering was measured using 1 μM and 2 μM ribozyme, and no systematic increases were observed in the R_g values with increasing ribozyme concentration. 20 exposures of 1 s (12 keV) were collected and

averaged for each sample. Blanks consisted of either the concentrator flow-through (2 mM Mg^{2+}) or were prepared from the ingredients of the ribozyme folding solution, by substituting TE buffer for the ribozyme (10 mM Mg^{2+}). The data were circularly averaged at the beamline and further analyzed using IgorPro (WaveMetrics) and Gnom (184).

The R_g values were determined using Gnom and verified by the Guinier approximation (167, 184). All Gnom solutions had scores of 0.594 or higher, representing reasonable fits, as defined by the program. The R_g values at 2 mM Mg^{2+} are averages of results from 2–3 independently prepared samples measured during three beam times. The data at 10 mM Mg^{2+} were obtained during a single beamtime and the reported R_g values are averages of samples measured at 1 and 2 μM ribozyme. For Kratky analysis, the scattering profiles were normalized by the forward scattering (I_0) values obtained through Guinier analysis.

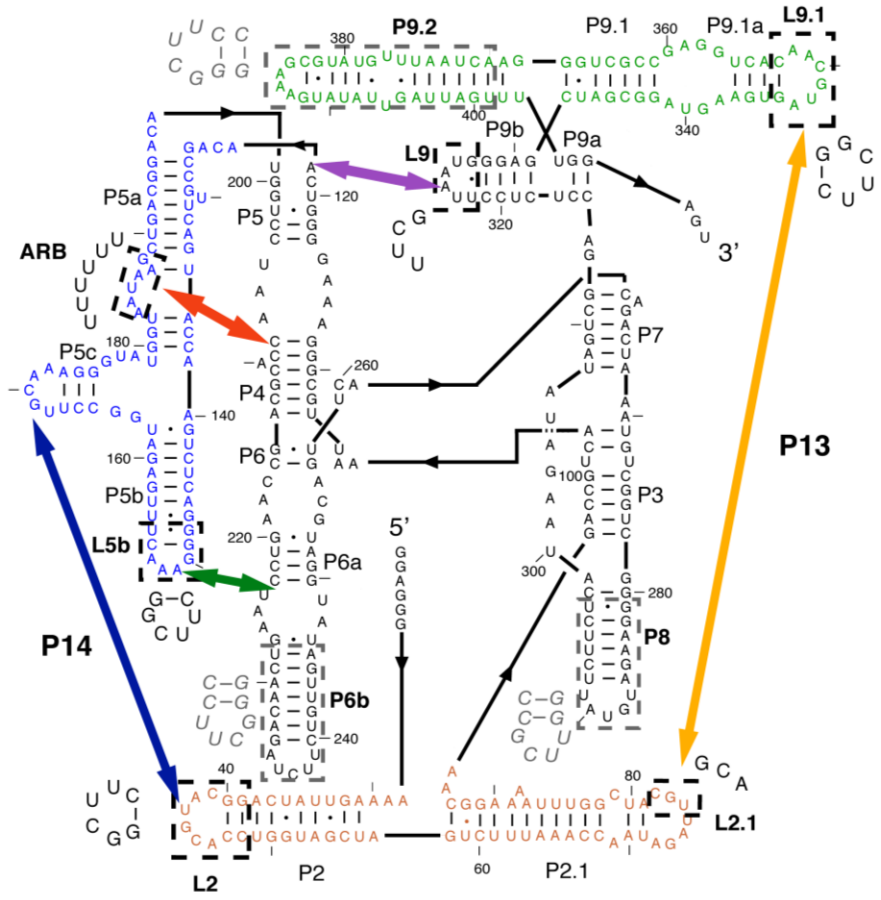


Figure 2.1: Tertiary contact and truncation mutants of the *Tetrahymena* group I intron ribozyme.

Peripheral elements are colored and bold arrows mark the long-range peripheral tertiary contacts. Paired regions (P) and loops that were mutated in this study (L) are labeled based on group I intron nomenclature in Ref. (132). The mutated regions are enclosed in dashed boxes and labeled in bold, with sequence substitutions indicated nearby. Sequences that were deleted to construct the helix truncation mutants (see Figure 2.6) are enclosed in gray dashed boxes and the replacement nucleotides are shown in gray italic font.

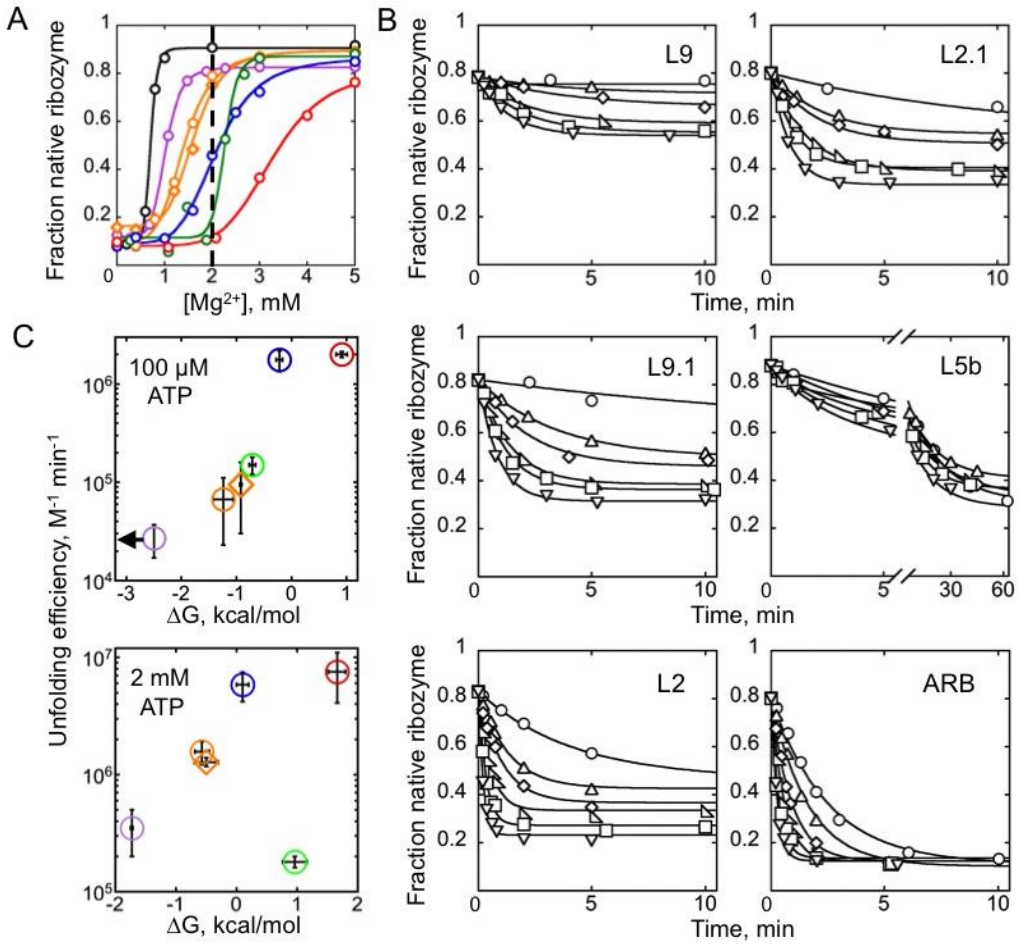


Figure 2.2: CYT-19-mediated ribozyme unfolding.

(*A*) Mg^{2+}-dependent folding of wild-type (WT; black) and mutant ribozymes. The mutants are designated by the tertiary contact component that is mutated: L9 (purple), L9.1 (orange circles), L2.1 (orange diamonds), L5b (green), L2 (blue), and ARB (red). For clarity, only representative isotherms are shown, with cumulative data shown in Figure A1. The dashed line at 2 mM Mg^{2+} marks the condition used for unfolding experiments with CYT-19. The maximal measured fractions of native ribozyme were <1 and varied between ribozyme preparations (Figure A1), presumably because of a small and variable amount of damaged RNA (140). (*B*) Unfolding of the native ribozyme by

CYT-19 (2 mM ATP•Mg^{2+}). Representative time courses in the presence of CYT-19 buffer only (○) or 100 nM (△), 200 nM (◇), 300 nM (◺), 400 nM (□), or 500 nM (▽) CYT-19 are shown. (*C*) Correlation of CYT-19-mediated unfolding rate constants with ribozyme stability in the presence of 100 μM and 2 mM ATP•Mg^{2+}. Horizontal error bars indicate standard errors from two or more determinations of ribozyme stability, and vertical error bars represent standard errors from linear fits through cumulative data of CYT-19-dependent unfolding rates (Figure A2*B*, A3; see *Materials and Methods*). In cases where the concentration dependences displayed upward curvature, the error bars correspond to the upper and lower limits of the unfolding efficiencies. For the L5b mutant, the reported efficiencies are derived from the fast phase of unfolding. The mutants are slightly less stable in 2 mM ATP•Mg^{2+} than in 100 μM ATP•Mg^{2+}, most likely reflecting sequestration of Mg^{2+} ions by residual free ATP. The L5b mutant is the most sensitive to this change due to its steep Mg^{2+} concentration dependence. The arrow indicates that the ΔG value for the L9 mutant is an upper limit because the amplitude of unfolding was too small under these conditions to measure the transition accurately.

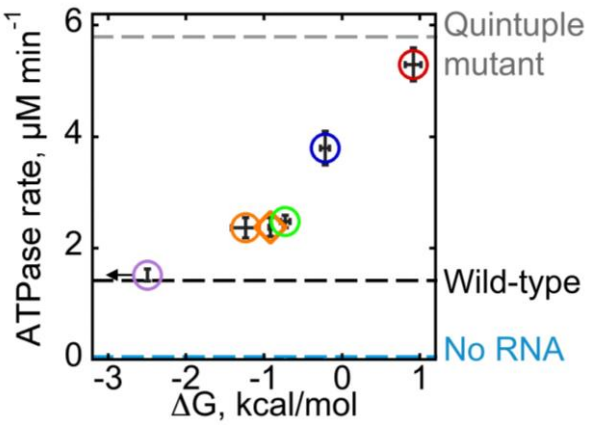


Figure 2.3: Relationship between ATP hydrolysis rates and ribozyme stability.

The ATPase rates wee measured at 100 μM ATP, 500 nM CYT-19, 200 nM ribozyme, and 2 mM Mg^{2+}. The color scheme for ribozyme mutants is the same as in Figure 2.1. Dashed lines indicate the ATPase stimulation by the wild-type ribozyme (black; (1.37 ± 0.06) μM min^{-1}) and the quintuple mutant (gray; (5.6 ± 0.4) μM min^{-1}), as well as the RNA-independent ATPase rate (cyan, (0.05 ± 0.02) μM min^{-1}). Vertical error bars indicate standard error from at least four determinations.

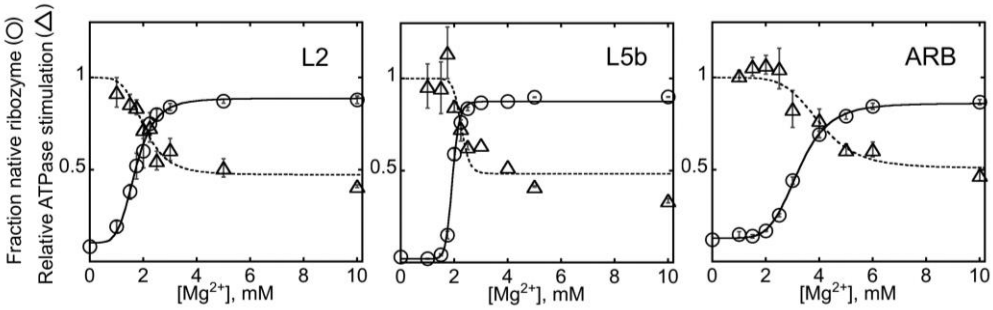


Figure 2.4: Correspondence between the ATPase stimulation (triangles) and the fraction of native ribozyme at equilibrium (circles) for the L2, L5b and ARB mutants.

Circles indicate the ATPase rates and triangles indicate the fractions of native ribozyme at equilibrium. The ATPase rates were normalized by the rates measured at the same Mg^{2+} concentration with the quintuple mutant to account for tertiary folding-independent effects of Mg^{2+} on the ATPase activity (Figure A7*A*). The Mg^{2+} concentration dependences of ATPase activity are reasonably well fit by the Hill equation (dashed curves), with the Hill coefficient fixed to match that of the folding isotherm for each mutant (solid curves) and the starting value fixed to 1, with minor deviation present for the L5b mutant.

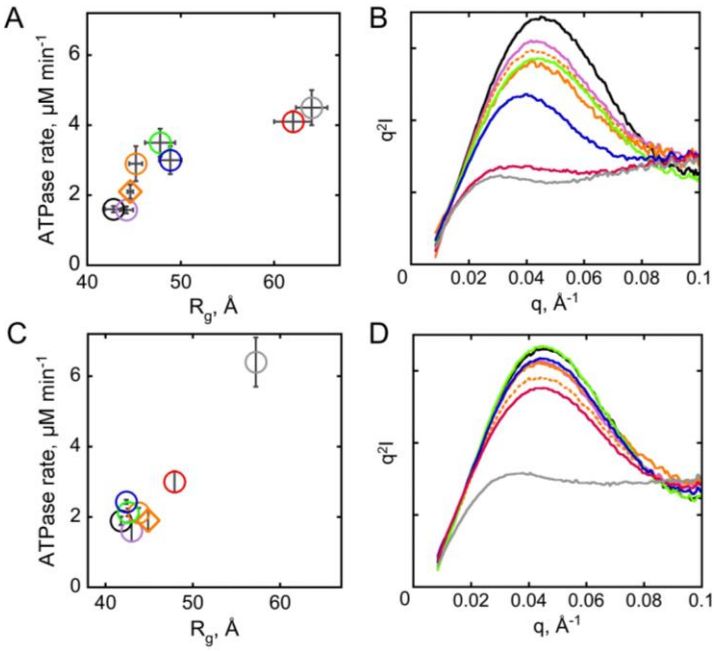


Figure 2.5: ATPase stimulation tracks with ribozyme compactness.

(*A*) Relationship between the R_g value of the equilibrium ribozyme population and the ATPase stimulation by ribozymes equilibrated at 2 mM Mg^{2+}. The R_g values were determined by P(R) analysis using Gnom (184). The error bars indicate standard errors from 2–3 independent measurements. (*B*) Kratky plots of ribozymes equilibrated at 2 mM Mg^{2+}. Scattering profiles are shown for the wild-type ribozyme (black), the quintuple mutant (gray), and the single tertiary contact mutants (color scheme as in Figure 2.1, with the profiles of the P13 mutants L9.1 and L2.1 shown as a solid and a dotted line, respectively). Representative Kratky plots acquired during the same beam time are shown. (*C*) Clustering of the R_g values and the ATPase rates at 10 mM Mg^{2+}. The R_g values represent averages from measurements performed during a single beam time using 1 μM and 2 μM ribozyme. (*D*) Kratky plots of ribozymes equilibrated at 10 mM Mg^{2+}.

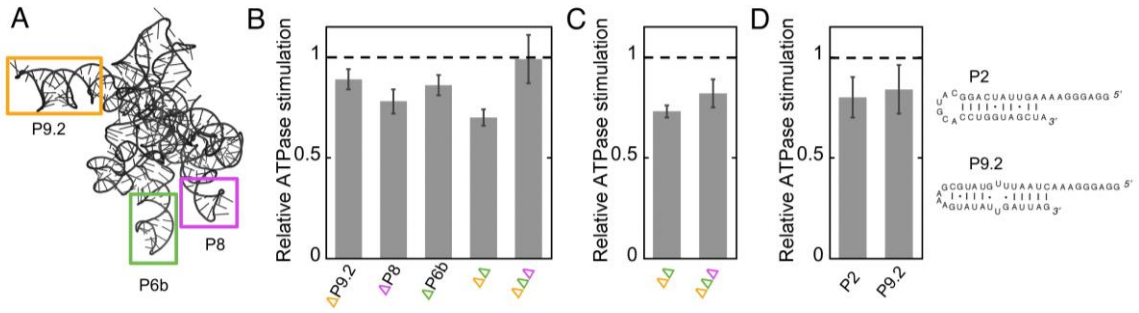


Figure 2.6: ATPase stimulation by exposed ribozyme helices.

(*A*) Ribozyme model with helices P6b, P8 and P9.2 highlighted. In the truncation mutants, the helices were replaced with shorter stems and UUCG tetraloops (see Figure 2.1). (*B*) ATPase stimulation by helix truncation mutants under conditions identical to those used for tertiary contact mutant ribozymes (100 μM ATP, 500 nM CYT-19, 200 nM native ribozyme, 2 mM Mg^{2+}). The ATPase rates were normalized by the rates measured with the wild-type ribozyme in the same experiment. The error bars indicate standard errors from 5–6 measurements. ΔΔ and ΔΔΔ denote double (ΔP9.2+ΔP6b) and triple truncations, respectively. Two independent preparations of each ribozyme were measured to ensure that the small differences were not artifacts due to differences between preparations. (*C*) ATPase stimulation by the double and triple truncation mutants at 10 mM Mg^{2+}. Error bars indicate standard errors from 4 measurements using two ribozyme preparations. (*D*) ATPase stimulation by isolated RNA helices, corresponding to ribozyme helices P2 and P9.2. Shown are ATPase rates relative to the ATPase stimulation by the same molar concentration (200 nM) of wild-type ribozyme measured in the same experiment (100 μM ATP, 500 nM CYT-19, 2 mM Mg^{2+}).

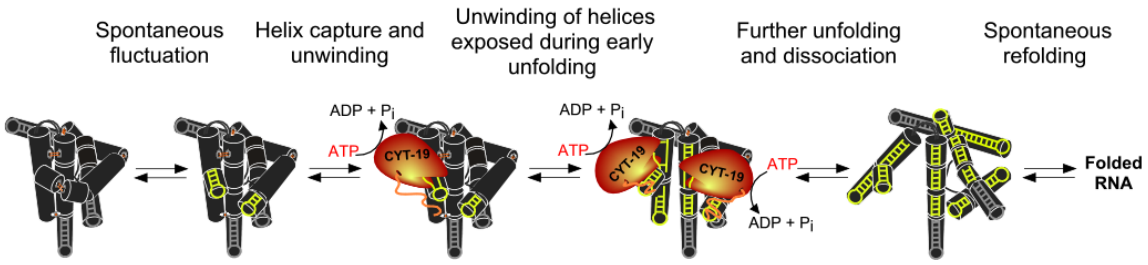


Figure 2.7: Model for CYT-19-mediated unfolding of structured RNA.

The unfolding process is illustrated for the *Tetrahymena* ribozyme, shown schematically with black cylinders representing base-paired regions. The long-range peripheral tertiary contacts are shown as orange lines. Gray hairpins indicate the helices that are exposed and accessible for CYT-19 in the native structure but are not productive sites for unfolding. From left to right, unfolding is initiated as a spontaneous local disruption, which exposes helical segments that can be bound by CYT-19 (highlighted in yellow). ATP-dependent unwinding can induce further unfolding, exposing additional helices that can also be unwound by one or more CYT-19 monomers. The unfolding process is ultimately followed by refolding, giving formation of a new structure or a return to the initial structure.

Chapter 3: ATP utilization during DEAD-box protein-mediated refolding of the misfolded group I intron ribozyme

INTRODUCTION

One of the most common functions of DEAD-box and related RNA helicases is to promote rearrangements of RNA folding intermediates, and understanding how DEAD-box proteins utilize ATP to resolve misfolded RNA structures is essential for further characterization of their mechanisms as RNA chaperones. CYT-19, which functions in group I intron folding in *N. crassa* mitochondria, was the first DEAD-box protein shown to accelerate refolding of a misfolded RNA and was shown to act as a true chaperone, as it is not required for the activity of the refolded RNA (22, 85, 126). CYT-19 functions non-specifically and can efficiently refold the well-characterized misfolded conformer of the non-cognate *Tetrahymena* group I intron ribozyme (22, 85, 127). Spontaneous refolding of this misfolded RNA, which is structurally almost identical to the native conformer and appears to only differ in core topology, is known to require transient global disruption of tertiary structure and a long-range core helix (145, 146). This requirement for extensive unfolding, together with the demonstrated importance of ATP-dependent unwinding for DEAD-box protein-mediated group I and group II intron rearrangements (129), suggest that CYT-19 promotes refolding by disrupting the misfolded structure.

The mechanism of CYT-19-mediated disruption of the misfolded ribozyme has not been completely understood. For instance, it is unknown how many ATP-dependent steps are required for CYT-19-mediated unfolding and whether unfolding occurs in a well-defined or randomly determined number of steps. In the model for ribozyme unfolding that is supported by work in Chapter 2 and Pan et al. (submitted), CYT-19

disrupts the ribozyme structure non-specifically, by binding and unwinding exposed secondary structure and without directly disrupting tertiary interactions. This and previous work for CYT-19 and related DEAD-box proteins (128, 129) suggest that remodeling of the RNA structure is accomplished at least in part by the same fundamental mechanisms as unwinding of isolated helices. Studies with oligonucleotide helices showed that DEAD-box proteins disrupt secondary structure in an energy-efficient manner, with only a single cycle of ATP binding and hydrolysis required for complete unwinding of a short helix, of lengths typical of the helices in the ribozyme (13, 17, 95) (see Chapter 1). Despite mechanistic similarities, the ATP requirement is likely to be considerably greater for rearrangements of large, structured RNAs, such as the misfolded *Tetrahymena* ribozyme, than for isolated helices. In addition to multiple helices being disrupted during unfolding of the ribozyme, some local rearrangements may not be useful for unfolding, and even at productive sites multiple disruption attempts may be required.

Here we quantified the ATP utilization by CYT-19 for refolding of the misfolded group I intron ribozyme and investigated the processes that contribute to ATP consumption. We show that hundreds of ATP molecules are hydrolyzed during refolding of each misfolded ribozyme molecule. Most of these hydrolysis events arise from CYT-19 interactions with non-productive sites, including surface helices that are constitutively exposed in both the misfolded and the native ribozyme. After accounting for these non-productive interactions, we show that ATP utilization is relatively constrained at approx. 150–200 ATP molecules per ribozyme across a range of CYT-19 and ATP concentrations, but increases with increasing stability of the misfolded ribozyme. Importantly, previous studies of ribozyme folding indicate that only a fraction of unfolding events result in native folding, suggesting that the large ATP utilization value

for refolding reflects repeated global disruptions, each of which consumes only a handful of ATP molecules. Overall, this work indicates that DEAD-box protein-mediated refolding of the misfolded group I intron is an energy-intensive process, with non-productive interactions and spontaneous reformation of RNA structure being a major factor in increasing ATP consumption.

RESULTS

CYT-19 hydrolyzes hundreds of ATP molecules during refolding of the misfolded ribozyme

To measure ATP utilization for refolding of the misfolded *Tetrahymena* ribozyme, we determined the ATPase and refolding activities of CYT-19 under identical conditions. We first established conditions of strong CYT-19 and ATP-dependent acceleration of refolding beyond the spontaneous rate, using a misfolded ribozyme concentration that was comparable to the concentration of CYT-19. The misfolded ribozyme was formed by a 5 min incubation at 25 ˚C in the presence of 10 mM Mg^{2+} and refolding was initiated by diluting the Mg^{2+} to 2 mM Mg^{2+}. After establishing the dependences of the refolding rate on ATP and CYT-19 concentrations (Figure 3.1*A, B*), we performed the initial ATP utilization measurements with 500 nM CYT-19 and 500 μM ATP, which provided substantial, ~5-fold acceleration, allowing us to attribute most of the observed refolding to ATP-dependent CYT-19 activity. To quantify the ATP utilization per refolding event, we divided the initial rate of ATP hydrolysis in the presence of misfolded ribozyme by the initial rate of ribozyme refolding. At 500 nM CYT-19 and 500 μM ATP, this approach yielded ~350 ATP molecules hydrolyzed per refolded ribozyme (Figure 3.1*C, D*). Even with extensive rearrangements required for refolding of the misfolded ribozyme, this was a strikingly large number, considering that

unwinding of a helix is known to require only a single ATP (13, 17). In the following sections, we dissect the measured ATP utilization into productive and non-productive events and narrow down the number of required ATP-dependent steps.

CYT-19 interactions with native helices are a source of futile ATP hydrolysis

Not all ATP-dependent rearrangements by CYT-19 are likely to be useful for unfolding of the misfolded ribozyme. Previous work demonstrated a general preference of CYT-19 for exposed helices, including constitutively exposed surface helices in the native ribozyme (Ref. (147), Chapter 2 and Pan et al., submitted). However, as shown in Chapter 2, ATP-dependent interactions with the constitutively exposed helices P9.2, P8 and P6b, and likely other constitutively exposed sites, do not efficiently promote unfolding of the native ribozyme, with transiently exposed internal helices providing much more productive interaction sites for unfolding. The misfolded and native ribozymes are structurally similar, and the protruding surface helices of the native ribozyme are also present in the misfolded ribozyme, likely engaging CYT-19 at the same rates as in the native ribozyme and contributing to non-productive ATP hydrolysis. Comparison of the initial ATPase rates measured with the native vs. misfolded ribozyme revealed only approx. 50 % greater ATPase stimulation by the misfolded ribozyme (Figure 3.2*A*). Thus, as much as two thirds of ATP hydrolysis events with the misfolded ribozyme could be a result from the same nonproductive interactions as those CYT-19 forms with the native ribozyme.

To address if CYT-19 interacted with native surface helices in the misfolded ribozyme and whether these interactions were non-productive for refolding, we compared the refolding activities of CYT-19 using the misfolded wild-type ribozyme and a mutant with truncated P9.2 and P6b helices (see Chapter 2). In the absence of CYT-19, the

misfolded conformer of the truncation mutant refolded to the native state with the same rate constant as the misfolded wild-type ribozyme, consistent with the same misfolded conformation being formed and with the earlier result that helix truncations have no effect on ribozyme stability (Figure 3.2*B*, see Chapter 2). Interestingly, with increasing CYT-19 concentrations, the refolding rates diverged between the mutant and wild-type ribozymes (Figure 3.2*B*). The wild-type ribozyme displayed a clear non-linear CYT-19 concentration dependence under our conditions, with the refolding rates increasing only weakly up to ~400 nM CYT-19 before displaying a steeper dependence at ≥ 400 nM CYT-19[1]. In contrast, the truncation mutant showed essentially linear CYT-19 concentration dependence across the same range and was refolded up to ~2-fold faster than the wild-type ribozyme at low, < 400 nM, CYT-19 concentrations. This result suggested that CYT-19 was engaging in non-productive interactions with the constitutively exposed P6b and P9.2 helices, which may suppress refolding by reducing the amount of free protein available to interact with productive sites in the misfolded ribozyme. Lowering the ribozyme concentration from 200 nM to 50 nM also eliminated upward curvature in CYT-19 concentration dependence at <~400 nM CYT-19, as would be expected if most CYT-19 was free to interact with productive sites, rather than being sequestered at native surface helices. Overall, these data indicate that CYT-19 forms non-productive interactions with native, constitutively exposed helices in the misfolded ribozyme. Further, the result that these interactions efficiently inhibit refolding at low ratios of CYT-19 to the ribozyme, suggests that CYT-19 binds the surface helices

[1] Linear CYT-19-concentration dependence was previously observed at 5 mM Mg^{2+}, 2 mM ATP (Ref. 22). The shape of the CYT-19 concentration dependence appears to depend on reaction conditions, as the curvature was also less pronounced in our experiments at 2 mM ATP, 2 mM Mg^{2+}(data not shown). The origins of these differences remain to be addressed.

preferentially to productive sites that are unique to the misfolded ribozyme, likely because productive sites only become exposed transiently.

Quantifying ATP utilization for CYT-19 interactions with unique features of the misfolded ribozyme

CYT-19 shows greater ATPase activity with the misfolded than native ribozyme (Figure 3.2*A*), and the extent of this enhancement can provide a measure of ATPase events that are unique to the misfolded ribozyme. During a refolding reaction, the differences in ATPase stimulation by the native and misfolded ribozymes are expected to give rise to a burst in ATPase activity, corresponding to the disappearance of the misfolded ribozyme and accumulation of the native ribozyme. To test if a burst was present in the ATPase kinetics, we monitored the ATPase activity for times long enough for all the misfolded ribozyme to refold to the native state. Indeed, we observed a gradual decrease in the ATPase rate during refolding, approaching the rate that was measured in parallel with the pre-folded native ribozyme (Figure 3.3*A, B*). Less pronounced curvature over long times was also observed with the native ribozyme, which is not unfolded detectably by CYT-19; however, this curvature can be explained by ATP depletion and accumulation of ADP, which can inhibit DEAD-box proteins (77). An appreciable burst was absent at low CYT-19 concentrations (Fig. 3.3C), consistent with the above result that the same non-productive sites dominate CYT-19 interactions with the native and misfolded ribozymes in the absence of sufficient excess of CYT-19.

To test if the burst in ATPase activity indeed represented refolding, we determined how rapidly the ATPase rate declined towards the levels observed with the native ribozyme. Subtracting the ATPase progress curves that were obtained with the native ribozyme from those collected during refolding yielded single exponential curves with rate constants within ~30 % of the refolding rate constants measured by ribozyme

activity (Figure 3.3*D, E*, Table 3.1). The good agreement between the ATPase and refolding kinetics suggested that the ATPase activity indeed tracked with ribozyme refolding.

The amplitudes of the subtracted ATPase curves represent ATP utilization for interactions that CYT-19 forms only with the misfolded ribozyme. At 500 nM CYT19 and 500 μM ATP, this approach yields 157±11 ATP molecules hydrolyzed per refolded ribozyme. Figure 3.3*F* shows the ATP utilization values derived from the burst amplitudes at a series of CYT-19 and ATP concentrations that give at least 4–5-fold acceleration of refolding. The ATP utilization values appear to plateau at ~150–200 ATP molecules per refolded ribozyme, suggesting that a large, but relatively well constrained amount of ATP is used during refolding.

ATP utilization increases with increasing stability of the misfolded ribozyme

We next addressed how the global stability of the misfolded ribozyme affected ATP utilization for refolding. At the level of secondary structure disruption by DEAD-box proteins, increased stability of RNA helices is known to increase ATP utilization, presumably because upon local disruption the base-pairs next to the protein binding site are less likely to separate spontaneously, lowering the probability of complete unwinding (181). Similarly, it is likely that more stable ribozyme structures require more ATP-dependent rearrangements for refolding, because spontaneous disruptions contribute less to the unfolding process – either locally at the level of base-pairs flanking the CYT-19-mediated local disruption sites or in the form of more extensive, cooperative rearrangements.

We used two approaches to address the roles of global stability: varying the Mg^{2+}-concentration, which is critical for stabilization of both secondary and tertiary RNA

structure, and using destabilizing tertiary contact mutations of the ribozyme. In both cases, we used conditions under which most refolding is CYT-19-dependent (at least 4.9–fold acceleration by CYT-19 relative to spontaneous refolding, Table 3.2) and quantified the ATP utilization using an equivalent approach to the one described in the previous section. Here, we calculated the enhancement in ATPase stimulation by the misfolded vs. native ribozyme by subtracting the ATPase rate measured with the native ribozyme from that measured with the misfolded ribozyme at each condition and divided this subtracted rate by the initial rate of refolding.

In order to reduce background ATPase activity with native sites, we varied the Mg^{2+} concentration using the double truncation mutant (ΔP9.2+ΔP6b). The rate of CYT-19-mediated refolding decreased by 5.6-fold from 2 mM to 5 mM Mg^{2+}; however, the enhancement in ATPase stimulation by the misfolded vs. native ribozyme decreased less strongly, by 2.4-fold. The decrease in ATPase stimulation by the misfolded ribozyme with increasing Mg^{2+} concentration was expected based on the result in Chapter 2 that more stable structures provide fewer interaction opportunities for the DEAD-box protein and was presumably caused by decreased dynamics of the misfolded ribozyme, consistent with the lower rate constants of spontaneous refolding at higher Mg^{2+} concentrations (Table 3.2). Together, the greater magnitude of the change in refolding rates than the change in ATPase rates leads to a modest, ~2–3-fold increase in ATP utilization from 2 mM to 5 mM Mg^{2+} (Figure 3.4*A, B*). Changes in ATP utilization with ribozyme stability were also observed when comparing tertiary contact mutants to the wild-type ribozyme. While the accessible range of stabilities was limited by the fact that for highly destabilized ribozymes CYT-19 also efficiently unfolds the native conformer (Ref.(127), Chapter 2 and data not shown), it is clear that even for the two most stable tertiary contact mutants the amount of ATP utilized for refolding is considerably lower than for the

misfolded wild-type ribozyme. Together, these results indicate that CYT-19-mediated refolding of more stable misfolded structures requires more ATP-dependent rearrangements, consistent with an increase in futile activity.

DISCUSSION

Here, we measured ATP utilization for refolding of the misfolded group I intron ribozyme by the DEAD-box protein CYT-19. Our initial measurements revealed a strikingly large value of hundreds of ATP molecules consumed per ribozyme refolding event; however, further dissection suggests that a much smaller number is used for productive rearrangements that contribute to unfolding and subsequent refolding of the misfolded ribozyme. We show that one source of this futile activity is CYT-19 interactions with helices that are exposed on the surface of both the native and misfolded ribozymes. These interactions do not contribute substantially to refolding, yet account for most of ATP hydrolysis with the misfolded ribozyme. After subtracting the contributions from these non-productive interactions, ATP utilization appears to be relatively well constrained at ~150–200 ATP molecules per refolded ribozyme. The large number suggests that even interactions with helices that are uniquely exposed in the misfolded ribozyme are still dominated by futile disruption attempts at either local or global level, and may include disruptions of sites that are not productive for promoting global unfolding. Regarding the futility of refolding, important insights came from previous studies of ribozyme folding, which demonstrated that initial folding of the ribozyme kinetically favors the formation of misfolded ribozyme over native ribozyme, by approx. an order of magnitude at 25 ˚C (140). The implication for CYT-19-mediated refolding is that even after successful global unfolding of the ribozyme tertiary structure, the majority of the ribozyme may immediately reform the misfolded structure rather than folding to

the native state. Assuming a partitioning factor of 10 % for refolding to the native state upon CYT-19-mediated global disruption, the ATP requirement per global unfolding event would be only $1/10^{th}$ of the total ATP utilization for refolding, or ~15–20 ATP molecules per ribozyme. This considerably smaller number may in turn reflect single disruptions of a series of ribozyme helices, or repeated disruption events at a single or a handful of helices, with the number of active disruptions per unfolding event increasing with ribozyme stability. Further research will be required to determine which parts of ribozyme structure are actively disrupted by CYT-19 during unfolding and whether unwinding of a defined set of helices or even just a single helix underlie CYT-19-mediated refolding. However, given the non-specific activity of CYT-19 and the requirement for extensive unfolding, we favor a scenario in which CYT-19 disrupts multiple helices in the course of unfolding, likely following a stochastic mechanism in which the protein simply interacts with helices exposed during spontaneous disruptions of tertiary structure or during prior steps of CYT-19-assisted unfolding (see Chapter 2).

To our knowledge, ATP utilization has not been previously measured for a chaperone-mediated rearrangement of a structured RNA; however ATP consumption has been quantified for a series of protein chaperones, providing a broader perspective on the efficiency of ATP utilization by ATP-dependent remodelers. The DnaK-DnaJ-GrpE chaperone system (Hsp70), which mechanistically probably the most closely resembles DEAD-box proteins, consumes thousands of ATP molecules for protein refolding under most tested conditions, but has also been suggested to refold misfolded luciferase with hydrolysis of just 5 ATP molecules under some conditions (185). The chaperonin GroEL-GroES hydrolyzes 14 ATP molecules per cycle of protein encapsulation and conformational rearrangements, and multiple cycles are generally required to allow protein refolding to the native state (120). Finally, ClpX, the unfoldase component of the

bacterial ClpXP protease, requires several hundred ATP molecules to unfold the I27 domain of the titin protein with an attached degradation tag, but ATP utilization is drastically reduced for protein substrates with unstable structures in the vicinity of the tag (186-188). Most ATP appears to be consumed by ClpX during futile attempts to capture the degradation tag sequence, which is required for initiation of unfolding (189). These examples illustrate the diversity in energy efficiencies of ATP-dependent remodelers, suggesting that a high energetic cost, exacerbated by futile cycles of global or local unfolding, is general to chaperone-mediated structural rearrangements. However, although protein chaperones provide a useful conceptual reference, it will be important to further explore the ATP-dependent mechanisms of RNA remodelers. Questions of particular interest include whether adaptations exist that increase the efficiency of ATP-dependent rearrangements for cognate substrates and how ATP consumption varies between DEAD-box proteins that function as specific vs. general chaperones. The effects of RNA size, types of RNA misfolding, as well as roles of local vs. global stability are also potentially important in determining the ATP requirement. We hope that this study will provide a useful initial framework to further characterize the ATP utilization by DEAD-box proteins and other ATP-dependent RNA chaperones.

MATERIALS AND METHODS

The ribozyme, CYT-19 and radiolabeled ribozyme substrate were prepared as described in Chapter 2. Refolding of the misfolded ribozyme was monitored essentially as described (127, 147). The misfolded ribozyme was prepared by incubating 2 μM ribozyme with 10 mM $Mg(OAc)_2$ and 50 mM Na-MOPS (pH 7.0) for 5 min at 25 ˚C, and the ribozyme was stored on ice until the measurement. Refolding reactions were initiated by addition of the ribozyme and CYT-19 (or equivalent volume of CYT-19 buffer) to a

reaction mix containing the desired concentrations of $Mg(OAc)_2$ and $ATP•Mg^{2+}$, as well as 50 mM (final) Na-MOPS (pH 7.0). Refolding was monitored at 25 ˚C and the fraction of native ribozyme at different folding times was determined by moving aliquots to a folding quench and measuring the burst of substrate cleavage activity, as described in Chapter 2 and references therein. The burst amplitudes were determined from 1 min time points for 200 nM ribozyme reactions (57 nM in the cleavage reaction), and a 2 min time point was representative of the slower burst in reactions using 50 nM ribozyme (14 nM in the cleavage reaction). The '0 time points' were obtained by removing a reaction aliquot ~15 s after transferring the misfolded ribozyme to the reaction mix and immediately prior to adding CYT-19, as in Ref. (127). In later replicates, as well as all experiments with truncation mutant and tertiary contact mutant ribozymes, the 0 time points were collected by directly quenching the misfolded ribozyme, and reactions were initiated by adding the misfolded ribozyme to the reaction mix immediately after CYT-19 (as in experiments with the native ribozyme in Chapter 2). No significant deviations between the two approaches were observed. The refolding curves were normalized by the total ribozyme concetration (200 nM).

ATPase assays were performed under conditions identical to those of the refolding reactions, except for the addition of trace ^{32}P-γATP. ATPase activity during refolding reactions was monitored for times long enough to give complete refolding of the misfolded ribozyme, as determined by ribozyme activity-based assays. ATPase stimulation by the native ribozyme was always measured in parallel to the refolding reaction. The hydrolysis products were separated by thin-layer chromatography and quantified using a phosphorimager and ImageQuant. The small amounts of contaminating $^{32}P_i$ present in the ^{32}P-γATP stocks (0.4–2.3 %) were subtracted prior to multiplying the fraction of ^{32}P-γATP by the total ATP concentration.

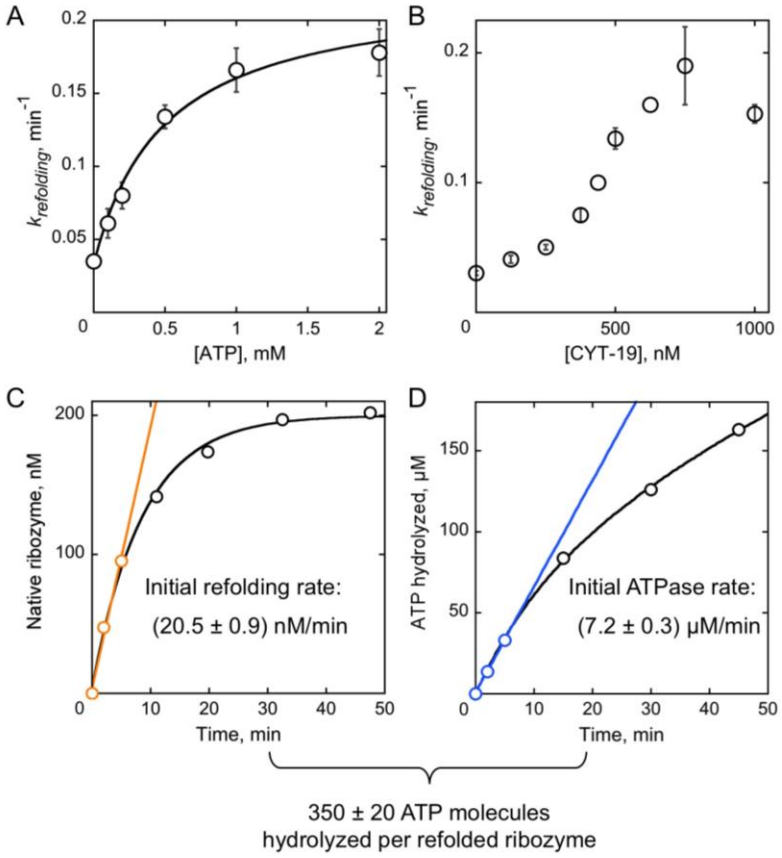


Figure 3.1: CYT-19-mediated refolding of the misfolded *Tetrahymena* ribozyme.

(*A, B*) ATP and CYT-19 concentration dependences of the refolding rate constant of the misfolded ribozyme. All measurements were performed with 200 nM misfolded ribozyme, 2 mM $Mg(OAc)_2$, 50 mM Na-MOPS (pH 7.0) and 10 % (v/v) CYT-19 buffer at 25 ˚C. The concentration of ATP was varied using 500 nM CYT-19 and the CYT-19 concentration was varied in the presence of 500 µM ATP. Averages and standard errors from two or more determinations are plotted, except for single determinations at 440 nM and 625 nM CYT-19 in *B*. (*C, D*), Initial rates of the refolding and ATPase activities of CYT-19 in the presence of the misfolded ribozyme. Representative data obtained with 500 nM CYT-19 and 500 µM ATP are shown, and the numbers indicate averages and

standard errors from 8 (ATPase) and 10 (refolding) determinations. In both cases the initial rates were estimated by fitting the data from the first 5 minutes to a linear equation (see below for alternative quantification methods). The refolding curves (*C*) were normalized to the total ribozyme concentration (200 nM). The value at the bottom indicates the ATP utilization value per refolded ribozyme derived by dividing the average initial ATPase rate by the initial rate of refolding.

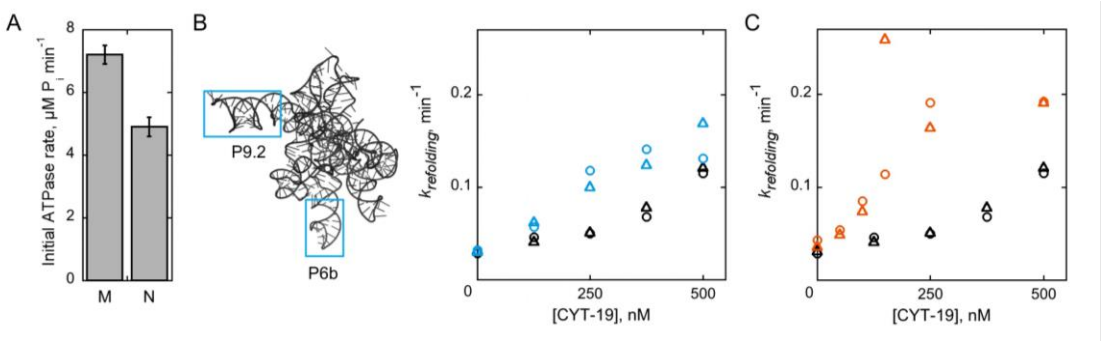


Figure 3.2: Non-productive CYT-19 interactions with exposed native helices in the misfolded ribozyme.

(*A*) Initial ATPase rates measured in the presence of the misfolded and native ribozymes at 500 nM CYT-19, 200 nM ribozyme, 500 μM ATP•Mg^{2+}, 2 mM $Mg(OAc)_2$. Averages and standard errors from 8 determinations are shown. (*B*) CYT-19 concentration dependence for refolding of the misfolded wild-type (black) and ΔP9.2+ΔP6b truncation mutant (blue) ribozymes. The P9.2 and P6b helices are highlighted in a tertiary structure model of the *Tetrahymena* ribozyme; see Figure 2.1 for sequence changes in the ΔP9.2+ΔP6b truncation mutant. The circles and triangles indicate refolding rate constants from two determinations. For the wild-type ribozyme, only the data collected in the same experiments as the data for the truncation mutant are shown; however, these rate constants are consistent with the overall averages shown in Figure 3.1*A*. The reactions were performed with 200 nM ribozyme, 500 μM ATP•Mg^{2+} and 2 mM $Mg(OAc)_2$. (*C*) CYT-19 concentration dependence for refolding of 200 nM (black) and 50 nM (orange) misfolded wild-type ribozyme. The 200 nM ribozyme data are reproduced from *B* for comparison, with the data for 50 nM ribozyme collected during the same experiments.

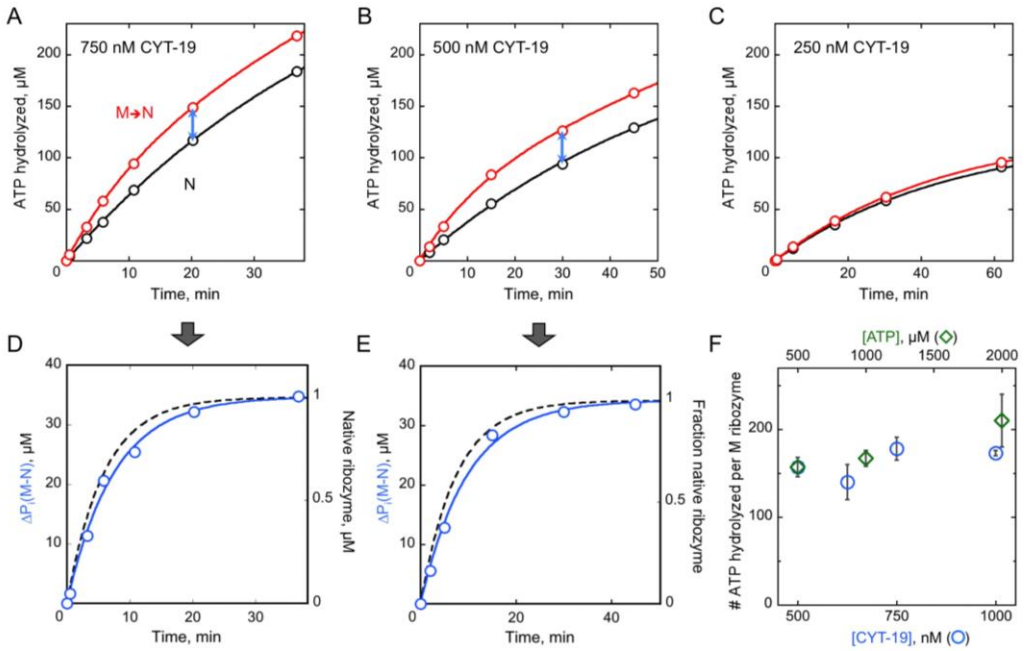


Figure 3.3: ATP utilization derived from changes in ATPase activity during the refolding reaction.

(*A*–*C*) ATPase activity monitored over the full time course of refolding of the misfolded ribozyme (red). The black curves show ATPase activity measured in parallel for the pre-formed native ribozyme. Representative time courses for 750 nM, 500 nM and 250 nM CYT-19 (500 μM ATP•Mg^{2+}, 200 nM ribozyme, 2 mM Mg^{2+}) are shown. (*D, E*) The blue curves represent the differences between ATP hydrolyzed at different times of the refolding reaction and in the presence of native ribozyme (blue arrows in *A, B*). The subtracted curves were fit by a single-exponential equation. The dashed black lines represent the average refolding kinetics measured under the same conditions by ribozyme activity (see Figure 3.1*B* and Table 3.1). No appreciable burst was observed in the presence of 250 nM CYT-19 (*C*) (see text). (*F*) Dependence of ATP utilization on CYT-19 and ATP concentrations. The ATP utilization values per ribozyme were determined from the amplitudes of the subtracted ATPase curves (*D, E*), divided by the ribozyme

concentration (200 nM). Averages and standard errors from at least two determinations are shown. CYT-19 concentration was varied in the presence of 500 μM ATP and ATP concentration was varied in the presence of 500 nM CYT-19. Only conditions at which CYT-19 provided at least ~4.5-fold acceleration of refolding are included.

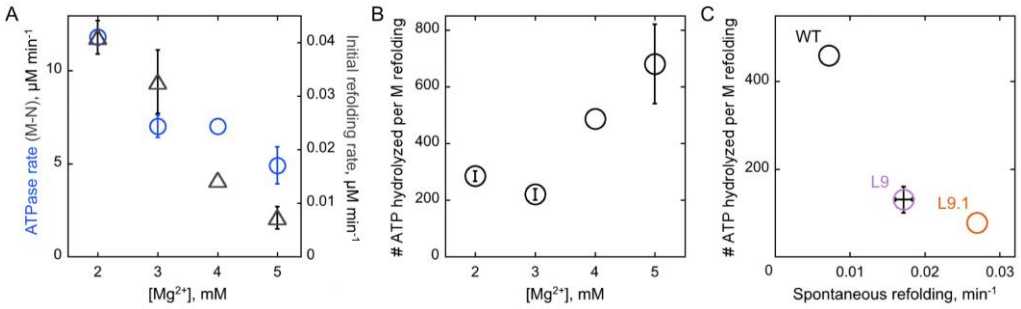


Figure 3.4: Effects of ribozyme stability on ATP utilization for refolding.

(*A*) Mg^{2+}-concentration dependence of ATP utilization for refolding of the misfolded ΔP9.2+ΔP6b ribozyme. Blue circles indicate the differences between the initial ATPase rates measured with the misfolded and native ribozymes, and the gray triangles indicate the initial rates of refolding. Averages and standard errors from two determinations are shown, except for single determinations at 4 mM Mg^{2+}. The measurements were performed with 200 nM misfolded ribozyme, 1 μM CYT-19 and 2 mM ATP•Mg^{2+} at 25 ˚C. (*B*) ATP utilization values derived from initial rates in *A*. (*C*) ATP utilization for refolding of misfolded wild-type and tertiary contact mutant ribozymes. The tertiary contact mutants are the same as those used in Chapter 2 (see Figure 2.1). ATP utilization was determined analogously to the truncation mutant in *A, B*, from the differences in ATPase stimulation by the misfolded and native ribozyme, divided by the initial rate of refolding. The ATP utilization values are plotted against the rate constants of spontaneous refolding, which provide a measure of stability, as they reflect the energetic difference between the misfolded structures and the less structured transition state ensembles for refolding to the native structure. The measurements were performed with 200 nM ribozyme, 500 nM CYT-19, 2 mM ATP•Mg^{2+} at 5 mM Mg^{2+}. Note that the larger ATP utilization value for the wild-type ribozyme than in Figure 3.3 is expected due to the higher Mg^{2+} concentration used in this experiment. Results from two

determinations for the L9 mutant and single determinations for the wild-type and L9.1 ribozymes are shown.

	Ribozyme refolding rate $k_{refolding.}$, min^{-1}	Rate of decrease in ATPase activity $k_{\Delta M\text{-}N}$, min^{-1}	% Difference $k_{refolding} - k_{\Delta M\text{-}N}$
[CYT-19], nM *(500 μM ATP)*			
500	0.134 ± 0.008	0.104 ± 0.007	22 %
625	0.16	0.12 ± 0.02	25 %
750	0.17 ± 0.03	0.16 ± 0.03	6 %
1000	0.147 ± 0.007	0.148 ± 0.003	- 0.7 %
[ATP], μM *(500 nM CYT-19)*			
1000	0.166 ± 0.015	0.16 ± 0.02	6 %
2000	0.184 ± 0.019	*0.16 ± 0.03**	11 %

* In three out of six measurements at this condition, noise in the ATPase data precluded determination of a rate constant. Average and standard deviation of the rate constants from the three remaining measurements are shown.

Table 3.1: Consistency between refolding rate constants determined by ribozyme activity and ATPase assays.

The rate constants were determined from single-exponential fits to changes in the fraction of native ribozyme determined by ribozyme activity assay or changes in ATPase activity at different times of the refolding reaction (see Figure 3.3*D, E*). The values are averages and standard errors from at least two determinations.

$[Mg^{2+}]$, mM	Spontaneous refolding, k_{spont}, min^{-1}	Refolding with 1 μM CYT-19 (2 mM ATP) $k_{+CYT-19}$, min^{-1}	Acceleration $k_{+CYT-19}/ k_{spont}$
2	0.0424 ± 0.0016	0.207 ± 0.013	4.9
3	0.0256 ± 0.0004	0.163 ± 0.006	6.3
4	0.0147	0.072	4.9
5	0.0062 ± 0.0005	0.037 ± 0.001	6.0

Table 3.2: CYT-19-dependent acceleration of refolding of the misfolded truncation mutant ribozyme at 2–5 mM Mg^{2+}.

Refolding of the misfolded truncation mutant ribozyme was monitored using 200 nM ribozyme, 1 μM CYT-19, 2 mM ATP•Mg^{2+} at 25 ˚C. Averages and standard errors from 2 determinations are shown, except single determinations at 4 mM Mg^{2+}.

Ribozyme	Spontaneous refolding, k_{spont}, min^{-1}	Refolding with 0.5 μM CYT-19 (2 mM ATP) $k_{+CYT-19}$, min^{-1}	Acceleration $k_{+CYT-19}/ k_{spont}$
WT	0.0072	0.0403	5.6
L9	0.0172 ± 0.0011	0.085 ± 0.013	4.9
L9.1	0.027	0.223	8.3

Table 3.3: CYT-19-dependent acceleration of refolding of misfolded wild-type and tertiary contact mutant ribozymes at 5 mM Mg^{2+}.

Measurements were performed with 200 nM ribozyme, 500 nM CYT-19, 2 mM ATP•Mg^{2+} and 5 mM $Mg(OAc)_2$ at 25 ˚C. The values are from single determinations for wild-type and L9.1 mutant ribozymes and two determinations for the L9 mutant ribozyme.

The text and figures in Chapter 4 were originally published by the National Academy of Sciences:
Mallam, A.L., Jarmoskaite, I., Tijerina, P., Del Campo, M., Seifert, S., Guo, L., Russell, R., Lambowitz, A.M. (2011) Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *Proc Natl Acad Sci U S A* 108(30):12254-9.

A.L.M., I.J., P.T., M.D.C., R.R., and A.M.L. designed research; A.L.M., I.J., P.T., S.S., and L.G. performed research; A.L.M., I.J., M.D.C., R.R., and A.M.L. analyzed data; and A.L.M., I.J., P.T., M.D.C., S.S., L.G., R.R., and A.M.L. wrote the paper.

Anna Mallam designed and performed experiments and analysis for Mss116.
Inga Jarmoskaite and Pilar Tijerina performed experiments and analysis for CYT-19.
Unpublished supporting activity data in Appendix B (Figure B8 and Table B5) were collected by Inga Jarmoskaite and Serena Chen.

Chapter 4: Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail

INTRODUCTION

DEAD-box proteins comprise the largest family of helicases and play critical roles in all aspects of RNA metabolism, including RNA splicing, translation, ribosome assembly, RNA degradation, and RNA transport (2, 190). While their functions vary broadly, all DEAD-box proteins have a conserved helicase core consisting of two RecA-like domains (denoted domains 1 and 2) separated by a flexible linker and operate by a mechanism involving conformational changes within this core (2, 36). These conformational changes couple cycles of ATP-binding and hydrolysis to RNA binding and release by the core, enabling DEAD-box proteins to promote local RNA unwinding and remodeling of structured RNAs and RNA-protein complexes (1).

Structural and biochemical studies have given important insights into how DEAD-box proteins unwind RNA (2, 36). The two helicase core domains are separated from each other in an open state in the absence of substrates, but they interact to form a compact, closed state upon binding of ATP and RNA. In this closed state, the interface

between the two core domains forms a catalytic site for ATP hydrolysis and an RNA-binding cleft, which can accommodate a short region of a duplex strand (54). The tight binding of the RNA strand within this cleft results in a bend that is incompatible with partner-strand base pairing, leading to local RNA unwinding (13, 17, 19). ATP hydrolysis and dissociation of the P_i product are then necessary to release the bound single-stranded RNA (17, 83, 95).

While the helicase core is central to the biochemical activities of DEAD-box proteins, most also have substantial extensions or additional domains at their N- and/or C-termini (36). These extensions vary widely in size, composition, and function, but many are thought to interact with RNA or protein components of RNP-complexes to direct the DEAD-box proteins to desired sites of action. A common type of extension is a basic C-terminal sequence (C-tail), which is typically predicted to lack a defined structure, and in some cases, truncation of the C-tail has been shown to affect activities of DEAD-box proteins (14, 16, 33, 97). Unstructured extensions, such as C-tails, are absent from the crystal structures of DEAD-box proteins reported to date, likely reflecting their flexibility or disorder, and thus there is little direct information about how they function together with the helicase core.

As general RNA chaperones, the DEAD-box proteins Mss116p of *Saccharomyces cerevisiae* and CYT-19 of *Neurospora crassa* are important model systems for studying the mechanisms of DEAD-box proteins (Figure 4.1*A*). These proteins interact non-specifically with RNA substrates and disrupt stable inactive structures that limit productive RNA folding (85, 127, 129). Mss116p is required for the efficient splicing of 13 different mitochondrial (mt) group I and group II introns and is also involved in translational activation and RNA end processing (121). It consists of a helicase core that is flanked on one side by an N-terminal extension (NTE) and on the other by a

functionally important α-helical C-terminal extension (CTE) and a basic tail (C-tail; Figure 4.1*A*). The related *N. crassa* DEAD-box protein CYT-19 plays similar roles in mt RNA metabolism and is to a large degree functionally interchangeable with Mss116p *in vitro* and *in vivo* (85, 121, 126, 129).

High-resolution X-ray crystal structures have been obtained for Mss116p in complex with a single-stranded U_{10}-RNA oligonucleotide and three different ATP analogs (AMP-PNP, ADP-BeF_3^-, and ADP-AlF_4^-) (55). These structures show the entire closed-state helicase core (residues 88-505) and CTE (residues 506-596), and they reveal that the CTE packs against the core, stabilizing it and functioning in RNA unwinding (Figure 4.1*B*). The helicase core induces a bend in the bound RNA, similar to that seen in other DEAD-box proteins, while the CTE produces a second bend resulting in RNA crimping. Because the C-tail was not present in the crystallized construct of Mss116p and the NTE was not visible, their structures and how they might contribute to RNA-chaperone function have remained unclear.

Based on sequence similarity and limited proteolysis studies, CYT-19 is predicted to have a helicase core, CTE, and basic C-tail similar to those of Mss116p (Figure 4.1*A*). The C-tails of Mss116p and CYT-19 enhance RNA binding, and they have been suggested to contribute to RNA-chaperone activities by forming additional, non-specific interactions with RNA substrates (16, 33). This interaction could potentially tether the core to structured RNAs, either rigidly or with sufficient flexibility to allow it to sample neighboring regions and binding orientations while remaining near the substrate. However, the ability of C-tails to bind separate sites on nucleic acid substrates and the degree of flexibility between the C-tail and the helicase core upon RNA binding have not been determined for these or any other DEAD-box protein.

Here we use small-angle X-ray scattering (SAXS) to obtain solution structures of full-length Mss116p and CYT-19 and a series of deletion mutants both in the open state and in the closed state in complex with an adenosine nucleotide and single-stranded U_{10}-RNA. Our results reveal the locations of the NTE and C-tail; show surprisingly that the two core domains of Mss116p adopt a preferred relative orientation in the open state; and enable us to visualize the transition to a compact closed state upon binding of substrates. An analysis of Mss116p and CYT-19 in complex with larger chimeric oligonucleotides containing duplex DNA extensions indicates a mechanism in which the C-tail flexibly tethers the core to large, physiological RNA substrates, while the core samples surrounding regions and unwinds short RNA duplexes non-specifically.

RESULTS

Experimental strategy

For both Mss116p and CYT-19, we collected SAXS data for full-length proteins and for constructs lacking the basic C-tails (Figure 4.1*A*). Additionally, for Mss116p we used versions that lack the NTE, which is much smaller and not conserved in CYT-19. All of these truncations retain the helicase core and CTE regions present in the previous X-ray crystal structures of Mss116p (Figure 4.1*B*). Below, we first present data for Mss116p by itself and in complex with nucleic acid substrates (Figure 4.1*C*), and then we present corresponding data for CYT-19.

SAXS reconstructions reveal the spatial organization of full-length Mss116p

The SAXS profiles of Mss116p constructs in solution are shown in Figure 4.2*A*. Particle molecular weights were estimated from the scattering intensity at zero angle ($I(0)$) when calibrated against protein and RNA standards and indicated that all of the constructs are monomeric and free from aggregates under these solution conditions

(Table B1). Values for the radius of gyration (R_g) ranged from 32.8 Å for the minimal Mss116p/ΔNTE+ΔC-tail construct to 38.0 Å for full-length Mss116p (Table B1) and were in good agreement with the hydrodynamic radii of the proteins measured by size-exclusion chromatography (SEC) (Table B2).

The scattering profiles can be represented in real space by a distance distribution function, $P(r)$, revealing two maxima for each protein construct (Figure 4.2*B*). This pattern indicates the presence of a dumb-bell-like scattering macromolecule, as expected for the two separated helicase core domains in the open state (36). The first maximum, at ~30 Å for each construct, corresponds to the size of the domains, and the second at ~50 Å corresponds to their separation (191). Estimates of the maximum particle diameter, D_{max}, for the different constructs vary from 115-135 Å (Table B1) and suggest that, in the absence of ligands, Mss116p adopts an elongated conformation in solution.

To obtain more specific information on solution structures of the helicase core and to evaluate the positions of the C-tail and NTE, we generated particle envelopes from the SAXS data using DAMMIN and GASBOR (192, 193) and atomic models using BUNCH (194) (Figure 4.2, Figure B1*A* and Table B3). For the latter method, we used the domain structures determined by X-ray crystallography and modeled the C-tail and NTE as unstructured extensions, as predicted from their sequences (33) and supported by far ultra-violet circular dichroism (far-UV CD) spectra (Figure B2 and Table B4). Reconstructed envelopes for all of the protein constructs revealed well-separated lobes corresponding to the two domains of the helicase core (Figure 4.2), and the scattering data were well-described by single molecular models calculated using BUNCH (χ = 0.67-2.13; Figure 4.2*A* and Table B3).

In addition to the core domains, the SAXS envelope for full-length Mss116p included two clear protrusions that were not seen in the NTE and/or C-tail deletion

mutants (Figure 4.2*C*). The locations of these protrusions indicate that the C-tail extends outward from domain 2 and the NTE extends from domain 1 on the opposite face of the protein. Although previous structures of DEAD-box proteins in the absence of ligands have suggested that the two core domains are separated from each other and free to adopt different relative orientations, our reconstructions indicate that Mss116p core domains are restricted sufficiently that the extensions appear as discrete protrusions. This result implies that the two domains exist in a preferred relative orientation, even in the absence of substrates (see Discussion).

The best-fit molecular models for those constructs containing the C-tail depict the tail in a strongly bent conformation. Although the full conformational ensemble of the C-tail cannot be determined from these data, they raised the possibility of a conformational bias that predisposes the C-tail to being less than fully extended. To explore this question further, we performed rigid-body modeling with conformational sampling. Starting from the simplest model in which the C-tail and NTE are fully unstructured, an ensemble optimization method (EOM) was used to select a group of conformers that best describes the scattering profile for each protein construct from a large pool of conformers that randomly samples the conformational space of unstructured polypeptides (195). The optimized ensembles for proteins that include the C-tail have R_g and D_{max} distributions with lower averages than those of the pool of random structures (Figure B3), suggesting that the C-tail is biased to be more compact than expected for a fully random chain. In contrast, the ensembles for the mutants lacking the C-tail have R_g and D_{max} distributions similar to or slightly larger than those of the corresponding random pools (Figure B3). Thus, we conclude that the C-tail has a conformational bias that predisposes it to be in proximity to domain 2.

The C-tail is positioned to bind RNA extending from the helicase core

We next used SAXS to study the closed-state conformation of Mss116p by forming complexes with U_{10}-RNA substrate and the non-hydrolyzable ATP analog ADP-BeF_x (Figure 4.3). Mss116p binds tightly to U_{10}-RNA in the presence of ADP-BeF_x under the SAXS buffer conditions (Figure B4), and the formation of a stable complex was confirmed by SEC (Figure B5 and Table B2). The R_g values for the U_{10}-RNA bound proteins range from 26.3-33.8 Å, indicating a decrease of at least 4 Å upon formation of the closed state (Table B1). The experimental scattering profile and R_g values measured for closed-state Mss116p/ΔNTE+ΔC-tail agree with those calculated from the crystal structure of the equivalent complex (Figure 4.3*A* and Table B3), indicating that the closed complex of Mss116p in solution is similar to that in the crystal.

The $P(r)$ functions for the U_{10}-RNA complexes are bell-like and are typical of compact globular particles (191), indicating that the two core domains are no longer separated but instead behave as a single structural module (Figure 4.3*B*), as seen in the crystal structure (Figure 4.1*B*) (55). Therefore, we used this structure as a constraint in BUNCH reconstructions. Models generated using DAMMIN, GASBOR and BUNCH gave good internal agreement, and the SAXS envelopes accommodated the two interacting core domains in the closed state (Figure 4.3*C*-*F* and Figure B1*B*). Importantly, the reconstructions indicate that the NTE and C-tail protrude from the helicase core in positions similar to those in the open, unbound state, with no indication that the C-tail interacts with the RNA-binding cleft or other regions of the core (Figure 4.3). The latter result provides evidence against models in which the C-tail communicates allosterically with the core. Rather, the disposition of the C-tail adjacent to domain 2 suggests that it is poised to interact with nucleic acid extending from the RNA-binding cleft of the core, a hypothesis that is probed further by experiments below.

The C-tail interacts with longer nucleic acid substrates. The position of the C-tail extending adjacent to the U_{10}-RNA bound to the core is consistent with previous models in which the C-tails of Mss116p and CYT-19 interact with nearby RNA regions (16, 33). To test this model and to probe the degree of flexibility between RNA-binding interactions of the C-tail and the helicase core, we examined the solution structures of versions of Mss116p with or without the C-tail in complexes with larger chimeric substrates that contain a short ssRNA linked to a DNA duplex extension. These substrates were based on similar ones used in previous DEAD-box protein studies (13, 16, 22) and were composed of an 11-nt ssRNA covalently linked at either its 5'-end (RNA-DNA duplex 1) or 3'-end (RNA-DNA duplex 2) to a 10-bp hairpin DNA duplex (Figure 4.1*C*). DNA extensions do not compete effectively with ssRNA for binding to the helicase core, which relies on interactions with 2'-OH groups of RNA, but can interact functionally with the C-tail, as demonstrated previously by enhancements in RNA unwinding (16). Thus, both substrates are expected to bind with the ssRNA in the RNA-binding cleft of the core and the rigid duplex extending out from one or the other end, arriving at very different spatial positions depending on its polarity. SAXS data and *ab-initio* reconstructions for the substrates by themselves (Figure B6) are consistent with them adopting the predicted secondary structure shown in Figure 4.1*C*.

SAXS profiles were determined for constructs of Mss116p with or without the C-tail in complex with ADP-BeF_x and either RNA-DNA duplex 1 or RNA-DNA duplex 2 (denoted substrates 1 and 2, respectively; Figure 4.4*A*). The complexes were isolated by SEC prior to SAXS measurements to ensure sample homogeneity (Figure B5). Although Mss116p and other DEAD-box proteins can potentially function as oligomers on large substrates (196), the SEC and SAXS measurements indicate that Mss116p binds both substrates as a monomer under these conditions. R_g values were approximately 36 Å and

33 Å for complexes in the presence and absence of the C-tail, respectively, while $P(r)$ functions displayed characteristics of globular species (Figure 4.4*B* and Table B1). We used the multiphase bead-modeling program MONSA to reconstruct *ab-initio* models that include two phases, protein and nucleic acid, by accounting for the different scattering intensities of these components (192, 197). Several independent runs gave reproducible models with an average NSD of less than 0.65 that described the scattering data well (χ = 0.7-1.5) (Table B3).

The reconstructions for the minimal helicase core construct Mss116p/ΔNTE+ΔC-tail indicate that the polarity of the duplex region dictates its position and orientation. For RNA-DNA duplex 1, the DNA duplex region of the substrate protrudes at the 5' end of the ssRNA, close to the CTE of Mss116p and along the long axis of the protein (Figure 4.4*C*). This position is near the region occupied by the C-tail in the intact protein in solution (see Figure 4.2 and Figure 4.3 above), and indeed the reconstruction of the same duplex substrate in complex with the Mss116p construct that includes the C-tail shows additional volume that corresponds to protein in this region, indicating interaction of the C-tail with the duplex extension (Figure 4*E*). In contrast, the duplex region of RNA-DNA duplex 2 extends from the 3' end of the ssRNA, closer to the center of the helicase core and nearly perpendicular to the long axis of the protein (Figure 4*D*). Strikingly, the reconstruction of this complex with Mss116p including the C-tail indicates that the changes in position and orientation of the duplex extension are matched by the C-tail, such that it remains associated with the duplex (Figure 4*F*), presumably interacting by electrostatic interactions with the negatively charged nucleic acid. These results indicate that the C-tail and core bind different sites of large substrates and that the C-tail remains anchored to the extensions while moving relative to the core over a wide region of space.

CYT-19 behaves similarly to Mss116p in solution

To explore whether the behavior of Mss116p is characteristic of other DEAD-box protein RNA chaperones, we performed additional SAXS experiments with CYT-19, which performs similar functions to Mss116p (85). CYT-19 lacks a significant NTE but is similar to Mss116p in that the helicase core is followed by a homologous CTE and a C-tail, which is predicted to be unstructured (Figure 4.1*A*). The C-tail of CYT-19 is shorter than that of Mss116p (~49 residues based on partial proteolysis results compared to 68 residues for Mss116p), but it is more basic and has a higher proportion of arginine residues (16, 33).

We performed SAXS measurements on two versions of CYT-19, a wild-type construct and one in which the C-tail was removed by deleting the C-terminal 49 amino acid residues (Figure 4.5 and Figure B7) (16). As for Mss116p, the *ab-initio* and rigid-body SAXS reconstructions indicate that CYT-19 in the absence of ligands has an extended structure with two distinct globular domains (Figure 4.5*A*-*C*) and that this open conformation compacts to a closed globular structure upon binding of ADP-BeF_x and U_{10}-RNA (Figure 4.5*D*-*F*). In both the closed and open conformations, BUNCH atomic models show that the C-tail emerges from domain 2 in a position to form additional RNA contacts similar to that of Mss116p. Importantly, the ternary complexes with the large substrates RNA-DNA duplex 1 and 2 gave scattering profiles that are well described by MONSA models in which CYT-19 binds as a monomer and the C-tail interacts with the duplex extension regardless of its polarity (Figure 4.5*G*-*I*). Thus, the position and flexibility of the C-tail are properties common to Mss116p and CYT-19.

DISCUSSION

Here, we used SAXS to obtain the first solution structures of full-length DEAD-box proteins bound to nucleic acid substrates. These structures give important insights into the architectures of the RNA-chaperone proteins Mss116p and CYT-19 and their interactions with RNA. They show directly that the two domains of the helicase core undergo a large-scale, global compaction to a 'closed' conformation upon binding RNA and adenosine nucleotide in solution, supporting models proposed from crystallography and fluorescence measurements (36). Further, they reveal the locations of the basic tail and NTE in Mss116p, and provide insight into the function of the basic tail in binding to large substrates.

The SAXS results indicate that the NTE, which is attached to domain 1 of Mss116p, is on the opposite face of the protein from the bound RNA. It apparently does not contact RNA but may interact with protein partners, such as the mt RNA polymerase (198). The C-tail is attached to the CTE of domain 2. It does not appear to interact with the helicase core but is instead positioned adjacent to domain 2, where it extrudes into the solvent and is available to bind to structured RNAs.

The protrusions corresponding to the NTE and C-tail provide landmarks on each of the two core domains that enable us to infer their relative orientation in the SAXS solution structures in both the open and closed states. Surprisingly, we find that the two core domains of Mss116p adopt a preferred relative orientation even in the open state in the absence of substrates. This result was not expected from previous X-ray crystallographic studies of DEAD-box proteins in the absence of substrates, which showed that the two core domains could exist in multiple relative orientations and suggested that they were free to move independently of each other via the flexible linker (36). The preferred orientation of the two domains found for Mss116p could reflect that

the linker between them is less flexible than was believed previously and/or that there are transient or unstable interactions between the domains in the open state. While it is not known if all DEAD-box proteins behave similarly in this respect, pre-alignment of the two core domains prior to substrate binding would be advantageous in reducing the entropic cost of binding ATP and RNA at sites formed at the domain interface.

Importantly, the SAXS data of RNA-protein complexes provides information about the structure, function, and flexibility of the C-tail, which could not be obtained by X-ray crystallography. In complexes with nucleic acid substrates that include rigid duplex DNA extensions at either the 5' or 3' end of a short ssRNA segment, the C-tail interacts with the extension regardless of the polarity. Because the ssRNA portion of the substrate can bind the core in only one orientation, the attached duplex segments extend from the core in very different directions. The ability of the tail to contact both extensions indicates that it is sufficiently flexible to interact with neighboring parts of large substrates over a wide arc extending from the core (Figure 4.6*A*).

In the context of large RNAs that are the physiological substrates for these DEAD-box proteins, we envision the basic tail as being fixed by non-specific, electrostatic interactions with the RNA and then functioning as a tether to allow the helicase core to interact with nearby single-stranded and double-stranded regions of the RNA (Figure 4.6*B*). These regions can include segments that are adjacent in linear sequence, as well as those that are non-contiguous but close in tertiary structure (23). Binding of the core to ssRNA segments could enhance folding by sequestering these segments, stabilizing intermediates or preventing formation of kinetically trapped species, while binding to dsRNA segments disrupts kinetic traps as the strands are separated in an ATP-dependent unwinding reaction. The two strands are then released at

different times, allowing them to find new partners and the RNA to refold (22, 127, 153, 199).

Our model suggests that RNA binding by the DEAD-box protein C-tail plays multiple roles in RNA-chaperone activity. First, the interactions of the C-tail increase the affinity for large RNA substrates, thereby enhancing activity. For CYT-19, the C-tail has been shown to increase the efficiency for unwinding RNA duplexes with extensions, implying that the C-tail binds the extension and that this interaction is maintained up to and through the transition state for duplex unwinding (16). Second, the C-tail may remain bound to the RNA substrate after the core has undergone a cycle of RNA binding and release, thus allowing the protein to perform multiple rounds of unwinding while remaining tethered to the structured RNA. Under these circumstances, the flexibility of the tail relative to the core, as shown here, would be critical to allow the core to change positions and orientations during large-scale RNA conformational transitions of complex RNAs, such as group I and group II introns. An analogous mechanism may apply to DEAD-box proteins that use ancillary domains to target specific RNAs or RNA-protein complexes, such as the bacterial DbpA/YxiN DEAD-box proteins (28, 29, 200, 201). In these cases, the RNA-unwinding activity of the helicase core would be limited to the spatial proximity of the specific tethering site. A final possibility is that the unstructured tails participate in disruption of RNA structure, as many positively-charged proteins have been shown to possess RNA-chaperone activity, presumably by binding preferentially to unstructured regions of RNA (119). However, for CYT-19, Mss116p, and other DEAD-box proteins that have been tested, the principal unwinding activity arises from the helicase core, as only low levels of unwinding are observed in the absence of ATP, and much of this activity presumably reflects ATP-independent unwinding by the core (13, 128).

While tethering by the C-tail is likely to be broadly important for the RNA-chaperone activity, the available evidence suggests that its functional contribution can vary between DEAD-box proteins. The activity of CYT-19 appears to depend strongly upon the C-tail, as its removal compromises the RNA binding and RNA unwinding of extended duplexes by 7-20-fold (16). For Mss116p, however, deletion of the C-tail and a small neighboring region of the CTE has smaller effects, causing only a two- to five-fold decrease in splicing efficiency for group II introns aI5γ and bI1 *in vitro* and a modest decrease in levels of spliced *COX1* and *COB* mRNAs in strains containing multiple group I and II introns *in vivo* (33). Thus, it is possible that the C-tail plays a greater role in the functions of CYT-19 than Mss116p. Mss116p is highly expressed (202) and non-specific binding by the core in the absence of adenosine nucleotide appears to be stronger than for CYT-19 (33). Further, Mss116p and other DEAD-box proteins appear to function as oligomers under some conditions, raising the possibility that one protomer can provide an anchor for RNA binding by another protomer, analogous to the role of the C-tail (5, 14). These differences may lead to less dependence by Mss116p on the basic C-tail for RNA binding and unwinding.

Flexible basic tails are a common feature of DEAD-box proteins and are also found in the related SWI/SNF family of superfamily 2 helicase proteins, members of which interact with DNA and function in chromatin remodeling (1, 36, 190, 203-205). Interestingly, both of these groups of protein bind non-specifically to nucleic acid structures and promote local structural rearrangements. Basic polypeptide extensions are also found in other classes of RNA-binding proteins, raising the possibility that these tails are a hallmark of proteins that bind non-specifically to RNAs of diverse structure. Consistent with this hypothesis is the observation that an unstructured, basic tail in the La protein has recently been shown to participate in binding to diverse RNAs (206). The

unstructured and pliable nature of these extensions, as indicated by the results here, may be critical to allow the flexibility necessary for them to interact productively with the diverse shapes and binding surfaces presented by structured RNAs.

MATERIALS AND METHODS

All experiments and data analysis with Mss116 were designed and performed by Anna Mallam.

Protein expression and purification

Full-length Mss116p (excluding its mitochondrial (mt) targeting sequence (residues 1-36)) was cloned in pMAL-c2t (a derivative of pMAL-c2x; New England Biolabs; Ipswich, MA) and expressed as an N-terminal fusion to maltose-binding protein (MBP), as described (21). Derivatives of pMAL-Mss116p that express the deletion proteins Mss116p/ΔNTE (deletion of N-terminal extension; residues 37-87), Mss116p/ΔC-tail (deletion of C-terminal tail; residues 597-664), and Mss116p/ΔNTE+ΔC-tail (deletion of both the N-terminal extension and C-terminal tail) were created as described (207). The proteins were expressed in *Escherichia coli* Rosetta 2 (EMD Biosciences) grown in ZYP-5052 auto-inducing medium for 24 h at 22 °C and purified at 4 °C, as described (207). Purification steps included (i) removal of nucleic acids by polyethyleneimine precipitation; (ii) isolation of MBP-Mss116p by amylose affinity chromatography (New England Biolabs); (iii) removal of the MBP tag by digestion with tobacco etch virus (TEV) protease; and (iv) isolation of Mss116p by affinity chromatography, using a heparin-Sepharose column (GE Healthcare). Proteins were dialyzed into SAXS buffer (20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT)). Protein concentrations were determined by Bradford Assay (Bio-Rad Laboratories) and by absorbance at 280 nm. Extinction

coefficients were calculated from the protein sequence by using the ExPASy Proteomics Server ProtParam tool (208).

Full-length CYT-19 and CYT-19/ΔC-tail (residues 578-626) were expressed and purified as MalE fusions using a similar strategy, as described (16). Prior to SAXS experiments, CYT-19 and CYT-19/ΔC-tail were concentrated in a 30-kDa MWCO concentrator (Millipore) and re-diluted into a buffer suitable for SAXS measurements (20 mM Tris-HCl (pH 7.5), 200 mM KCl, 5 mM $MgCl_2$, 20% glycerol, 0.4 mM EDTA, 0.2 mM DTT).

SAXS sample preparation and characterization

The oligonucleotides U_{10}-RNA, RNA-DNA duplex 1 and RNA-DNA duplex 2 (Figure 4.1*C*) were synthesized by Integrated DNA Technologies and reconstituted in sterile, double-deionized water to 1 mM. The chimeric substrates RNA-DNA duplex 1 and RNA-DNA duplex 2 were annealed by heating to 94 ºC for 1 min and slowly cooling to room temperature over 1 h. Oligonucleotide purity was confirmed by size-exclusion chromatography (SEC) (Figure B5). The equilibrium dissociation constant for the complex between Mss116p/ΔNTE and a fluorescein-labeled U_{10}-RNA (FAM-U_{10}-RNA) was measured under the SAXS buffer conditions (Figure B4). This K_d was in the nM range and indicates tight binding of ssRNA at the concentrations of protein necessary for SAXS measurements (30-70 μM). Complexes (~50 μM) for analysis by SAXS were assembled from a mixture of 1:1.2-1.5 protein:oligonucleotide in the presence of 0.5-1.0 mM ADP-BeF_x and were formed at room temperature for at least 10 min. ADP-BeF_x was prepared as described (55).

All proteins and complexes were purified to homogeneity by SEC on a BioLogic DuoFlow or an AKTA FPLC using a Superdex 200 10/300 GL analytical gel filtration

column equilibrated in the SAXS buffer (20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 5 mM $MgCl_2$, 1 mM DTT for Mss116p or the same buffer with 200 mM KCl for CYT-19). Complex formation was assessed by the A_{260}/A_{280} ratio (Figure B5). The lack of dissociation observed when purified complexes were re-analyzed by SEC confirmed complex stability (Figure B5). The relative elution volumes of the Mss116p SAXS samples were compared to those of molecular weight standards to estimate their hydrodynamic radii, R_H. The relative elution volume was calculated as:

$$K_{AV} = \frac{V_e - V_o}{V_g - V_o} \qquad \text{(Equation 1)}$$

where V_e is the elution volume, V_o is the void volume determined by the elution of Blue Dextran 2000 (Mr = 2,000 kDa), and V_g is the geometric column volume determined by the elution of free tyrosine (Mr = 0.13 kDa). A standard curve was plotted of K_{AV} versus log(R_H) (Figure B5). Molecular weight standards were thyroglobulin (R_H = 85 Å), apoferritin (R_H = 61 Å), β-amylase (R_H = 50.4 Å), alcohol dehydrogenase (R_H = 46 Å), bovine serum albumin (R_H = 35.5 Å), carbonic anhydrase (R_H = 24.3 Å), and cytochrome c (R_H = 17 Å).

Proteins and complexes were concentrated by centrifugation at 4 °C in a 30-kDa MWCO concentrator (Millipore) and concentrations were determined as above. SAXS sample concentrations were 1-3 mg/ml. Care was taken to subject proteins and complexes to only one freeze-thaw cycle before SAXS measurements. Additional complexes with U_{10}-RNA and ADP-BeF_x were assembled at the synchrotron in the presence of excess RNA and nucleotide. In these cases, the flow through from centrifugation in a 30-kDa MWCO concentrator was used as a buffer blank. Scattering profiles and R_g values for these complexes were in good agreement with those purified by SEC.

Far ultra-violet circular dichroism

All measurements were performed in SAXS buffer using a thermostatically controlled 0.01-cm path-length cuvette at 25 °C and a Jasco J-815 spectrometer. Scans were taken between 175 and 260 nm at a scan rate of 0.5 nm s^{-1} with 30 accumulations. Sample concentrations were 1.1-2.5 mg/ml. Spectra were analyzed by the CDSSTR analysis program (209), using the DichroWeb online circular dichroism analysis website (210).

SAXS data collection

Synchrotron radiation X-ray scattering data were collected at the Advanced Photon Source beamlines 12-ID-C and 18-ID-D and recorded on two-dimensional CCD detectors. Twenty exposures of 1 s each were acquired for each sample at a sample-to-detector distance of 2.0 m and over a range of momentum transfer $0.007 < q < 0.35$ $Å^{-1}$, $q = 4\pi \sin(\theta)/\lambda$. No measurable radiation damage was observed under these conditions. Scattering data were radially averaged to produce one-dimensional profiles of scattering intensity *vs.* q. Data and a buffer blank were collected for three different concentrations of each sample to check for concentration-dependent scattering effects, such as aggregation or interparticle interference. The scattering intensity at zero angle ($I(0)$) was calibrated against known concentrations of protein and RNA standards (cytochrome c and U_{10}-RNA, respectively).

SAXS data analysis

Analysis of the SAXS data was performed with IGOR-Pro (WaveMetrics) and ATSAS (version 2.4) software. Background-subtracted scattering data were subject to an initial inspection to determine sample monodispersity. Plots of $\log[I(q)]$ versus $\log(q)$ displayed a flat region in the lowest q regime, consistent with the presence of

monodisperse particles (211). The zero-angle scattering intensity, $I(0)$, and the radius of gyration, R_g were evaluated using the Guinier approximation for the scattering intensity, $I(q)$, at very low q angles ($qR_g < 1.3$) (Table B1) (212):

$$I(q) = I(0)\exp(\frac{-q^2 R_g^{\ 2}}{3}) \qquad \text{(Equation 2)}$$

Guinier plots of $\ln[I(q)]$ versus q^2 were inspected for linearity, and samples at different concentrations were checked for consistent values of $I(0)/C$ and R_g to confirm sample monodispersity. $I(0)$ is related to the number of scattering particles per unit volume; particle molecular weights of samples with homogeneous components (MM_p) were estimated using the equation:

$$MM_p = (I(0)_p / c_p) \times \frac{MM_{st}}{I(0)_{st} / c_{st}} \qquad \text{(Equation 3)}$$

where $I(0)_p$ and $I(0)_{st}$ are the scattering intensities at zero angle, MM_p and MM_{st} are the molecular masses and c_p and c_{st} are the concentrations measured (in g/L) for the sample under study and the protein or RNA standard, respectively (Table B1) (213). The good agreement of these values with those calculated from the primary sequence indicates that the samples are monomeric, free from aggregates, and lack long-range particle interactions under the experimental conditions. In most cases, the scattering data measured for the highest-concentration sample were of high quality and were used in subsequent structural analysis. Where a slight upturn at very low q in plots of $\log[I(q)]$ versus $\log(q)$ was observed for high-concentration samples, the high q values for these data were merged with low q values measured at a lower concentration to obtain a high-quality dataset for the whole range of q. Final scattering profiles were analyzed with the indirect transform program AUTOGNOM to obtain the maximum particle dimension, D_{max}, the distance distribution function, $P(r)$, and a value for the R_g calculated from the

entire scattering profile (214). The latter were in good agreement with those calculated using the Guinier approximation, which demonstrates the internal consistency of the data (Table B1) (211).

***Ab-initio* shape reconstructions**

Three-dimensional reconstructions of the SAXS data using *ab-initio* modeling were performed for the constructs of Mss116p and CYT-19 in the free and U_{10}-RNA-bound states. Nucleic-acid components give rise to a significantly greater scattering intensity compared to an equivalent amount of protein because of their higher electron density (192, 197). In this analysis, U_{10}-RNA-protein-ADP-BeF_x complexes (which consist of < 5% nucleic acid) were assumed to be single-component systems of uniform electron densities. Low-resolution particle envelopes were determined by using the simulated annealing procedures implemented by the programs DAMMIN (for $q_{max}<8/R_g$) and GASBOR (192, 193, 215). In DAMMIN, a single-phase dummy atom protein model is generated from an initial spherical search volume of densely packed beads that can be either protein or solvent. GASBOR uses an assembly of dummy residues that form a polypeptide chain-compatible model to encourage shape reconstructions with protein-like properties. Multiple independent runs were performed for each analysis, and these resulted in reproducible results with good fits to the experimental data (Table B3). Reconstructions were averaged using the DAMAVER suite to confirm the uniqueness and reliability of the solution (216). This method generates the normalized spatial discrepancy (NSD), a quantitative measure of the similarity between sets of three-dimensional points, which was used to compare models of the same resolution. A low value for NSD (~1) indicates a good agreement between models (Table B3) (217). The final DAMMIN model represents a refinement of the average of 10 individual

reconstructions against the experimental data (Figure 4.2, Figure 4.3 and Figure 4.5), while the most typical GASBOR model with the lowest average NSD compared to the others is shown in Figure B1. Low-resolution envelopes were also reconstructed using DAMMIN from SAXS data measured for RNA-DNA duplex 1 and RNA-DNA duplex 2, as described above (Figure B6).

Homology modeling of CYT-19

A structural model of CYT-19 was generated from the crystal structure of Mss116p using the program MODELLER (218) and the published sequence alignment (33). The sequence alignment with Mss116p suggests that the CTE of CYT-19 extends to residue 590, leaving a C-tail of 36 amino acid residues. However, a somewhat larger segment of 49 residues is sensitive to proteolysis (16), indicating either that the final helix of the CTE is also protease sensitive or that the C-tail is somewhat longer than predicted from the sequence alignment. The CYT-19/ΔC-tail mutant was constructed based upon the proteolysis data and therefore has a 49-amino-acid residue C-terminal truncation, which includes an ~13 residue segment that could belong either to the CTE or C-tail. The disposition of this ~13 residue segment – helical or unstructured – does not significantly affect the modeling comparisons of the SAXS data or the conclusions.

Molecular modeling

The high-resolution crystal structure of the helicase core of Mss116p (Figure 4.1*B*) and the homology model of CYT-19 were used as additional restraints to generate molecular models of the open and closed states. Multi-domain atomic models of Mss116p and CYT-19 were reconstructed from the SAXS data by the program BUNCH, which employs a combination of rigid-body and *ab-initio* modeling (194). The protein domains with known structure are used to generate a theoretical scattering pattern, while

the unknown regions in each construct are represented by an appropriate number of dummy residues. For open-state models, an additional flexible linker was introduced between the helicase domains 1 and 2 (Figure 4.1*A*). BUNCH calculates the relative orientation of the components that is most consistent with the SAXS data. Results from multiple runs gave reproducible results with low average NSD values (Table B3). The low discrepancies between experimental data and computed scattering curves of the final BUNCH models also suggest that a single, unique model is sufficient to describe the SAXS data (Table B3). The molecular models produced using BUNCH were overlaid with the shape reconstructions from DAMMIN using SUPCOMB (217).

Multiphase modeling

For accurate SAXS reconstructions of DEAD-box proteins bound to large nucleic acid substrates, it is necessary to take into account the difference in scattering intensities of protein and nucleic acid. Therefore, the program MONSA, a multiphase version of DAMMIN, was used to obtain *ab-initio* models of protein complexes that include RNA-DNA duplex 1 or RNA-DNA duplex 2 (192, 197). MONSA uses simulated annealing to reconstruct a model of two phases, protein and nucleic acid, from an assembly of beads inside a defined search volume. Within this volume, each bead is assigned to solvent, protein or nucleic acid. The program was applied to simultaneously fit three scattering curves of protein, substrate, and complex of protein and substrate to find the best distribution of beads that correspond to interconnected phases representing the protein and nucleic acid. In this analysis, SAXS data from the equivalent protein-U_{10}-RNA complex were used to represent the closed-state protein structure. Several independent runs gave reproducible models with an average NSD of less than 0.65 that described the

scattering data well (Table B3). The most typical model is shown with the lowest NSD compared to all other models as calculated using the DAMAVER suite.

Rigid-body modeling with conformational sampling

SAXS data can be used to investigate the relative compactness of a macromolecule by analysis methods that involve rigid-body modeling with conformational sampling (195). The ensemble optimization method (EOM) was used to characterize the degree to which Mss116p is extended in solution (195). This method uses the available crystal-structure coordinates and a polypeptide chain of dummy residues to represent the structured and unstructured regions of the protein, respectively. A large pool of conformations is generated in which the unstructured regions are modeled by consecutively adding amino acids with phi and psi angles that are randomly drawn from a library of unstructured loops in the protein data bank. From this pool of 10,000 conformations, an optimized ensemble of 100 conformations is selected that best describes the SAXS data. We used the EOM to model the scattering profiles of the constructs of Mss116p in open and U_{10}-RNA-bound states. The R_g and D_{max} distributions for this optimized ensemble were then compared to those of the random pool (Figure B3).

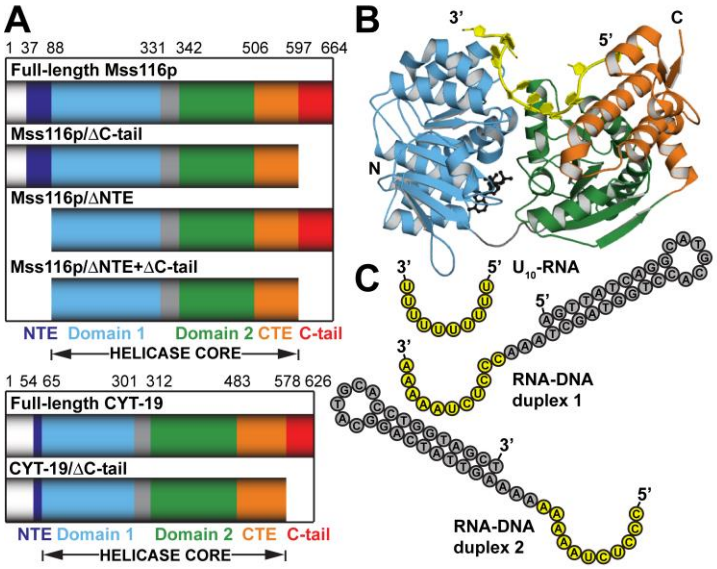


Figure 4.1: DEAD-box proteins Mss116p and CYT-19 and nucleic acid substrates.

(*A*) Schematic representations of the domain architectures of Mss116p, CYT-19, and deletion mutants. Mss116p consists of a mt targeting sequence that is cleaved *in vivo* and is absent in the constructs used here (white); an N-terminal extension, which corresponds to the N-terminus of the mature proteins (NTE; dark blue); a helicase core of two RecA-like domains (domain 1 and domain 2; light blue and green, respectively), which are joined by a flexible linker (gray); a structured C-terminal extension (CTE; orange); and a basic hydrophilic tail (C-tail; red). CYT-19 consists of the same elements but with a very short NTE. (*B*) The crystal structure of the closed-state helicase core and the CTE of Mss116p (PDB code 3I6I (55)) with domains colored as in (*A*). The bound single-stranded U_{10}-RNA and the ATP-analogue ADP-BeF_3^- are shown in yellow and black, respectively. (*C*) Schematic representations of nucleic acid substrates. RNA and DNA nucleotides are shown in yellow and gray, respectively, and nucleic-acid secondary structure was predicted using RNAfold (219).

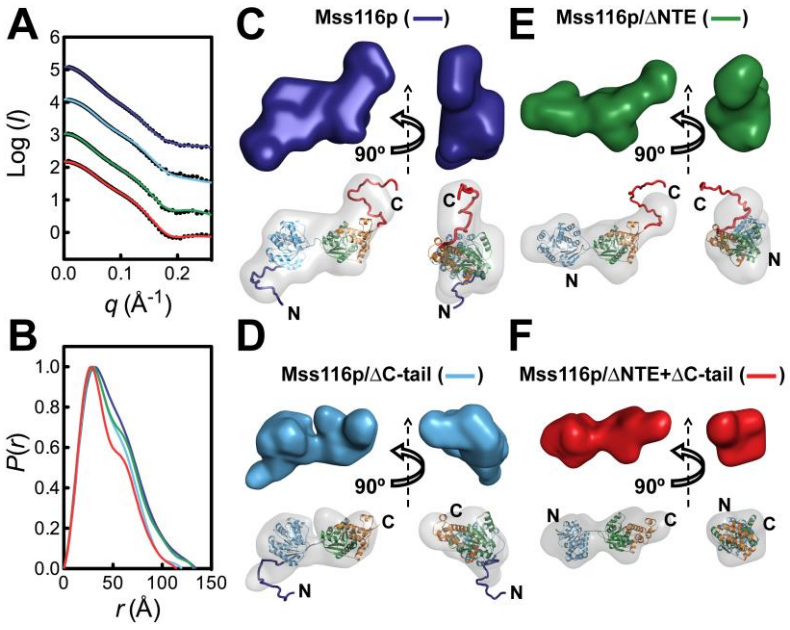


Figure 4.2: SAXS analysis of Mss116p in the open state without substrates.

SAXS data are shown for full-length Mss116p (dark blue), Mss116p/ΔC-tail (light blue), Mss116p/ΔNTE (green), and Mss116p/ΔNTE+ΔC-tail (red). (*A*) Scattering profiles, which are displaced along the logarithmic axis for visualization, are shown as the logarithm of the scattering intensity, I (black dots), as a function of the momentum transfer, $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength. The solid curves overlaying the SAXS data are the expected scattering profiles of the corresponding BUNCH models (see below). (*B*) Normalized distance distribution functions calculated from the scattering profiles using the program AUTOGNOM (214). (*C-F*) *Ab-initio* and rigid-body SAXS reconstructions of the open state of full-length Mss116p (*C*), Mss116p/ΔC-tail (*D*), Mss116p/ΔNTE (*E*), and Mss116p/ΔNTE+ΔC-tail (*F*). Low-resolution envelopes calculated by DAMMIN are shown separately (above) and superposed onto atomic models determined by BUNCH (below). In this and other

figures, protein domains are colored as in Figure 4.1 and views are rotated by 90 ° about the vertical axis for each model.

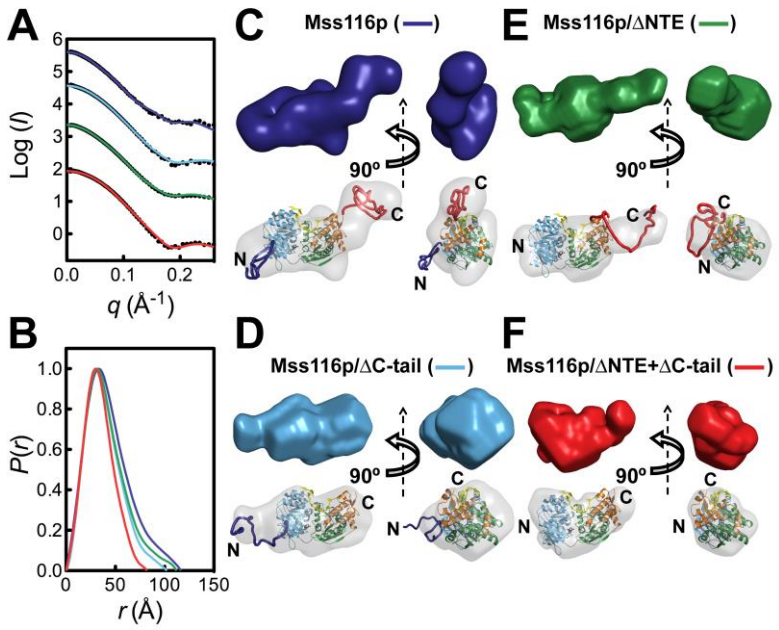


Figure 4.3: SAXS analysis of Mss116p ternary complexes with ssRNA and a non-hydrolyzable ATP analog.

(*A*) Scattering profiles and (*B*) normalized distance distribution functions for full-length Mss116p (dark blue), Mss116p/ΔC-tail (light blue), Mss116p/ΔNTE (green), and Mss116p/ΔNTE+ΔC-tail (red) bound to U_{10}-RNA and ADP-BeF_x. In (A), the scattering profiles are shown as black dots, and the solid curves overlaying the data are the expected scattering profiles of the corresponding BUNCH models, except for the Mss116p/ΔNTE+ΔC-tail complex where the overlay is the expected scattering profile calculated from the corresponding X-ray crystal structure using CRYSOL (220). (*C-F*) SAXS reconstructions of full length Mss116p (*C*), Mss116p/ΔC-tail (*D*), Mss116p/ΔNTE (*E*) and Mss116p/ΔNTE+ΔC-tail (*F*) in the closed state. Low-resolution envelopes calculated by DAMMIN (above) are colored as in (*A*) and atomic models (below) are shown aligned inside the DAMMIN envelope (gray).

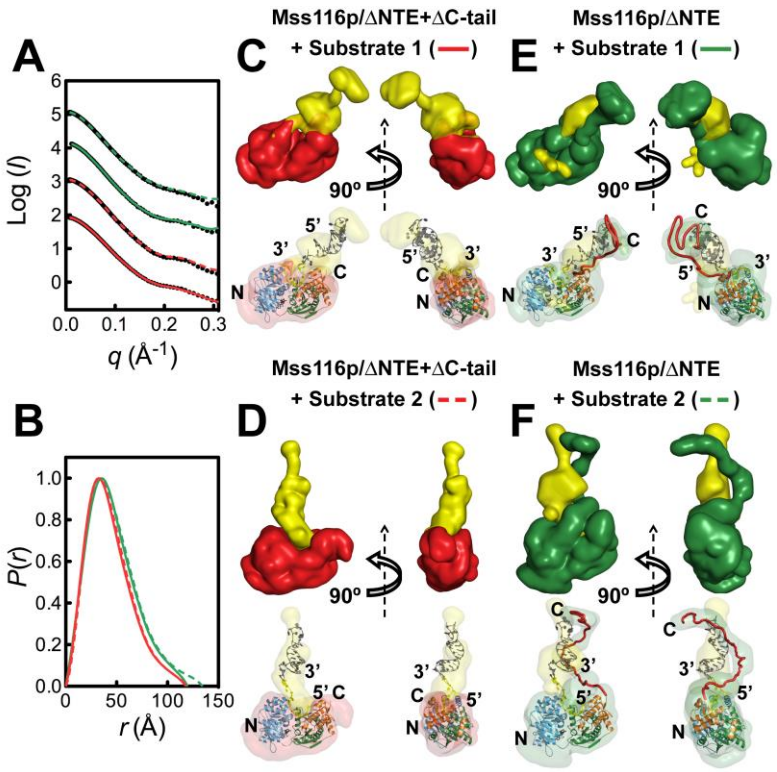


Figure 4.4: Binding modes of Mss116p to large nucleic acid substrates.

(*A*) Scattering profiles and (*B*) normalized distance distribution functions for Mss116p/ΔNTE (green) and Mss116p/ΔNTE+ΔC-tail (red) bound to ADP-BeF_x and either RNA-DNA duplex 1 (solid lines) or RNA-DNA duplex 2 (dashed lines). In (*A*), the scattering profiles are shown by black dots, and the colored lines represent the fit of the *ab-initio* model of the complex obtained by MONSA. (*C-F*) *Ab-initio* multiphase reconstructions from the SAXS data in (*A*) of complexes of Mss116p/ΔNTE+ΔC-tail with either RNA-DNA duplex 1 (substrate 1; *C*) or RNA-DNA duplex 2 (substrate 2; *D*) and complexes of Mss116p/ΔNTE with either RNA-DNA duplex 1 (*E*) or RNA-DNA duplex 2 (*F*). Two-phase models of protein (colored as above) and nucleic acid (yellow) were reconstructed by MONSA (top). The ssRNA regions of these substrates were assumed to bind within the helicase core in the orientation observed in the crystal

structure of the closed state of Mss116p (Figure 4.1*B*). This information was used to place manually atomic models for protein and nucleic acid inside the corresponding SAXS envelopes (bottom).

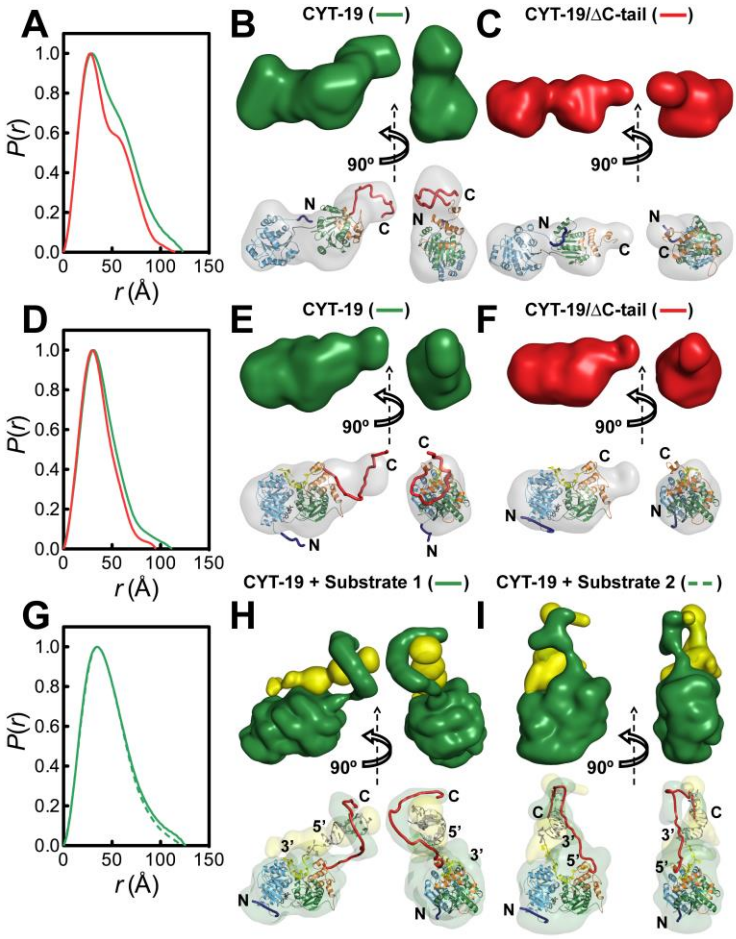


Figure 4.5: SAXS analysis of CYT-19.

(*A-C*) SAXS data for full-length CYT-19 (green) and CYT-19/ΔC-tail (red) in the absence of ligands. (*A*) Normalized distance distribution functions. (*B*) and (*C*) low-resolution envelopes calculated by DAMMIN (above) and BUNCH atomic models (below), which are colored as in Figure 4.1 and aligned inside the DAMMIN envelope (gray). BUNCH models were generated using a homology model of CYT-19 that is based upon its sequence similarity to Mss116p (see *SI Materials and Methods*). (*D-F*) SAXS data for CYT-19 bound to U_{10}-RNA and ADP-BeF_x, shown in the same arrangement as in (*A-C*). For the minimal CYT-19/ΔC-tail complex, the DAMMIN envelope in (*F*) is

aligned to the homology model for CYT-19. (*G*-*I*) SAXS data for full-length CYT-19 bound to large nucleic acid substrates. (*G*) Normalized distribution functions for CYT-19-ADP-BeF_x bound to RNA-DNA-duplex 1 (solid green line) and RNA-DNA-duplex 2 (dashed green line). (*H* and *I*) *Ab-initio* multiphase reconstructions of CYT-19 in complex with RNA-DNA duplex 1 (substrate 1) and RNA-DNA duplex 2 (substrate 2), respectively. Two-phase models of protein (green) and nucleic acid (yellow) were constructed by MONSA (top) and atomic models for protein and nucleic acid were manually placed inside the corresponding SAXS envelopes (bottom).

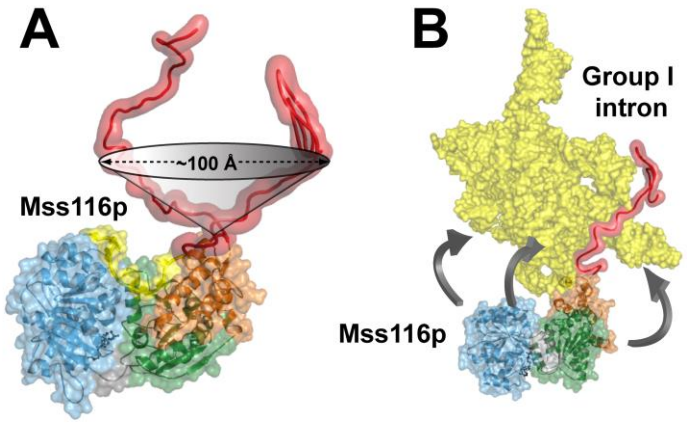


Figure 4.6: The tethering range observed for the basic tail of Mss116p when bound to nucleic acid substrates.

(*A*) Range of motion of Mss116p's flexibly attached C-tail. The cone shows the lower limit for the region of space over which the C-terminal tail of Mss116p can bind nucleic acid, as indicated by MONSA reconstructions of complexes with chimeric substrates containing duplex DNA extensions. (*B*) Model of Mss116p interacting via the C-tail with the *Tetrahymena* group I intron ribozyme (221). The model shows that, when anchored by the flexible C-tail, the helicase core of Mss116p can act at numerous sites over a wide region of the RNA.

Chapter 5: CYT-19 delays compaction, but does not alter folding outcomes of the unfolded *Tetrahymena* ribozyme

The well-characterized folding landscape of the *Tetrahymena* ribozyme provides a unique opportunity for a comprehensive analysis of DEAD-box protein involvement at different stages of RNA folding, from the earliest compaction events to rearrangements of highly structured intermediates and the native structure. CYT-19 roles have been extensively studied in disruption of and accelerating transitions between the native and misfolded states, which dominate the structural ensemble under standard in vitro conditions. However, potential effects of CYT-19 on early steps of ribozyme folding, either during Mg^{2+}-induced folding of the denatured RNA or upon transient formation of globally unfolded states during structural transitions, remained to be addressed. Intriguingly, the work in Chapter 2 revealed extensive ATP-dependent interactions of CYT-19 with non-compact ribozyme states, suggesting that CYT-19 may remodel these states in ways that bias folding towards different pathways than preferentially traversed during spontaneous folding.

Studies of ribozyme folding in the absence of chaperones showed that after adding Mg^{2+} ions the ribozyme rapidly undergoes electostatic compaction (~10 ms), followed by a second stage of compaction that involves formation of long-range tertiary contacts (Figure 5.1) (165, 222). The collapsed state then folds preferentially to the long-lived misfolded state, while approx. 10 % of the ribozyme enters a slower pathway leading directly to the native conformation. These alternative pathways of ribozyme folding appear to be coupled tightly to the topology at the single-stranded junctions between the core domains and the stability of the adjacent long-range core helix P3 (145, 146, 223).

An intriguing possibility is that by interacting with the unfolded ribozyme and early intermediates CYT-19 may alter the partitioning between pathways that lead to the long-lived misfolded and native states, potentially decreasing the fraction of ribozyme that becomes trapped in the misfolded state. Possible mechanisms through which CYT-19 may affect early folding is, first, simply by binding ribozyme helices before the collapse, which may prevent rapid compaction, providing the ribozyme with more time to explore different topologies while in an extended state. CYT-19 may also actively rearrange secondary structure in extended intermediates and, for example, by unwinding the core helix P3, bias the unfolded ribozyme to the native or misfolded topology. Alternatively, CYT-19 may simply slow compaction of the non-compact early intermediates without influencing the topology.

Here we examined the roles of CYT-19 in compaction and early folding events of the *Tetrahymena* ribozyme by time-resolved small-angle X-ray scattering and ribozyme activity-based assays. SAXS has been widely used to follow conformational transitions of RNA and protein molecules, including seminal studies of the compaction of the *Tetrahymena* ribozyme (165, 222, 224). While SAXS is best suited for studies of large-scale conformational changes, such as compaction of the unfolded ribozyme, complementing SAXS studies with ribozyme activity-based assays allows to also track transitions between native and globally similar intermediate states. The work presented here suggests that CYT-19 can form complexes with the unfolded ribozyme and delay compaction. However, ribozyme activity assays indicate that this slowed compaction has little or no effect on the fraction of ribozyme that misfolds. Analogous results were obtained for a ribozyme mutant with strongly destabilized tertiary structure. Overall, this work indicates that CYT-19 does not bias early ribozyme folding towards the native or misfolded states, consistent with a chaperone mechanism in which CYT-19 promotes

folding primarily by disrupting the misfolded intermediate, allowing the ribozyme to refold along intrinsic folding pathways.

RESULTS

Delayed ribozyme compaction in the presence of CYT-19

To probe the roles of CYT-19 in initial folding of the ribozyme, we first established conditions that allowed to monitor changes in ribozyme compaction by SAXS without significant contributions of free CYT-19 to the scattering signal. The scattering intensity, expressed as forward scattering intensity I_0, is a function of concentration (c), scattering contrast ($\Delta\rho$; defined as difference in electron density between the molecule and surrounding buffer), and molecular weight (MW):

$$I_0=\kappa c(\Delta\rho)^2(MW)^2,$$

where κ is a proportionality constant (224). The ribozyme is expected to dominate the scattering signal because of the much higher scattering contrast of nucleic acids than proteins (2–3-times stronger in water due to the high electron density of the phosphodiester backbone) (224, 225), and the 2x greater molecular weight of the ribozyme than CYT-19 (126 kDa vs. 64 kDa, with no evidence for oligomerization of either molecule). With a quadratic dependence of the scattering intensity on both the scattering contrast and molecular weight and only a linear dependence on concentration, CYT-19 is expected to give negligible signal even when used in several-fold excess over the ribozyme. This was indeed confirmed experimentally, with 14-fold and 8-fold stronger scattering by the ribozyme than a 2.5-fold or 4-fold higher concentration of CYT-19, respectively (Figure 5.2*A* and Table 5.1). Calculations based on the measured intensities indicated a 2.8-fold greater scattering contrast for the ribozyme, consistent with reported values in aqueous solutions (Appendix C) (225).

To study how CYT-19 affected initial folding of the unfolded ribozyme, we first performed static measurements in the absence of Mg^{2+}, conditions at which the ribozyme contains secondary, but not tertiary structure. Under our conditions, the ribozyme was partially electrostatically relaxed (Figure 5.1), due to the presence of 70 mM monovalent metal ions in the buffer (20 mM Na^+ from 50 mM Na-MOPS (pH 7.0) and 50 mM K^+ from 10% (v/v) CYT-19 buffer) (182). The shapes of the scattering profiles for the unfolded ribozyme were consistent with previous studies and lacked a prominent peak at low q values (141, 157). Adding CYT-19 to the unfolded ribozyme drastically changed both the shape of the scattering profile and the scattering intensity (Figure 5.2*A*, Table 5.1). These changes were clearly non-additive, i.e., they were greater than the negligible change in scattering intensity and the height of the peak in the Kratky plot that would result from simply adding the individual scattering profiles measured separately for CYT-19 and the ribozyme. The non-additive increase in scattering intensities of CYT-19-ribozyme mixtures indicates an increase in the molecular weight of the scattering particles, consistent with complex formation between CYT-19 and the ribozyme. Note that despite weak scattering when free in solution, CYT-19 can contribute much more strongly to the scattering signal when bound to the ribozyme, a consequence of the quadratic dependence of the scattering intensity on particle molecular weight (see derivation in Appendix C). The enhancement in forward scattering intensity in the presence of 2.5 μM CYT-19 and 1 μM ribozyme vs. ribozyme alone was consistent with the estimate for 1:1 binding stoichiometry (Appendix C) with more CYT-19 apparently bound at 4 μM (Table 5.1).

To test if CYT-19 could also form complexes with unfolded ribozyme in the presence of Mg^{2+} (under conditions at which folding was to be monitored), we performed measurements with the quintuple mutant of the ribozyme and CYT-19 at 5 mM Mg^{2+}. As

noted in Chapter 2, the quintuple mutant contains mutations of all five peripheral tertiary interactions and is unable to undergo tertiary collapse after addition of Mg^{2+}, remaining in a non-compact state (165). Very similar scattering profiles and the same enhancement in scattering intensity as with the unfolded ribozyme in the absence of Mg^{2+} were indeed observed upon addition of CYT-19, although the peak at low values of the scattering vector q suggested some aggregation occurring when CYT-19 and the ribozyme were mixed in the presence of Mg^{2+} (Figure 5.2*B*). Nearly identical profiles were observed also with added ATP and AMPPNP, and complexes also formed in the presence of the C-tail deletion variant of CYT-19, indicating that binding was not mediated only by electrostatic interactions with the basic C-tail (Figure 5.2*B*). Importantly, a CYT-19-dependent increase in scattering intensity and substantial changes in the shapes of the scattering profiles only were detected with the unfolded ribozyme, but not the native and misfolded ribozymes at 5 mM Mg^{2+} (Figure 5.2*C, D*). Thus, the results suggest that CYT-19 can form complexes with the unfolded ribozyme, but does not associate stably with the folded states.

To test if CYT-19 binding to the unfolded ribozyme affected the rate of compaction, we performed time-resolved SAXS measurements. Using a remotely controlled syringe pump, we initiated folding by mixing the unfolded ribozyme with solutions containing Mg^{2+} ions and CYT-19, with nucleotides included as applicable (Figure 5.3). The reaction was then exposed to synchrotron radiation at different times during the first several minutes of folding. Consistent with rapid spontaneous collapse, in the absence of CYT-19 the ribozyme adopted a compact conformation within the dead time of the instrument (~1s) (Figure 5.4). However, in the presence of CYT-19, we consistently observed one or more time points corresponding to the first several seconds of the folding reaction and the scattering profiles at these early time points resembled in

shape the scattering profiles of CYT-19 bound to the unfolded ribozyme (Figure 5.4). Thus, CYT-19 appeared to delay compaction of the ribozyme by stabilizing a non-compact conformation. This delay was more evident in the presence of ATP, as indicated by a greater number of time points that appeared to be intermediate between the unfolded and compact states. At later time points, a compact ribozyme state with a SAXS profile that resembled the profile of the misfolded ribozyme accumulated both in the presence and absence of CYT-19, suggesting dissociation of the protein after initial interactions with the unfolded ribozyme. Quantitative analysis of the time-resolved SAXS results was complicated by apparent aggregation, as indicated by the pronounced peaks at lowest q values. Aggregation was likely caused by non-specific RNA-protein interactions that were facilitated by the high concentrations required for SAXS.

Using lower CYT-19 and ribozyme concentrations to limit aggregation, we probed the CYT-19 concentration dependence of the delay in compaction observed above (Figure 5.5). In the presence of ATP, there was a clear dependence on CYT-19 concentration, with virtually no delay (on the accessible time scale) detected with 0.5 μM CYT-19, several intermediate time points observed at 0.875 nM CYT-19, and the most pronounced delay was observed at 1.25 μM CYT-19 (Figure 5.5*A*), although aggregation also appeared to increase substantially. In the absence of ATP, aggregation also appeared to be more prominent at 1.25 μM than 0.875 μM CYT-19, with a similar number of intermediate time points detected at each concentration. The delay in compaction with 1.25 μM CYT-19 was somewhat less pronounced than in the presence of ATP (Figure 5.5*B*).

We next addressed how nucleotides that are known to weaken or tighten RNA binding affected the ability of CYT-19 to slow compaction. ADP, which binds anticooperatively with RNA (83, 226), essentially eliminated the delay in compaction on

the observed time scale, as well as aggregation, suggesting that CYT-19 either did not bind the unfolded ribozyme or dissociated rapidly after Mg^{2+} addition (Figure 5.6). In contrast, AMPPNP, which tightens RNA binding, gave a similar effect as ATP and no nucleotide, suggesting that CYT-19 bound the unfolded ribozyme, but dissociated on the time scale of seconds after Mg^{2+} addition (Figure 5.6). The differences between ADP and other nucleotides and the generally stronger delay observed with ATP than no nucleotide suggest that nucleotide-dependent binding by the CYT-19 core is at least in part responsible for the delay in ribozyme compaction. Overall, the SAXS results suggest that CYT-19 can slow ribozyme compaction of the ribozyme from < 1 s to several seconds by forming complexes with the unfolded ribozyme, but eventually allows accumulation of compact ribozyme that resembles the misfolded state, presumably following protein dissociation.

The presence of CYT-19 during compaction does not affect further ribozyme folding

To test if CYT-19 binding to the unfolded ribozyme early in folding affected the rate of subsequent folding to the native state, we turned to ribozyme activity-based assays, analogous to those used in Chapters 2 and 3 (136). At different times after mixing the ribozyme with Mg^{2+} ions in the presence or absence of CYT-19 and nucleotide, aliquots were moved to a folding quench and the fraction of native ribozyme was determined based on the fraction that rapidly cleaved the oligonucleotide substrate (Figure 5.7*A*). These experiments were performed using high protein and ribozyme concentrations comparable to those used in SAXS. The ribozyme activity assays allowed to, first, determine if CYT-19 binding during early folding affected the fraction of the ribozyme that rapidly folds to the native state without misfolding. Second, for any inactive ribozyme that accumulated upon compaction, measuring the rate constant of

folding to the native state would reveal if this state was the same misfolded state as formed upon ribozyme compaction in the absence of CYT-19.

The results revealed no appreciable effects of CYT-19 on the initial fraction of native ribozyme, both in the presence and absence of nucleotides (Figure 5.7). The rate constant of folding to the native state was also unaffected relative to spontaneous folding when CYT-19 was added without nucleotide or with ADP or AMPPNP. In the presence of ATP, CYT-19 did accelerate folding, but this acceleration was identical to that of refolding of the misfolded ribozyme under the same conditions (Figure 5.7). We also tested if CYT-19 exerted more pronounced effects on folding when allowed to pre-bind the unfolded ribozyme before the addition of Mg^{2+}. No appreciable effects on the starting fraction of native ribozyme were observed under these conditions as well, and the folding rate was substantially changed only in the presence of ATP, which again gave the same acceleration as for refolding of the misfolded ribozyme (Figure 5.7). (Note that the nucleotides were added together with Mg^{2+} rather than during the preincubation; see below). Thus, the presence of CYT-19 at the initiation of ribozyme folding did not affect the intrinsic partitioning of the ribozyme between pathways that lead to the misfolded and native states and appears to allow the same misfolded conformation to accumulate.

CYT-19 interactions with an unfolded tertiary contact mutant ribozyme delay compaction, but do not affect further folding

In a final set of experiments, we asked if the folding behavior of the ribozyme was affected when CYT-19 was allowed to perform ATP-dependent rearrangements of the unfolded ribozyme prior to initiation of folding. Here, we used the A-rich bulge (ARB) mutant of the metal core–metal core receptor (MC/MCR) tertiary contact, which has highly unstable tertiary structure and remains in a predominantly unfolded state even at 2 mM Mg^{2+} (Chapter 2). Thus, the unfolded ribozyme can be pre-incubated with CYT-19

and ATP under conditions that previously were shown to give strong RNA-dependent ATPase activity (and thus presumably remodeling activity) (Chapter 2). (In the absence of Mg^{2+} and the very low Mg^{2+} concentrations required to accumulate unfolded wild-type ribozyme, CYT-19 displays only weak RNA-dependent ATPase activity; in addition, Mg^{2+} ions introduced with ATP•Mg^{2+} may affect folding, due to the steep Mg^{2+} dependence of the wild-type ribozyme (see Chapter 2)).

We performed both SAXS and ribozyme activity-based experiments in which the ARB ribozyme was incubated with or without CYT-19 and ATP at 2 mM Mg^{2+}, followed by shifting the Mg^{2+} concentration to 10 mM, which induces compaction and folding to the native state. Similarly to the compaction of the wild-type ribozyme without pre-incubation with CYT-19, the protein introduced a transient delay in compaction, with a misfolded-like state formed in less than 30 s after the Mg^{2+} shift to 10 mM. This delay was more pronounced with ATP than in the absence of nucleotide. Even stronger delay was observed with AMPPNP, as even ~30 s after the addition of 10 mM Mg^{2+} the ribozyme remained in a state consistent with the scattering by an unfolded ARB ribozyme in complex with CYT-19 (Figure 5.8).

Ribozyme activity assays indicated that even after the pre-incubation with CYT-19 and ATP or AMPPNP, the fraction of ribozyme that folded rapidly to the native state was unaffected relative to spontaneous folding. As expected, ATP did accelerate further folding, but this effect was indistinguishable whether CYT-19 and ATP were pre-incubated with the unfolded ribozyme or added at the same time as 10 mM Mg^{2+} (with the same result observed when CYT-19 was pre-incubated with the unfolded ribozyme and ATP was added at the initiation of folding) (Figure 5.9). Pre-incubation with AMPPNP also did not change the initial fraction of native ribozyme, although it appeared to slightly slow down the overall folding. Together, the results with the ARB ribozyme

support the above conclusion that CYT-19 interactions with the unfolded *Tetrahymena* ribozyme only transiently slow compaction, without substantially affecting the identity and further folding behavior of the compact ribozyme that forms following CYT-19 release.

DISCUSSION

Here we show that CYT-19 can delay Mg^{2+}-induced compaction of the *Tetrahymena* ribozyme by interacting with the unfolded ribozyme. However, once CYT-19 dissociates, typically within the first several seconds after addition of Mg^{2+}, folding of the ribozyme progresses along the same pathways as in the absence of CYT-19. Appreciable changes in ribozyme folding are absent even after CYT-19 pre-incubation with an unfolded mutant ribozyme under conditions shown previously to give strong RNA-dependent ATPase activity. Thus, despite interacting with and, in the presence of ATP, presumably locally rearranging the unfolded ribozyme, CYT-19 does not appear to directly alter the ribozyme features that determine whether the ribozyme misfolds or folds to the native conformation.

Our results suggest that CYT-19 forms complexes with the unfolded ribozyme rapidly, before Mg^{2+}-induced tertiary compaction, which is completed here with an upper limit of 1 s and was shown previously under similar conditions to take ~100 ms. Binding is at least in part mediated by the CYT-19 core, as indicated by the nucleotide effects in time-resolved SAXS experiments, as well as the formation of complexes with protein that lacks the basic C-tail. Once the ribozyme has formed a compact structure, complexes are no longer detected, based on the lack of enhancement in scattering intensity or changes in the shape of the scattering profile. This result is consistent with fewer interaction opportunities being present in the compact states and suggests that interactions of CYT-

19 with the exposed surface helices of the folded ribozyme that were demonstrated previously by ATPase activity are short-lived (see Chapter 2).

The finding that CYT-19 dissociates from the unfolded ribozyme within seconds and does not affect the fate of the ribozyme during further folding has important implications for understanding the general chaperone mechanism of CYT-19 in *Tetrahymena* ribozyme folding. Previous studies and kinetic simulations of CYT-19-mediated rearrangements of the ARB mutant ribozyme indicated that CYT-19 functioned primarily at the stage of disruption of the misfolded and native states, without affecting subsequent refolding of the partially unfolded ribozyme (127). Our results are consistent with this model and expand it to include interactions with globally unfolded ribozyme. Thus, analogously to some general protein chaperones, CYT-19 promotes native folding of the *Tetrahymena* ribozyme by iterative cycles of disruption of the misfolded intermediate, rather than directly guiding the unfolded ribozyme towards the native conformation (120).

MATERIALS AND METHODS

SAXS measurements

The protein and RNA were prepared as described in previous chapters. SAXS measurements were performed at the beamline 12-ID-C at Advanced Photon Source, using 2 m camera length. Static measurements were performed by injecting the samples into a temperature-controlled (25 ˚C) capillary and collecting 15–20 1s exposures at 12 keV.

Time-resolved measurements were performed using the remotely controlled mixer shown in Figure 5.3. For experiments with the wild-type ribozyme, solutions of the ribozyme (Line 1) and $Mg(OAc)_2$ with CYT-19 or CYT-19 buffer (Line 2) were prepared

at 2x final concentrations and injected into the corresponding lines prior to the measurement. 50 mM Na(MOPS) (pH 7.0) was included in both solutions, and the final concentrations after mixing were as follows: 5 mM $Mg(OAc)_2$, 0.5–1 μM ribozyme, 0.5–2.5 μM CYT-19, 10 % (v/v) CYT-19 buffer and 50 mM Na(MOPS) (pH 7.0). The reaction was initiated by simultaneously injecting equal volumes from each side into a T-junction leading to the capillary. The shutter of the X-ray beam was opened within estimated ~1 s after mixing. 10 exposures of 1 s were collected in 3 s intervals, followed by a series of longer time points, as indicated in the figure legends.

In experiments with the ARB mutant ribozyme, Line 1 included 2 μM ribozyme in 50 mM Na-MOPS (pH 7.0), 2.39 mM $Mg(OAc)_2$ (2 mM free Mg^{2+} after accounting for binding of ~1 Mg^{2+} ion per 2 nucleotides of RNA (182)), 10 % (v/v) CYT-19 buffer with or without 5 μM CYT-19, and 2 mM ATP nucleotide where indicated. Line 2 included 18 mM $Mg(OAc)_2$ which brought the final Mg^{2+} concentration to 10 mM upon mixing. Line 2 also included 10 % (v/v) CYT-19 buffer, 2 mM nucleotide and 50 mM Na-MOPS (pH 7.0) to maintain the same concentrations during pre-incubation and mixing.

Both during static and time-resolved experiments, the contents of the capillary were mixed by the attached syringe pump to avoid radiation damage. The blanks consisted of solutions prepared in parallel with the samples, with TE buffer substituted for the ribozyme and CYT-19 buffer substituted for protein.

SAXS analysis

The scattering profiles were radially averaged at the beamline and further analyzed with IgorPro (WaveMetrics). The individual exposures for static samples were averaged and the average of the exposures of the corresponding blank was subtracted.

Exposures that showed a strong time-dependent drift in intensity in static measurements, as well as exposures damaged by air bubbles in the capillary were not included in analysis. For time-dependent measurements, each exposure was analyzed individually, without averaging.

The R_g and I_0 values in Table 5.1 were determined by Guinier approximation, using the region $0.0002\ Å^{-2} \leq q^2 \leq 0.0004\ Å^{-2}$ for unfolded and quintuple mutant ribozymes and complexes and $0.00056\ Å^{-2} \leq q^2 \leq 0.001\ Å^{-2}$ for misfolded and native ribozymes. Note that because of the large size of the unfolded ribozyme and the complexes with CYT-19 and because of the limited number of data points at low q values, very few points satisfy the condition for Guinier fitting $q_{max}R_g<1.3$, where q_{max} is the highest q value used for the fit (167). We therefore estimated the R_g and I_0 values under a more flexible condition of $q_{max}R_g<1.7$. Thus, the reported values are for comparative purposes within groups of samples only and may deviate from absolute values. The low-q region affected by upward curvature due to aggregation or proximity to the beam stop was excluded from the fit ($q^2 < 0.0002\ Å^{-2}$ for unfolded ribozymes and $q^2 < 0.00056\ Å^{-2}$ for misfolded and native ribozymes in the presence of CYT-19).

We do not report R_g and I_0 values for time-resolved measurements, because their accurate determination was precluded by high noise levels inherent to non-averaged single exposures, variable curvature at low angles that was indicative of aggregation and heterogeneity, as well as mixing-induced fluctuations in scattering intensity.

Ribozyme activity assays

Ribozyme activity assays were performed as shown in Figures 5.7 and 5.9 and essentially as described in Chapter 2 and Ref. (136), with equal volumes (10 μL each) of the indicated solutions mixed to initiate the reaction. 50 mM Na-MOPS (pH 7.0) was

included in all solutions. In the reactions where the ribozyme was pre-incubated with CYT-19 or CYT-19 buffer prior to adding Mg^{2+} ions, the concentration of CYT-19 buffer was 10% (v/v) both in the pre-incubation and in the final folding reaction. In the experiments with the ARB ribozyme, the total concentration of Mg^{2+} in the preincubation was 2.39 mM to account for the Mg^{2+} ions expected to be sequestered from solution by the ribozyme (1 Mg^{2+} per 2 nucleotides (182), i.e., ~0.39 mM for 2 μM of 388 nt ribozyme in the pre-incubation). The pre-incubations of the ribozyme with CYT-19 or CYT-19 buffer were performed for 5 min at 25 ˚C.

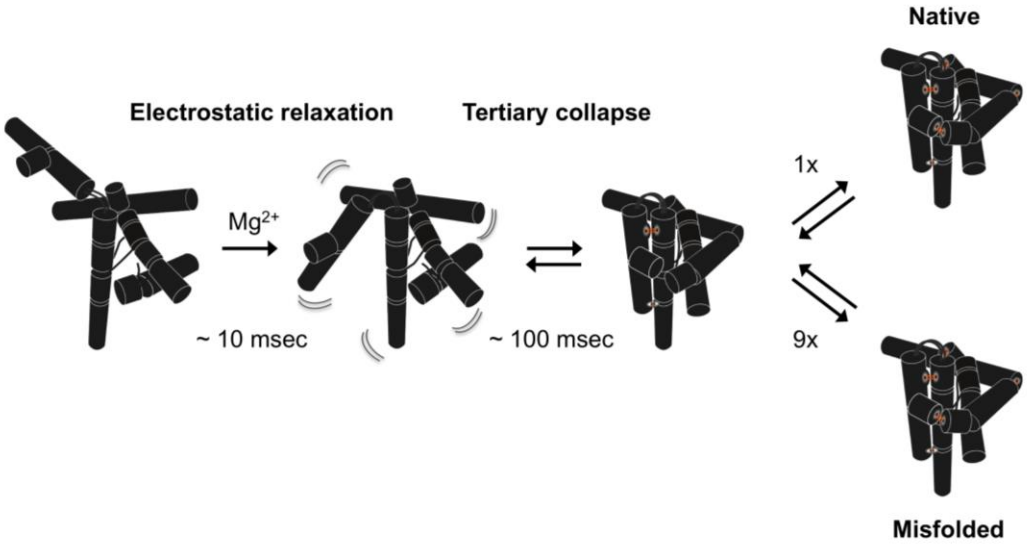


Figure 5.1: Schematic representation of the stages of Tetrahymena ribozyme compaction and folding.

After addition of Mg^{2+} ions, the ribozyme rapidly undergoes electrostatic relaxation, which is followed by slower tertiary collapse, involving the formation of several long-range tertiary contacts. Compaction is followed by slower folding to the native and misfolded states. The compaction and folding rates are from Refs. (140, 165, 222).

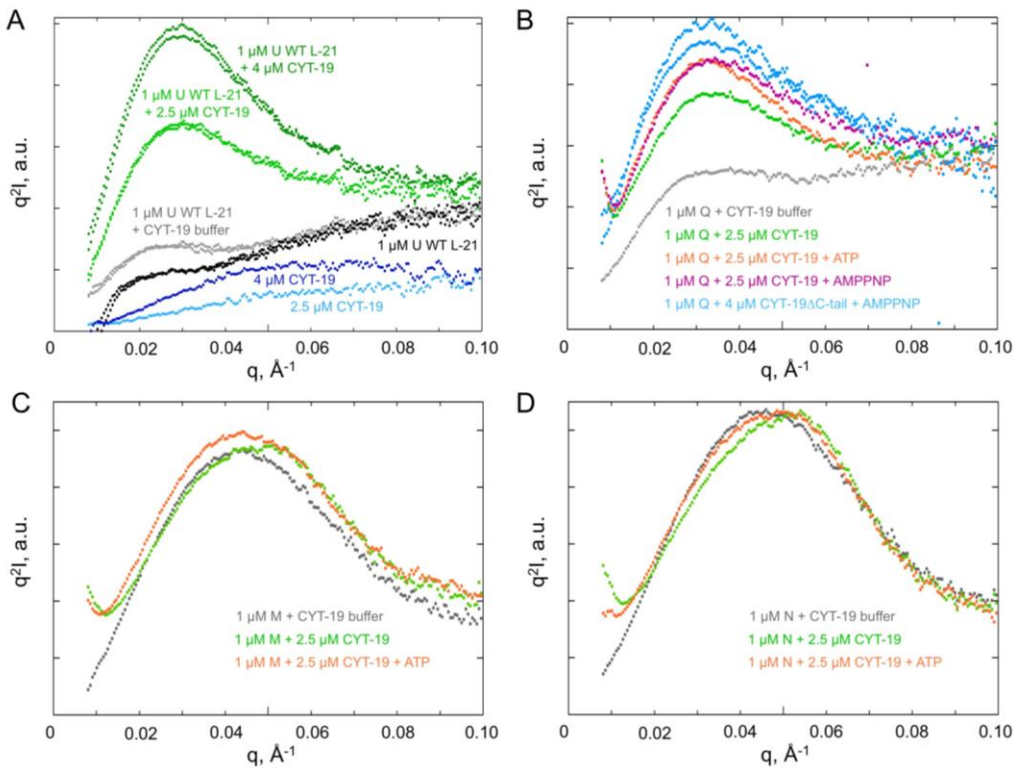


Figure 5.2: Kratky plots of the *Tetrahymena* ribozyme, CYT-19, and mixed ribozyme-CYT-19 solutions.

(*A*) Scattering by the unfolded ribozyme (U) and CYT-19 in the absence of Mg^{2+} ions. Here and below SAXS data are displayed as Kratky plots (q^2I vs q), a transformation useful for detecting changes in compactness of the scattering particle (167); q: scattering vector ($q=2\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength); I: intensity (in arbitrary units, a.u.). Results of two duplicate measurements are shown for all samples including the ribozyme and single measurements are shown for each concentration of CYT-19. All data in Figure 5.2 were obtained during a single beam time, allowing direct comparisons without normalization by intensity. All samples in *A–D* included 50 mM Na-MOPS (pH 7.0), as well as 1 mM Tris-HCl (pH 8.0) and 0.1 mM EDTA from the TE buffer in which ribozyme was stored, with other components indicated. The slight difference between the scattering profiles of the unfolded ribozyme in the presence and absence of CYT-19 buffer is expected due to increased electrostatic

relaxation upon addition of monovalent cations present in CYT-19 buffer (50 mM K^+; other ingredients included 2 mM Tris-HCl (pH 7.5), 5 % (v/v) glycerol, 0.1 mM EDTA and 0.02 mM DTT). (*B*) Scattering by the quintuple mutant (Q) in the presence of 5 mM Mg^{2+}. Results from single determinations are shown, with the exception of two determinations for the quintuple mutant with CYT-19/ΔC-tail. (*C, D*) Scattering by the misfolded (*C*) and native (*D*) ribozymes in the presence and absence of CYT-19 and ATP. The misfolded ribozyme data were collected during the 6^{th} minute (no CYT-19) or 4^{th} minute (CYT-19 ± ATP) after transfer of pre-formed misfolded ribozyme from 10 mM to 5 mM Mg^{2+}, which is expected to give refolding of less than 10 % of the misfolded ribozyme to the native state.

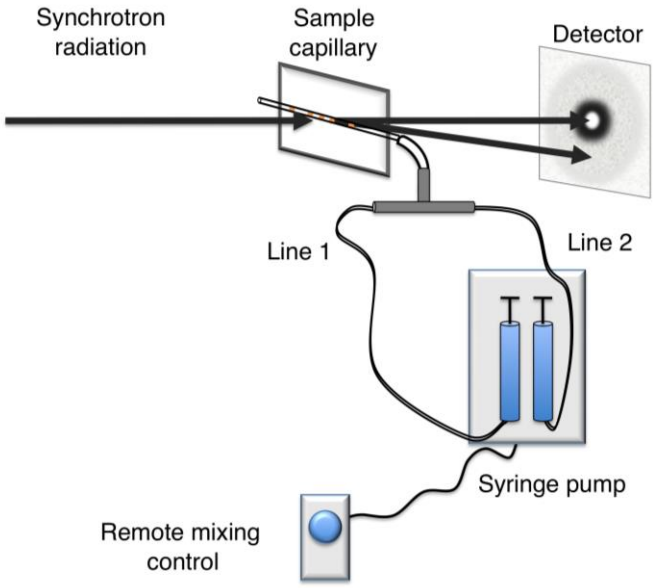


Figure 5.3: Time-resolved SAXS setup.

Time-resolved SAXS measurements were initiated by simultaneously injecting the contents of Line 1 (RNA) and Line 2 (Mg^{2+}) (with other ingredients as indicated for specific experiments) into the T-junction and the sample capillary. The capillary was exposed to X-rays in pre-determined intervals for 1s per exposure. The estimated delay between mixing and the first 1 s exposure to the X-ray beam was ~1s. The sample was mixed by the pump between exposures to minimize radiation-induced damage.

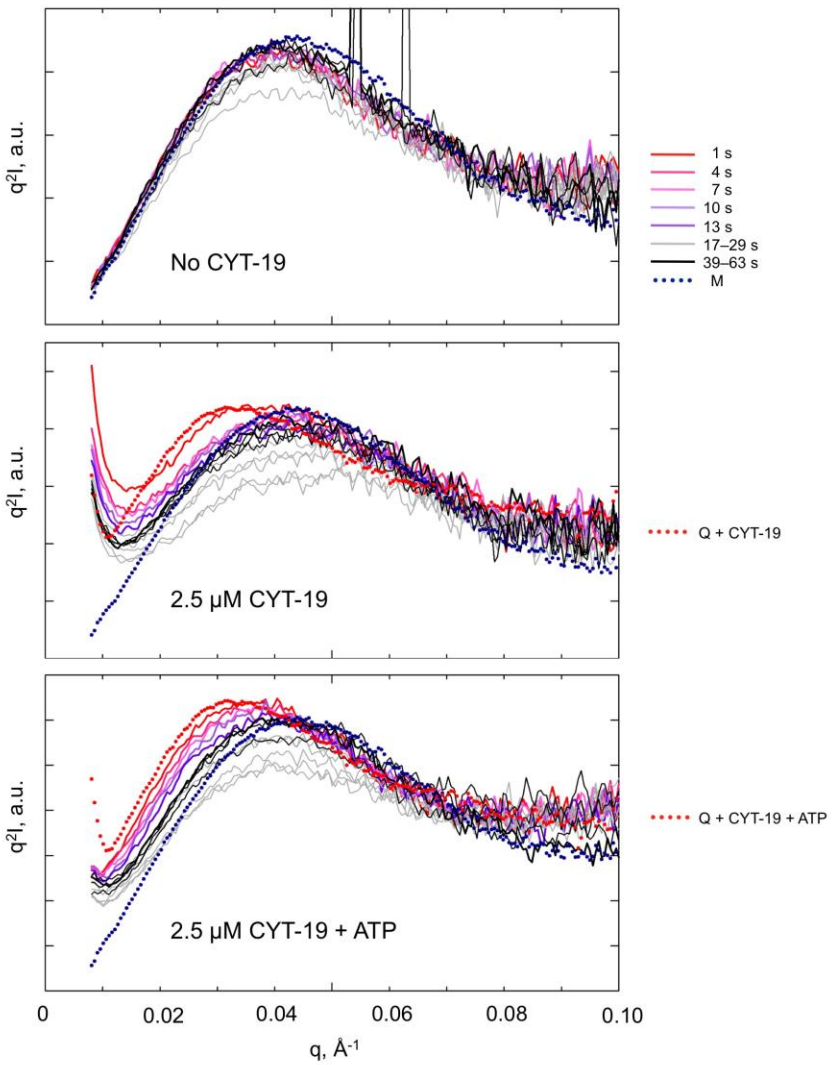


Figure 5.4: Ribozyme compaction in the presence and absence of CYT-19 and ATP.

Kratky plots of time points taken during the first minute after the addition of Mg^{2+} to the unfolded ribozyme (5 mM $Mg(OAc)_2$, 1 μM ribozyme, 2.5 μM CYT-19). The first five exposures, taken in ~3 second intervals are color-coded as indicated. Five subsequent exposures, also in intervals of 3 s, are shown in gray, while later exposures, in intervals of 13 s, are in black. Blue points indicate the scattering profile of the pre-formed misfolded ribozyme (M) at 5 mM $Mg(OAc)_2$. Red points indicate the scattering by the quintuple mutant (Q) at 5 mM $Mg(OAc)_2$ in the presence of the same concentration of CYT-19 (2.5

μM) and ATP (2 mM) as used in the time-resolved measurement. Quintuple mutant profiles are included as a reference, because this mutant has been shown to well represent the structure of the wild-type ribozyme prior to tertiary compaction (165). Here and further, the scattering profiles of the reference states (M and Q in this case) were multiplied by a constant to adjust intensities for comparing the shapes with those of the initial and late time points. Fluctuations in scattering intensity between time points, especially pronounced in the first ~30 s of the reaction, were observed in most time-resolved experiments and may be caused by incomplete mixing, in addition to expected changes due to dissociation of CYT-19.

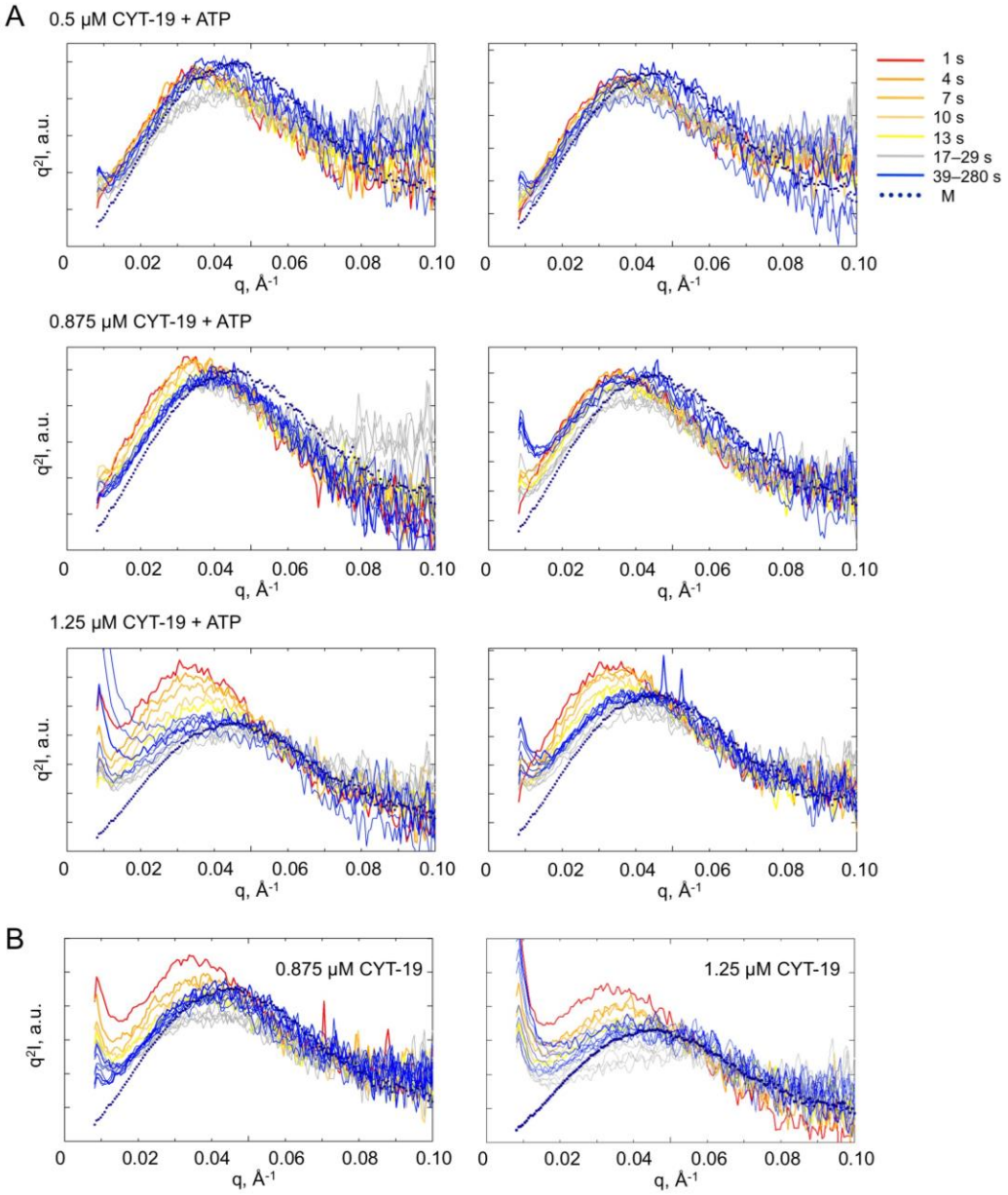


Figure 5.5: CYT-19-concentration dependence of the delay in ribozyme compaction.

(*A*) CYT-19 effects on ribozyme compaction in the presence of 2 mM ATP. Duplicate data sets for ribozyme compaction with three CYT-19 concentrations are shown (0.5 μM ribozyme, 5 mM $Mg(OAc)_2$). As in Figure 5.4, the time points are color-coded in a gradient for the first five exposures (in ~3 s intervals), five subsequent exposures (in 3 s intervals) are shown in gray, and later exposures — in blue. The blue dots indicate

scattering by the misfolded ribozyme. (*B*) CYT-19 effects on compaction in the absence of nucleotides (0.5 μM ribozyme, 5 mM $Mg(OAc)_2$). The peak at the lowest scattering vector q values suggests a substantial amount of aggregation, which was typically more pronounced in the absence of nucleotides.

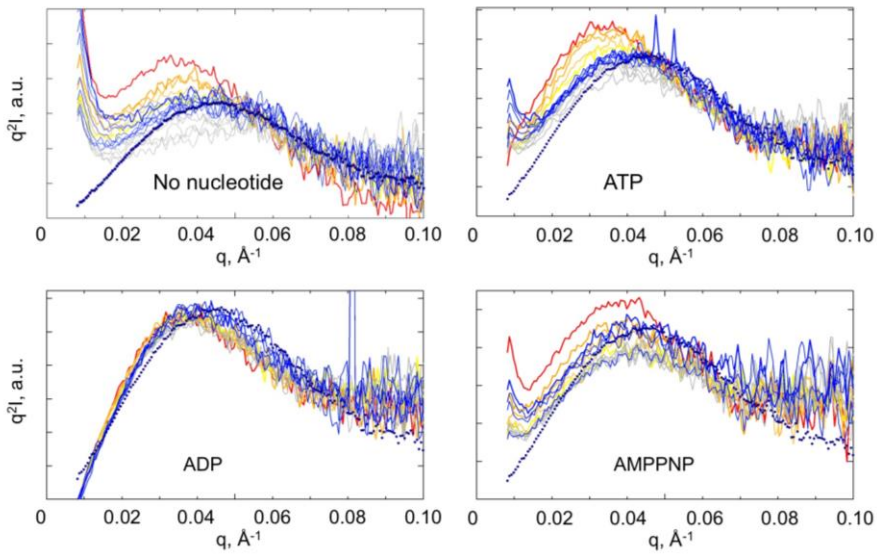
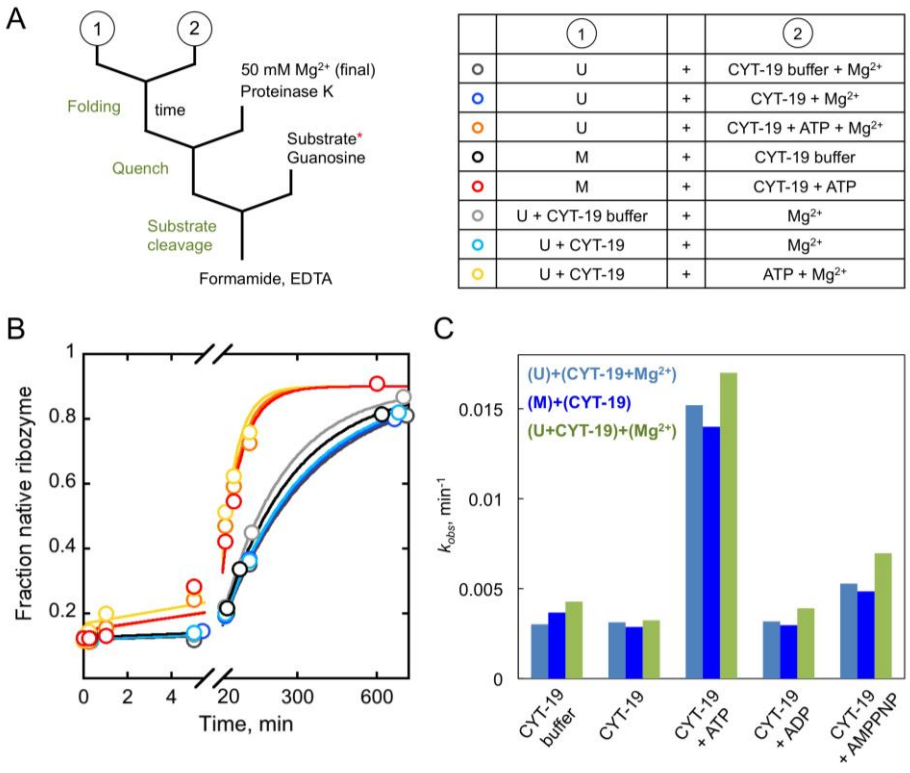


Figure 5.6: Nucleotide effects on CYT-19-induced delay in ribozyme compaction.

All reactions contained 1.25 μM CYT-19 and 0.5 μM ribozyme. The nucleotide concentration was 2 mM. The color scheme is as in Figure 5.5.



	1		2
○	U	+	CYT-19 buffer + Mg^{2+}
○	U	+	CYT-19 + Mg^{2+}
○	U	+	CYT-19 + ATP + Mg^{2+}
○	M	+	CYT-19 buffer
○	M	+	CYT-19 + ATP
○	U + CYT-19 buffer	+	Mg^{2+}
○	U + CYT-19	+	Mg^{2+}
○	U + CYT-19	+	ATP + Mg^{2+}

Figure 5.7: Probing the effects of CYT-19 on ribozyme folding by ribozyme activity.

(*A*) The reaction scheme of the ribozyme activity assay. The contents of solutions **1** and **2**, which were mixed at equal volumes to initiate the folding reaction, are indicated in the table on the right. All solutions contained 50 mM Na-MOPS (pH 7.0), the ribozyme concentration was 4 µM (2 µM final), and the CYT-19 concentration was 5 µM (2.5 µM final). U, M: unfolded, misfolded ribozyme. (*B*) Folding time courses. The colors are as indicated in the table in *A*. An axis break was introduced to display both the fractions of native ribozyme at early time points and the folding rates at later times. The data were fit by a single exponential equation. The 0 time points represent the fractions of native ribozyme determined after directly quenching mixture **1** (left side of the table). (*C*) The rate constants of accumulation of native ribozyme for the data shown in *B*, as well as in the presence of ADP and AMPPNP nucleotides. The colors correspond to reactions in which CYT-19 was added to the unfolded ribozyme either concomitantly with Mg^{2+}

(light blue) or 5 min before Mg^{2+} ions (green). Refolding rates for pre-formed misfolded ribozyme under the same conditions are shown in dark blue.

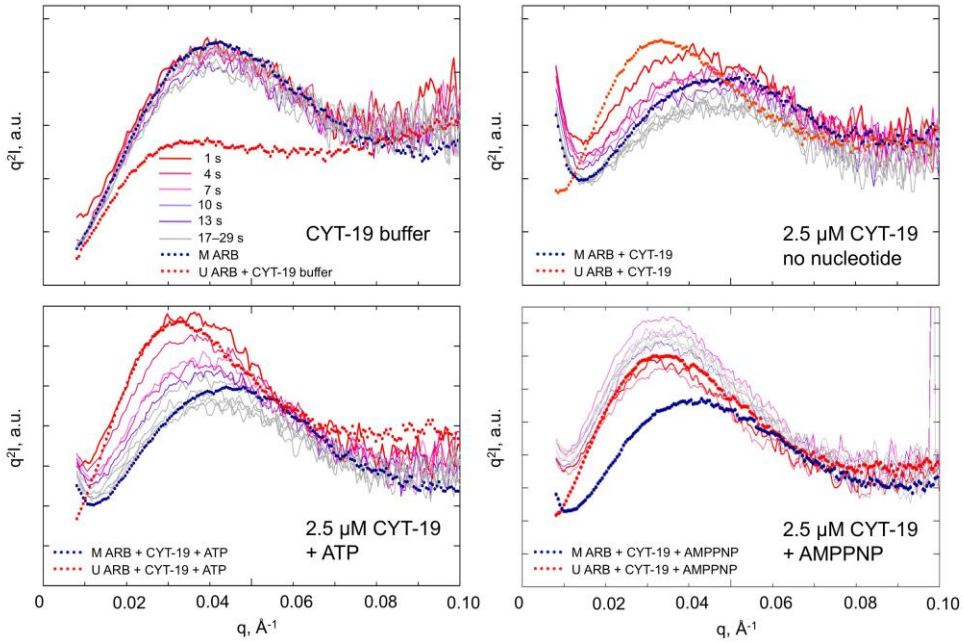


Figure 5.8: CYT-19 delays compaction of the ARB mutant ribozyme.

The unfolded ARB ribozyme was pre-incubated with CYT-19 or CYT-19 buffer at 2 mM Mg^{2+} with or without nucleotides before shifting the Mg^{2+} concentration to 10 mM Mg^{2+} to induce folding. The final ribozyme concentration was 1 µM (2 µM during pre-incubation), and the CYT-19 concentration was 2.5 µM final, and 5 µM in the pre-incubation (see *Materials and Methods*). Red dots indicate the scattering profiles of the ARB mutant ribozyme under the pre-incubation conditions (with or without CYT-19 and nucleotide, as indicated), representing the starting state ('0 time point') of the transition. The ribozyme and CYT-19 concentrations in these static measurements were 1 µM and 2.5 µM, respectively. Blue dots indicate the scattering profiles of the misfolded ARB ribozyme, prepared by incubation for 5 min at 10 mM Mg^{2+} and measured in the presence of the same CYT-19 and nucleotide concentrations as used in the folding reaction (1 µM ribozyme, 2.5 µM CYT-19, 2 mM nucleotide).

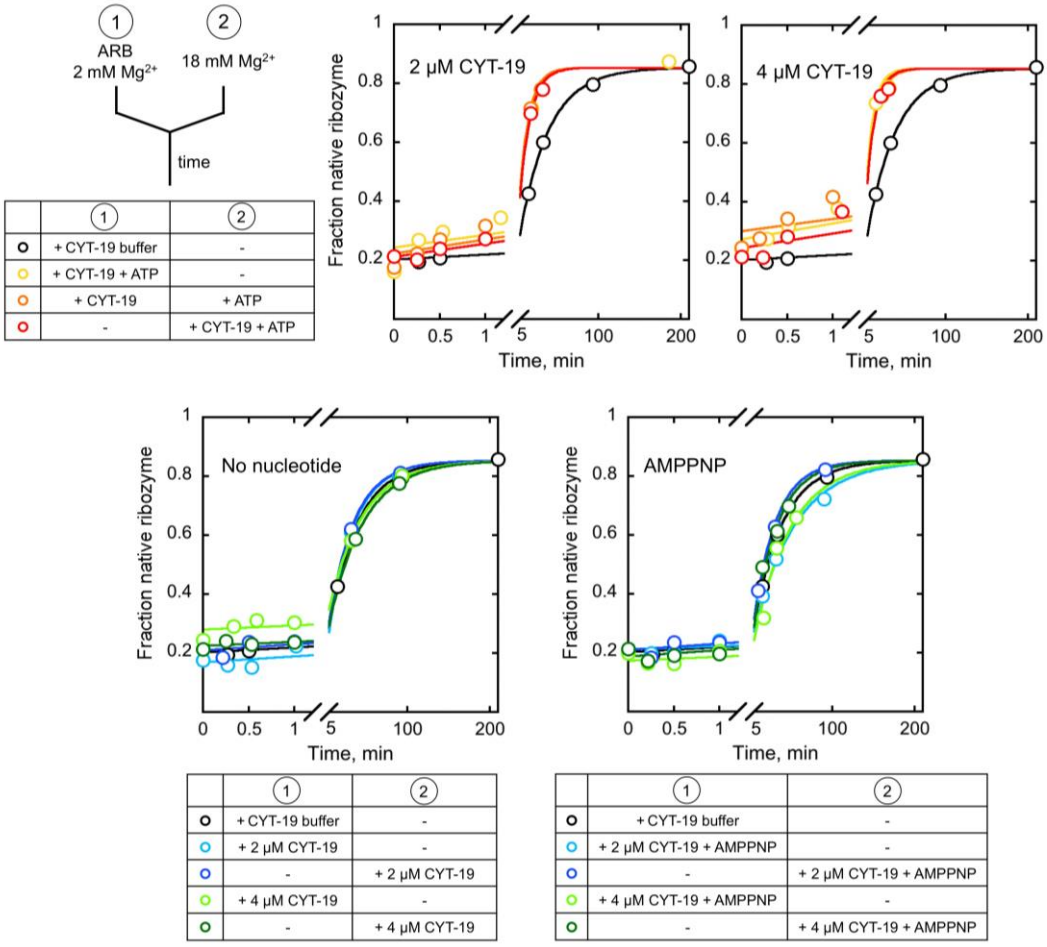


Figure 5.9: Monitoring folding of the ARB ribozyme by ribozyme activity.

The overall reaction scheme is as in Figure 5.7*A*, except for differences in the starting solutions **1** and **2**. The final Mg^{2+} concentration in the folding reaction was 10 mM Mg^{2+}. The tables indicate only the components included in addition to the ones shown in the scheme. 50 mM Na-MOPS (pH 7.0) was present in all solutions. The ribozyme concentration was 1 μM and nucleotide concentration was 2 mM. Data for folding in the presence and absence of CYT-19 and ATP or AMPPNP are shown for two CYT-19 concentrations.

	R_g, Å	I_0/ I_0(RNA)
1 µM U WT L-21	79 ± 3	**1**
2.5 µM CYT-19	34.5	0.073
4 µM CYT-19	33.1 ±1.8	0.13
1 µM U WT L-21 + 2.5 µM CYT-19	70.9 ± 0.5	1.9
1 µM U WT L-21 + 4 µM CYT-19	72.2 ± 0.7	2.8
1 µM Q + 5 mM Mg^{2+}	67.8	**1**
1 µM Q + 5 mM Mg^{2+} + 2.5 µM CYT-19	76.3	1.8
1 µM Q + 5 mM Mg^{2+} + 2.5 µM CYT-19 + ATP	73.0	1.9
1 µM Q + 5 mM Mg^{2+} + 2.5 µM CYT-19 + AMPPNP	71.1	1.8
1 µM Q + 5 mM Mg^{2+} + 4 µM CYT-19 ΔC-tail + AMPPNP	71 ± 2	2.1
1 µM M + 5 mM Mg^{2+}	43.6	**1**
1 µM M + 5 mM Mg^{2+} + 2.5 µM CYT-19	45.0	1.0
1 µM M + 5 mM Mg^{2+} + 2.5 µM CYT-19 + ATP	46.5	1.2
1 µM N + 5 mM Mg^{2+}	42.5	**1**
1 µM N + 5 mM Mg^{2+} + 2.5 µM CYT-19	44.3	1.0
1 µM N + 5 mM Mg^{2+} + 2.5 µM CYT-19 + ATP	45.1	1.1

Table 5.1: Forward scattering intensities and radii of gyration of the ribozyme, CYT-19, and ribozyme-CYT-19 complexes.

The values were derived from data in Figure 5.2. U: unfolded ribozyme; Q: quintuple mutant ribozyme; M: misfolded ribozyme; N: native ribozyme. The radius of gyration (R_g) and forward scattering intensity (I_0) values were estimated by Guinier approximation (see Materials and Methods). Scattering intensities are reported relative to scattering by the ribozyme alone. Values from single determinations or averages and standard errors from two determinations are shown.

Chapter 6 contains text and figures from the following review, co-written with Rick Russell and published by Annual Reviews in Biochemistry:

Jarmoskaite, I. and Russell, R. (2014) RNA helicase proteins as chaperones and remodelers *Annual Reviews in Biochemistry* 83 (expected final publication date: June 2, 2014).

Chapter 6: RNA helicase proteins as chaperones and remodelers

INTRODUCTION

Structured RNAs function throughout biology in processes such as translation, pre-mRNA splicing, RNA processing, chromosome end maintenance, and the regulation of gene expression**.** To function, these RNAs must fold to specific three-dimensional structures that include secondary structure (RNA helices) and tertiary structure, and they commonly include proteins to form ribonucleoprotein complexes (RNPs). The recent discovery of pervasive transcription of eukaryotic genomes has revealed further the large number of non-coding, and potentially structured, RNAs produced in cells (227-229).

The underlying physical properties of RNA result in a universal need for proteins to accelerate RNA structural transitions. It has been known for decades that RNAs face substantial challenges in folding efficiently to specific structures and in transitioning between structured conformations (105, 230, 231). Local secondary structure can be independently stable and long-lived, such that transitions requiring changes in secondary structure can require assistance in vivo (106). Further, because of the limited diversity of RNA side chains, non-native base pairs are likely to form during folding, resulting in kinetically trapped structures in vitro and in vivo. Indeed, even mRNAs, which must be unstructured over much of their length to be translated into protein, are likely to form local structures that require assistance from proteins.

While many diverse proteins possess RNA chaperone activity and have been shown to function as chaperones in some instances (119)**,** superfamily 2 (SF2) helicase

proteins stand out for their ubiquitous roles in manipulating structured RNAs in vivo (3, 149, 196). This expansive collection of proteins comprises multiple families, including many proteins that function on DNA instead of RNA. It also includes proteins that lack conventional helicase activity, despite the 'helicase' designation, which was coined decades ago from sequence similarity with the limited number of known helicases at the time (232). As described below, some SF2 proteins can unwind RNA locally but do not translocate significantly along RNA, while others, not covered here, translocate along RNA but do not unwind it. For simplicity, we refer to all of these proteins as helicases, despite recognizing that not all of them possess helicase activity.

A sizeable subset of these proteins, representing three families, function to manipulate structured RNAs. In parallel with the diversity inherent to structured RNAs and their transitions, SF2 helicases use diverse mechanisms to promote RNA rearrangements. Strikingly, emerging evidence suggests that the mechanisms track to a large extent with the family delineations. Here, we summarize the mechanistic capabilities of these three protein families, as deduced principally from structures and biochemical studies using simple, defined RNA substrates. We then survey biological processes that require these proteins, with emphasis on the specific RNA and RNP substrates, the conformational transitions and, if known, the mechanisms used to promote these transitions.

Helicase families involved in RNA chaperoning and RNP remodeling

Based on comparisons of their sequences and limited structural information, SF2 helicases have been divided into nine families and one smaller 'group' (18), refining and extending earlier classification schemes (5, 232). All SF2 proteins share a conserved core that binds a nucleoside triphosphate and a nucleic acid and consists of two domains (D1

and D2), each of which resembles RecA (Figure 6.1*A*). The core includes at least 12 identified sequence motifs, which are more highly conserved within families than between families**.** These motifs contact both ligands and play important roles in the processes mediated by these proteins, as has been reviewed elsewhere (18, 36, 190). In addition to the helicase core, many SF2 proteins include ancillary domains. For some families, these additional domains are conserved, while for others they are highly variable (18).

Most of the proteins known to function as RNA chaperones or RNP remodelers are found within two families, the DEAD-box and the DEAH/RHA families, and a few are found within the Ski2-like family. We therefore focus below on these three families, and representative schematics and structures are shown in Figure 6.1*A* and 6.1*B*, respectively. Although the DEAD-box and DEAH/RHA families have at times been described collectively as DExD/H-box proteins, recent analysis showed that these two families are no more closely related to each other than to other SF2 families, and they are correspondingly indicated here as two distinct families (18). This is an important change in viewpoint because the molecular properties and mechanisms of these two families appear to differ substantially from each other.

ATPase activity and RNA unwinding

As a step toward defining their mechanisms in RNA chaperoning and remodeling, the capabilities of individual SF2 helicases have been probed using simple, defined RNA substrates. Although these studies have revealed differences between families, there are also central properties that are shared by proteins in these families as well as in other SF2 families and even in the more distantly related SF1 helicases. A hallmark of these proteins is DNA- and/or RNA-dependent nucleotide hydrolysis activity, which is derived

from energetic coupling between these ligands. As illustrated most dramatically by structures of DEAD-box proteins, the two core domains have substantial flexibility relative to each other (47, 48, 50). Upon binding of the nucleotide, principally to D1, the domains have increased affinity for each other and are found in more intimate contact, at least in the presence of bound nucleic acid. Nucleotide hydrolysis and product release reset the protein to the open complex. The consequences of the cycle vary, as outlined below.

Directional RNA translocation and unwinding by DEAH/RHA and Ski2-like proteins

Conventional helicase mechanisms include translocation, or directional movement of the protein along one of the nucleic acid strands. This movement is achieved for SF1 and SF2 helicases at the level of a protein monomer and is directly linked to unwinding, because movement along one strand results in displacement of the complementary strand. The mechanistic steps involved in translocation have been described in detail for the SF1A DNA helicases PcrA and Rep (5). Crystallographic and biochemical approaches have provided high-resolution views of intermediates and revealed dynamics in the translocation process. When bound to a DNA that includes a 3' extension, these proteins form extensive contacts to single-stranded and double-stranded regions of the DNA, essentially encircling the ssDNA region. This encapsulation is made possible by domain insertions in D1 and D2 that provide additional DNA binding surfaces. Translocation occurs as ATP binding induces closure of the two core domains (D1 and D2), resulting in movement of D1 while D2 remains stationary on the DNA. ATP hydrolysis and product release allow a transient loosening of D2 from contacts with both ssDNA and D1, resulting in domain opening and movement of D2 by one nucleotide in the direction of

translocation, 3’ to 5’ for these proteins. Reformation of contacts by D2 then resets the core for ATP binding and another turn of the cycle.

A similar reaction cycle appears to operate in some SF2 helicases. The viral helicases NS3 and NPH-II (from hepatitis C and vaccinia virus respectively), which form their own group within superfamily 2, unwind RNA with moderate processivity by translocating from 3’ to 5’ (233-235). Although structural and biochemical data suggest some differences in the detailed mechanism of these proteins relative to the well-studied SF1 proteins, such as differences in the identities of amino acids that contact the nucleobases of unwound RNA and possible differences in chemical or kinetic step sizes, the basic features of translocation and unwinding are maintained. Thus, SF2 RNA helicases are capable of functioning much like SF1 DNA helicases. However, it should be noted that these viral helicases evolved to function in the context of viral genomes, where processivity and directionality are probably required for function. In contrast, helicase proteins that manipulate cellular RNAs are much less likely to encounter long, continuous helices, and one could envision that their mechanisms would be correspondingly different.

Compared to the SF1 and SF2 families above, there is less structural and mechanistic information for DEAH/RHA and Ski2-like proteins. However, such information is now accumulating at a rapid rate. The first crystal structure of a Ski2-like protein was Hel308, an archaeal DNA helicase (236, 237), and since then structures have been obtained for the *S. cerevisiae* RNA helicases Brr2, Mtr4, and Ski2 (238-242) (Figure 6.1*B*). The first crystal structure of a DEAH/RHA helicase was the *S. cerevisiae* protein Prp43 (243, 244), which functions to remodel the spliceosome and the pre-ribosome. Although it was not recognized by sequence analysis, the structures revealed that these protein families share two domains outside the core, referred to as a winged-

helix domain and a ratchet domain. These domains make key contacts with substrate RNA and are proposed to play key roles in RNA unwinding (245).

DEAH/RHA and Ski2-like proteins are thought to unwind RNA helices by a mechanism that is fundamentally similar to SF1 and viral SF2 helicases. Unwinding by Mtr4 of a 20-bp duplex requires a 3'-ssRNA extension, indicating that Mtr4 binds to the extension and unwinds RNA by translocating 3' to 5' (246)**,** and Mtr4 can unwind a helix as long as 36 bp when in complex with partner proteins of the TRAMP complex (247)**.** 3' overhangs were also shown to enhance unwinding by Brr2 (238). Similarly, directional unwinding is indicated for Prp22, a DEAH/RHA protein that functions in pre-mRNA splicing (248). The unwinding efficiencies appear to decline with increasing helix length, suggesting that the level of processivity is much lower than for SF1 helicases and lower than for the viral SF2 proteins noted above. However, the lower processivity probably does not reflect a difference in mechanism, but rather simply a change in the rate constant for further translocation relative to that for dissociation from the helix, which defines processivity (249). The mechanistic features of DEAH/RHA and Ski2-like helicases suggest a general requirement for a 3' single-stranded extension to allow for helicase loading, which most likely generates important constraints on their physiological substrates and specific functions.

Non-processive RNA unwinding by DEAD-box proteins

DEAD-box proteins make up the largest family of SF2 helicases and they function throughout biology to promote RNA rearrangements and RNP remodeling. Although DEAD-box proteins display RNA unwinding activity, they use a non-conventional mechanism. Early work using mammalian eIF4A protein showed that DEAD-box proteins can unwind RNA helices up to 15 base pairs but with efficiencies

that decrease with increases in helix length or stability (11), suggesting very low processivity. In addition, early studies showed that DEAD-box proteins do not require either 5' or 3' extensions, implying that they can interact directly with a duplex region and at least initiate unwinding without translocating (12, 14, 21, 23). This point was underscored by the finding that although DEAD-box proteins do not perform ATP-dependent unwinding of DNA helices, the *S. cerevisiae* protein Ded1 can unwind a DNA helix that is substituted in one strand with a small number of internal RNA nucleotides, implying that the unwinding is initiated and centered upon the RNA segment (19). The lack of translocation was highlighted by findings that complete unwinding of RNA helices can occur in a single cycle of ATP binding and hydrolysis (83, 95, 181). Further, complete unwinding is possible with the ATP analog ADP-BeF_3 (17), indicating that unwinding can precede ATP hydrolysis and the release of the hydrolysis products. These steps then promote release of the tightly bound RNA strand after unwinding, and they may also contribute to unwinding of longer or more stable helices (250).

Fluorescence studies and crystallography of the *S. cerevisiae* protein Mss116 have led to a detailed physical model for the nonprocessive RNA unwinding by DEAD-box proteins (64). This work showed that D2 alone can bind dsRNA without unwinding it, and it extended previous results indicating that D1 can bind ATP independently (60-63). Thus, in this model the two ligands bind initially to the separated domains, consistent with earlier structures indicating that the two RecA-like domains are flexible relative to each other in the absence of ligands (47, 48, 50, 51, 53, 251). Simultaneous binding of both ligands leads to closure of the domains (47, 48, 54-57, 64), providing a physical model for energetic coupling between binding of adenosine nucleotide and RNA (52, 66, 71, 78, 80, 83, 226). The domain closure results in the exclusion of one strand of dsRNA, bending of the other strand, and consequent local unwinding of the helix. Domain closure

also promotes ATP hydrolysis, which allows release of the tightly bound RNA strand after helix unwinding and resetting of the DEAD-box protein. Unlike the families described above, DEAD-box proteins lack conserved domains that surround the ssRNA and they lack a 'pin' element that may function as a wedge in unwinding (Figure 6.1*B*) (18). Either or both of these differences likely contribute to their lack of processivity.

In the context of this model, it remains possible that there are intermediates in which the domains are in contact but unwinding is not achieved, and that these intermediates accumulate under conditions that do not promote unwinding. For example, binding of a 154-mer RNA and AMP-PNP to the *B. subtilis* protein YxiN leads to favorable domain closure, as monitored by smFRET, but AMP-PNP does not give detectable unwinding of a model RNA duplex (52). Analogously, binding of an ATPase-deficient mutant of YxiN to a closely related RNA and AMP-PNP or ATP gives domain closure but not RNA unwinding (75). Alternatively, this domain closure may reflect binding of YxiN to a single-stranded segment within the large RNA, whereas domain closure may not be achieved upon binding dsRNA.

In addition to accelerating local unwinding of RNA, some DEAD-box proteins can accelerate the reverse reaction, the annealing of two strands to form a dsRNA helix (7, 21, 96, 97). Strand annealing activity does not depend on ATP (21, 97) and may reflect binding of the separated strands to the dsRNA binding site of D2 (64). A recent analysis suggested that DEAD-box proteins can be divided into two broad groups based on ATPase activity and corresponding RNA unwinding activity (10). Interestingly, Rok1, one of the proteins in the 'low' ATPase/unwinding group, is suggested to function by promoting an RNA annealing reaction in ribosome biogenesis (see below).

Unstructured, basic 'tails' of DEAD-box proteins as tethers

Despite their ability to interact directly with a duplex to initiate unwinding, several DEAD-box proteins are activated for unwinding by ssRNA extensions, which can be either 5' or 3' extensions (8, 10, 20, 21, 23, 73)**.** The puzzling ability of DEAD-box proteins to use extensions of either polarity led to them being described as bi-directional, but it is now clear that the extensions do not increase unwinding efficiency by providing starting points for translocation but instead provide binding sites for additional monomers or ancillary regions of the same monomer, which can tether the helicase core in proximity to a substrate helix and remain bound during the local RNA unwinding reaction (36)**.**

Several lines of evidence support this conclusion. The extension is not required to be contiguous with the duplex to be unwound; for Ded1, the extension can even be connected by a bridging streptavidin protein as long as it allows proximal binding of an additional protein molecule (23). For other DEAD-box proteins, the extension does not have to be single-stranded, as a dsRNA or DNA extension is at least as effective as ssRNA for the *Neurospora crassa* protein CYT-19 (22). Deletion of an unstructured C-terminal region of CYT-19 leaves the unwinding activity of the helicase core intact but abrogates the stimulation by DNA or RNA extensions, indicating that this 'C-tail' is responsible for binding the extension and tethering the core for unwinding of an adjacent helix (34, 155).

There is significant variation in the properties of these tethering interactions. Some DEAD-box proteins function as general RNA chaperones, and interactions of the C-tail are correspondingly versatile (252). There is no evidence for RNA sequence or structure specificity, and the C-tail is very flexible relative to the core, accounting for the enhanced RNA unwinding from extensions regardless of their polarity (34). On the other hand, some yeast DEAD-box proteins have C-terminal basic regions that may function

similarly but with less flexibility between the tail and the core, such that a preference for polarity is observed (10). The *Thermus thermophilus* protein Hera includes a bipartite extension with an RRM (RNA recognition motif) domain that binds ssRNA with a preference for guanosine and a short unstructured region that binds dsRNA (32, 37). Other DEAD-box proteins are targeted to specific substrates by interactions with ancillary domains (200, 253) or surfaces within the helicase core (40, 254).

Remodeling of RNP complexes

While much is known about the abilities of SF2 RNA helicases to unwind duplexes, the physiological substrates for many of these proteins include assembled proteins, and a few studies have probed the abilities of certain SF2 helicases to remodel simple, defined RNP complexes. The viral helicase NPH-II can remove several protein complexes from ssRNA or dsRNA in an ATP-dependent manner (101, 255), most likely by translocating 3' to 5' along one RNA strand and displacing the protein(s) as well as the complementary RNA strand, if present (256). Ded1 is also able to remove proteins from RNA (101, 102), despite unwinding RNA non-processively as described above. Interestingly, Ded1 displays selectivity, accelerating dissociation of some protein complexes but relying on spontaneous dissociation of the tight-binding U1A protein before unwinding the two RNA strands that form the U1A binding site (102). The physical origins of this selectivity remain to be elucidated (256).

RIBOSOME BIOGENESIS

Here and in the sections that follow, we outline the biological functions of SF2 helicase proteins in rearranging RNAs and remodeling RNPs. Ribosome biogenesis is an elaborate process that involves regulated cleavage, modification and folding of several large ribosomal RNAs, as well as extensive changes in RNP constitution throughout the

assembly process (172, 257). Large fractions of the DEAD-box proteins of *E. coli* and *S. cerevisiae* function in ribosome biogenesis (4 of 5 and 15 of 26, respectively), and in *S. cerevisiae* several DEAH/RHA helicases and a Ski2-like helicase also participate. In bacteria, the RNA helicases seem to be required only at low temperatures, presumably due to increased lifetimes of non-native RNA structures (258, 259). However, in eukaryotes most of the helicases involved in ribosome biogenesis are essential, likely reflecting the increased complexity of the eukaryotic ribosome and a greater need for regulation.

Ribosome biogenesis in *E. coli*

Ribosome biogenesis in *E. coli* involves four DEAD-box helicases: DbpA, SrmB, CsdA and RhlE. DbpA (YxiN in *B. subtilis*) was originally implicated based on specific stimulation of its ATPase activity by 23S rRNA, which later was narrowed down to hairpin 92 (hp92) of the peptidyltransferase center (24, 25). The recognition of hp92 is mediated by a C-terminal RRM domain, which is flexibly linked to the helicase core and can enhance unwinding of helices positioned either 3' or 5' from hp92 (26, 28, 200, 253). Functional and structural evidence supports a model in which DbpA rearranges RNA structure around hp92 in the 50S precursor while remaining tethered via the RRM domain. Despite this specific hp92 recognition, ΔdbpA mutants show no obvious defects, and defective ribosome assembly is only observed with a dominant negative DbpA mutant that lacks ATP-dependent unwinding activity but retains RNA binding (260, 261). Thus, in vivo, DbpA-mediated rearrangements appear to be redundant with other pathways but can be blocked by bound DbpA (261).

Recent evidence also suggests specific sites for SrmB. The absence of functional SrmB leads to accumulation of 40S intermediates that contain immature 23S rRNA and

reduced amounts of some ribosomal proteins (262). Tandem affinity purification revealed SrmB interactions with two ribosomal proteins, L4 and L24, which bind a 5' region of 23S rRNA (263). However, rRNA mutations that suppress the requirement of SrmB map to domain II of 23S rRNA, which is distant in the mature ribosome. Some suppressors map to a section of 5S rRNA with base complementary to domain II of 23S rRNA, suggesting that SrmB may prevent 5S rRNA annealing and promote native 23S rRNA folding and protein assembly within domain II (264).

CsdA, aka DeaD, is induced at low growth temperatures and several lines of evidence indicate that it functions as a general RNA chaperone in ribosome biogenesis and other RNA-related processes, including mRNA decay and translation (265, 266). CsdA functions at multiple stages of ribosome biogenesis and can suppress defects in ΔsrmB strains, suggesting functional overlap (267). Interestingly, mutations in CsdA can be suppressed by overexpression of another cold shock protein and general chaperone, CspA, – a small (70 aa), unrelated protein that can destabilize RNA helices and functions in transcription antitermination (268-270). CsdA contains a predicted RRM in its C-terminus that may contribute to general RNA recognition, possibly by recognizing a commonly occurring structural element (14).

RhlE appears to be largely but not completely redundant with CsdA. While growth defects are not observed in ΔrhlE cells, removal of RhlE and CsdA enhances the cold-sensitive phenotype of ΔcsdA cells and increases the accumulation of 50S precursors (271). However, unlike CsdA, RhlE cannot substitute for SrmB, and its overexpression even increases the accumulation of 50S precursors observed in ΔsrmB cells (271). Together the results suggest a complex network of functions for SrmB, CsdA, and RhlE, with extensive but incomplete overlap between them.

Ribosome biogenesis in *S. cerevisiae*

Nineteen yeast helicases have been implicated in ribosome biogenesis. Seventeen of them are essential, and several have been linked to distinct steps in small or large subunit biogenesis based on the sizes and compositions of ribosome intermediates that accumulate upon mutation or depletion of the helicase protein (Figure 6.2*A*) (272). While it can be difficult to separate direct from indirect effects, as some of the blocked steps may be downstream from those mediated directly by the helicase (172), these studies have provided important insights and constraints on the helicase functions in ribosome biogenesis. A distribution of the proposed roles of *S. cerevisiae* helicases in ribosome biogenesis is shown in Figure 6.2*B*. Below we discuss examples of proteins implicated in each of these roles, illustrating the mechanistic diversity of SF2 helicases, particularly within the DEAD-box family.

Several helicases have been suggested to regulate nucleolytic processing of pre-rRNA into mature rRNA by remodeling structures adjacent to cleavage sites. The DEAD-box protein Dbp3 promotes site A3 cleavage by RNase MRP during 25S rRNA maturation and was suggested to disrupt an adjacent hairpin to increase access of the RNase (273). The DEAH/RHA protein Prp43 (see below) is recruited near the site of Nob1-mediated 3' cleavage of 18S pre-rRNA by the cofactor Pfa1, which binds the OB-fold domain of Prp43 and stimulates its helicase activity (244, 274-276). 3' end processing of 5.8 S rRNA requires the Ski2-like helicase Mtr4 (277), which functions as part of the TRAMP complex (see below) and is also involved in general quality control during ribosome biogenesis (172).

The possibility of an interesting divergence of DEAD-box protein functions is introduced by recent studies of Rok1 in small subunit biogenesis. Rok1 promotes a structural change to facilitate cleavage at site A2 by the Nob1 nuclease. Interestingly, in

vitro experiments suggest that rather than disrupting structure, Rok1 specifically promotes the temporary formation of a helix that allows correct processing of pre-18S rRNA (98, 99). The specificity for this "pre-A2" helix is conferred by cofactor Rrp5, which stimulates Rok1-mediated annealing of pre-A2 by ~10-fold relative to other helices. Studies using model helices, noted above, showed that Rok1 has low ATPase and RNA unwinding activities relative to other DEAD-box proteins (10), suggesting an intriguing link between the capabilities and functions of Rok1. Nevertheless, Rok1 does retain a basal level of nucleotide-dependent RNA unwinding activity (10), and it is also suggested to function in snoRNP dissociation (see below).

Several helicases have been shown to function in rearrangements involving snoRNAs. snoRNAs are small structured RNAs that base pair with pre-rRNAs to guide site-specific cleavage, methylation and pseudouridylation, and can direct pre-rRNA folding (278). There are 75 snoRNAs in yeast, from structurally distinct box-C/D and H/ACA classes, with three essential ones (U3, U14, snR30) that guide cleavage events and 72 non-essential ones that specify nucleotide modifications (278). Three DEAD-box proteins have been implicated in dissociation of certain snoRNAs. Has1 is required primarily for dissociation of the essential U3 and U14 snoRNAs, although several other snoRNAs are also affected upon Has1 depletion (279), while snR30 dissociation requires Rok1 (100). Interestingly, dissociation of U14 requires Dbp4 in addition to Has1 (280), indicating that removal of a given snoRNA can require direct or indirect activities of more than one helicase. Several helicases have also been suggested to function in U3 snoRNP-guided cleavage and rearrangements of pre-rRNA. The DEAH/RHA helicases Dhr1 and Dhr2 and DEAD-box proteins Dbp8, Rrp3 and Rok1 coprecipitate U3 snoRNA and are required for U3-guided cleavages (281). One or more of these helicases may also function in U3-assisted formation of the central pseudoknot in the 18S rRNA (282).

Of the helicases in eukaryotic ribosome biogenesis, the DEAH/RHA protein Prp43p is particularly versatile, playing multiple roles while also functioning in pre-mRNA splicing (see below). In addition to promoting Nob1-mediated cleavage of 18S rRNA, as noted above, Prp43 co-precipitates with snoRNAs (primarily box-C/D) and rRNA sequences near snoRNA binding sites, and is required for binding or release of snoRNAs from the pre-ribosome (275, 283-285). The broad spectrum of RNA substrates indicates that Prp43 recognizes general features of structured RNAs, either intrinsically or through cofactors such as Pfa1 (276). Surprisingly, in vitro studies showed that Prp43 does not require a 3' extension for RNA unwinding, which would seem to be at odds with the 3'-5' directionality observed for other DEAH/RHA helicases (286, 287). Together, these results have led to the suggestion that Prp43 could function similarly to DEAD-box proteins by forming tethering interactions with structured RNA via a cofactor protein or an ancillary domain and disrupting RNA structure locally (288), which would be an important exception from the general differences in functional mechanisms between DEAH/RHA and DEAD-box proteins.

SPLICEOSOME ASSEMBLY AND FUNCTION

After ribosome biogenesis, eukaryotic pre-mRNA splicing is the process that requires the most RNA helicases [reviewed in (289, 290)]. Eight helicases function in splicing in *S. cerevisiae*, including three DEAD-box, one Ski2-like and four DEAH/RHA proteins. These helicases have been implicated in RNA secondary structure rearrangements and protein displacement, allowing the spliceosome to assemble, rearrange and dissociate in a controlled manner (Figure 6.3). Splicing helicases have also been shown to function in proofreading, counteracting the forward progression of slowly-reacting pre-mRNAs and intermediates, and promoting their dissociation (291). Below,

we summarize the current knowledge of helicase-mediated rearrangements of the *S. cerevisiae* spliceosome and discuss how the properties of each helicase family correspond to their roles in pre-mRNA splicing.

DEAD-box helicases: chaperoning spliceosome assembly

Two DEAD-box helicases, Sub2 and Prp5, are required for branchpoint (BP) recognition by U2 snRNP and promote rearrangements that allow base-pairing of the BP and U2 snRNA (104, 292). Genetic studies have implicated Sub2 in dissociation of the Mud2/Bbp protein heterodimer from the BP sequence, and consistent with a role in protein displacement, studies of the human homolog UAP56 suggest that ATPase activity, but not RNA unwinding, is required for BP exposure (104, 293). Prp5 promotes a transition between alternative conformations of the U2 snRNP, which involves direct or indirect disruption of the 8-bp helix IIc, formation of alternative helix IIa, and ejection of the Cus2 protein (294-296) (Figure 6.3). This transition increases the accessibility of the BP-recognition sequence of U2 snRNA, located in the loop of a conserved branchpoint interaction stemloop (BSL)(294, 297, 298). Upon initial recognition of the BP sequence by the BSL, Prp5 has also been implicated in unwinding the BSL stem to allow stable association of U2 snRNP with the BP (298) (Figure 6.3).

Following stable U2 snRNP binding, the Prp28 helicase is required for unwinding of the 6-bp helix between U1 snRNP and the 5' splice site (5'-ss), facilitating release of U1 snRNP and exposing the 5'-ss for base pairing with the U6 snRNP (299) (Figure 6.3). Although Prp28 and Prp5 have been described as lacking unwinding activity (294, 300), the reported experiments have used helices longer than the 10-15 bp that are efficiently unwound by other DEAD-box proteins. Overall, the localized rearrangements by all three DEAD-box helicases are consistent with the general capabilities of DEAD-box proteins

and in contrast with the generally more extensive remodeling events mediated by the helicases described below.

DEAH/RHA helicases: activation for catalysis, product release and snRNP recycling

The four spliceosomal DEAH/RHA-family helicases catalyze rearrangements that accompany or follow the chemical steps of splicing, beginning with activation of the spliceosome for the first step of splicing. Strikingly, these rearrangements generally involve more extensive base-pairing interactions and/or displacement of proteins than the changes promoted by DEAD-box proteins, and the current data suggest a strong correspondence of the functional roles of these proteins with the capabilities of the protein family as a whole, as established from crystallographic and biochemical studies described above.

Prp2 promotes dissociation of proteins from the BP region, exposing the BP adenosine for attack on the 5'ss (301, 302). Prp2 requires 23-33 nt of pre-mRNA in the 3' direction from the BP, suggesting 3' to 5' translocation (303). Along the way, Prp2 is thought to dislodge the BP-binding SF3a and SF3b proteins and a number of others, contributing to catalytic activation of the spliceosome (Figure 6.3) (301, 302, 304) (Figure 6.3). Following lariat formation, Prp16 promotes rearrangement of the spliceosome for the second catalytic step. The rearrangement involves refolding of U2 snRNP to the IIa state, destabilization of U2/U6 helix 1 (Figure 6.3), and displacement of several proteins (296, 305-308). Following splicing, Prp22 releases the mRNA by disrupting base pairs between the U5 snRNP and the mRNA and may destabilize protein-RNA contacts between the U5 snRNP and the exon-exon junction (309, 310) (Figure 6.3). Prp22 initiates translocation ~23 nts downstream from the splice junction, suggesting 3'-5' translocation (248, 311). Finally, Prp43 mediates the extensive

rearrangements that allow release of the intron lariat and recycling of the U2, U5 and U6 snRNPs (312-314). The activity of Prp43 in spliceosome disassembly is stimulated by cofactor Ntr1, which also enhances in vitro unwinding activity of the helicase (287, 314).

Ski2-like helicase Brr2: remodeling U4/U6 snRNP

Like the DEAH/RHA proteins, the Ski2-like helicase Brr2 promotes a large-scale rearrangement, acting early in the splicing cycle to disrupt an extensive base-pairing network between the U4 and U6 snRNAs (315, 316) (Figure 6.3). The disruption releases the 5'ss-recognition sequence within U6 snRNA and allows it to form catalytically important intra-U6 and U2/U6 interactions. The mechanism of U4/U6 snRNP dissociation is not fully understood, but recent data suggest that Brr2 binds the 3' end of the U4 snRNA and translocates 3' to 5' to unwind U4/U6 stem I and presumably stem II (317-319). This activity is tightly regulated by U5 snRNP components Prp8 and Snu114 (319). While Brr2 readily unwinds protein-free U4/U6 snRNA in vitro, it is unknown how proteins that coat the RNA in the U4/U6 snRNP affect unwinding (315, 317).

Proofreading by spliceosomal helicases

In analogy with GTPases in translation, several splicing helicases use ATP to enhance fidelity by proofreading the splicing process (291, 320). Productive steps are in kinetic competition with a non-productive branch in which a helicase uses ATP to reject the pre-mRNA substrate (321, 322). For a substrate with a non-consensus sequence, the reduced rate of the forward step increases flux through the non-productive branch, leading to rejection and eventual discard of the unspliced substrate or a splicing intermediate. In addition, the helicases may act preferentially on stalled splicing complexes because of their decreased stability and enhanced dynamics, analogous to

DEAD-box proteins that act as chaperones by resolving misfolded RNA structures [(127, 291); see *Self-splicing introns* below].

As expected from this mechanism, the proofreading checkpoints largely overlap with the step(s) mediated by a given helicase in the productive splicing cycle. Thus, Prp5 rejects pre-mRNAs with mutations in the branch site (323), Prp28 and Prp16 suppress splicing of 5'-ss mutants (324-327), and Prp22 primarily proofreads the 3'-ss (328). Some of the helicases may also act with some promiscuity on structures that become exposed due to their lower stability or to a lack of bound proteins. Thus, Prp43 can reject or discard stalled intermediates in at least two steps of the spliceosome cycle, in addition to disassembling the spliceosome to complete the native cycle (329, 330), while Prp22 can reject BP and 5'-ss mutants, although less efficiently than 3'-ss mutants (328).

SELF-SPLICING INTRONS: DEAD-BOX PROTEINS AS GENERAL RNA CHAPERONES

Group I and group II introns are autocatalytic RNAs that are broadly distributed in nature and function by splicing themselves out of precursor RNAs, with some retaining ancestral roles as mobile genetic elements (122). Several mitochondrial group I introns in *N. crassa* require the DEAD-box protein CYT-19 for efficient folding (85), and all of the *S. cerevisiae* mitochondrial group I and group II introns depend on Mss116, a CYT-19 homolog (121, 331). Group II introns in plant mitochondria and chloroplasts also require DEAD-box proteins, suggesting a widespread role (332, 333). The versatility of these DEAD-box proteins indicates that they are general RNA chaperones (85, 121), and biochemical studies of these proteins and introns have been instrumental in elucidating general features of the mechanisms by which DEAD-box proteins manipulate structured RNAs.

Model group I introns have been used to probe how DEAD-box proteins promote transitions away from misfolded structures. A ribozyme derivative of the *Tetrahymena thermophila* group I intron (from the large subunit rRNA) misfolds to a single, dominant conformation that is compact and includes extensive native structure but is thought to have a non-native topology (145, 157). This misfolded conformation undergoes a slow 'refolding' rearrangement to the native state that is rate limited by extensive unfolding (140, 145). CYT-19 accelerates refolding in an ATP-dependent manner, indicating that it can use ATP to accelerate unfolding of the misfolded ribozyme (22). The function of CYT-19 as a general chaperone, along with the non-cognate nature of the *Tetrahymena* ribozyme, would suggest that CYT-19 accelerates this refolding reaction without recognizing specific features of the misfolded ribozyme, and indeed CYT-19 can also unfold the native ribozyme (127). CYT-19 also accelerates refolding of a misfolded group I intron from the bacterium *Azoarcus* (116), and it promotes splicing of its cognate introns in vitro, most likely by accelerating refolding transitions (85). These refolding processes are known or suggested to require unfolding of tertiary structure, but DEAD-box proteins have not been shown to disrupt RNA tertiary contacts. Recent work indicates that CYT-19 disrupts tertiary structure in the *Tetrahymena* ribozyme using a 'helix capture' mechanism by trapping helical elements after they lose tertiary contacts spontaneously and then using RNA unwinding activity to separate the strands of the trapped helix (C. Pan et al., submitted).

Group I introns have also been valuable for probing how directionality in folding can be conferred upon the inherently non-specific activity of these DEAD-box proteins. For the *Tetrahymena* ribozyme, there is a correlation between global RNA stability and the efficiency of unfolding by CYT-19, with less stable conformations being unfolded more efficiently. The effect is that net folding proceeds from the less stable misfolded

conformation to the more stable native state despite the absence of specific recognition by CYT-19 (127). If the native structure is destabilized by mutations or by lowering Mg^{2+} concentration, it is unfolded efficiently by CYT-19 and instead of the native ribozyme accumulating, the misfolded conformation or unfolded conformations accumulate (127). Mss116, which has greater RNA unwinding activity than CYT-19, can readily unfold the native structure under a broader range of conditions (128, 129). Mss116 can also function as a chaperone to promote conformational transitions between or following the two steps of splicing for a model intron from yeast (334), most likely by disrupting structure non-specifically, with directionality conferred by the functionally irreversible dissociation of the intron from the ligated exons.

Group II introns also undergo slow folding transitions to the native state, and these transitions can be accelerated in vitro by DEAD-box proteins (154). Two yeast mitochondrial introns have been used most extensively in these studies. CYT-19 promotes native folding of the bI1 intron in an ATP-dependent manner and is dispensable after the native state is reached, establishing that it chaperones folding of the RNA (126). A second intron, aI5γ, and its ribozyme derivative (D135), fold slowly under near-physiological conditions (74, 128), with D135 folding through multiple pathways on time scales from minutes to hours (151, 335). Mss116 accelerates native folding of both the intron and D135 ribozyme (74, 128, 151, 152), and further work with D135 has shown that Mss116 accelerates two transitions, only one of which requires ATP (151, 153). The ATP-dependent step is thought to reflect disruption of non-native structure, which includes or is stabilized by exonic structure, analogous to the roles of DEAD-box proteins in folding of other group I and II introns (137, 153). This step may also include nucleotide-dependent dissociation of Mss116 from a folding intermediate (153). The ATP-independent step is earlier in folding and is thought to include compaction of the

largest domain of the intron, a transition that can limit folding for a subpopulation of the ribozyme (335, 336).

GENERAL RNA CHAPERONES IN TRANSLATION

Eukaryotic translation requires multiple helicases, largely for rearrangements of the 5' untranslated region (5'-UTR) in cap-dependent initiation. In this process, the pre-initiation complex (PIC) is recruited to the 5' cap through interactions with the eIF4F complex, followed by 5' to 3' scanning along the 5' UTR in search of the start codon (337). Both PIC loading and scanning, the latter process involving up to thousands of nucleotides, require disruption of RNA structures within the 5'UTR. Three helicases, eIF4A, Ded1 and DHX29, are thought to clear the way for the PIC by resolving these structures.

The DEAD-box protein eIF4A is a component of the eIF4F complex, which also includes eIF4G and the cap-binding protein eIF4E. Due to its localization to the cap, eIF4A is thought to promote PIC loading by disrupting RNA secondary structures in the vicinity of the 5' terminus (72, 337). A poor helicase on its own, eIF4A is enhanced for RNA unwinding by interactions with other initiation factors (92). eIF4G binds and modulates the orientation of the eIF4A core domains, and yeast eIF4G can bias eIF4A to unwind helices with 5' overhangs, potentially generating an overall 5' to 3' directionality for eIF4A-mediated rearrangements of the 5'UTR (39, 89, 91, 338). eIF4B binds to D2 of eIF4A and can interact with ssRNA via its RRM, potentially trapping ssRNA produced by eIF4A (39, 42, 337). eIF4B may also promote RNA unwinding by tethering eIF4A to RNA, analogous to the ancillary RNA-binding domains and extensions found in other helicases (42).

Following PIC loading, eIF4A is also thought to disrupt downstream 5'UTR structures to facilitate scanning. Consistent with this model, eIF4A and eIF4B crosslink to the mRNA up to 52 nt downstream from the cap (339). The much higher intracellular concentration of eIF4A than the other initiation factors and ribosomes suggests that eIF4A may carry out some functions independently, and that multiple eIF4A monomers can act per mRNA (337, 340). However, Ded1p and DHX29 are also involved in translation and are more potent helicases, suggesting that eIF4A may not be the primary helicase during scanning (337). In addition to cap-dependent translation initiation, eIF4G and 4A are required for initiation at Type 1 and Type 2 IRES elements in viral mRNAs, where they induce conformational rearrangements necessary for PIC loading and initiation (341, 342).

Both the helicase and RNP remodeling activities of Ded1 are likely to play roles in translation initiation. Ded1 interacts with eIF4G, and this interaction can both stimulate and repress translation initiation (93). While the role of Ded1 in repression is ATP-independent, ATP is required for Ded1-mediated release of mRNAs from stress granules, which reverses repression (93). As a helicase, Ded1 is thought to disrupt 5'UTR helices to promote PIC loading and scanning of the 5'UTR, although these functions remain to be established experimentally (337).

The DEAH/RHA helicase DHX29 is required for efficient translation of mRNAs with highly structured 5'UTRs and is also important for overall translation (343, 344). DHX29 interacts with the 40S ribosome in the PIC near the mRNA entry channel and promotes disruption of double-stranded segments, allowing these sequences to enter the ribosome (344). DHX29 may unwind these mRNA structures directly, or it may act indirectly by remodeling the entry channel of the ribosome (344).

Another helicase that promotes translation of specific groups of mRNAs is the DEAH/RHA helicase RNA helicase A (RHA), which interacts with mRNAs containing the 150 nt post-transcriptional control element (PCE) (345, 346). RHA also interacts with other structures in 5'UTRs, including G-quadruplexes (347), and it presumably destabilizes these structures to give the observed increases in polysome loading and translation.

Finally, Vasa (DDX4) promotes translation of mRNAs encoding proteins that are involved in embryogenesis and germline development in animals (348). Vasa interacts with the initiation factor eIF5B, which mediates ribosomal subunit joining (349, 350), and the role of the helicase appears to be to promote eIF5B loading to the PIC, perhaps by removing repressors from the eIF5B binding site (349, 350).

Although helicase functions have been best characterized in translation initiation, a handful of RNA helicases function at later steps. The Ski2-like protein Slh1 associates with translating ribosomes and is important for translation (351). The DEAD-box protein Dbp5 and its cofactors Gle1 and IP6 function in translation termination (148, 352). Based on genetic data, Dbp5 is suggested to facilitate binding of release factors eRF1 and 3, likely by promoting rearrangements of the ribosome (353). The translating ribosome may also disrupt mRNA structures directly during elongation, analogous to the prokaryotic ribosome (354, 355), limiting the requirement for helicases past the initiation stage.

RNA DECAY

Stable RNA structures and bound proteins can inhibit enzymatic RNA degradation, and unsurprisingly RNases are often found to recruit RNA helicases (356). The DEAD-box helicase RhlB is a component of the *E. coli* degradosome and promotes degradation of RNAs with stable secondary structures by the exonuclease PNPase (357,

358). Interactions with the degradosome have also been reported for other DEAD-box poteins in *E. coli*, and DEAD-box components are found in the degradosomes of other bacteria (266, 356).

The eukaryotic exosome primarily utilizes helicases from two regulatory complexes: Ski2 helicase of the cytoplasmic Ski complex and the closely related Ski2-like helicase Mtr4 of the nuclear complex TRAMP (359). Both helicases appear to unwind structured RNAs and channel ssRNA into the exosome for degradation (247, 360). Directional 3'-5' unwinding by these Ski2-like helicases is thought to promote coupling of unwinding and degradation, which also proceeds 3' to 5'. An interesting interplay has been observed between Mtr4 and the poly(A) polymerase component of TRAMP. The polymerase promotes Mtr4 loading on RNAs by adding oligo(A) tails, and once bound, Mtr4 inhibits the polymerase, thus controlling the length of the tail (247, 361). The RHAU helicase can also transiently interact with the exosome and promote degradation of AU-rich element-containing mRNAs, likely by disrupting RNA structures or protein binding (362).

OTHER RNA CHAPERONE AND REMODELING ACTIVITIES

Transcription

It has long been known that the DEAD-box proteins p68 and p72 function in transcription by recruiting transcription regulators to RNA polymerase II, including the tumor suppressor p53 (363). This function appears to be carried out at the level of DNA, but recent results suggest a general RNA chaperone function for the *S. cerevisiae* homolog Dbp2. Dbp2 was shown to promote clearance of cryptic RNA transcripts from chromatin, apparently by remodeling RNA co-transcriptionally and promoting binding of decay factors (364). Dbp2 probably remodels the RNA indiscriminately, as it also

promotes loading of export factors to native transcripts (365). Whether p68 and/or p72 play similar roles remains unknown, but there is evidence for their localization to regions of elongation in transcription (363).

Nuclear export

Several DEAD-box proteins have been implicated in the RNP remodeling events that occur as mRNA is exported through the nuclear pore complex (NPC). Dbp5 is localized to the cytoplasmic face of the NPC and is required for dissociation of several export factors, contributing to the directionality of mRNA export (103, 366). The ATPase activity of Dbp5, which is required for protein displacement, is regulated by the activator Gle1 and co-activator IP6, while the nuclear pore component Nup159 promotes recycling of nucleotide-free Dbp5 after ATP hydrolysis (254). Inside the nucleus, UAP56 (Sub2) plays an incompletely understood, ATP-dependent role in mRNA export as a component of the TREX complex (367, 368). Furthermore, multiple human helicases are used by viruses during export of viral RNAs, including DEAD-box proteins DDX3 (the human homolog of Ded1) and DDX1, as well as the DEAH/RHA helicase RHA (369).

Editing

Illustrating the diversity of SF2 helicase-dependent processes, RNA helicases have been implicated in small RNA-guided pre-mRNA editing that occurs in the mitochondria of some protozoa (370). During editing, the 5' region of a short guide RNA (gRNA) pairs with pre-mRNA, while the partially complementary 3' region serves as template for uridine insertions and deletions by the editosome. Editing creates new complementarity between the pre-mRNA and the 5' region of another gRNA, resulting in a 3'-5' cascade of editing events. In *Trypanosoma brucei*, the sequential binding and dissociation of gRNAs is promoted by the DEAD-box helicase REH1, which appears to

dissociate gRNAs after editing (371). A second helicase, REH2 (related to DEAH/RHA and Ski2-like families), associates with the editosome and may promote RNA rearrangements during editing (372).

CONCLUSIONS

It is now clear that SF2 helicase proteins function throughout RNA biology, playing a diverse set of roles and using a diverse set of mechanisms. The list of helicases that function to rearrange or remodel RNAs or RNPs is large and growing. Still, despite remarkable progress in recent years, we have very limited knowledge of the interaction sites for helicases on their biological RNA and RNP substrates, and we know little about the structural rearrangements that they promote. Mechanistically, the available data suggest a general model in which DEAD-box proteins act locally, disrupting or in some cases promoting RNA structure without significant translocation, while DEAH/RHA and Ski2-like proteins load onto ssRNA and translocate 3'-5', disrupting RNA structures and displacing proteins as they translocate. In the coming years, it will be fascinating to continue to learn about the structures and dynamics of the RNA and RNP substrates and the mechanisms used by helicases to manipulate them.

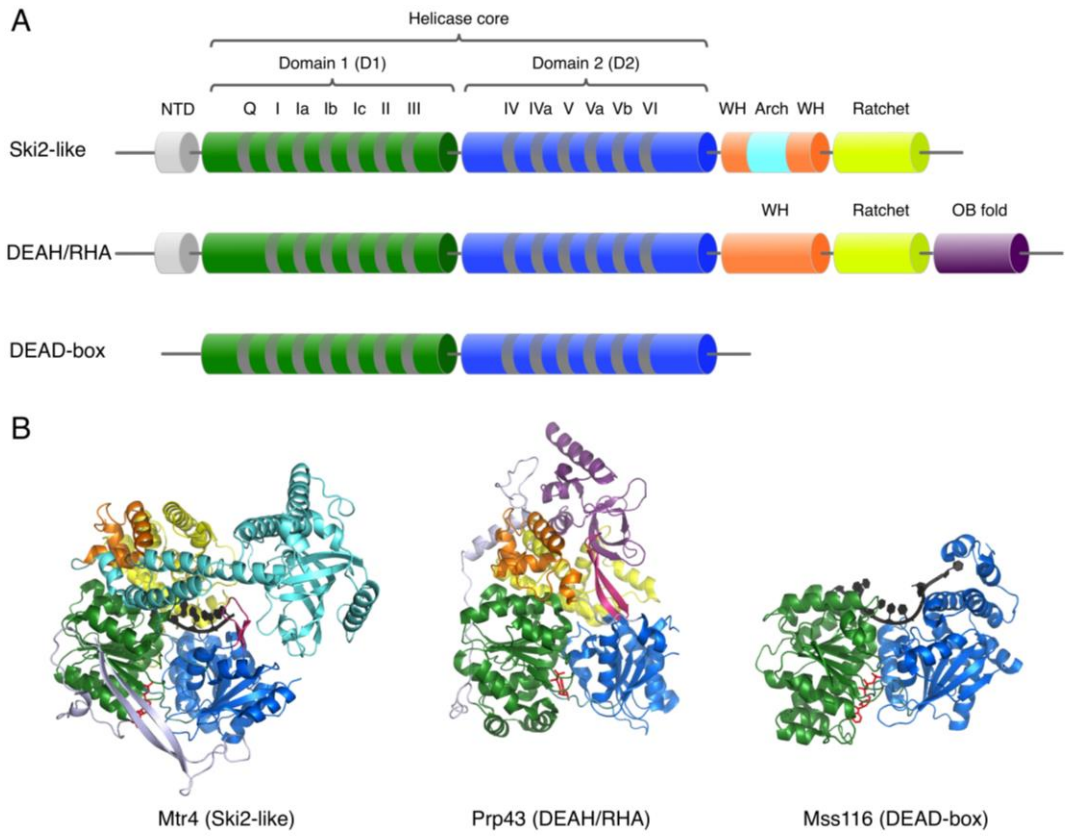


Figure 6.1: Superfamily 2 (SF2) families involved in RNA chaperoning and RNP remodeling.

(*A*) Arrangement of conserved structural domains. Conserved motifs of the helicase core are shown in dark gray. The domain arrangements of DEAH/RHA and Ski2-like families shown are based on Mtr4 and Prp43 structures, respectively (domains not to scale). The winged helix (WH) and ratchet domains are conserved in DEAH/RHA and Ski2-like helicases, and the oligonucleotide/oligosaccharide-binding fold (OB-fold) domain is conserved in DEAH/RHA helicases, whereas the arch domain is present in only a subset of Ski2-like helicases. The N-terminal domains (NTD), shown for Mtr4 and Prp43, are variable. Some individual proteins from all three families include other, nonconserved domains that are not shown here. (*B*) Crystal structures of Ski2-like (Mtr4) (241),

DEAH/RHA (Prp43) (243) and DEAD-box (Mss116) (55) helicases. Domains are colored as in panel *A*. The nucleotide (ADP for Mtr4 and Prp43 and AMP-PNP for Mss116) is shown in red, and cocrystallized single-stranded RNA (ssRNA) bound to Mtr4 and Mss116 is shown in black. The β hairpin within D2 in Ski2-like and DEAH/RHA helicases that is thought to function as a pin during RNA unwinding is highlighted in pink.

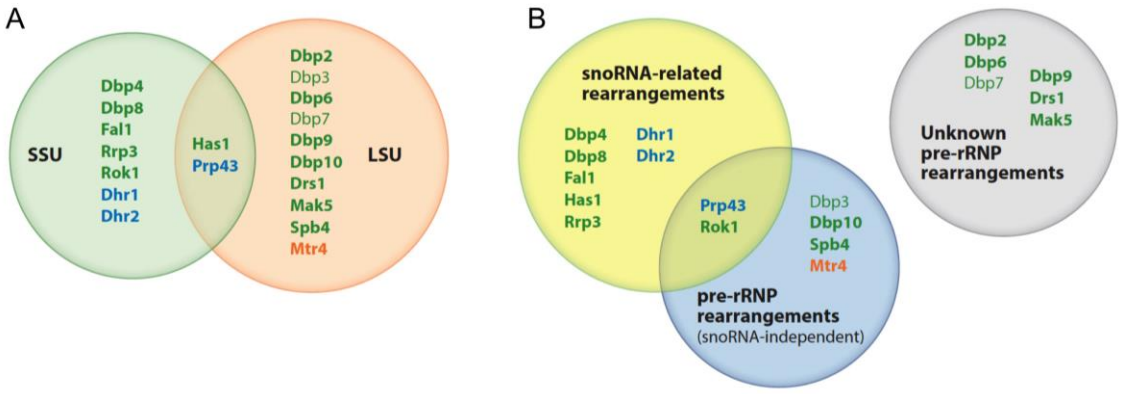


Figure 6.2: Helicases that function in eukaryotic ribosome biogenesis.

(*A*) *Saccharomyces cerevisiae* helicases that function in biogenesis of the small ribosomal subunit (SSU) and large ribosomal subunit (LSU). Essential helicases are shown in bold. DEAD-box, DEAH/RHA, and Ski2-like helicases are shown in green, blue, and orange, respectively. (*B*) Known and proposed roles of RNA helicases in ribosome biogenesis. Small nucleolar RNA (snoRNA)-related rearrangements include promoting binding and dissociation of snoRNAs as well as rearrangements associated with snoRNA-guided cleavage events. Helicases indicated as promoting snoRNA-independent rearrangements facilitate pre-rRNA cleavage at specific snoRNA-independent processing sites.

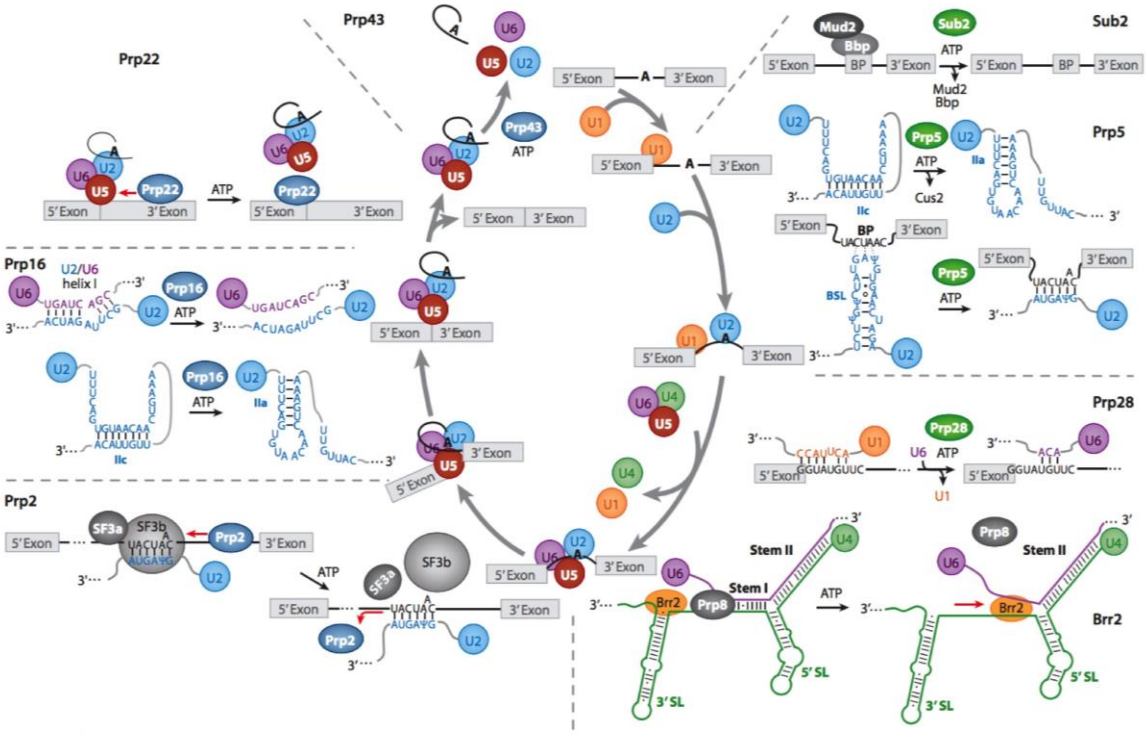


Figure 6.3: Helicase-mediated rearrangements in splicing.

DEAD-box proteins are shown in green, DEAH/RHA proteins in blue, and Ski2-like proteins in orange. Abbreviations: BP, branchpoint (shown as an A in the cycle); Bbp, branchpoint-binding protein (also known as Msl5); BSL, branchpoint stemloop; SL, stemloop. The direction of translocation by Prp2, Prp22, and Brr2 is indicated by red arrows.

Appendix A

Appendix A contains supporting figures for Chapter 2.

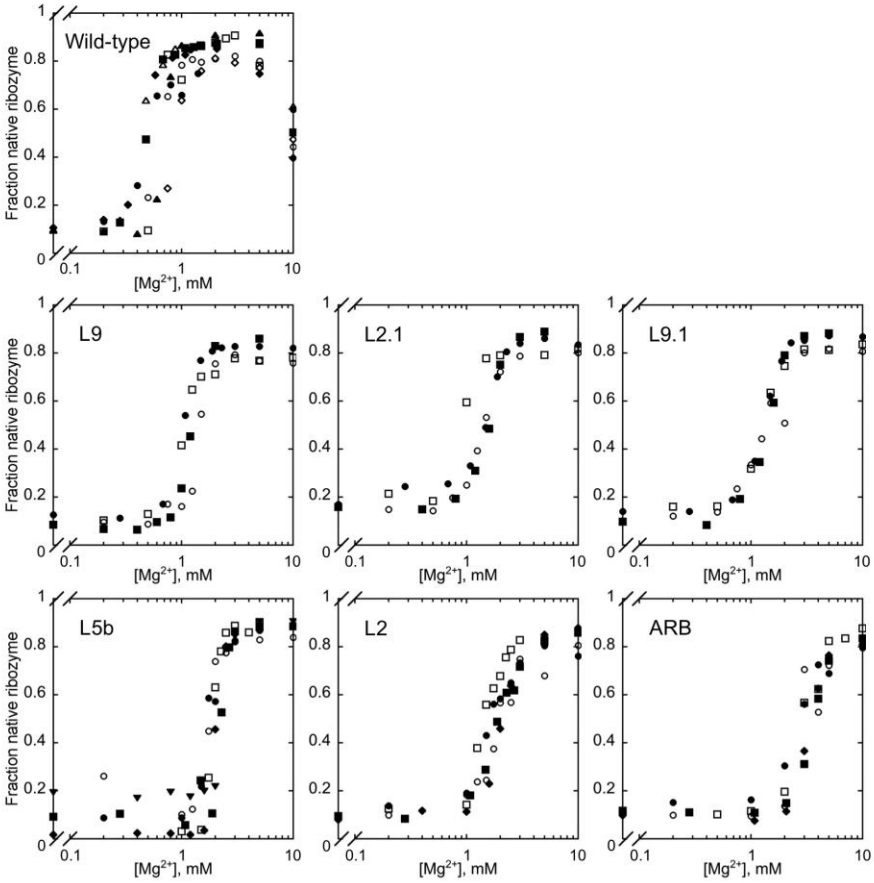


Figure A1: Cumulative isotherms for Mg^{2+}-dependent ribozyme folding.

The ribozymes were folded for ~24 h at 25 ˚C, in the presence of 50 mM Na-MOPS, pH 7.0, 10% CYT-19 buffer (v/v), and 2 mM ATP•Mg^{2+}. Different symbols indicate data from individual experiments. Note that at 10 mM Mg^{2+}, the wild-type-ribozyme does not reach equilibrium during the ~24 h incubation due to the slow refolding of the long-lived misfolded intermediate, thus reducing the fraction of native ribozyme at this Mg^{2+} concentration (140).

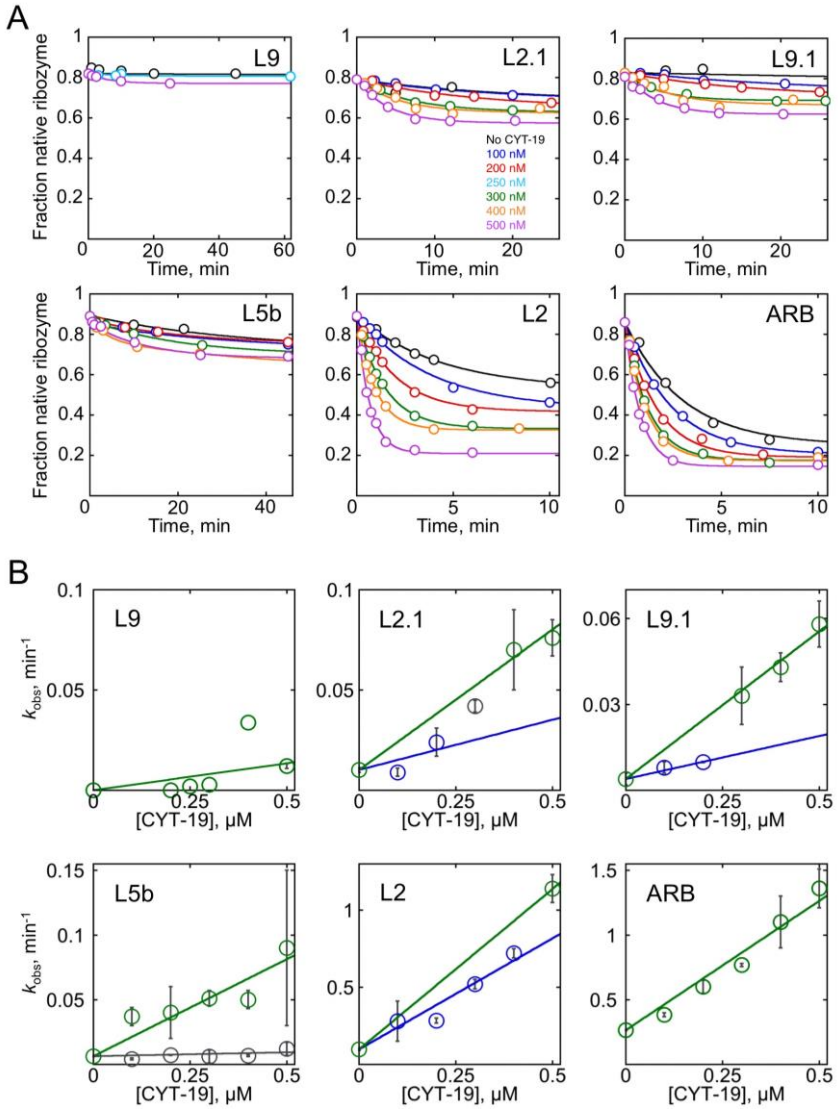


Figure A2: Unfolding of tertiary contact mutant ribozymes by CYT-19 at 2 mM Mg^{2+}, 100 μM ATP•Mg^{2+}.

(*A*) Representative time courses of native ribozyme unfolding in the absence of CYT-19 or with 100–500 nM CYT-19. Colors correspond to different CYT-19 concentrations, as shown in the top middle panel. (*B*) Dependences of unfolding rate constants on CYT-19 concentration. In cases where systematic deviations from linearity were observed, upper

limits (slope of green lines) and lower limits (slope of blue lines) of CYT-19 efficiency were obtained for the range of 0–500 nM CYT-19. For the L5b mutant, the gray data points indicate the rate constants of the slow second phase of unfolding, which are indistinguishable from the rate constant of spontaneous unfolding, indicating loss of CYT-19 activity on long time scales. Error bars indicate standard error from at least two independent determinations.

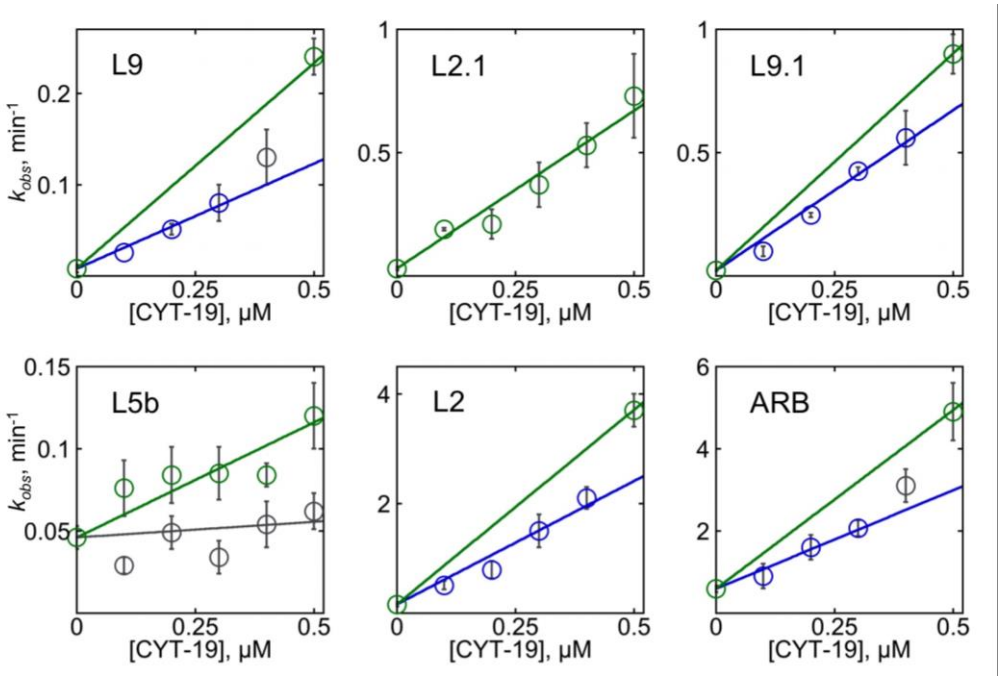


Figure A3: Ribozyme unfolding rate dependences on CYT-19 concentration at 2 mM Mg^{2+}, 2 mM ATP•Mg^{2+}.

Blue and green lines represent lower and upper limits of CYT-19 efficiency for the 0–500 nM range, respectively.

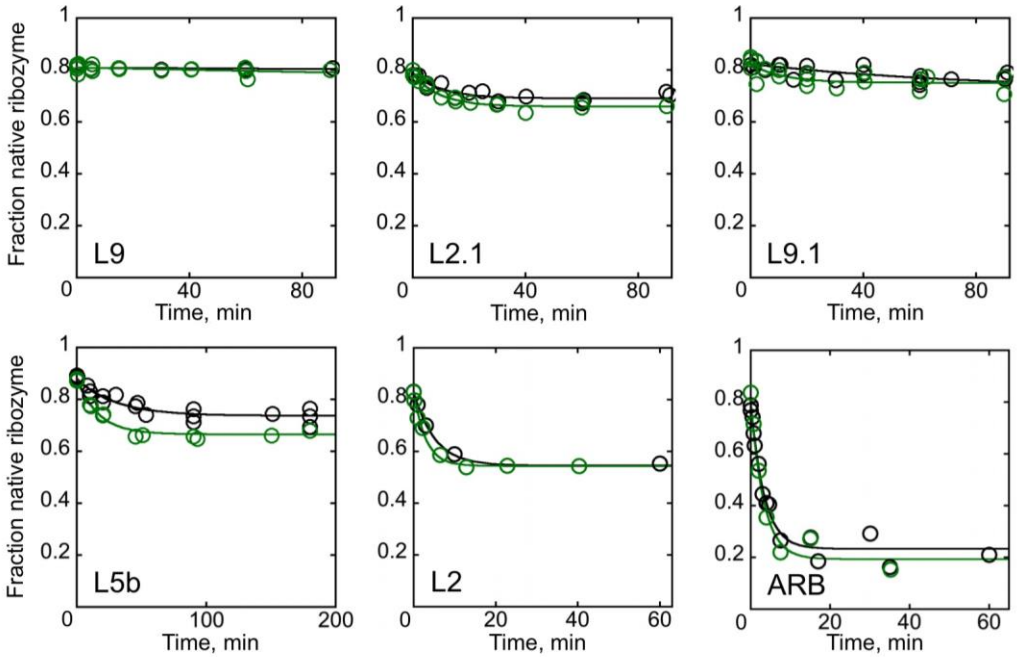


Figure A4: Ribozyme unfolding in the absence of ATP.

Unfolding time courses are shown at 2 mM Mg^{2+} in the absence of CYT-19 (black) or with 500 nM CYT-19 (green). With the exception of the L2 mutant, the plots show cumulative data from two or more independent experiments.

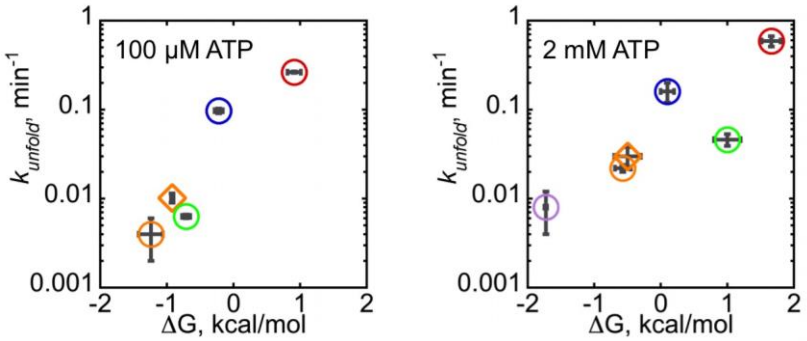


Figure A5: Correlation between native ribozyme stability and the rate constant of spontaneous unfolding at 2 mM Mg^{2+}.

The unfolding rate constants were determined from the progress curves in Figures 2.2*B* and A2*A* as described in *Materials and Methods*. Both the stabilities and the rate constants of unfolding shift slightly from 100 μM to 2 mM ATP•Mg^{2+} as a result of small changes in free Mg^{2+} concentration. Error bars indicate standard error from at least two determinations. The L9 mutant is not included in the first plot because the unfolding amplitude was too small for accurate determination of the rate constant.

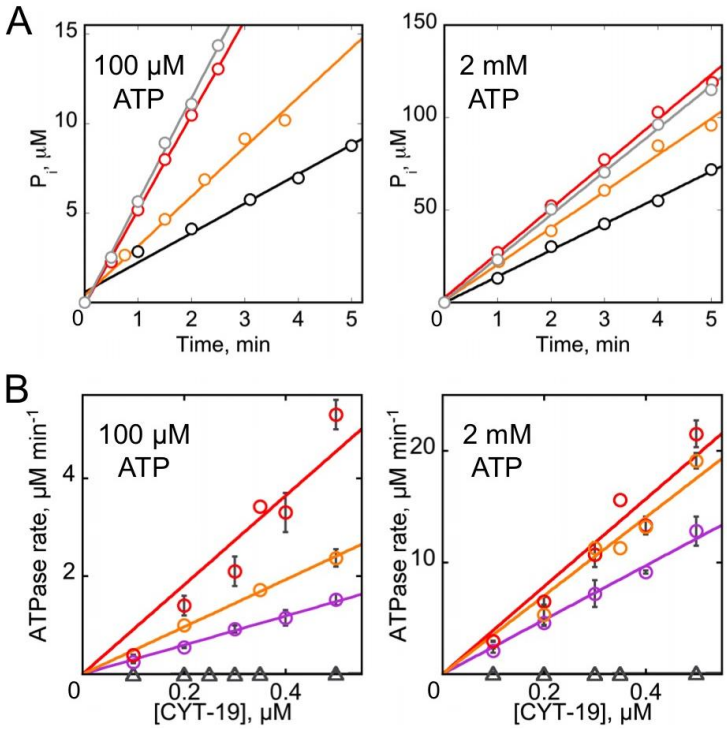


Figure A6: ATPase stimulation by tertiary contact mutants of the *Tetrahymena* ribozyme.

(*A*) Representative time courses of ATP hydrolysis in the presence of wild-type (black), L9.1 (orange), ARB (red) and quintuple mutant (gray) ribozymes (200 nM ribozyme, 500 nM CYT-19, 2 mM Mg^{2+}). Background P_i that was present in the γ-^{32}P-ATP stock (2.5 % in this experiment) has been subtracted from all data points. (*B*) CYT-19 concentration dependence of ATP hydrolysis in the presence of representative ribozyme mutants: L9 (purple), L9.1 (orange), and ARB (red) (200 nM ribozyme, 2 mM Mg^{2+}). Triangles indicate ATP hydrolysis in the absence of RNA.

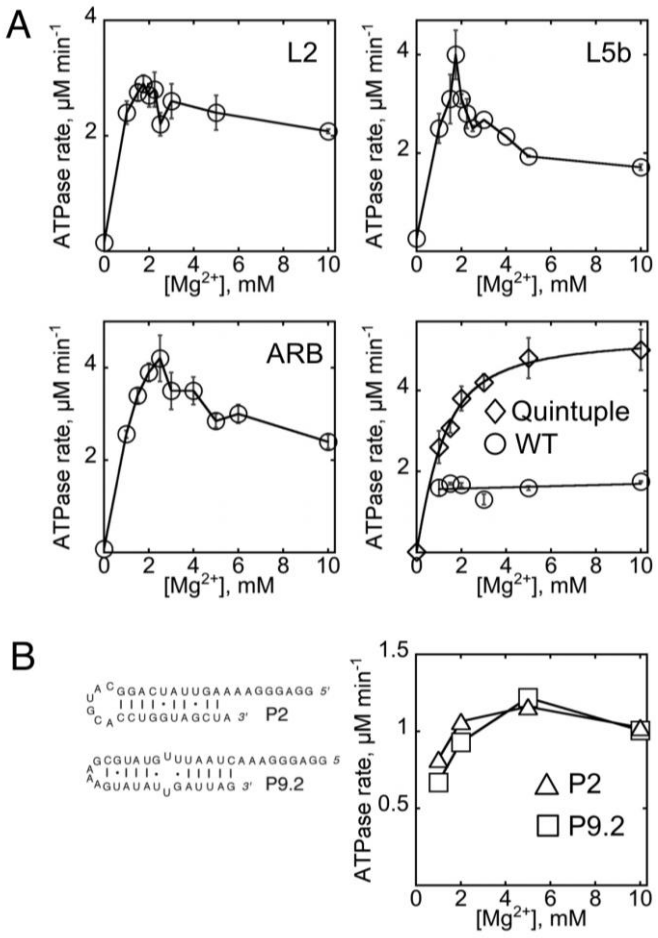


Figure A7: Mg^{2+}-concentration dependence of ATPase stimulation by the wild-type and mutant ribozymes and oligonucleotide helices.

(*A*) The ribozymes were incubated for ~24 h at 0–10 mM Mg^{2+}, and the ATPase stimulation was measured in the presence of 200 nM ribozyme, 500 nM CYT-19 and 100 μM ATP (see *Materials and Methods*). With the exception of the quintuple mutant, the data points are connected only to guide the eye. The quintuple mutant data were fit to a quadratic binding equation, and the fitting parameters were used during normalization of the data for the other mutants to extrapolate the ATPase rates to each Mg^{2+} concentration. The normalization accounted for the tertiary-folding independent increase in ATP

hydrolysis rates with increasing Mg^{2+} concentrations, which was observed both in ribozyme mutants and with oligonucleotide RNAs (181). (*B*) The oligonucleotide helices (200 nM final), corresponding to secondary structure elements in the ribozyme, were incubated for 5 min at 10 mM Mg^{2+} and 25˚C and transferred to the final Mg^{2+} concentrations indicated. ATPase rates were determined with 500 nM CYT-19 and 100 μM ATP as above.

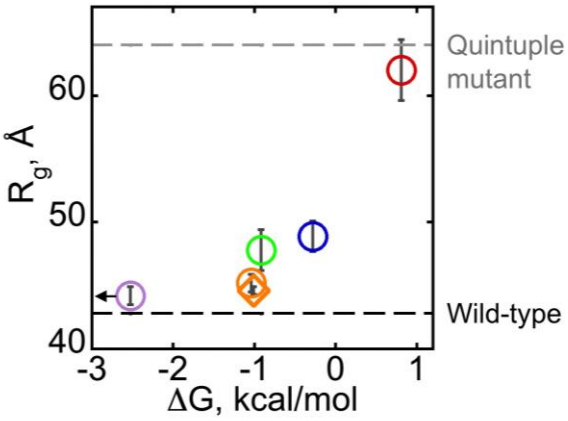


Figure A8: Decreased native ribozyme stability corresponds to lower average compactness of ribozyme mutants at equilibrium (2 mM Mg^{2+}).

The color scheme is as in Figure 2.1, with the diamond indicating the L2.1 mutant to distinguish it from the L9.1 mutant of the P13 tertiary contact. Error bars indicate standard error from multiple determinations. The stabilities on the X-axis were derived from folding equilibria at 2 mM Mg^{2+} in the absence of ATP. These values are the same within error of those measured with 100 μM ATP, see, e.g., Fig. 2*C*.

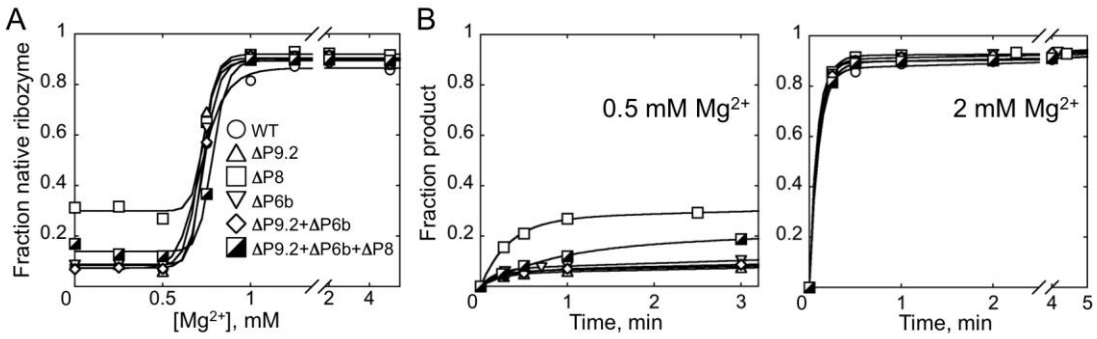


Figure A9: Probing the effects of helix truncations on native ribozyme stability and activity.

(*A*) [Mg^{2+}]-dependences for native folding determined by the discontinuous ribozyme activity assay (see *Materials and Methods*). (*B*) Time courses of substrate cleavage by ribozymes equilibrated below and above the Mg^{2+} folding midpoint (0.5 mM and 2 mM Mg^{2+}, respectively). The observed increase in the fraction of the native ΔP8 and triple mutant ribozymes at 0 mM Mg^{2+} in panel *A* is most likely a result of an increased fraction that rapidly folds to the native state in the folding quench, perhaps from stabilization of the core helix P3 due to a C-G base pair that was introduced at the P8 truncation site (146). While this behavior introduces some uncertainty to the stabilities of these mutants, it does not affect the overall conclusion that any effects on native ribozyme stability are modest and that the presence of unfolded ribozyme is not expected to contribute significantly to ATPase stimulation by these mutants at 2 mM Mg^{2+}.

Appendix B

Appendix B contains supporting data and figures for Chapter 4.

All experiments and analysis for Mss116 were performed by Anna Mallam, except for ATPase assays in Table B5, performed by Inga Jarmoskaite and Serena Chen.

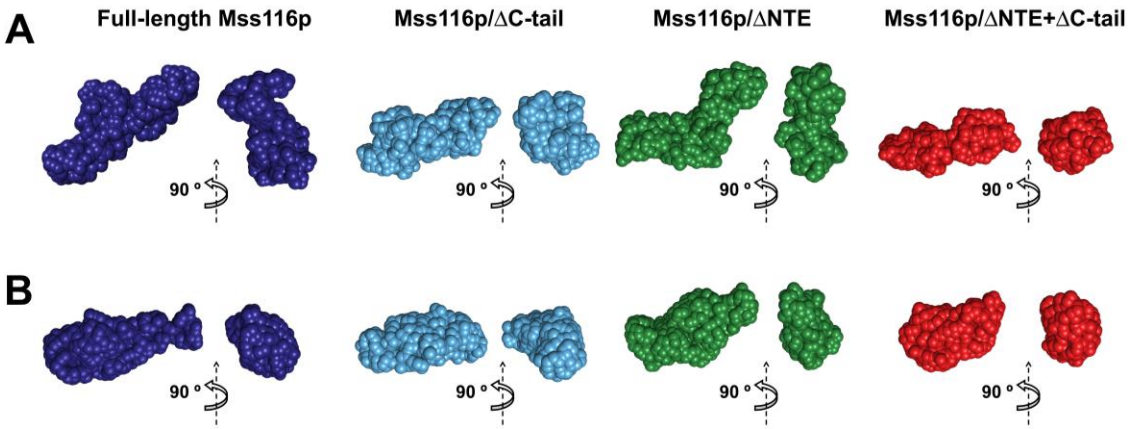


Figure B1: Reconstructions of SAXS data for Mss116p using modeling with chain-like dummy residues.

Low-resolution envelopes calculated from the SAXS data by GASBOR are shown for full-length Mss116p (dark blue), Mss116p/ΔC-tail (light blue), Mss116p/ΔNTE (green) and Mss116p/ΔNTE+ΔC-tail (red) for (*A*) proteins in the absence of ligands and (*B*) proteins bound to ADP-BeF_x and U_{10}-RNA. Views rotated by 90 ° about the vertical axis are shown for each model.

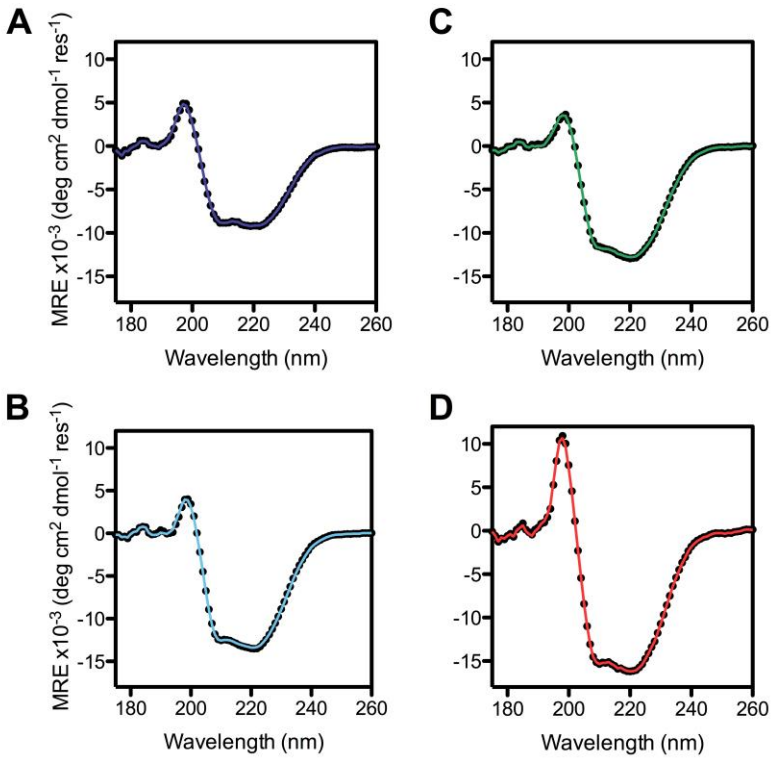


Figure B2: Far ultra-violet circular dichroism spectra of full-length Mss116p and deletion proteins.

Spectra are shown in units of mean residue ellipticity (MRE). Secondary structure content was calculated from the experimental far-UV CD data by the DichroWeb on-line server (210) using the CDSSTR analysis program with the SP175 reference data set (209). The results of this analysis are shown in Table B4. The fit of the theoretical far-UV CD spectra from the analysis to the experimental data (black circles) is shown for (*A*) full-length Mss116p (blue), (*B*) Mss116p/ΔC-tail (light blue), (*C*) Mss116p/ΔNTE (green) and (*D*) Mss116p/ΔNTE+ΔC-tail (red). Normalized-root-mean-square-deviation (NRMSD) values were less than 0.025 for all fits.

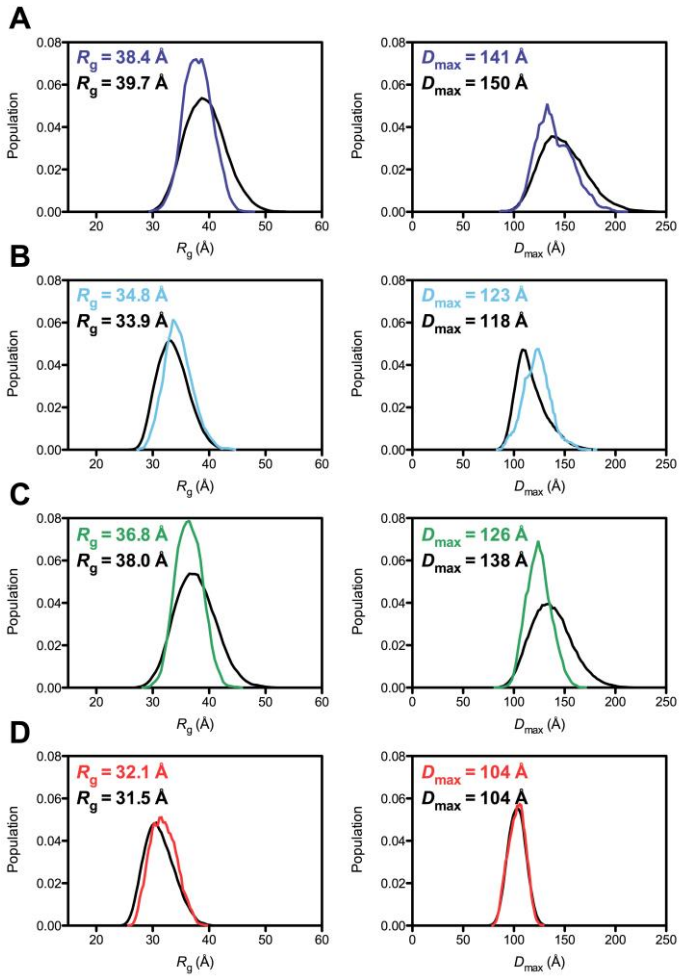


Figure B3: Size and shape distributions of optimized ensembles for full-length Mss116p and deletion proteins.

An optimized ensemble of the 100 models that best describe each scattering profile was selected from a pool of 10,000 random conformers using the EOM. The distributions for the R_g (left panel) and D_{max} (right panel) of the models in the optimized ensembles for (*A*) full-length Mss116p (blue), (*B*) Mss116p/ΔC-tail (light blue), (*C*) Mss116p/ΔNTE (green) and (*D*) Mss116p/ΔNTE+ΔC-tail (red) are shown compared to those of the corresponding random pool (black). The average values for the R_g and D_{max} of the optimized ensembles and the random pool are shown in equivalent colors.

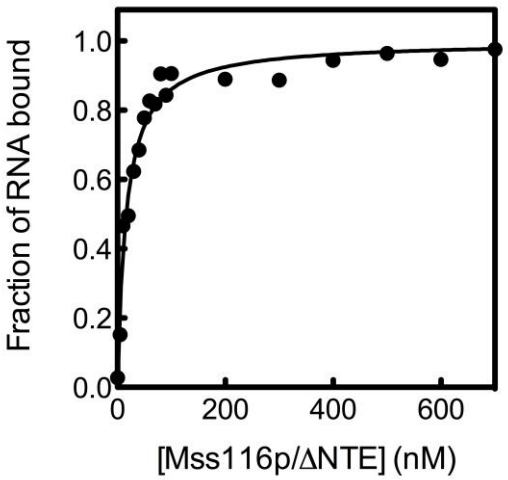


Figure B4: Equilibrium binding of Mss116p to U_{10}-RNA.

Fluorescein-labeled FAM-U_{10}-RNA (20 nM) was incubated with increasing concentrations of Mss116p/ΔNTE in the presence of ADP-BeF_x (0.5 mM) and bovine serum albumin (0.1 mg/ml) in the SAXS buffer of 20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$ for 30 min at 22 °C. The fraction of bound RNA was measured by monitoring the change in fluorescence at 510 nm after excitation at 480 nm. Incubations of longer than 30 min gave no additional change in fluorescence, indicating that this time was sufficient to reach equilibrium. The solid line represents the fit of the data to a one-site binding model to give an upper limit for the value of the dissociation constant, K_d, of ~20 nM.

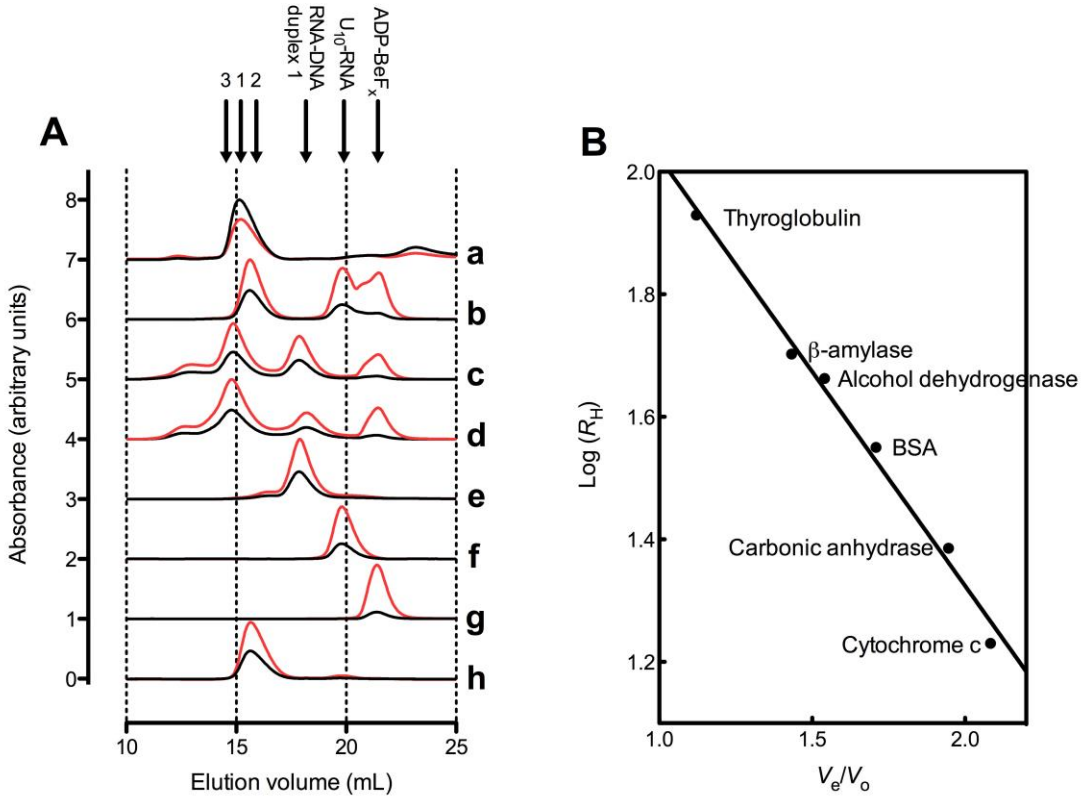


Figure B5: Purification and characterization of Mss116p-nucleic acid complexes using size exclusion chromatography.

SEC was performed using a Superdex 200 10/300 GL column (GE Healthcare) and a BioLogic DuoFlow chromatography system (Bio-Rad) in the SAXS buffer of 20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$. (*A*) SEC data are shown for Mss116p/ΔNTE+ΔC-tail and were measured using absorbance at 280 nm (black) and 260 nm (red). Example elution profiles for (a) free protein in the absence of ligands (elution volume, V_e = 15.1 ml; peak position 1), (b) U_{10}-RNA-protein-ADP-BeF_x complex (V_e = 15.6 ml; peak position 2), (c) RNA-DNA-duplex 1-protein-ADP-BeF_x complex (V_e = 14.8 ml; peak position 3), (d) RNA-DNA-duplex 2-protein-ADP-BeF_x complex (V_e = 14.8 ml; peak position 3), (e) RNA-DNA duplex 1 (V_e = 17.9 ml), (f) U_{10}-

RNA (V_e = 19.8 ml) and (g) ADP-BeF_x (V_e = 21.4 ml). Fractions corresponding to the center of the elution peak of the complexes were collected for analysis by SAXS. The stability of the complexes was confirmed by a reanalysis of the isolated complex by SEC to check for dissociation, an example of which is shown in (h) for U_{10}-RNA-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x (V_e = 15.6 ml; peak position 2). The absence of peaks corresponding to either U_{10}-RNA or ADP-BeF_x indicated that the complex remains intact. The ratio of A_{260}/A_{280}, which is approximately 0.5 for free protein and 2 for protein-nucleic acid complexes, was also used as an indicator of complex formation. (*B*) A calibration curve for the SEC column used in these experiments showing the logarithm of the hydrodynamic radius, R_H, as a function of the relative elution volume, V_e/V_o, where V_o is the void volume of the column. The hydrodynamic radii for the free proteins and complexes of full-length Mss116p and deletion proteins, which are shown in Table B2, were determined by a comparison of their relative elution volumes to those of standard proteins with known hydrodynamic radii (Equation S1).

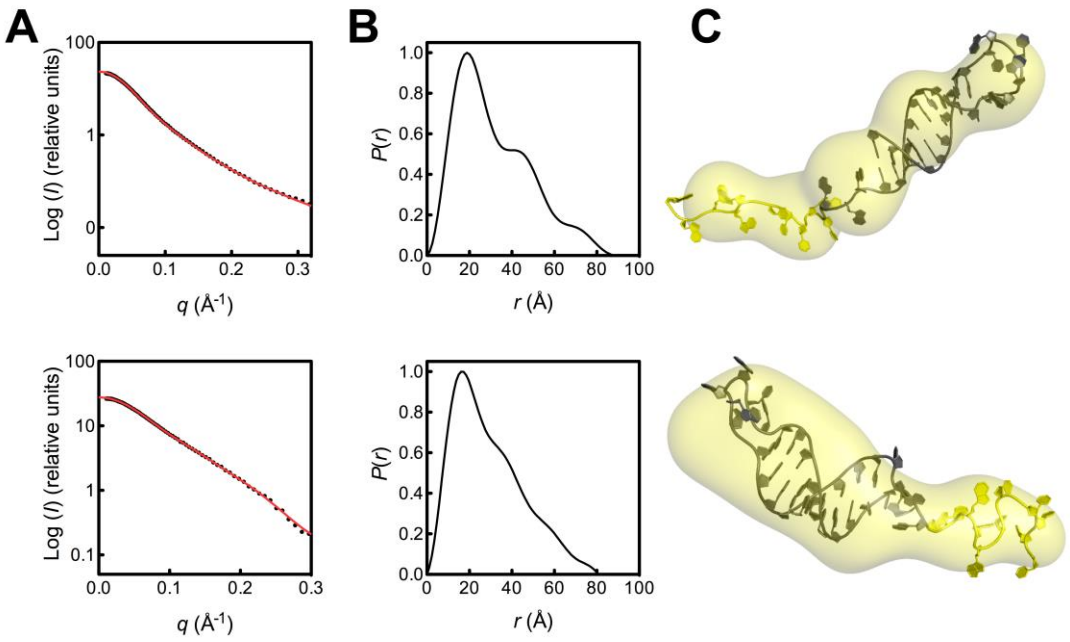


Figure B6: SAXS data for chimeric nucleic acid substrates.

(*A*) Scattering profiles, (*B*) normalized distance distribution functions, and (*C*) low-resolution DAMMIN reconstructions for RNA-DNA duplex 1 (top panels) and RNA-DNA duplex 2 (bottom panels). The red solid line in (*A*) represents the fit of the simulated scattering curve for the DAMMIN envelopes shown in (*C*) to the experimental data (χ = 1.02 and 0.43 for RNA-DNA duplex 1 and RNA-DNA duplex 2, respectively). Atomic models for each substrate, colored according to the sequences in Figure 4.1*C*, were generated using the nucleic acid structure prediction program RNABuilder (373, 374), and were manually placed inside the low-resolution SAXS envelopes.

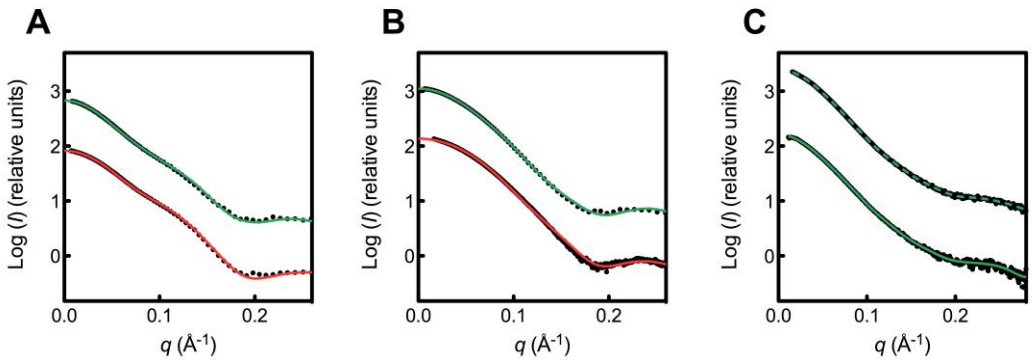


Figure B7: Scattering data for CYT-19.

Scattering profiles were measured for proteins (*A*) in the absence of ligands, (*B*) bound to ADP-BeF_x and U_{10}-RNA, and (*C*) bound to ADP-BeF_x and larger nucleic acid substrates. SAXS data are shown as the logarithm of the scattering intensity, I (black dots), as a function of the momentum transfer, $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength. In (*A*) and (*B*), the solid curves represent the expected scattering profiles of the corresponding BUNCH atomic models or, in the case of U_{10}-RNA-CYT-19/ΔC-tail-ADP-BeF_x, the homology model of the closed-state CYT-19. In (*C*), the fits of the multiphase *ab-initio* model obtained by MONSA for complexes that contain RNA-DNA duplex 1 (solid line) and RNA-DNA duplex 2 (dashed line) are displayed. In panels (*A*-*C*), fits shown in green are for full-length CYT-19 and those shown in red are for CYT-19/ΔC-tail.

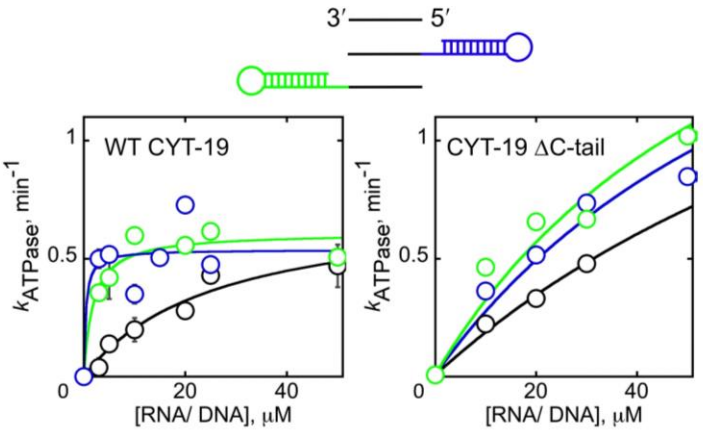


Figure B8: Duplex extensions enhance the efficiency of ATPase stimulation more strongly for wild-type CYT-19 than for protein lacking the C-tail.

ATPase assays were performed with wild-type CYT-19 (left) or CYT-19/ΔC-tail (right). The RNA-dependent ATPase activity was measured in the presence of duplex 1 (blue), duplex 2 (green) or an 11 nt RNA oligonucleotide corresponding to the single-stranded RNA region of duplex 1 and duplex 2 constructs (CCCUCUAAAAA). The measurements were performed at 100 μM ATP, 5 mM $MgCl_2$, 100 mM KCl, 50 mM Na-MOPS (pH 7.0), 25 ˚C, using 500 nM WT CYT-19 or 250 nM CYT-19/ΔC-tail. The ATPase assays were performed with trace ^{32}P-γ-ATP and analyzed by thin-layer chromatography.

Sample	Molecular weight determined from sequence /kDa	Abs $I(0)/c^a$ /cm² mg⁻¹	Molecular weight determined from SAXS data[b] /kDa	R_g (Guinier analysis) /Å	R_g (GNOM analysis) /Å	D_{max} /Å
Free protein and nucleic acid						
Full-length Mss116p	72.5	69.1 ± 2.4	73.4 ± 3.0	38.0 ± 0.2	38.9 ± 0.1	135 ± 2
Mss116p/ΔC-tail	64.4	55.9 ± 3.0	59.4 ± 3.5	34.2 ± 0.3	34.7 ± 0.2	120 ± 2
Mss116p/ΔNTE	66.1	56.2 ± 1.5	59.7 ± 2.1	36.9 ± 0.2	38.1 ± 0.3	132 ± 2
Mss116p/ΔNTE+ΔC-tail	58.1	50.3 ± 2.9	53.4 ± 3.4	32.8 ± 0.1	33.7 ± 0.1	115 ± 2
Full-length CYT-19	63.9	-	68*	36.6 ± 0.1	36.9 ± 0.1	123 ± 2
CYT-19/ΔC-tail	58.2	-	59*	32.6 ± 0.1	33.2 ± 0.2	115 ± 2
RNA-DNA duplex 1	12.4	62.7 ± 0.3	13.4 ± 1.0	24.6 ± 0.2	24.7 ± 0.1	85 ± 2
RNA-DNA duplex 2	12.4	60.8 ± 0.5	13.0 ± 1.0	23.1 ± 0.3	23.6 ± 0.2	80 ± 2
Cytochrome c (horse heart)	12.3	11.6 ± 0.3	-	15.2 ± 0.2	-	-
U_{10}-RNA	3.0	14.0 ± 1.1	-	12.6 ± 0.8	-	-
Protein-nucleic acid complexes						
U_{10}-RNA-full-length Mss116p-ADP-BeF_x	75.9	77.6 ± 1.5	-	33.8 ± 0.1	34.0 ± 0.2	117 ± 2
U_{10}-RNA-Mss116p/ΔC-tail-ADP-BeF_x	67.9	66.7 ± 2.4	-	29.9 ± 0.1	29.8 ± 0.3	102 ± 2
U_{10}-RNA-Mss116p/ΔNTE-ADP-BeF_x	69.6	67.2 ± 3.3	-	30.3 ± 0.1	31.8 ± 0.1	112 ± 2
U_{10}-RNA-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	61.5	60.5 ± 1.8	-	26.3 ± 0.2	25.9 ± 0.1	82 ± 2
U_{10}-RNA-full-length CYT-19-ADP-BeF_x	67.4	-	-	30.4 ± 0.1	31.0 ± 0.1	109 ± 2
U_{10}-RNA-CYT-19/ΔC-tail-ADP-BeF_x	61.7	-	-	27.4 ± 0.1	27.5 ± 0.1	95 ± 2
RNA-DNA duplex 1-Mss116p/ΔNTE-ADP-BeF_x	79.0	92.9 ± 4.9	-	36.0 ± 0.4	36.1 ± 0.1	120 ± 2
RNA-DNA duplex 1-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	70.9	91.5 ± 5.3	-	33.2 ± 0.2	34.0 ± 0.1	118 ± 2
RNA-DNA duplex 2-Mss116p/ΔNTE-ADP-BeF_x	79.0	83.7 ± 1.7	-	36.1 ± 0.2	37.6 ± 0.1	135 ± 2
RNA-DNA duplex 2-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	70.9	81.3 ± 3.4	-	33.1 ± 0.1	34.3 ± 0.1	119 ± 2
RNA-DNA duplex 1-full-length CYT-19-ADP-BeF_x	76.8	-	-	36.2 ± 0.1	36.6 ± 0.1	126 ± 2
RNA-DNA duplex 2-full-length CYT-19-ADP-BeF_x	76.8	-	-	35.2 ± 0.1	35.2 ± 0.1	120 ± 2

Table B1: Structural parameters for DEAD-box proteins calculated from SAXS data.

[a] The forward scattering intensity, $I(0)$, normalized against protein concentration, c, measured in mg/ml. Error estimates are the standard errors calculated from linear regression.

[b] Particle molecular weights of samples with homogeneous components were estimated using Equation S3 and standards of cytochrome c and U_{10}-RNA for protein and nucleic acid samples, respectively. The errors were estimated by propagation of the error calculated for $I(0)/c$.

* Molecular weights for CYT-19 samples were determined from SAXS using the calculated Porod volume (375).

The error estimates for R_g values calculated by Guinier and GNOM analyses are the standard error for linear regression and the standard deviation evaluated by GNOM, respectively. Errors for the maximum particle dimension, D_{max}, are the estimated range for the optimum solution.

Sample	R_H [a] /Å	R_g [b] /Å
Free protein		
Full-length Mss116p	35.0	38.0 ± 0.2
Mss116p/ΔC-tail	32.0	34.2 ± 0.3
Mss116p/ΔNTE	34.3	36.9 ± 0.2
Mss116p/ΔNTE+ΔC-tail	31.1	32.8 ± 0.1
Protein-nucleic acid complexes		
U_{10}-RNA-full-length Mss116p-ADP-BeF_x	33.2	33.8 ± 0.1
U_{10}-RNA-Mss116p/ΔC-tail-ADP-BeF_x	30.7	29.9 ± 0.1
U_{10}-RNA-Mss116p/ΔNTE-ADP-BeF_x	31.8	30.3 ± 0.1
U_{10}-RNA-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	28.5	26.3 ± 0.2
RNA-DNA duplex 1-Mss116p/ΔNTE-ADP-BeF_x	35.4	36.0 ± 0.4
RNA-DNA duplex 1-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	32.8	33.2 ± 0.2
RNA-DNA duplex 2-Mss116p/ΔNTE-ADP-BeF_x	36.0	36.1 ± 0.2
RNA-DNA duplex 2-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	33.2	33.1 ± 0.1

Table B2: Hydrodynamic radii and R_g values for Mss116p SAXS samples

[a] The hydrodynamic radius, R_H, was calculated using SEC (Equation S1 and Figure B5). The error in the value for R_H is estimated to be approximately ±5 %, which is based upon the standard error in the calibration curve shown in Figure B5*B*.

[b] Radii of gyration are shown here for comparison and are taken from Table B1.

Sample	Modeling program					
	DAMMIN or MONSA Single or multi-phase *ab-initio* bead modeling, respectively		GASBOR *Ab-initio* modeling with chain-like dummy residues		BUNCH Rigid-body modeling	
	χ	NSD	χ	NSD	χ	NSD
Free protein and RNA						
Full-length Mss116p	0.98	0.82 ± 0.04	1.09	1.44 ± 0.06	0.83	1.53 ± 0.06
Mss116p/ΔC-tail	1.67	0.92 ± 0.05	0.72	1.26 ± 0.02	0.67	0.96 ± 0.16
Mss116p/ΔNTE	0.45	0.81 ± 0.07	0.52	1.09 ± 0.03	1.03	0.96 ± 0.09
Mss116p/ΔNTE+ΔC-tail	0.94	0.67 ± 0.05	1.00	1.15 ± 0.02	2.13	0.62 ± 0.03
Full-length CYT-19	1.06	0.72 ± 0.04	-	-	1.12	0.69 ± 0.18
CYT-19/ΔC-tail	0.47	0.71 ± 0.19	-	-	1.42	0.77 ± 0.13
RNA-DNA duplex 1	1.01	0.71 ± 0.04	-	-	-	-
RNA-DNA duplex 2	0.43	0.65 ± 0.01	-	-	-	-
U_{10} RNA-full-length Mss116p-ADP-BeF_x	0.54	0.82 ± 0.15	0.59	1.25 ± 0.05	1.19	1.04 ± 0.20
U_{10} RNA-Mss116p/ΔC-tail-ADP-BeF_x	0.38	0.72 ± 0.13	0.52	1.09 ± 0.03	1.43	0.47 ± 0.05
U_{10} RNA-Mss116p/ΔNTE-ADP-BeF_x	1.32	0.70 ± 0.03	2.23	1.20 ± 0.06	1.86	0.46 ± 0.18
U_{10} RNA-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	0.55	0.92 ± 0.03	0.66	0.98 ± 0.01	1.66	-
U_{10}-RNA-full-length CYT-19-ADP-BeF_x	1.15	0.76 ± 0.12	-	-	0.96	0.27 ±0.01
U_{10}-RNA-CYT-19/ΔC-tail-ADP-BeF_x	2.03	0.69 ± 0.10	-	-	-	-
RNA-DNA duplex 1-Mss116p/ΔNTE-ADP-BeF_x*	0.67	0.64 ± 0.02	-	-	-	-
RNA-DNA duplex 1-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x*	1.24	0.46 ± 0.02	-	-	-	-
RNA-DNA duplex 2-Mss116p/ΔNTE-ADP-BeF_x*	1.48	0.63 ± 0.09	-	-	-	-
RNA-DNA duplex 2-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x*	1.44	0.47 ± 0.03	-	-	-	-
RNA-DNA duplex 1-full-length CYT-19-ADP-BeF_x*	2.24	0.63 ± 0.04	-	-	-	-
RNA-DNA duplex 2-full-length CYT-19-ADP-BeF_x*	2.12	0.72 ± 0.05	-	-	-	-

Table B3: Parameters for the *ab-initio* and rigid-body models constructed from SAXS data

For *ab-initio* modeling, DAMMIN was used for SAXS reconstructions of homogeneous samples, GASBOR for reconstructions of homogeneous protein samples and MONSA for reconstructions of complexes that contain significant amounts of nucleic acid (indicated by an *). U_{10}-RNA-protein-ADP-BeF_x complexes (which comprise less than 5 % nucleic acid) were assumed to be single-component systems of uniform electron densities. Reconstructions from individual runs were compared and averaged using the DAMAVER suite of programs (216) to obtain a value for the average normalized spatial discrepancy (NSD) and an estimate of the standard-deviation error between models. The NSD gives an indication of the stability and uniqueness of the shape reconstructions, and values close to unity indicate that the individual solutions are very similar. The final DAMMIN models represent a refined averaged structure, while final GASBOR and MONSA models represent the most typical structure with the lowest NSD value compared to the others. Values for χ are for the fit of the calculated scattering profile of the final model to the experimental data.

For rigid-body modeling, the final BUNCH models shown in Figure 4.2, Figure 4.3 and Figure 4.5 represent the most typical structures from multiple runs that have the lowest average NSD value when compared to all the others. The value for χ represents the difference between the calculated scattering profile of the final model and the experimental data. In the case of the U_{10}-RNA-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x complex where the crystal structure is already known, the χ-value is the difference between the predicted scattering profile of the crystal structure calculated using CRYSOL (220) and the experimental scattering data of the complex. This low χ value is a good indication that the complex formed in solution resembles that seen in the crystal structure of Mss116p. In addition, the R_g value calculated by CRYSOL for the crystal structure is

25.8 Å, which is in good agreement with the value of 26.3 Å calculated from the SAXS data for the equivalent complex by Guinier analysis (Table B1).

Construct	Residues	% helix	% sheet	% turns and unordered
Full-length Mss116p	37-664	33 (36)	14 (14)	53 (50)
Mss116p/ΔC-tail	37-597	43 (41)	13 (15)	44 (44)
Mss116pΔNTE	88-664	41 (40)	14 (15)	45 (45)
Mss116pΔNTE+ΔC-tail	88-597	51 (45)	12 (17)	37 (38)
Crystal structure	88-596	45	17	38

Table B4: Secondary structure analysis of far ultra-violet circular dichroism spectra for Mss116p

The far-UV CD spectra shown in Figure B2 were analyzed using the CDSSTR algorithm with the SP175 data set implemented by the DichroWeb on-line server for deconvolution of circular dichroism spectra (209, 210). Normalized-root-mean-square-deviation (NRMSD) values between the calculated and experimental CD data were less than 0.025 in all cases. Protein concentrations ranged from 1.1-2.5 mg/ml and far-UV CD data from 175-260 nm were used in the analysis. The secondary structure content of the crystal structure of Mss116p (PDB code 3I5X) was computed using the PROMOTIF program (376). The numbers in brackets indicate the expected secondary-structure content assuming that residues 37-87 (NTE) and 597-664 (C-terminal basic tail) are unstructured, and residues 88-596 adopt the same secondary structure as seen in the crystal structure. The good agreement between the % of turns and unordered regions calculated by analysis of the far-UV CD spectra and that expected if the NTE and C-tail extensions are assumed to be disordered is consistent with these extensions existing as unstructured polypeptide chains. Conditions: 20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% (v/v) glycerol, 1 mM DTT, 5 mM $MgCl_2$ at 25 °C.

	k_{cat}/K_m enhancement relative to ssRNA	
	Duplex 1	Duplex 2
WT CYT-19	8x	4x
CYT-19/ΔC-tail	2x	2x
WT Mss116	8x	12x
Mss116/ΔC-tail	3x	3x

Table B5: Duplex extensions provide greater enhancement of RNA-dependent ATPase activity in wild-type than ΔC-tail proteins.

k_{cat}/K_m values were determined from linear fits to the subsaturating region of the oligonucleotide concentration dependences of ATPase activity. The values are reported relative to the k_{cat}/K_m value for ssRNA without duplex extensions, corresponding to the ssRNA portion of duplex 1 and duplex 2 (Figure 4.1). ATPase assays with CYT-19 were performed at 100 μM ATP, 5 mM $MgCl_2$, 100 mM KCl, 50 mM Na-MOPS (pH 7.0), 25 ˚C, using 500 nM WT CYT-19 or 250 nM CYT-19/ΔC-tail protein. For Mss116, the conditions were adjusted to capture the subsaturating region for all oligonucleotides and were as follows: 500 nM ATP, 5 mM $MgCl_2$, 200 mM KCl, 50 mM Na-MOPS (pH 7.0), and 200 nM Mss116 (WT or ΔC-tail). The ATPase assays were performed with trace ^{32}P-γ-ATP and analyzed by thin-layer chromatography.

Appendix C

Calculation of difference in scattering contrast between ribozyme and CYT-19 (based on data in Table 5.1)

Forward scattering intensity: $I_0=\kappa c(\Delta\rho)^2(MW)^2$

κ: proportionality constant
c: concentration
Δρ: scattering contrast
MW: molecular weight

$$\frac{I_{0,\ RNA}}{I_{0,\ CYT\text{-}19}} = \frac{\kappa c_{RNA}(\Delta\rho_{RNA})^2(MW_{RNA})^2}{\kappa c_{CYT\text{-}19}(\Delta\rho_{CYT\text{-}19})^2(MW_{CYT\text{-}19})^2} = \frac{c_{RNA}(\Delta\rho_{RNA})^2(2\times MW_{CYT\text{-}19})^2}{4\times c_{RNA}(\Delta\rho_{CYT\text{-}19})^2(MW_{CYT\text{-}19})^2} = \frac{(\Delta\rho_{RNA})^2\times 2^2}{4\times(\Delta\rho_{CYT\text{-}19})^2}$$

$$\frac{I_{0,\ RNA}}{I_{0,\ CYT\text{-}19}} = \frac{1}{0.13} = 7.7 \quad \rightarrow \quad \frac{\Delta\rho_{RNA}}{\Delta\rho_{CYT\text{-}19}} = \sqrt{7.7} = 2.8$$

Derivation of the enhancement in forward scattering intensity for an RNA-protein complex vs. a non-interacting mixture

$$I_0=\kappa c(\Delta\rho)^2(MW)^2$$

Non-interacting mixture:

$$I_{0,\ mixture}=\kappa((c_{RNA}(\Delta\rho_{RNA})^2(MW_{RNA})^2)+\underbrace{(c_{CYT\text{-}19}(\Delta\rho_{CYT\text{-}19})^2(MW_{CYT\text{-}19})^2)}_{negligible})$$

A complex with 1:1 stoichiometry :

$$I_{0,\ complex}=\kappa c_{complex}(\Delta\rho_{RNA}MW_{RNA}+\Delta\rho_{CYT\text{-}19}MW_{CYT\text{-}19})^2$$

$$I_{0,\ complex}=\kappa c_{complex}((\Delta\rho_{RNA}MW_{RNA})^2+2(\Delta\rho_{RNA}MW_{RNA})\ (\Delta\rho_{CYT\text{-}19}MW_{CYT\text{-}19})+(\Delta\rho_{CYT\text{-}19}MW_{CYT\text{-}19})^2)$$

Enhancement for a complex vs. a non-interacting mixture (assuming that all RNA and CYT-19 are in complex):

$$I_{0,\ complex}-I_{0,\ mixture}= 2\kappa c_{complex}\ (\Delta\rho_{RNA}MW_{RNA})\ (\Delta\rho_{CYT\text{-}19}MW_{CYT\text{-}19})$$

Relative enhancement of scattering by a 1:1 complex vs. ribozyme alone :

$$\frac{2\kappa c_{complex}\ (\Delta\rho_{RNA}MW_{RNA})\ (\Delta\rho_{CYT\text{-}19}MW_{CYT\text{-}19})}{\kappa c_{RNA}\ (\Delta\rho_{RNA}MW_{RNA})\ (\Delta\rho_{RNA}MW_{RNA})} = \frac{2\ (\Delta\rho_{CYT\text{-}19}MW_{CYT\text{-}19})}{(2.8\times\Delta\rho_{CYT\text{-}19})(2\times MW_{CYT\text{-}19})} = 0.36$$

(greater for >1 CYT-19 per ribozyme)

References

1. Linder P (2006) Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res* 34(15):4168-4180.
2. Jankowsky E & Fairman M (2007) RNA helicases--one fold for many functions. *Curr Opin Struct Biol* 17(3):316-324.
3. Linder P & Jankowsky E (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12(8):505-516.
4. Linder P*, et al.* (1989) Birth of the D-E-A-D box. *Nature* 337(6203):121-122.
5. Singleton MR, Dillingham MS, & Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76:23-50.
6. Lohman TM, Tomko EJ, & Wu CG (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol* 9(5):391-401.
7. Rossler OG, Straka A, & Stahl H (2001) Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res* 29(10):2088-2096.
8. Huang Y & Liu ZR (2002) The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. *J Biol Chem* 277(15):12810-12815.
9. Chamot D, Colvin KR, Kujat-Choy SL, & Owttrim GW (2005) RNA structural rearrangement via unwinding and annealing by the cyanobacterial RNA helicase, CrhR. *J Biol Chem* 280(3):2036-2044.
10. Garcia I, Albring MJ, & Uhlenbeck OC (2012) Duplex destabilization by four ribosomal DEAD-box proteins. *Biochemistry* 51(50):10109-10118.
11. Rogers GW, Jr., Richter NJ, & Merrick WC (1999) Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J Biol Chem* 274(18):12236-12244.
12. Rogers GW, Jr., Lima WF, & Merrick WC (2001) Further characterization of the helicase activity of eIF4A. Substrate specificity. *J Biol Chem* 276(16):12598-12608.
13. Chen Y*, et al.* (2008) DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc Natl Acad Sci U S A* 105(51):20203-20208.
14. Bizebard T, Ferlenghi I, Iost I, & Dreyfus M (2004) Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* 43(24):7857-7866.
15. Iost I & Dreyfus M (2006) DEAD-box RNA helicases in Escherichia coli. *Nucleic Acids Res* 34(15):4189-4197.

16. Grohman JK, *et al.* (2007) Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry* 46(11):3013-3022.

17. Liu F, Putnam A, & Jankowsky E (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A* 105(51):20209-20214.

18. Fairman-Williams ME, Guenther UP, & Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol*.

19. Yang Q, Del Campo M, Lambowitz AM, & Jankowsky E (2007) DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell* 28(2):253-263.

20. Rozen F, *et al.* (1990) Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol Cell Biol* 10(3):1134-1144.

21. Halls C, *et al.* (2007) Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol* 365(3):835-855.

22. Tijerina P, Bhaskaran H, & Russell R (2006) Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *Proc Natl Acad Sci U S A* 103(45):16698-16703.

23. Yang Q & Jankowsky E (2006) The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* 13(11):981-986.

24. Fuller-Pace FV, Nicol SM, Reid AD, & Lane DP (1993) DbpA: a DEAD box protein specifically activated by 23s rRNA. *EMBO J* 12(9):3619-3626.

25. Tsu CA, Kossen K, & Uhlenbeck OC (2001) The Escherichia coli DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA. *RNA* 7(5):702-709.

26. Diges CM & Uhlenbeck OC (2001) Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J* 20(19):5503-5512.

27. Karginov FV, Caruthers JM, Hu Y, McKay DB, & Uhlenbeck OC (2005) YxiN is a modular protein combining a DEx(D/H) core and a specific RNA-binding domain. *J Biol Chem* 280(42):35499-35505.

28. Wang S, *et al.* (2006) The domain of the Bacillus subtilis DEAD-box helicase YxiN that is responsible for specific binding of 23S rRNA has an RNA recognition motif fold. *RNA* 12(6):959-967.

29. Wang S, Overgaard MT, Hu Y, & McKay DB (2008) The Bacillus subtilis RNA helicase YxiN is distended in solution. *Biophys J* 94(1):L01-03.

30. Linden MH, Hartmann RK, & Klostermeier D (2008) The putative RNase P motif in the DEAD box helicase Hera is dispensable for efficient interaction with RNA and helicase activity. *Nucleic Acids Res* 36(18):5800-5811.

31. Rudolph MG & Klostermeier D (2009) The Thermus thermophilus DEAD box helicase Hera contains a modified RNA recognition motif domain loosely connected to the helicase core. *RNA* 15(11):1993-2001.

32. Steimer L*, et al.* (2013) Recognition of two distinct elements in the RNA substrate by the RNA-binding domain of the T. thermophilus DEAD box helicase Hera. *Nucleic Acids Res* 41(12):6259-6272.

33. Mohr G*, et al.* (2008) Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. *J Mol Biol* 375(5):1344-1364.

34. Mallam AL*, et al.* (2011) Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *Proc Natl Acad Sci U S A* 108(30):12254-12259.

35. Klostermeier D & Rudolph MG (2009) A novel dimerization motif in the C-terminal domain of the Thermus thermophilus DEAD box helicase Hera confers substantial flexibility. *Nucleic Acids Res* 37(2):421-430.

36. Hilbert M, Karow AR, & Klostermeier D (2009) The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol Chem* 390(12):1237-1250.

37. Klostermeier D (2013) Rearranging RNA structures at 75°C? toward the molecular mechanism and physiological function of the thermus thermophilus DEAD-box helicase hera. *Biopolymers* 99(12):1137-1146.

38. Oberer M, Marintchev A, & Wagner G (2005) Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev* 19(18):2212-2223.

39. Marintchev A*, et al.* (2009) Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* 136(3):447-460.

40. Andreou AZ & Klostermeier D (2013) The DEAD-box helicase eIF4A: paradigm or the odd one out? *RNA Biol* 10(1):19-32.

41. Schutz P*, et al.* (2008) Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* 105(28):9564-9569.

42. Sun Y*, et al.* (2012) The eukaryotic initiation factor eIF4H facilitates loop-binding, repetitive RNA unwinding by the eIF4A DEAD-box helicase. *Nucleic Acids Res*.

43. Noble KN, *et al.* (2011) The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes Dev* 25(10):1065-1077.

44. Montpetit B, *et al.* (2011) A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* 472(7342):238-242.

45. Buchwald G, *et al.* (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc Natl Acad Sci U S A* 107(22):10050-10055.

46. Barbosa I, *et al.* (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* 19(10):983-990.

47. Bono F, Ebert J, Lorentzen E, & Conti E (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126(4):713-725.

48. Andersen CB, *et al.* (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313(5795):1968-1972.

49. Tritschler F, *et al.* (2009) Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. *Mol Cell* 33(5):661-668.

50. Caruthers JM, Johnson ER, & McKay DB (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A* 97(24):13080-13085.

51. Story RM, Li H, & Abelson JN (2001) Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proc Natl Acad Sci U S A* 98(4):1465-1470.

52. Theissen B, Karow AR, Kohler J, Gubaev A, & Klostermeier D (2008) Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc Natl Acad Sci U S A* 105(2):548-553.

53. Shi H, Cordin O, Minder CM, Linder P, & Xu RM (2004) Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc Natl Acad Sci U S A* 101(51):17628-17633.

54. Sengoku T, Nureki O, Nakamura A, Kobayashi S, & Yokoyama S (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* 125(2):287-300.

55. Del Campo M & Lambowitz AM (2009) Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol Cell* 35(5):598-609.

56. von Moeller H, Basquin C, & Conti E (2009) The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat Struct Mol Biol* 16(3):247-254.

57. Collins R*, et al.* (2009) The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J Biol Chem* 284(16):10296-10300.

58. Lorsch JR & Herschlag D (1998) The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry* 37(8):2194-2206.

59. Collins R*, et al.* (2009) The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J Biol Chem* 284(16):10296-10300.

60. Benz J, Trachsel H, & Baumann U (1999) Crystal structure of the ATPase domain of translation initiation factor 4A from Saccharomyces cerevisiae--the prototype of the DEAD box protein family. *Structure* 7(6):671-679.

61. Rudolph MG, Heissmann R, Wittmann JG, & Klostermeier D (2006) Crystal structure and nucleotide binding of the Thermus thermophilus RNA helicase Hera N-terminal domain. *J Mol Biol* 361(4):731-743.

62. Napetschnig J*, et al.* (2009) Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19. *Proc Natl Acad Sci U S A* 106(9):3089-3094.

63. Schütz P*, et al.* (2010) Comparative structural analysis of human DEAD-box RNA helicases. *PLoS One* 5(9).

64. Mallam AL, Del Campo M, Gilman B, Sidote DJ, & Lambowitz AM (2012) Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* 490(7418):121-125.

65. Tanner NK, Cordin O, Banroques J, Doere M, & Linder P (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* 11(1):127-138.

66. Cordin O, Tanner NK, Doère M, Linder P, & Banroques J (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* 23(13):2478-2487.

67. Rozen F, Pelletier J, Trachsel H, & Sonenberg N (1989) A lysine substitution in the ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. *Mol Cell Biol* 9(9):4061-4063.

68. Blum S*, et al.* (1992) ATP hydrolysis by initiation factor 4A is required for translation initiation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 89(16):7664-7668.

69. Pause A & Sonenberg N (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J* 11(7):2643-2654.

70. Pause A, Methot N, & Sonenberg N (1993) The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol* 13(11):6789-6798.

71. Iost I, Dreyfus M, & Linder P (1999) Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. *J Biol Chem* 274(25):17677-17683.

72. Svitkin YV*, et al.* (2001) The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* 7(3):382-394.

73. Rocak S, Emery B, Tanner NK, & Linder P (2005) Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. *Nucleic Acids Res* 33(3):999-1009.

74. Solem A, Zingler N, & Pyle AM (2006) A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol Cell* 24(4):611-617.

75. Karow AR & Klostermeier D (2009) A conformational change in the helicase core is necessary but not sufficient for RNA unwinding by the DEAD box helicase YxiN. *Nucleic Acids Res* 37(13):4464-4471.

76. Banroques J, Doere M, Dreyfus M, Linder P, & Tanner NK (2010) Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J Mol Biol* 396(4):949-966.

77. Lorsch JR & Herschlag D (1998) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* 37(8):2180-2193.

78. Polach KJ & Uhlenbeck OC (2002) Cooperative binding of ATP and RNA substrates to the DEAD/H protein DbpA. *Biochemistry* 41(11):3693-3702.

79. Cordin O, Tanner NK, Doere M, Linder P, & Banroques J (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* 23(13):2478-2487.

80. Banroques J, Cordin O, Doère M, Linder P, & Tanner NK (2008) A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins. *Mol Cell Biol* 28(10):3359-3371.

81. Aregger R & Klostermeier D (2009) The DEAD box helicase YxiN maintains a closed conformation during ATP hydrolysis. *Biochemistry* 48(45):10679-10681.

82. Henn A, Cao W, Hackney DD, & De La Cruz EM (2008) The ATPase cycle mechanism of the DEAD-box rRNA helicase, DbpA. *J Mol Biol* 377(1):193-205.

83. Cao W*, et al.* (2011) Mechanism of Mss116 ATPase reveals functional diversity of DEAD-box proteins. *J Mol Biol* 409(3):399-414.

84. Tarn WY & Chang TH (2009) The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA Biol* 6(1):17-20.

85. Mohr S, Stryker JM, & Lambowitz AM (2002) A DEAD-box protein functions as an ATP-dependent RNA chaperone in group I intron splicing. *Cell* 109(6):769-779.

86. Karow AR, Theissen B, & Klostermeier D (2007) Authentic interdomain communication in an RNA helicase reconstituted by expressed protein ligation of two helicase domains. *FEBS J* 274(2):463-473.

87. Nielsen KH*, et al.* (2009) Mechanism of ATP turnover inhibition in the EJC. *RNA* 15(1):67-75.

88. Ballut L*, et al.* (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* 12(10):861-869.

89. Hilbert M, Kebbel F, Gubaev A, & Klostermeier D (2011) eIF4G stimulates the activity of the DEAD box protein eIF4A by a conformational guidance mechanism. *Nucleic Acids Res* 39(6):2260-2270.

90. Ozes AR, Feoktistova K, Avanzino BC, & Fraser CS (2011) Duplex unwinding and ATPase activities of the DEAD-box helicase eIF4A are coupled by eIF4G and eIF4B. *J Mol Biol* 412(4):674-687.

91. Rajagopal V, Park EH, Hinnebusch AG, & Lorsch JR (2012) Specific domains in yeast translation initiation factor eIF4G strongly bias RNA unwinding activity of the eIF4F complex toward duplexes with 5'-overhangs. *J Biol Chem* 287(24):20301-20312.

92. Rogers GW, Richter NJ, Lima WF, & Merrick WC (2001) Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J Biol Chem* 276(33):30914-30922.

93. Hilliker A, Gao Z, Jankowsky E, & Parker R (2011) The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Mol Cell* 43(6):962-972.

94. Noble KN*, et al.* (2011) The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes Dev* 25(10):1065-1077.

95. Henn A*, et al.* (2010) Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase, DbpA. *Proc Natl Acad Sci U S A* 107(9):4046-4050.

96. Uhlmann-Schiffler H, Jalal C, & Stahl H (2006) Ddx42p--a human DEAD box protein with RNA chaperone activities. *Nucleic Acids Res* 34(1):10-22.

97. Yang Q & Jankowsky E (2005) ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* 44(41):13591-13601.

98. Young CL, Khoshnevis S, & Karbstein K (2013) Cofactor-dependent specificity of a DEAD-box protein. *Proc Natl Acad Sci U S A* 110(29):E2668-2676.

99. Lamanna AC & Karbstein K (2011) An RNA conformational switch regulates pre-18S rRNA cleavage. *J Mol Biol* 405(1):3-17.

100. Bohnsack MT, Kos M, & Tollervey D (2008) Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase. *EMBO Rep* 9(12):1230-1236.

101. Fairman M*, et al.* (2004) Protein displacement by DExH/D "RNA helicases" without duplex unwinding. *Science* 304(5671):730-734.

102. Bowers HA*, et al.* (2006) Discriminatory RNP remodeling by the DEAD-box protein DED1. *RNA* 12(5):903-912.

103. Tran EJ, Zhou Y, Corbett AH, & Wente SR (2007) The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. *Mol Cell* 28(5):850-859.

104. Kistler AL & Guthrie C (2001) Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev* 15(1):42-49.

105. Herschlag D (1995) RNA chaperones and the RNA folding problem. *J Biol Chem* 270(36):20871-20874.

106. Russell R (2008) RNA misfolding and the action of chaperones. *Front Biosci* 13:1-20.

107. Treiber D, Rook M, Zarrinkar P, & Williamson J (1998) Kinetic intermediates trapped by native interactions in RNA folding. *Science* 279(5358):1943-1946.

108. Lindahl T, Adams A, & Fresco JR (1966) Renaturation of transfer ribonucleic acids through site binding of magnesium. *Proc Natl Acad Sci U S A* 55(4):941-948.

109. Aubert M, Scott JF, Reynier M, & Monier R (1968) Rearrangement of the conformation of *Escherichia coli* 5S RNA. *Proc Natl Acad Sci U S A* 61(1):292-299.

110. Walstrum SA & Uhlenbeck OC (1990) The self-splicing RNA of *Tetrahymena* is trapped in a less active conformation by gel purification. *Biochemistry* 29(46):10573-10576.

111. Woodson SA & Cech TR (1991) Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the *Tetrahymena* intervening sequence RNA. *Biochemistry* 30(8):2042-2050.

112. Pan J & Woodson SA (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J Mol Biol* 280(4):597-609.

113. Russell R & Herschlag D (1999) New pathways in folding of the *Tetrahymena* group I RNA enzyme. *J Mol Biol* 291(5):1155-1167.

114. Zhang L, Bao P, Leibowitz MJ, & Zhang Y (2009) Slow formation of a pseudoknot structure is rate limiting in the productive co-transcriptional folding of the self-splicing Candida intron. *RNA* 15(11):1986-1992.

115. Duncan CD & Weeks KM (2008) SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* 47(33):8504-8513.

116. Sinan S, Yuan X, & Russell R (2011) The *Azoarcus* group I intron ribozyme misfolds and is accelerated for refolding by ATP-dependent RNA chaperone proteins. *J Biol Chem* 286(43):37304-37312.

117. 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(11):931-938.

118. Chadalavada DM, Senchak SE, & Bevilacqua PC (2002) The folding pathway of the genomic hepatitis delta virus ribozyme is dominated by slow folding of the pseudoknots. *J Mol Biol* 317(4):559-575.

119. Rajkowitsch L*, et al.* (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* 4(3):118-130.

120. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, & Hartl FU (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* 82:323-355.

121. Huang H*, et al.* (2005) The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc Natl Acad Sci U S A* 102(1):163-168.

122. Lambowitz AM & Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* 3(8):a003616.

123. Nielsen H & Johansen SD (2009) Group I introns: Moving in new directions. *RNA Biol* 6(4):375-383.

124. Treiber DK & Williamson JR (1999) Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* 9(3):339-345.

125. Russell R (2008) RNA misfolding and the action of chaperones. *Front Biosci* 13:1-20.

126. Mohr S, Matsuura M, Perlman P, & Lambowitz A (2006) A DEAD-box protein alone promotes group II intron splicing and reverse splicing by acting as an RNA chaperone. *Proc Natl Acad Sci U S A* 103(10):3569-3574.

127. Bhaskaran H & Russell R (2007) Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* 449(7165):1014-1018.

128. Del Campo M*, et al.* (2007) Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol Cell* 28(1):159-166.

129. Del Campo M*, et al.* (2009) Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. *J Mol Biol* 389(4):674-693.

130. Cate JH*, et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273(5282):1678-1685.

131. Golden B, Gooding A, Podell E, & Cech T (1998) A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* 282(5387):259-264.

132. Lehnert V, Jaeger L, Michel F, & Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem Biol* 3(12):993-1009.

133. Michel F & Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216(3):585-610.

134. Adams PL, Stahley MR, Kosek AB, Wang J, & Strobel SA (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* 430(6995):45-50.

135. Zaug AJ, Grosshans CA, & Cech TR (1988) Sequence-specific endoribonuclease activity of the *Tetrahymena* ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes. *Biochemistry* 27(25):8924-8931.

136. Wan Y, Mitchell D, & Russell R (2009) Catalytic activity as a probe of native RNA folding. *Methods Enzymol* 468:195-218.

137. Potratz JP & Russell R (2012) RNA catalysis as a probe for chaperone activity of DEAD-Box helicases. *Methods Enzymol* 511:111-130.

138. Zarrinkar PP & Williamson JR (1994) Kinetic intermediates in RNA folding. *Science* 265(5174):918-924.

139. Sclavi B, Sullivan M, Chance MR, Brenowitz M, & Woodson SA (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279(5358):1940-1943.

140. Russell R & Herschlag D (2001) Probing the folding landscape of the *Tetrahymena* ribozyme: commitment to form the native conformation is late in the folding pathway. *J Mol Biol* 308(5):839-851.

141. Russell R*, et al.* (2002) Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci U S A* 99(1):155-160.

142. Laederach A, Shcherbakova I, Jonikas MA, Altman RB, & Brenowitz M (2007) Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc Natl Acad Sci U S A* 104(17):7045-7050.

143. Shcherbakova I, Mitra S, Laederach A, & Brenowitz M (2008) Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol* 12(6):655-666.

144. 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(5):367-370.

145. Russell R*, et al.* (2006) The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* 363(2):531-544.

146. Mitchell D, 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(15):2670-2686.

147. Tijerina P, Bhaskaran H, & Russell R (2006) Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *Proc Natl Acad Sci U S A* 103(45):16698-16703.

148. Gross T*, et al.* (2007) The DEAD-box RNA helicase Dbp5 functions in translation termination. *Science* 315(5812):646-649.

149. Pan C & Russell R (2010) Roles of DEAD-box proteins in RNA and RNP Folding. *RNA Biol* 7(6):667-676.

150. Halls C*, et al.* (2007) Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol* 365(3):835-855.

151. Potratz JP, Del Campo M, Wolf RZ, Lambowitz AM, & Russell R (2011) ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing in vitro and in vivo. *J Mol Biol* 411(3):661-679.

152. Zingler N, Solem A, & Pyle AM (2010) Dual roles for the Mss116 cofactor during splicing of the ai5γ group II intron. *Nucleic Acids Res* 38(19):6602-6609.

153. Karunatilaka KS, Solem A, Pyle AM, & Rueda D (2010) Single-molecule analysis of Mss116-mediated group II intron folding. *Nature* 467(7318):935-939.

154. Russell R, Jarmoskaite I, & Lambowitz AM (2013) Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol* 10(1):44-55.

155. Grohman J*, et al.* (2007) Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry* 46(11):3013-3022.

156. Cate J*, et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273(5282):1678-1685.

157. Russell R, Millett I, Doniach S, & Herschlag D (2000) Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* 7(5):367-370.

158. 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(5):1067-1077.

159. Benz-Moy TL & Herschlag D (2011) Structure-function analysis from the outside in: long-range tertiary contacts in RNA exhibit distinct catalytic roles. *Biochemistry* 50(40):8733-8755.

160. Shcherbakova I & Brenowitz M (2005) Perturbation of the hierarchical folding of a large RNA by the destabilization of its scaffold's tertiary structure. *J Mol Biol* 354(2):483-496.

161. Treiber D & Williamson J (2001) Concerted kinetic folding of a multidomain ribozyme with a disrupted loop-receptor interaction. *J Mol Biol* 305(1):11-21.

162. Pan J & Woodson S (1999) The effect of long-range loop-loop interactions on folding of the *Tetrahymena* self-splicing RNA. *J Mol Biol* 294(4):955-965.

163. Celander DW & Cech TR (1991) Visualizing the higher order folding of a catalytic RNA molecule. *Science* 251(4992):401-407.

164. Sattin B, Zhao W, Travers K, Chu S, & Herschlag D (2008) Direct measurement of tertiary contact cooperativity in RNA folding. *J Am Chem Soc* 130(19):6085-6087.

165. Das R*, et al.* (2003) The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the *Tetrahymena* ribozyme. *J Mol Biol* 332(2):311-319.

166. Liu F, Putnam AA, & Jankowsky E (2013) DEAD-box helicases form nucleotide-dependent long-lived complexes with RNA. *Biochemistry*.

167. Doniach S (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev* 101(6):1763-1778.

168. Shi H, O'Brien CA, Van Horn DJ, & Wolin SL (1996) A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens. *RNA* 2(8):769-784.

169. Fuchs G, Stein AJ, Fu C, Reinisch KM, & Wolin SL (2006) Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. *Nat Struct Mol Biol* 13(11):1002-1009.

170. Reinisch KM & Wolin SL (2007) Emerging themes in non-coding RNA quality control. *Curr Opin Struct Biol* 17(2):209-214.

171. Staley JP & Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92(3):315-326.

172. Strunk BS & Karbstein K (2009) Powering through ribosome assembly. *RNA* 15(12):2083-2104.

173. Grohman JK*, et al.* (2013) A guanosine-centric mechanism for RNA chaperone function. *Science* 340(6129):190-195.

174. Grossberger R*, et al.* (2005) Influence of RNA structural stability on the RNA chaperone activity of the *Escherichia coli* protein StpA. *Nucleic Acids Res* 33(7):2280-2289.

175. Hartl FU, Bracher A, & Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475(7356):324-332.

176. Taipale M, Jarosz DF, & Lindquist S (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* 11(7):515-528.

177. Böcking T, Aguet F, Harrison SC, & Kirchhausen T (2011) Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating. *Nat Struct Mol Biol* 18(3):295-301.

178. Taipale M*, et al.* (2012) Quantitative analysis of Hsp90-client interactions reveals principles of substrate recognition. *Cell* 150(5):987-1001.

179. Zaug AJ, Grosshans CA, & Cech TR (1988) Sequence-specific endoribonuclease activity of the *Tetrahymena* ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes. *Biochemistry* 27(25):8924-8931.

180. Russell R, Tijerina P, Chadee AB, & Bhaskaran H (2007) Deletion of the P5abc peripheral element accelerates early and late folding steps of the *Tetrahymena* group I ribozyme. *Biochemistry* 46(17):4951-4961.

181. Chen Y, *et al.* (2008) DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc Natl Acad Sci U S A* 105(51):20203-20208.

182. Bai Y, *et al.* (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J Am Chem Soc* 129(48):14981-14988.

183. Pabit SA, *et al.* (2010) Counting ions around DNA with anomalous small-angle X-ray scattering. *J Am Chem Soc* 132(46):16334-16336.

184. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25:495-503.

185. Sharma SK, De los Rios P, Christen P, Lustig A, & Goloubinoff P (2010) The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat Chem Biol* 6(12):914-920.

186. Matouschek A (2003) Protein unfolding--an important process in vivo? *Curr Opin Struct Biol* 13(1):98-109.

187. Kenniston JA, Baker TA, Fernandez JM, & Sauer RT (2003) Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* 114(4):511-520.

188. Kenniston JA, Burton RE, Siddiqui SM, Baker TA, & Sauer RT (2004) Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J Struct Biol* 146(1-2):130-140.

189. Baker TA & Sauer RT (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* 1823(1):15-28.

190. Cordin O, Banroques J, Tanner NK, & Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* 367:17-37.

191. Svergun DI & Koch MHJ (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* 66:1735-1782.

192. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879-2886.

193. Svergun DI, Petoukhov MV, & Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946-2953.

194. Petoukhov MV & Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89(2):1237-1250.

195. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, & Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129(17):5656-5664.

196. Jankowsky E (2011) RNA helicases at work: binding and rearranging. *Trends Biochem Sci* 36(1):19-29.

197. Svergun DI & Nierhaus KH (2000) A map of protein-rRNA distribution in the 70 S Escherichia coli ribosome. *J Biol Chem* 275(19):14432-14439.

198. Markov DA*, et al.* (2009) Identification of proteins associated with the yeast mitochondrial RNA polymerase by tandem affinity purification. *Yeast* 26(8):423-440.

199. Del Campo M*, et al.* (2009) Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. *J Mol Biol* 389(4):674-693.

200. Hardin JW, Hu YX, & McKay DB (2010) Structure of the RNA binding domain of a DEAD-box helicase bound to its ribosomal RNA target reveals a novel mode of recognition by an RNA recognition motif. *J Mol Biol* 402(2):412-427.

201. Karow AR & Klostermeier D (2010) A structural model for the DEAD box helicase YxiN in solution: localization of the RNA binding domain. *J Mol Biol* 402(4):629-637.

202. Ghaemmaghami S*, et al.* (2003) Global analysis of protein expression in yeast. *Nature* 425(6959):737-741.

203. Lopez-Ramirez V, Alcaraz LD, Moreno-Hagelsieb G, & Olmedo-Alvarez G (2011) Phylogenetic Distribution and Evolutionary History of Bacterial DEAD-Box Proteins. *J Mol Evol*.

204. Hargreaves DC & Crabtree GR (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* 21(3):396-420.

205. Rajkowitsch L*, et al.* (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* 4(3):118-130.

206. Kucera NJ, Hodsdon ME, & Wolin SL (2011) An intrinsically disordered C terminus allows the La protein to assist the biogenesis of diverse noncoding RNA precursors. *Proc Natl Acad Sci U S A* 108(4):1308-1313.

207. Del Campo M & Lambowitz AM (2009) Crystallization and preliminary X-ray diffraction of the DEAD-box protein Mss116p complexed with an RNA oligonucleotide and AMP-PNP. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 65(Pt 8):832-835.

208. Wilkins MR*, et al.* (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531-552.

209. Sreerama N & Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287(2):252-260.

210. Whitmore L & Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89(5):392-400.

211. Jacques DA & Trewhella J (2010) Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein Sci* 19(4):642-657.

212. Guinier A (1939) Diffraction of X-rays of very small angles–application to the study of ultramicroscopic phenomena. *Ann. Phys.* 12:161-237.

213. Mylonas E & Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J Appl Crystallogr* 40:S245-S249.

214. Petoukhov MV, Konarev PV, Kikhney AG, & Svergun DI (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J Appl Crystallogr* 40:S223-S228.

215. Franke D & Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342-346.

216. Volkov VV & Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860-864.

217. Kozin MB & Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34:33-41.

218. Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779-815.

219. Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31(13):3429-3431.

220. Svergun D, Barberato C, & Koch MHJ (1995) CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768-773.

221. Lehnert V, Jaeger L, Michel F, & Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. *Chem Biol* 3(12):993-1009.

222. Russell R*, et al.* (2002) Rapid compaction during RNA folding. *Proc Natl Acad Sci U S A* 99(7):4266-4271.

223. Mitchell Dr (2013) Characterization of folding and misfolding of the *Tetrahymena thermophila* group I ribozyme. PhD (The University of Texas at Austin).

224. Lipfert J & Doniach S (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu Rev Biophys Biomol Struct* 36:307-327.

225. Koch M, Vachette P, & Svergun D (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* 36(2):147-227.

226. Lorsch JR & Herschlag D (1998) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* 37(8):2180-2193.

227. Okazaki Y*, et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420(6915):563-573.

228. Kapranov P*, et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316(5830):1484-1488.

229. Nagalakshmi U*, et al.* (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320(5881):1344-1349.

230. Sigler PB (1975) An analysis of the structure of tRNA. *Annu Rev Biophys Bioeng* 4(00):477-527.

231. Karpel RL*, et al.* (1975) Acceleration of RNA renaturation by nucleic acid unwinding proteins. *Brookhaven Symp Biol* (26):165-174.

232. Gorbalenya AE & Koonin EV (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology* 3(3):419-429.

233. Jankowsky E, Gross CH, Shuman S, & Pyle AM (2000) The DExH protein NPH-II is a processive and directional motor for unwinding RNA. *Nature* 403(6768):447-451.

234. Pang PS, Jankowsky E, Planet PJ, & Pyle AM (2002) The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *EMBO J* 21(5):1168-1176.

235. Serebrov V & Pyle AM (2004) Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature* 430(6998):476-480.

236. Büttner K, Nehring S, & Hopfner KP (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol* 14(7):647-652.

237. Richards JD*, et al.* (2008) Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *J Biol Chem* 283(8):5118-5126.

238. Pena V*, et al.* (2009) Common design principles in the spliceosomal RNA helicase Brr2 and in the Hel308 DNA helicase. *Mol Cell* 35(4):454-466.

239. Zhang L*, et al.* (2009) Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nat Struct Mol Biol* 16(7):731-739.

240. Jackson RN*, et al.* (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J* 29(13):2205-2216.

241. Weir JR, Bonneau F, Hentschel J, & Conti E (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc Natl Acad Sci U S A* 107(27):12139-12144.

242. Halbach F, Rode M, & Conti E (2012) The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA* 18(1):124-134.

243. He Y, Andersen GR, & Nielsen KH (2010) Structural basis for the function of DEAH helicases. *EMBO Rep* 11(3):180-186.

244. Walbott H*, et al.* (2010) Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO J* 29(13):2194-2204.

245. Johnson SJ & Jackson RN (2013) Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol* 10(1):33-43.

246. Bernstein J, Patterson DN, Wilson GM, & Toth EA (2008) Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome. *J Biol Chem* 283(8):4930-4942.

247. Jia H, Wang X, Anderson JT, & Jankowsky E (2012) RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc Natl Acad Sci U S A* 109(19):7292-7297.

248. Schwer B (2008) A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell* 30(6):743-754.

249. Lohman TM & Bjornson KP (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* 65:169-214.

250. Henn A, Bradley MJ, & De La Cruz EM (2012) ATP Utilization and RNA Conformational Rearrangement by DEAD-Box Proteins. *Annu Rev Biophys*.

251. Cheng Z, Coller J, Parker R, & Song H (2005) Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA* 11(8):1258-1270.

252. Jarmoskaite I & Russell R (2011) DEAD‑box proteins as RNA helicases and chaperones. *Wiley Interdisciplinary Reviews: RNA* 2(1):135–152.

253. Kossen K, Karginov FV, & Uhlenbeck OC (2002) The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA. *J Mol Biol* 324(4):625-636.

254. Folkmann AW, Noble KN, Cole CN, & Wente SR (2011) Dbp5, Gle1-IP6 and Nup159: a working model for mRNP export. *Nucleus* 2(6):540-548.

255. Jankowsky E, Gross CH, Shuman S, & Pyle AM (2001) Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. *Science* 291(5501):121-125.

256. Jankowsky E & Bowers H (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res* 34(15):4181-4188.

257. Shajani Z, Sykes MT, & Williamson JR (2011) Assembly of bacterial ribosomes. *Annu Rev Biochem* 80:501-526.

258. Jagessar KL & Jain C (2010) Functional and molecular analysis of Escherichia coli strains lacking multiple DEAD-box helicases. *RNA* 16(7):1386-1392.

259. Lehnik-Habrink M*, et al.* (2013) DEAD-Box RNA helicases in Bacillus subtilis have multiple functions and act independently from each other. *J Bacteriol* 195(3):534-544.

260. Elles LM & Uhlenbeck OC (2008) Mutation of the arginine finger in the active site of Escherichia coli DbpA abolishes ATPase and helicase activity and confers a dominant slow growth phenotype. *Nucleic Acids Res* 36(1):41-50.

261. Sharpe Elles L, Sykes M, Williamson J, & Uhlenbeck O (2009) A dominant negative mutant of the E. coli RNA helicase DbpA blocks assembly of the 50S ribosomal subunit. *Nucleic Acids Res* 37(19):6503-6514.

262. Charollais J, Pflieger D, Vinh J, Dreyfus M, & Iost I (2003) The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in Escherichia coli. *Mol Microbiol* 48(5):1253-1265.

263. Trubetskoy D, Proux F, Allemand F, Dreyfus M, & Iost I (2009) SrmB, a DEAD-box helicase involved in Escherichia coli ribosome assembly, is specifically targeted to 23S rRNA in vivo. *Nucleic Acids Res* 37(19):6540-6549.

264. Proux F, Dreyfus M, & Iost I (2011) Identification of the sites of action of SrmB, a DEAD-box RNA helicase involved in Escherichia coli ribosome assembly. *Mol Microbiol* 82(2):300-311.

265. Jones PG, Mitta M, Kim Y, Jiang W, & Inouye M (1996) Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in Escherichia coli. *Proc Natl Acad Sci U S A* 93(1):76-80.

266. Iost I, Bizebard T, & Dreyfus M (2013) Functions of DEAD-box proteins in bacteria: current knowledge and pending questions. *Biochim Biophys Acta* 1829(8):866-877.

267. Charollais J, Dreyfus M, & Iost I (2004) CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Res* 32(9):2751-2759.

268. Jiang W, Hou Y, & Inouye M (1997) CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J Biol Chem* 272(1):196-202.

269. Awano N*, et al.* (2007) Complementation analysis of the cold-sensitive phenotype of the Escherichia coli csdA deletion strain. *J Bacteriol* 189(16):5808-5815.

270. Phadtare S (2011) Unwinding activity of cold shock proteins and RNA metabolism. *RNA Biol* 8(3):394-397.

271. Jain C (2008) The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *RNA* 14(2):381-389.

272. Martin R, Straub AU, Doebele C, & Bohnsack MT (2013) DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol* 10(1):4-18.

273. Weaver PL, Sun C, & Chang TH (1997) Dbp3p, a putative RNA helicase in Saccharomyces cerevisiae, is required for efficient pre-rRNA processing predominantly at site A3. *Mol Cell Biol* 17(3):1354-1365.

274. Pertschy B*, et al.* (2009) RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1. *J Biol Chem* 284(50):35079-35091.

275. Bohnsack MT*, et al.* (2009) Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell* 36(4):583-592.

276. Lebaron S*, et al.* (2009) The ATPase and helicase activities of Prp43p are stimulated by the G-patch protein Pfa1p during yeast ribosome biogenesis. *EMBO J* 28(24):3808-3819.

277. de la Cruz J, Kressler D, Tollervey D, & Linder P (1998) Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. *EMBO J* 17(4):1128-1140.

278. Watkins NJ & Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* 3(3):397-414.

279. Liang XH & Fournier MJ (2006) The helicase Has1p is required for snoRNA release from pre-rRNA. *Mol Cell Biol* 26(20):7437-7450.

280. Kos M & Tollervey D (2005) The Putative RNA Helicase Dbp4p Is Required for Release of the U14 snoRNA from Preribosomes in Saccharomyces cerevisiae. *Mol Cell* 20(1):53-64.

281. Granneman S, Bernstein KA, Bleichert F, & Baserga SJ (2006) Comprehensive mutational analysis of yeast DEXD/H box RNA helicases required for small ribosomal subunit synthesis. *Mol Cell Biol* 26(4):1183-1194.

282. Colley A, Beggs JD, Tollervey D, & Lafontaine DL (2000) Dhr1p, a putative DEAH-box RNA helicase, is associated with the box C+D snoRNP U3. *Mol Cell Biol* 20(19):7238-7246.

283. Lebaron S*, et al.* (2005) The splicing ATPase prp43p is a component of multiple preribosomal particles. *Mol Cell Biol* 25(21):9269-9282.

284. Combs DJ, Nagel RJ, Ares M, & Stevens SW (2006) Prp43p is a DEAH-box spliceosome disassembly factor essential for ribosome biogenesis. *Mol Cell Biol* 26(2):523-534.

285. Leeds NB, Small EC, Hiley SL, Hughes TR, & Staley JP (2006) The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis. *Mol Cell Biol* 26(2):513-522.

286. Tanaka N & Schwer B (2006) Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function. *Biochemistry* 45(20):6510-6521.

287. Tanaka N, Aronova A, & Schwer B (2007) Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev* 21(18):2312-2325.

288. Guenther UP & Jankowsky E (2009) Helicase multitasking in ribosome assembly. *Mol Cell* 36(4):537-538.

289. Cordin O & Beggs JD (2013) RNA helicases in splicing. *RNA Biol* 10(1):83-95.

290. Chang TH, Tung L, Yeh FL, Chen JH, & Chang SL (2013) Functions of the DExD/H-box proteins in nuclear pre-mRNA splicing. *Biochim Biophys Acta* 1829(8):764-774.

291. Semlow DR & Staley JP (2012) Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem Sci* 37(7):263-273.

292. Ruby SW, Chang TH, & Abelson J (1993) Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev* 7(10):1909-1925.

293. Shen H*, et al.* (2008) Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev* 22(13):1796-1803.

294. O'Day CL, Dalbadie-McFarland G, & Abelson J (1996) The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem* 271(52):33261-33267.

295. Perriman R, Barta I, Voeltz GK, Abelson J, & Ares M (2003) ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc Natl Acad Sci U S A* 100(24):13857-13862.

296. Perriman RJ & Ares M (2007) Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev* 21(7):811-820.

297. Abu Dayyeh BK, Quan TK, Castro M, & Ruby SW (2002) Probing interactions between the U2 small nuclear ribonucleoprotein and the DEAD-box protein, Prp5. *J Biol Chem* 277(23):20221-20233.

298. Perriman R & Ares M (2010) Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol Cell* 38(3):416-427.

299. Staley JP & Guthrie C (1999) An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3(1):55-64.

300. Strauss EJ & Guthrie C (1994) PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing in vitro. *Nucleic Acids Res* 22(15):3187-3193.

301. Warkocki Z*, et al.* (2009) Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol* 16(12):1237-1243.

302. Lardelli RM, Thompson JX, Yates JR, & Stevens SW (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* 16(3):516-528.

303. Liu HL & Cheng SC (2012) The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. *Mol Cell Biol* 32(24):5056-5066.

304. Ohrt T*, et al.* (2012) Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* 18(6):1244-1256.

305. Ohrt T*, et al.* (2013) Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* 19(7):902-915.

306. Mefford MA & Staley JP (2009) Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* 15(7):1386-1397.

307. Hilliker AK, Mefford MA, & Staley JP (2007) U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev* 21(7):821-834.

308. Tseng CK, Liu HL, & Cheng SC (2011) DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* 17(1):145-154.

309. Schwer B & Gross CH (1998) Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J* 17(7):2086-2094.

310. Wagner JD, Jankowsky E, Company M, Pyle AM, & Abelson JN (1998) The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J* 17(10):2926-2937.

311. Tanaka N & Schwer B (2005) Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-box splicing factor Prp22. *Biochemistry* 44(28):9795-9803.

312. Arenas JE & Abelson JN (1997) Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A* 94(22):11798-11802.

313. Martin A, Schneider S, & Schwer B (2002) Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J Biol Chem* 277(20):17743-17750.

314. Tsai RT*, et al.* (2005) Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev* 19(24):2991-3003.

315. Laggerbauer B, Achsel T, & Lührmann R (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proc Natl Acad Sci U S A* 95(8):4188-4192.

316. Raghunathan PL & Guthrie C (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* 8(15):847-855.

317. Mozaffari-Jovin S*, et al.* (2012) The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev* 26(21):2422-2434.

318. Hahn D, Kudla G, Tollervey D, & Beggs JD (2012) Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev* 26(21):2408-2421.

319. Nielsen KH & Staley JP (2012) Spliceosome activation: U4 is the path, stem I is the goal, and Prp8 is the keeper. Let's cheer for the ATPase Brr2! *Genes Dev* 26(22):2461-2467.

320. Zaher HS & Green R (2009) Fidelity at the molecular level: lessons from protein synthesis. *Cell* 136(4):746-762.

321. Hopfield JJ (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* 71(10):4135-4139.

322. Ninio J (1975) Kinetic amplification of enzyme discrimination. *Biochimie* 57(5):587-595.

323. Xu YZ & Query CC (2007) Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol Cell* 28(5):838-849.

324. Yang F*, et al.* (2013) Splicing proofreading at 5' splice sites by ATPase Prp28p. *Nucleic Acids Res* 41(8):4660-4670.

325. Burgess S, Couto JR, & Guthrie C (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell* 60(5):705-717.

326. Konarska MM & Query CC (2005) Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev* 19(19):2255-2260.

327. Villa T & Guthrie C (2005) The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes Dev* 19(16):1894-1904.

328. Mayas RM, Maita H, & Staley JP (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* 13(6):482-490.

329. Mayas RM, Maita H, Semlow DR, & Staley JP (2010) Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc Natl Acad Sci U S A* 107(22):10020-10025.

330. Koodathingal P, Novak T, Piccirilli JA, & Staley JP (2010) The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Mol Cell* 39(3):385-395.

331. Séraphin B, Simon M, Boulet A, & Faye G (1989) Mitochondrial splicing requires a protein from a novel helicase family. *Nature* 337(6202):84-87.

332. Köhler D, Schmidt-Gattung S, & Binder S (2010) The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of Arabidopsis thaliana. *Plant Mol Biol* 72(4-5):459-467.

333. Asakura Y, Galarneau E, Watkins KP, Barkan A, & van Wijk KJ (2012) Chloroplast RH3 DEAD box RNA helicases in maize and Arabidopsis function in splicing of specific group II introns and affect chloroplast ribosome biogenesis. *Plant Physiol* 159(3):961-974.

334. Bifano A & Caprara M (2008) A DExH/D-box protein coordinates the two steps of splicing in a group I intron. *J Mol Biol* 383(3):667-682.

335. Fedorova O, Solem A, & Pyle AM (2010) Protein-facilitated folding of group II intron ribozymes. *J Mol Biol* 397(3):799-813.

336. Waldsich C & Pyle AM (2008) A kinetic intermediate that regulates proper folding of a group II intron RNA. *J Mol Biol* 375(2):572-580.

337. Aitken CE & Lorsch JR (2012) A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol* 19(6):568-576.

338. Schütz P*, et al.* (2008) Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* 105(28):9564-9569.

339. Lindqvist L, Imataka H, & Pelletier J (2008) Cap-dependent eukaryotic initiation factor-mRNA interactions probed by cross-linking. *RNA* 14(5):960-969.

340. von der Haar T & McCarthy JE (2002) Intracellular translation initiation factor levels in Saccharomyces cerevisiae and their role in cap-complex function. *Mol Microbiol* 46(2):531-544.

341. Kolupaeva VG, Lomakin IB, Pestova TV, & Hellen CU (2003) Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. *Mol Cell Biol* 23(2):687-698.

342. de Breyne S, Yu Y, Unbehaun A, Pestova TV, & Hellen CU (2009) Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. *Proc Natl Acad Sci U S A* 106(23):9197-9202.

343. Pisareva VP, Pisarev AV, Komar AA, Hellen CU, & Pestova TV (2008) Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. *Cell* 135(7):1237-1250.

344. Abaeva IS, Marintchev A, Pisareva VP, Hellen CU, & Pestova TV (2011) Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. *EMBO J* 30(1):115-129.

345. Hartman TR*, et al.* (2006) RNA helicase A is necessary for translation of selected messenger RNAs. *Nat Struct Mol Biol* 13(6):509-516.

346. Ranji A, Shkriabai N, Kvaratskhelia M, Musier-Forsyth K, & Boris-Lawrie K (2011) Features of double-stranded RNA-binding domains of RNA helicase A are necessary for selective recognition and translation of complex mRNAs. *J Biol Chem* 286(7):5328-5337.

347. Chakraborty P & Grosse F (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)* 10(6):654-665.

348. Lasko P (2013) The DEAD-box helicase Vasa: evidence for a multiplicity of functions in RNA processes and developmental biology. *Biochim Biophys Acta* 1829(8):810-816.

349. Carrera P*, et al.* (2000) VASA mediates translation through interaction with a Drosophila yIF2 homolog. *Mol Cell* 5(1):181-187.

350. Johnstone O & Lasko P (2004) Interaction with eIF5B is essential for Vasa function during development. *Development* 131(17):4167-4178.

351. Daugeron MC, Prouteau M, Lacroute F, & Séraphin B (2011) The highly conserved eukaryotic DRG factors are required for efficient translation in a manner redundant with the putative RNA helicase Slh1. *Nucleic Acids Res* 39(6):2221-2233.

352. Bolger TA, Folkmann AW, Tran EJ, & Wente SR (2008) The mRNA export factor Gle1 and inositol hexakisphosphate regulate distinct stages of translation. *Cell* 134(4):624-633.

353. Tieg B & Krebber H (2013) Dbp5 - from nuclear export to translation. *Biochim Biophys Acta* 1829(8):791-798.

354. Takyar S, Hickerson RP, & Noller HF (2005) mRNA helicase activity of the ribosome. *Cell* 120(1):49-58.

355. Qu X*, et al.* (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* 475(7354):118-121.

356. Hardwick SW & Luisi BF (2013) Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control. *RNA Biol* 10(1):56-70.

357. Py B, Higgins CF, Krisch HM, & Carpousis AJ (1996) A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* 381(6578):169-172.

358. Coburn GA, Miao X, Briant DJ, & Mackie GA (1999) Reconstitution of a minimal RNA degradosome demonstrates functional coordination between a 3' exonuclease and a DEAD-box RNA helicase. *Genes Dev* 13(19):2594-2603.

359. Schneider C & Tollervey D (2013) Threading the barrel of the RNA exosome. *Trends Biochem Sci* 38(10):485-493.

360. Halbach F, Reichelt P, Rode M, & Conti E (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* 154(4):814-826.

361. Jia H*, et al.* (2011) The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* 145(6):890-901.

362. Tran H, Schilling M, Wirbelauer C, Hess D, & Nagamine Y (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol Cell* 13(1):101-111.

363. Fuller-Pace FV (2013) The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim Biophys Acta* 1829(8):756-763.

364. Cloutier SC, Ma WK, Nguyen LT, & Tran EJ (2012) The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. *J Biol Chem* 287(31):26155-26166.

365. Ma WK, Cloutier SC, & Tran EJ (2013) The DEAD-box Protein Dbp2 Functions with the RNA-Binding Protein Yra1 to Promote mRNP Assembly. *J Mol Biol*.

366. Lund MK & Guthrie C (2005) The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol Cell* 20(4):645-651.

367. Taniguchi I & Ohno M (2008) ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56. *Mol Cell Biol* 28(2):601-608.

368. Strässer K*, et al.* (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 417(6886):304-308.

369. Fullam A & Schröder M (2013) DExD/H-box RNA helicases as mediators of anti-viral innate immunity and essential host factors for viral replication. *Biochim Biophys Acta* 1829(8):854-865.

370. Kruse E, Voigt C, Leeder WM, & Göringer HU (2013) RNA helicases involved in U-insertion/deletion-type RNA editing. *Biochim Biophys Acta* 1829(8):835-841.

371. Li F, Herrera J, Zhou S, Maslov DA, & Simpson L (2011) Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. *Proc Natl Acad Sci U S A* 108(9):3542-3547.

372. Hernandez A*, et al.* (2010) REH2 RNA helicase in kinetoplastid mitochondria: ribonucleoprotein complexes and essential motifs for unwinding and guide RNA (gRNA) binding. *J Biol Chem* 285(2):1220-1228.

373. Flores SC & Altman RB (2010) Turning limited experimental information into 3D models of RNA. *RNA* 16(9):1769-1778.

374. Flores SC, Wan Y, Russell R, & Altman RB (2010) Predicting RNA structure by multiple template homology modeling. *Pac Symp Biocomput*:216-227.

375. Porod G (1982) *General Theory* (Academic Press, London).

376. Hutchinson EG & Thornton JM (1996) PROMOTIF - A program to identify and analyze structural motifs in proteins. *Protein Science* 5(2):212-220.