The Dissertation Committee for Hari Prakash Bhaskaran Certifies that this is the approved version of the following dissertation:

Mechanistic Studies of CYT-19 and Related DExD/H-box Proteins on Folding of the *Tetrahymena* Group I Ribozyme

Committee:

Rick Russell, Supervisor

David W. Hoffman

Kenneth A. Johnson

Alan M. Lambowitz

Scott W. Stevens

Mechanistic Studies of CYT-19 and Related DExD/H-box Proteins on Folding of the *Tetrahymena* Group I Ribozyme

by

Hari Prakash Bhaskaran, B. V. Sc.; M. S.

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, 2008

Dedication

To my mother and father

Acknowledgements

I would like to take this opportunity to thank the people without whom this thesis would not exist. First, I would like to thank my advisor, Rick Russell for everything that I learnt in his lab including planning, designing, analyzing and interpreting experiments with utmost care. I came into the lab with little or no experience in molecular biology and he taught me so much. I also thank Rick for being instrumental in identification of my specific goal oriented research project taking into account, my strengths and weaknesses.

I would also like to thank Pilar Tijerina for teaching me how to maintain a good lab notebook and for putting up with all the mistakes that I made during my learning period in the lab. I would like to thank Dr. Ken Johnson for his stimulating lectures during his kinetics class; Dr. Alan Lambowitz for valuable discussions during group meetings and Dr. Scott Stevens & Dr. Lambowitz for writing me excellent recommendation letters. I thank Travis Johnson for the fun and insightful science conversations. I thank Amanda Chadee, Yaqi Wan, David Mitchell and Brian Canon for putting up with me and helping me with lab duties. Last but not least, I would like to thank my friends, Chad Mckee and Jamie Vernon for helping me realize that there is also life outside of the lab.

Mechanistic Studies of CYT-19 and Related DExD/H-box Proteins on Folding of the *Tetrahymena* Group I Ribozyme

Publication No._____

Hari Prakash Bhaskaran, Ph.D.

The University of Texas at Austin, 2008

Supervisor: Rick Russell

DExD/H-box proteins are a diverse class of proteins that are implicated in RNA and RNP remodeling. They have sequence homology to DNA helicases and share conserved ATPase domains, suggesting that they use the energy of ATP binding and hydrolysis to mediate conformational rearrangements in RNAs. In the past, the action of DExD/H-box proteins has been characterized primarily on simple model substrates such as small RNA duplexes. It is not known how DExD/H-box proteins manipulate structured RNA, what determines target specificity and what molecular events follow their action. Here, using the well-characterized *Tetrahymena* group I intron ribozyme, I performed kinetic and thermodynamic studies to understand the mechanism of CYT-19 and related DExD/H-box proteins. CYT-19 has been shown previously to facilitate the folding of several group I and group II introns. I demonstrated that CYT-19 acts as a chaperone, accelerating the re-folding of a long-lived misfolded species of the *Tetrahymena* group I ribozyme to its native state. Further characterization of this reaction gave insights into how CYT-19 achieves this action; CYT-19 partially unfolds the misfolded ribozyme and allows it to

fold again along the same pathway that exists in the absence of CYT-19. In addition to acting on the misfolded state, CYT-19 also acts on the native state, but this action is largely obscured under stabilizing conditions for the native state because the action is inefficient under such conditions. However, under conditions where the native state is destabilized, the native ribozyme was indeed shown to be partially unfolded by CYT-19. By acting on either species, CYT-19 sets up a steady state of unfolding, and the distribution is shifted from equilibrium to kinetic control, increasing the relative populations of conformations that are kinetically preferred during folding. The efficiency of action seems to correlate with the stability of the ribozyme. These activities are not restricted to CYT-19; the DExD/H-box proteins Mss116p and Ded1 were demonstrated to possess similar activities. Together, these studies give important insights into the mechanisms of action for this ubiquitous class of proteins and have implications for all structured RNAs in cells.

Table of Contents

List of Tables

List of Figures

Chapter 1: Overview of misfolding and conformational transitions in RNAs and the role of RNA chaperones

1.1 Introduction

A unifying theme among structured, biological molecules is that form follows function. RNAs rival proteins in achieving a splendid array of structures that underlie their diverse cellular functions. Structured RNAs display a variety of sizes and shapes, from the small, L-shaped tRNAs to the large globular rRNAs that comprise the interlocking pieces of the ribosomal subunits[1]. They perform a multitude of functions, from simple binding by base-pairing to the formation of intricate active sites that are capable of accelerating chemical reactions by many orders of magnitude[2-4]. In order to become functional molecules, the structured RNAs have to fold from linear polymers into specific three-dimensional structures. This is a challenge for RNAs because they have an enormous propensity to form alternative structures that can persist on a timescale long enough to interfere with their biological functions[5].

Misfolding of RNAs has been documented extensively and appears to be pervasive. Essentially every large RNA studied *in vitro* has been found to form functionally compromised, alternative structures[6]. However, the extent of misfolding appears to be lower *in vivo*, and this discrepancy has been attributed largely to the association of proteins *in vivo*[6,7]. Indeed, many non-specific RNA binding proteins have been shown to facilitate formation of correct structures *in vitro*, strongly suggesting that proteins also assist with RNA folding in cells[8,9]. These proteins, termed RNA chaperones,

have been proposed to disrupt non-native structures in RNAs to allow folding to correct structures[7,8,10]. Attractive candidates for RNA chaperoning are the DExD/H-box proteins (pronounced "DEAD box" proteins after the major subgroup with sequence D-E-A-D). These proteins have been implicated in promoting ATP-dependent conformational changes of RNA or RNP structures[11-13].

Many cellular processes involve remodeling of RNA and RNP structures by DExD/H-box proteins; among these are translation initiation, ribosome biogenesis, nuclear export of RNAs, 5' and 3' RNA processing, and RNA decay. Because DExD/H-box proteins act in a gamut of cellular processes, it is essential to understand how they are targeted to act on specific RNAs and how they mediate RNA remodeling. Is the mechanism, for example, purely mediated by unwinding of RNA duplexes, or are there other specific actions such as displacement of proteins or disruption of tertiary contacts during remodeling of structure? Investigations to address these questions will provide insight into the mechanisms of this ubiquitous class of proteins, increasing the understanding of the similarities between them and how they act in the context of different processes.

1.2 Folding of structured RNAs

All of the information for folding of RNAs, as for proteins, is contained in the primary sequence. Although RNAs and proteins share this fundamental property, RNA folding encounters both thermodynamic and kinetic challenges that are distinct from those of proteins. First, with the limited information of only four bases, specifying the

correctly folded structure by stabilizing it over all other structures, such as partially folded and alternative structures, can be extremely difficult. Second, the secondary structures of RNAs, short helical segments, have substantial local stability (Figure 1.1). Thus the formation of secondary structure can be independent of the overall structure[14]. This is in marked contrast to protein secondary structural elements, which are typically unstable without the tertiary interactions they make with the rest of the protein[14] (Figure 1.1). Thus, the inherent local stability of RNA gives RNA a greater propensity to form misfolded or alternative secondary structures that can persist for a long time.

Proteins and RNAs differ with respect to the strategies they employ for folding. Protein folding is facilitated by an increase in entropy of surrounding water that accompanies interactions between hydrophobic amino acids. The folding process for many proteins is "all or none", where the tertiary structure forms in conjunction with the formation of secondary structure due to the latter's instability in the absence of enforcing tertiary contacts. However, due to the high local stability, RNA secondary structure can independently exist such that its formation can precede formation of tertiary interactions. This type of folding is called hierarchical folding because secondary structure can form prior to the formation of tertiary structure.

As first pointed out by Levinthal for protein folding, it would not be possible for RNAs to sample every possible conformer during folding to their final structures[15]. Instead, RNAs must fold through intermediates. As a first step, global collapse can reduce the astronomical number of possible conformations to fewer partially folded structures. Collapse during RNA folding refers to compaction that results from the

shielding of the negatively charged phosphodiester backbone by the positively charged counter ions during *in vitro* folding. Upon addition of Mg^{2+}, there may be non-specific relaxation of RNA from denatured states to compact but partially denatured states[16,17]. Following electrostatic relaxation, there could be formation of native contacts[17], some of which may be disrupted during rearrangement of the collapsed RNA to form the final structure.

In contrast to the model of global collapse and subsequent rearrangement, an alternative model was proposed for group II intron folding, where the rate-limiting steps involve slow conformational searches[18]. However, this could be as a result of misfolding, which may have gone unnoticed presumably because folding from it to subsequent conformational search was not rate-limiting.

Binding of site-specific metal ions has been suggested to represent a slow 'folding' step for RNAs that fold rapidly and without detectable misfolding, such as the catalytic domain of RNAse P from *Bacillus subtilis*[19]. In this case, the rate constant for folding decreased upon urea addition and increased with increasing magnesium concentration[20], suggesting the presence of a rate-limiting step that involves magnesium binding rather than folding from a kinetically-trapped species. Magnesium-dependent folding has also been demonstrated for the catalytic core of the bI5 group I intron[21]. Despite the proposal and observation, in some cases, of hierarchical assembly, where formation of secondary structure precedes formation of tertiary contacts, there are exceptions. For example, in the case of the P5abc sub-domain from a group I intron from *Tetrahymena thermophila*, the secondary structure that forms in solution in the absence

of magnesium differs at 15 positions in the 56-nucleotide RNA upon comparison to the structure that forms in the presence of magnesium. The latter structure more closely resembles the solved crystal structure[22]. This suggests that ordered folding with the formation of secondary structure prior to the formation of tertiary structure does not always hold true and that some folding events may occur concurrently, or that secondary structure could be rearranged upon formation of tertiary structure. Nevertheless, formation of intermediates can help RNA folding by reducing the number of possible conformations; however, if the intermediates are long-lived, this can potentially lead to interference of normal function. *In vitro* studies of RNA folding are rife with examples of misfolded structures[23].

1.3 Misfolding and formation of alternative structures in RNAs

The earliest evidence for misfolding comes from studies of tRNAs[24,25]. Many tRNAs have been found to form functionally compromised, alternative structures. For example, a misfolded state of leucyl tRNA requires several weeks to convert to the native state[24], suggesting the presence of a kinetic trap during folding. Tryptophanyl tRNA was also found to exist in two forms, only one of which was capable of being charged with the correct amino acid, tryptophan, and the two forms could be interconverted by changing divalent ion concentration or pH[25]. By systematically examining thermal denaturation profiles for various tRNAs, such as $tRNA^{Tyr}$, $tRNA^{Phe}$, $tRNA^{Val}$ and $tRNA^{fMet}$, Cole *et. al.* proposed an inactive species that forms at low temperature and low

salt conditions and has an extended secondary structure, in contrast to the clover leaf conformation of the active structure[26,27].

These studies suggested that misfolding is prevalent even in the small structured tRNA, about 76 nucleotides in length, and implies that larger structured RNAs may be vulnerable to misfolding to a greater degree. This idea was confirmed by later experiments with larger RNAs. The slightly larger 5S RNA (120 nucleotides) was clearly shown to exist in two different conformations, which interconverted slowly and with a high activation energy[28,29]. Novel structures were proposed for the two conformations based on comparative sequence analysis, enzymatic degradation, and chemical modification patterns[30].

In certain cases, formation of alternative secondary structures may control gene expression, and this is seen in many riboswitches, which are RNA control elements that directly bind metabolite substrates and change conformation to regulate gene expression[31-34]. A classic example of alternative RNA structure is the bacterial *trp* operon mRNA, which has a *cis* element that adopts two different structures based on the availability of tryptophan[35]. Similarly, large-scale structural reorientation occurs in S-adenosylmethionine (SAM) riboswitch upon SAM binding[36].

With the discovery of ribozymes, distinguishing native states from misfolded states became more straightforward because of the former's enzymatic activity. Altman and Takada found that the catalytic subunit of ribonuclease P from *Escherichia coli* reaches the catalytically active state in a folding transition with a high activation energy barrier[37]. The RNA demonstrated a distinct lag in cleavage, which was eliminated by

longer periods of incubation at a certain temperature before performing the cleavage reaction. The time required for incubation was also considerably shortened by increasing the incubation temperature. Investigation of variants of the self-cleaving, viral 'Hammerhead' ribozyme using steady-state substrate cleavage reactions isolated a particular variant that was incapable of cleaving the substrate to completion[38]. Purification of uncleaved substrate and addition to this ribozyme variant resulted in the same incomplete reaction, strongly suggesting the presence of the same misfolded species. *Tetrahymena* group I ribozyme, obtained through *in vitro* transcription and gel purification, was also found to possess less activity, and the activity was restored by heating and cooling in the presence of magnesium ions[39]. The smallest known ribozyme, the hairpin ribozyme, can form alternative structures as demonstrated by the biphasic kinetics in which the fast phase represents cleavage by active ribozymes and the slow phase represents rearrangement of inactive species[40,41]. Later studies using single molecule methods revealed a much more complex picture with multiple distinct substrate docked states in the hairpin ribozyme[42].

Some of the largest RNAs must undoubtedly be those that form the ribosomal subunits, and it is not surprising that misfolding has been observed during formation of ribosomal subunits. Although several proteins associate with the 16S rRNA to form the functional 30S ribosomal subunit, reconstitution to the active form *in vitro* requires an activation step, suggesting the existence of misfolding[43]. Studies have shown significant changes in kethoxal reactivity of many residues of the 16S rRNA in the inactive 30S subunit, and the original reactivity was restored when the subunit was reactivated[44].

These results strongly suggested the presence of alternative conformations of the 30S ribosomal subunit. More recent work on the 30S ribosomal subunit indicates that ribosomal assembly involves binding of multiple proteins with potential for several rearrangement steps during this assembly landscape[45]. Other RNAs that show misfolding include the Col E1 primer RNA and the *E.coli* α mRNA pseudoknot. The isolated pseudoknot from *E.coli* α mRNA was found to occur in two forms based on differential migration during electrophoresis[46]. The interconversion between the two forms proceeds with a high energy of activation and depends on the concentrations of Mg^{2+}, K^{+} or H^{+} ions. The results discussed have provided evidence for misfolding in RNAs from *in vitro* studies. Although results from *in vivo* folding studies can be hard to interpret due to the difficulty in isolating variables, recent evidence does provide some insight into the nature of RNA folding *in vivo*.

1.4 *In vitro* versus *In vivo* folding

In vivo folding may or may not be the same as *in vitro* magnesium-induced folding. In some cases, it is plausible that misfolding observed *in vitro* originates from RNA handling during *in vitro* studies. Typically, RNAs are made by *in vitro* transcription using a template DNA, which has the great advantage of yielding very large quantities of RNAs in the lab. After subjecting the RNAs to denaturing conditions, they are stored in EDTA, which chelates magnesium and prevents acceleration of non-specific cleavage by magnesium and other cations. Prior to use in experiments, the RNAs are renatured using magnesium. Such a treatment could result in RNA structures that differ from those found

in vivo, where RNAs fold co-transcriptionally[6]. Nevertheless, folding studies of isolated RNAs *in vitro* can provide valuable information on fundamental principles, without interference from a multitude of factors that can complicate interpretation of results from such studies *in vivo*. On the other hand, *in vivo* folding studies are necessary to validate results from *in vitro* studies and to place folding in the context of cellular gene expression.

It has been suggested that the speed of transcription relative to folding could play a very important role in determining whether RNAs that misfold *in vitro* will also misfold *in vivo*[47]. If transcription occurs much faster than folding, folding may be essentially the same in both cases because folding is initiated after the entire RNA is synthesized. Using the hairpin ribozyme, Fedor and colleagues demonstrated that the kinetics of folding were the same *in vitro* and *in vivo* and that a ribozyme mutant that slowed kinetics *in vitro* also slowed kinetics *in vivo*[48-50]. This suggests that the rate-limiting steps are likely the same in both cases and that folding is presumably slower than transcription *in vivo*. In contrast to the hairpin ribozyme, the L-21*Sca*I version of the *Tetrahymena* group I ribozyme showed differences between *in vitro* and *in vivo* folding. The folding of this RNA was at least an order of magnitude faster *in vivo* than folding *in vitro*[51,52]. This difference could be due to host-associated polymerase transcription, which could involve pausing, or interaction of trans-acting factors, such as maturases, that prevent misfolding *in vivo*. The ribozyme also lacks the flanking exon sequences normally found *in vivo*, and these extensions may also contribute to efficient splicing observed *in vivo*[53,54].

Proteins can help facilitate folding *in vivo* by potentially avoiding formation of misfolded structures. The folded structures of human telomerase RNA showed several differences in the chemical modification pattern[55] *in vivo* compared to that folded *in vitro*, presumably due to association of proteins. The differences suggest that protein association can avoid formation of certain structures *in vivo*. Transcriptional pausing by proteins can also facilitate folding during transcription. If the formation of a long transcript is slow during the elongation phase, the 5' end could form structures that compete with formation of a downstream alternative non-native structure. The effect of pausing during transcription and the effect of a trans-acting protein factor were described by Sosnick, Pan and colleagues[56,57]. First, they demonstrated that in a circularly permutated RNA (RNA in which the 5' and 3' ends are covalently connected and new 5' and 3' ends are created at a different location), both magnesium induced and co-transcriptional folding introduces kinetic traps. Addition of the transcription elongation factor, NusA, increased the correct folding of certain domains due to NusA's ability to induce pausing during co-transcription and potentially avoid formation of kinetic traps[56]. A subsequent study demonstrated that pausing indeed prevented formation of misfolded structures in a particular domain, apparently by accelerating folding of that domain[57]. Although details of the mechanisms need to be investigated, these studies clearly demonstrate that co-transcriptional folding can differ from magnesium-induced folding. However, co-transcriptional folding alone cannot account for differences between *in vitro* and *in vivo* folding, as demonstrated in a study by Mahen *et. al*[58]. Here, kinetic traps were demonstrated in cell free transcription system, but no such traps were observable in yeast

cells, suggesting that the *in vitro* transcription system may lack factors that contribute to efficient folding *in vivo*.

1.5 Structural transitions in cellular RNAs

Many cellular processes involving RNAs are dynamic in nature. Formation and assembly of functional RNA and RNP complexes involve formation of favorable base pairing, disruption of protein-RNA conglomerates, disruption of non-native interactions, and other conformational changes. Many of these structural transitions are thermodynamically unfavorable and require activation energy. The same features that allow formation of stable misfolded species in RNA may also account for the difficulty in undergoing these transitions. For example, in eukaryotic splicing, many steps involve transient RNA-RNA and RNA-protein contacts that have to be disrupted before proceeding to subsequent steps[59] (see section 1.10.2).

The folding of ribosomal RNAs involves several steps, many of which may have to undergo extensive structural transitions. Some of these transitions can be facilitated by other RNAs that are separate from the molecule being modified. Trans-acting RNAs called snoRNAs (small nucleolar RNAs) are proposed to base-pair to complementary sequences on the pre-ribosomal RNA and facilitate RNA modification and ribosome assembly. For example, the U3 snoRNA interacts with the 5' end of the nascent small ribosomal subunit RNA to help in formation of a highly conserved structure, the 5' end pseudoknot of the small subunit[60,61]. The formation of this pseudoknot is likely to be

preceded by a thermodynamically unfavorable step[62], necessitating assistance from co-factors.

1.6 DExD/H-box proteins as ATP dependent RNPases

In vivo, RNAs are generally associated with proteins, either transiently or persistently. Many of the conformational dynamics in RNAs and RNPs are thought to be facilitated by RNA binding proteins that have low sequence specificity[12]. DExD/H-box proteins are major players in this regard, necessary for both the assembly and remodeling of ribonucleoproteins, such as the spliceosome and the ribosome[12]. They are also crucial for proper functioning of many fundamental processes of the cell, including translation initiation and RNA decay[63-65]. During their initial discovery, many of these proteins had a common amino acid sequence "DEAD" in one of their motifs, and so, the acronym was designated to represent this group of proteins. Later analysis revealed that a subset of proteins contained DEAH motifs and other amino acids in place of alanine; thus, the general name of DExD/H-box proteins came to represent the whole family[66]. Sequence analysis also revealed another important feature. Six residues, including GK in motif I and DE in motif II (Figure 1.2), were invariant among DExD/H-box proteins and DNA helicases from viruses and bacteria. This observation immediately suggested that DExD/H-box proteins could possess helicase activities[67]. Although the role of every motif in DExD/H-box proteins was not clearly defined, all of these motifs were highly conserved within "DEAD"-box proteins from different organisms and showed similar spacing between them[68]. This led to the speculation that the "core" of these proteins

possesses a common mechanism of action. Later findings showed that the core was flanked by less conserved N- and C-terminal extensions[69], and these have been implicated in binding specificity toward their targets.

At present, all DExD/H-box proteins are classified as a separate family of "RNA helicases" under the super family 2 (SF2); DNA helicases are classified under SF1 and along with SF3, which comprise mostly viral hexameric helicases, they represent the three helicase superfamilies[70]. Two other smaller families of helicases are the DNA-B like and Rho families that mostly contain viral and bacterial hexameric helicases[71]. Although such classifications are useful to understand the relationships between these proteins, it should be noted that DExD/H-box proteins are likely to possess activities distinct from canonical DNA helicases because structured RNAs are more complex than long stretches of double-stranded polynucleotides. Attesting to this view are recent demonstrations that many DExD/H-box proteins are non-processive, suggesting local action rather than persistent unwinding[72,73].

1.7 Domains and motifs in DExD/H-box proteins

All members of the DExD/H-box protein family have two "RecA-like domains", so called because of the domain's structural homology to RecA protein. RecA is a crucial enzyme involved in the process of homologous recombination that generates genetic diversity[74]. RecA aids in both pairing and unwinding of DNA strands. The RecA domain consists of five α-helices and five β-sheets arranged in an alternating fashion (Figure 1.3). The β-sheets are surrounded by α-helices in three dimensional space. The RecA-like

domains that form the conserved core region are about 400 amino acids in length and contain eight motifs based on sequence conservation (Figures 1.2). DEAD-box proteins have an additional motif termed the Q-motif. Generally, the conserved motifs are considered to have specific functions, although some of the functions may overlap[64]. Motif I and II in domain 1 are ATP binding and hydrolysis motifs, whereas motifs Ia and Ib in domain 2 are RNA binding motifs. Motif III has been proposed to couple the energy of ATP binding and hydrolysis to RNA unwinding. The motifs of domain 2, IV and V, have been proposed to act similarly to motifs Ia and Ib, respectively, by binding RNA. In addition to having RecA-like domains, all SF1 and SF2 family members have additional extensions or insertions[64] (Figure 1.3).

Motif I, also known as the Walker A motif, has the consensus AxTGoGKT. Mutations in these residues strongly affect ATP binding and hydrolysis, suggesting that they form interactions with the magnesium coordinated phosphate backbone[75,76]. In one study, where the N-terminal glycine or the C-terminal lysine was mutated, only the lysine mutation affected the ability of dATP to crosslink to the protein, suggesting that the lysine plays an important role in ATP binding[76]. However, the glycine was found to be important not for ATP binding, but for ATP hydrolysis and helicase activities[77]. Motif I makes interactions with motif Q and motifs II and III[64]. Pause and Sonenberg demonstrated that all four motifs (I, II, III and VI) reduced ATP binding or hydrolysis and affected RNA binding activities, suggesting that these activities are inter-dependent[75].

Motif I is followed by motifs Ia and Ib, whose consensus sequences are PTRELA and TPGR respectively. Motifs Ia and Ib along with motifs IV and V, which are structurally similar to Ia and Ib[78], are implicated in RNA binding[79]. However, their exact roles are yet to be clarified.

Motif II or the Walker B motif gives the family its name. The highly conserved 'DEAD' or 'DEAH' motifs are present in many known DExD/H-box proteins; however, the alanine can be replaced, as in the human splicing factor UAP56, by other residues such as cysteine. Motif II makes functional interactions with the magnesium coordinated phosphate of ATP. Mutating the glutamic acid residue to alanine results in substantial loss of ATP hydrolysis activity[75] and, consequently, RNA unwinding activity. This strongly suggests that the two activities are coupled[80]. Motif I and motif VI interact and show co-variation. The last D residue of DEAD interacts with the first H residue of motif VI and the last H residue makes an interaction with the Q residue of motif VI[64].

Motif III is proposed to couple ATP hydrolysis and RNA unwinding activities. Conserved residues in motif III (also called SAT motif because of the residues S-A-T) are responsible for flexibility of this region and are proposed to transmit the molecular changes accompanying ATP hydrolysis to conformational changes in RNA[81]. In a DEAD-box protein, eIF4A, when the amino acids SAT were converted to AAA, the resulting protein was less efficient in unwinding activity but not RNA binding and ATP binding activities[75]. Schwer and Meszaros found that individual alanine substitutions in 'S' or 'T' of SAT motif in splicing factor Prp22 resulted in proteins defective in duplex

unwinding and mRNA release activities although they retained their ability to hydrolyze ATP[82].

Motifs IV and V are not well characterized. In the yeast DExD/H-box protein, Dhh1P, motif V was proposed to interact with motif I and the newly discovered Q motif, bringing these two domains together[83].

Motif VI has the consensus sequence HRIGRTGR in DEAD-box proteins and QRxGRxGR in DEAH-box proteins. The sequences are always associated with the specific Walker B motifs. This co-variation could imply evolutionary interrelationships between the two domains and could be important for their prescribed activities. Any mutation in motif VI of the DEAD-box protein, eIF4A, affects the helicase activity. However, only some mutations affect ATP binding and hydrolysis, consistent with the view that the motif is involved in coupling of ATP binding and hydrolysis to helicase activity[84].

Recently, a novel motif was discovered in Patrick Linder's lab[85,86]. This motif, called the Q motif due to the highly conserved glutamine residue, is only present in DEAD-box proteins and forms a cap structure in the RecA-like domain 1. The Q motif is proposed to act as a sensor for the bound nucleotide and modulate RNA binding activity[86]. The motifs listed above are the ones that have been characterized at this time. However, like the discovery of the Q motif, there may be other motifs that are yet to be discovered.

1.8 Three-dimensional structure of DExD/H-box proteins

Atomic level structural information is now available for a number of DExD/H-box proteins. These structures provide unprecedented detail on the orientation of the two domains and on the various motif interactions with substrate polynucleotide and NTP analogs. However, all of the known crystal structures consist only of the core region, and this may reflect the difficulty of the extensions to crystallize. The first crystal structure of any helicase belonging to either families (possessing DExx in motif II) emerged in 1996 from Dale Wigley's lab[87]. The structure was that of the DNA helicase from *Bacillus thermophilus* bacteria, PcrA. The striking feature of this structure is the helicase having a strong resemblance to the DNA recombination factor, RecA[87]. In addition to having two RecA-like domains, the structure also shows the presence of insertions in both domains. Later, two structures of *Escherichia coli* Rep helicase, one in complex with ssDNA and ADP, and the other without ADP, were solved[88]. The latter structure also shows two forms of Rep in the asymmetric unit, the open and the closed conformation and suggests that Rep can undergo very large conformational changes to allow for transient clamping down of the Rep monomer on DNA in accordance with the "active rolling" model of translocation[89]. This model requires that the protein be a multimer, at least a dimer, to be a functional helicase. The open conformation of Rep also closely resembles the structure of PcrA. Later, Velankar *et. al.* solved the crystal structure of PcrA with either ATP or a substrate resembling ADP[90]. Both structures were in monomeric form, suggesting that dimerization is not a requirement, supporting the "inchworm" model.

The first crystal structure from the SF2 family was that of a DEAH-box protein, the Hepatitis C virus (HCV) NS3 helicase. Its nucleotide binding pocket shows striking resemblance to the DNA helicases, Rep and PcrA of the SF1 family[91]. The structure shows the helicase in complex with a single stranded DNA. Two RecA-like domains are observed. In contrast to Rep/PcrA, however, only one insertion is present. The structure shows DNA bound in the groove between the first two domains and the third domain. The insertions, typical of DNA helicases, are absent in putative RNA helicases. However, the insertions may be functionally analogous to the N- and C-terminal extensions of RNA helicases[64].

The first DEAD-box protein to be crystallized was the N-terminal domain of eIF4A. The structure shows unequivocal resemblance to NS3 and PcrA, suggesting a common mode of ATP binding[92]. Other DExD/H-box proteins that have been crystallized are bstDEAD, UAP56, mjDEAD, eIF4AIII and eIF4A[78,93-95]. Most DExD/H-box proteins crystallized have been found to be monomeric without DNA or RNA bound. The crystal structure of mjDEAD is the only structure that shows dimeric molecules, although it is unclear if dimerization is required for biological activity[78]. The first crystal structure of a DExD/H-box protein involved in pre-mRNA splicing was that of UAP56[95]. The structure in the presence or absence of ADP suggests a conformational change that could be functionally important. Caruthers *et. al*. solved the structure of the full length eIF4A. The structure shows the protein as a dumbbell shaped molecule with a flexible linker[94]. The core structure of bstDEAD is similar to that of other DEAD-box proteins. However, there is a “closed” domain that sterically blocks binding of ATP, suggesting that this could be

an important regulatory mechanism[93]. ATP binding in the cleft region between the two domains may help orient the two RecA-like domains relative to each other[96]. Thus, ATP binding and hydrolysis may cause a major conformational change in the RNA[97].

Despite the apparent similarity to conventional DNA helicases, it has been difficult to determine how local separation of RNA duplexes is achieved by the action of DExD/H-box proteins[98]. Sengoku *et. al*. provided the first crystal structure of a DEAD-box protein, (*Drosophila* protein VASA) in complex with RNA and an ATP analog[99]. The structure shows how local unwinding may be mediated by sharp bending of the RNA strand. Additionally, it shows that general specificity without preference to a particular RNA sequence can be achieved through predominant interactions of VASA with the RNA occurring through 2' hydroxyl groups and phosphates of RNA and not the bases.

Although the crystal structures are useful starting points for understanding the basic mechanism of DExD/H-box proteins, which could be common to many because of a conserved core, they provide few answers to mechanistic questions concerning the dynamic role of DExD/H-box proteins in mediating conformational rearrangements in RNAs and in what roles, if any, the less conserved extensions of these proteins are involved. Mechanistic studies, on the other hand, have provided extensive information about many basic features of DExD/H-box proteins such as unwinding activity[73,100-106], nucleotide usage[84,85,105-108], cycles of nucleotide binding and induction of conformational change[109,110]. Kinetic studies have also provided evidence for the role of extensions for at least one protein, DbpA, whose 76-residue carboxyl terminal domain is proposed to act as targeting domain and make specific interactions with the substrate[111-114]. The further

combination of biochemical and structural studies may provide more insight to the mechanism and targeting of these proteins to their RNA substrates.

1.9 Specificity of DExD/H-box proteins toward their targets

In general, DExD/H-box proteins are considered to be non-specific motors that disrupt structures independent of sequence. On the other hand, the participation of many DExD/H-box proteins in specific biological processes implies that they act in the context of defined RNA/RNP complexes, suggesting specific recognition of defined structural motifs. *In vitro* studies, however, do not provide strong evidence for structural targeting or high affinity binding. This is presumably because the targets are poorly defined in biochemical studies. For example, a structured RNA may transiently fold to a conformation that acts as a target, but such structures are, by necessity, difficult to populate for *in vitro* studies. Purification of DExD/H-box proteins by over-expression in bacteria yields large amount of proteins for biochemical studies, but they may lack the necessary co-factors that normally associate with them *in vivo*. The association of protein co-factors has been reported to stimulate the activity of one DExD/H-box protein eIF4A[115,116] (Table 1.1). From the protein perspective, the N and C terminal extensions of DExD/H-box proteins have long been implicated to act as targeting domains. However, the evidence is clear only for one protein so far, the *E.coli* protein DbpA and its *B.subtilis* homolog, YxiN. It remains to be seen how the various DExD/H-box proteins that act in different processes are targeted to act on certain RNAs or RNPs.

eIF4A is a DEAD-box protein known to interact with several factors at the 5' end of the eukaryotic mRNA to facilitate initiation of translation. Its helicase activity is proposed to disrupt secondary structure and to help the formation of the initiation complex. eIF4A forms part of the larger eIF4F complex which contains a scaffolding protein, eIF4G and a cap-binding protein, eIF4E. eIF4A makes direct interactions with the scaffolding protein, eIF4G[117]. eIF4A is stimulated by eIF4G with its two interacting sites and the sites differ in the way they stimulate the RNA-dependent ATPase rates[115,116]. The ATPase activity of eIF4A has also been known to be stimulated by another factor, eIF4B[108]. The stimulation of eIF4A by other factors suggests that its activity might be localized to its specific target, the 5' end of mRNAs for example, by these factors.

The proposed role of N and C terminal extensions is to provide specificity by interacting with their specific RNA and RNP targets and stimulating the activity of DExD/H-box proteins, although it is unclear if the substrate-dependent stimulation of activity is universally true. For one DExD/H-box protein, DbpA, the activity is stimulated by a specific fragment of the 23S ribosomal RNA[111-113] (Table 1.1). To investigate the specific portion of 23S rRNA responsible for this stimulation, Tsu *et. al.* carefully made sub-fragments from this RNA and monitored ATPase activity. The results provide strong evidence that hairpin 92 with at least one extension either on the 5' or 3' end is sufficient to provide maximum stimulation of the ATPase activity[114]. The protein domain that is responsible for the specific stimulation was localized to the C-terminus, and the transfer of this segment from a DbpA ortholog (Yxin) in *B.subtilis* to the non-specific core domain of DbpA resulted in a chimeric protein which possessed specific 23S rRNA

stimulated activity[118]. The carboxyl-terminal domain of DbpA has been crystallized and it resembles the RRM (RNA recognition motif) present in many RNA binding proteins[119].

Pre-mRNA splicing involves several proteins of the DExD/H-box family which may mediate various structural transitions. The strongest evidence for specific RNA-dependent stimulation during splicing is provided by Prp16 and Prp5 (Table 1.1). The second transesterification step in splicing is facilitated by Prp16. Several studies have shown that its ATPase activity is stimulated by U2 and U6 RNAs, suggesting that some features of these RNAs are recognized specifically by Prp16[120]. Prp5 is a DExD/H-box protein that has a role in assembly of the pre-spliceomal complexes. Specifically, Prp5 has been proposed to bring U2 snRNP to the branchpoint[65]. Comparison of ATPase activity in the presence of the full length U2 snRNA versus other snRNAs or non-specific RNAs showed that the activity was stimulated only by the U2 snRNA, suggesting a specific interaction between the protein and the RNA[121].

Recently, the processivity of HCV NS3 was shown to be enhanced by *E.coli* single stranded binding protein (SSB) on DNA substrates[122]. NS3 is unusual among DExD/H-box proteins because it can bind both RNA and DNA substrates[123]. It remains to be seen whether NS3 can also have such protein stimulated processive action on RNAs.

1.10 Involvement of DExD/H-box proteins in RNA-mediated cellular processes

Genetic, mutational, and biochemical studies have established clear roles for many DExD/H-box proteins in RNA mediated processes. Many of these proteins are essential, since mutations of the residues have resulted in unviable organisms. The ability

to modify structures in a non-specific manner makes them versatile enzymes of RNAs and RNPs. Almost every cellular process that involves RNA conformational changes requires DExD/H-box proteins. The processes include, but are not limited to, transcription, ribosome assembly, prokaryotic and eukaryotic RNA decay, pre-mRNA splicing, and nuclear export of tRNAs and mRNAs in eukaryotes. Besides playing indispensable roles in molecular processes, DExD/H-box proteins are also imperative for proper functioning of cellular organelles, such as mitochondria and chloroplasts. In addition, clear evidence is emerging for multiple roles of some DExD/H-box proteins.

1.10.1 Transcription

Transcription in eukaryotes is an elaborate process involving association of several transcription factors, both general and specific, with the 5' and 3' ends of open reading frames (ORFs) of DNA. Although the precise roles of several DExD/H-box proteins involved in transcription remain to be defined, current evidence suggests that many are multifunctional, interacting with multiple substrates, such as nascent mRNAs, transcription factors, and potentially promoters on DNA[124]. Examples of DExD/H-box proteins involved are Dhx9, Ddx20, Ddx5 and Ddx17, and these have been implicated to function as transcriptional repressors or activators[124]. RNA helicase A (RHA) (or Dhx9) belonging to the DExH family, has been shown to act as a mediator for the interaction of a co-activator, CBP with RNA pol II[125]. The DEAD-box protein, Ddx20 or Gemin 3, has roles in snRNP assembly and in both transcription activation and repression[124] .

1.10.2 Pre-mRNA splicing

In eukaryotes, mature mRNA is processed from pre-mRNA by the spliceosome. Pre-mRNA splicing involves five snRNPs containing five snRNAs and over 200 proteins[126]. For formation of a functional complex and for splicing to occur, the pre-mRNA and snRNPs have to undergo precise rearrangements that result ultimately in a lariat intron and a spliced exon. The traditional step-wise view of this assembly process involves association of U1 and U2 to the 5' splice site and the branch point, respectively, followed by association of the tri-snRNP, U4/U6.U5. This view has been challenged recently with the isolation of penta-snRNP in yeast[127]. The penta-snRNP contains all five known snRNPs preassembled[128] and is capable of being fully functional[129,130]. Instead of the cascade hypothesis of stepwise assembly, a holoenzyme hypothesis has been postulated[131]. Regardless, rearrangement of snRNPs with mRNAs requires both formation and disruption of base pairs (Figure 1.4) and association and dissociation of trans-acting protein factors that are the hallmark of spliceosomes[128]. Many of these actions are performed by DExD/H-box proteins that interact with the spliceosome transiently. In yeast, where pre-mRNA splicing has been studied extensively[101], both DEAD-box and DEAH-box proteins are involved, and they are called precursor RNA processing (or Prp) factors. The association of U1 to the 5' splice site is the only step not known to require any DExD/H-box proteins. The association of U2 to the branch point is facilitated by Prp5. This protein has been shown to possess U2 snRNA stimulated ATPase activity; therefore, Prp5 may act in a substrate targeted manner. Another protein, sub2p, is also required. The breaking of U4/U6.U5 and the separation of the extensively

base-paired U4/U6, together with replacement of U1 by U6, requires Prp28[132]. The first transesterification step is facilitated by Prp2[133] and the second step by Prp16[134,135]. Prp22 has been implicated in release of the mature mRNA[136,137] (exon) and Prp43 is considered a recycling factor to help dislodge protein factors associated with snRNAs from the lariat after splicing[138,139].

1.10.3 Nuclear export of tRNAs and mRNAs

In eukaryotes, RNAs originate in the nucleus and must be exported from the nucleus and into the cytoplasm. So far, mRNA export is known to involve primarily two DExD/H-box proteins, one on the nuclear side and the other on the cytoplasmic side of the nuclear pore complex[140]. On the nuclear side, the Yeast DECD protein, Sub2p, and its mammalian counterpart, UAP56 are required[141,142]. UAP56 was shown to interact with an mRNA export factor, Aly, and to be present with Aly in spliced mRNA in complex with other proteins[143]. The data suggest that UAP56 recruits Aly for mRNA export. Similarly, Sub2p was shown to associate with an export factor, Yra1[144], which is essential in yeast. In addition to spliced mRNA, intronless mRNAs were also exported, suggesting that splicing is not a prerequisite for Sub2p mediated mRNA export[145]. Analogous to UAP56/Sub2p mediated export on the nuclear side, the cytoplasmic export is performed by the DExD/H-box protein, Dbp5. Dbp5 was found exclusively on the cytoplasmic side of the nuclear pore[146] and its mutation resulted in accumulation of mRNAs in the nucleus[147].

1.10.4 Ribosome assembly

The ribosome is an enormous RNP complex that requires folding of extremely long RNAs (and, thus, is prone to formation of kinetic traps[148]) and precise positioning of ribosomal proteins during folding. It is not surprising that many DEAD-box proteins are involved in facilitating the assembly process. Of the 24 DExD/H-box proteins discovered in yeast, 14 have been found to be involved in ribosome assembly process[64,149]. In *Escherichia coli*, three DEAD-box proteins, DbpA, DeaD and SrmB are all essential for ribosome assembly[150]. DbpA specifically interacts with a fragment of 23S RNA and is presumably involved in its assembly[111-114], and SrmB is implicated during steps that involve rearrangement during ribosome assembly[151].

1.10.5 Translation initiation

Translation is a major step during gene regulation that is primarily executed by the ribosome, although several other factors are indispensable. The process is generally divided into 3 phases: initiation, elongation and termination. Eukaryotic initiation is more complicated than prokaryotic initiation, and DExD/H-box proteins have been shown to play crucial roles by presumably disrupting structure and facilitating binding of the ribosomal subunit[117]. The eukaryotic translation initiation factor, eIF4A has been extensively characterized[117]. It contains only the core and is one of the simplest DExD/H-box proteins, structure-wise. In fact, the family of DEAD-box protein was identified based on sequence analysis of eIF4A from different species[68]. Initial biochemical characterization demonstrated that eIF4A can act non-processively[152]. Several factors are

implicated in enhancing the activity of eIF4A. eIFB was shown to clearly stimulate the helicase activity of eIF4A[152]. Other factors that could possibly stimulate the activity are eIf4H[153] and eIF4G [115,116] (discussed in section 1.9). Binding of substrates, ATP and RNA, can cause changes in affinity and conformation of DExD/H-box proteins. Lorsch and Herschlag demonstrated that cycles of ATP/ADP binding are coupled to affinity for single stranded RNA and showed how such an activity can induce conformational changes in eIF4A[109,110]. Like eIF4A, Ded1 from yeast is also essential for translation initiation[154] although its exact role is not yet defined. Ded1 has also been implicated in splicing and ribosomal assembly from proteomic characterization[127,155].

1.10.6 RNA decay

RNA decay is one of the most important modes of gene regulation in prokaryotes, and it plays crucial roles in eukaryotes[150]. In *E.coli*, the 3' to 5' degradosome complex contains a DEAD-box protein, Rh1B and is part of a larger degradosome complex with an endonuclease and exonuclease[64,156]. This illustrates one way in which specificity can be achieved, i.e., by making the DExD/H-box protein part of the degradation machinery. There is evidence that Rh1B can also have activity independent of the degradosome complex[157]. In metazoans, the DEAD-box protein eIF4AIII is involved in nonsense-mediated decay (NMD) and is deposited in the exon junction complex (EJC) following splicing. eIF4AIII is associated with three other proteins in this complex, and the entire complex appears to undergo conformational changes during assembly and disassembly of

the EJC[158]. Other DExD/H-box proteins in eukaryotic exosomes, that have not been well characterized, include DEAH-box, Suv3 and Ski2 proteins.

1.10.7 Organelle function

Mitochondria and chloroplasts have their own genome encoding RNA and proteins. Both contain group I introns, which are self-splicing RNAs. *Neurospora crassa* DEAD-box protein CYT-19 and related yeast protein Mss116p are nuclear-encoded proteins and act as chaperones for the folding of a diverse set of group I intron splicing in mitochondria[107,159]. In yeast, group II intron splicing requires another DEAD-box protein, Mrh4, in addition to Mss116p[159,160].

The functions of CYT-19 have been well characterized biochemically. CYT-19 facilitates folding *in vivo* and accelerates splicing *in vitro*, suggesting that it functions as a chaperone by resolving kinetic traps[107]. Additional evidence for chaperone action comes from structural changes induced in the secondary structure of a well-characterized misfolded species of *Tetrahymena* group I ribozyme[107]. CYT-19 has also been shown to interact non-specifically with the ribozyme and unwind loosely associated duplexes[73]. It has also been shown to recognize the general RNA structure with the help of its C-terminal domain[161].

1.10.8 Multiple functions

Many DExD/H-box proteins are involved in multiple processes[100]. An example is the splicing factor, Sub2p, which is also required for mRNA export[144]. More recently, the

spliceosome disassembly factor Prp43 was shown to be involved in ribosome assembly[162-164].

1.11 Studies of DExD/H-box proteins on simple model substrates

The involvement of DExD/H proteins in diverse cellular processes immediately suggests the existence of more than one functional role for these proteins. Studies of DExD/H-box proteins on model RNA and protein substrates have indeed revealed the enormous potential of these proteins in modulating both RNA and RNP structure. Although DExD/H-box proteins are synonymous with RNA helicases, the latter name can be somewhat misleading because canonical helicase activity is not a hallmark of DExD/H-box proteins. Indeed, DExD/H-box proteins have been shown to possess activities that are not limited to and separate from helicase activity[72,165]. The varied functions, including processive helicase activity, emphasize the diversity of actions for these proteins. Comparative studies may shed light onto the features that direct individual DExD/H-box proteins to specific functions.

Studies on simple RNA duplexes and RNA-protein complexes demonstrate the capability of DExD/H-box proteins. The DExD/H-box proteins were thought to require single-stranded RNA targets with a defined polarity, either 5’ to 3’ or 3’ to 5’ for loading. However, eIF4A was shown to unwind in a non-polar fashion either from the 5’ end or from the 3’ end[11]. Recently, eIF4A has been shown to unwind blunt-ended RNA duplexes[152,153,166], strongly suggesting that the loading strand may not be a requirement. Additionally, the protein may be capable of local action[152].

Processivity is another property of many canonical DNA helicases; however, the RNA substrates of DExD/H-box proteins do not have long, extended duplexes. Therefore, there is no expectation of processive unwinding by DExD/H-box proteins. Consistent with this idea, eIF4A has been shown to unwind duplexes in a non-processive manner[152]. Again, this property is not universal as processive unwinding has been demonstrated for NPH-II[102] and HCV-NS3[104]. NPH-II and HCV-NS3 are involved in viral replication, and it remains to be seen if this processive action is a unique property of the DExD/H-box proteins involved in virus replication.

For many DExD/H-box proteins, non-processive local disruption of duplexes seems to be the general mode of action[72,73,167], and it is consistent with the recent crystal structure of a DEAD-box protein, VASA[99]. The structure shows how local unwinding can be achieved by contorting and inducing a bend in the structure (Figure 1.2). Strikingly, in addition to unwinding, annealing of simple duplexes has also been demonstrated for several DExD/H-box proteins[168]. Annealing and unwinding by Ded1 have been shown to be modulated by the ratio of ATP and ADP concentrations with higher ATP favoring unwinding and higher ADP favoring annealing[106].

DExD/H-box proteins have been demonstrated to displace simple protein-RNA complexes. The DExD/H-box protein, NPH-II, was shown to accelerate dissociation of U1A protein by more than three orders of magnitude without affecting the efficiency of unwinding of the duplex to which the protein was bound[169]. Active disruption of protein suggests that DExD/H-box proteins can act as "RNPases"[170]. However, it was unclear whether protein displacement activity was a result of duplex unwinding or an

independent activity. Using two model systems in which proteins are bound to single stranded RNAs, Fairman *et. al* showed that two DExD/H-box proteins, NPH-II and Ded1, are each capable of displacing proteins without unwinding duplexes, although, there appears to be some differences in their efficiencies[165]. Thus, DExD/H-box proteins can not only unwind/anneal duplexes but can also displace proteins due to their active translocation[171], although a recent study suggests a more localized action for Ded1[72]. The protein displacement activity has been documented *in vivo* for several DExD/H-box proteins, which includes displacement of U1snRNA by Prp28, of Mud2p by Sub2p and of Cus2p by Prp5[12].

The studies described above demonstrate the versatility of DExD/H-box proteins in remodeling RNAs and RNPs. Although such studies are tremendously useful in exploring how DExD/H-box proteins can affect simple duplex structures, induce changes in local structure, or modify proteins-RNA interaction, these activities need to be demonstrated in the context of a structured RNA, for large biological RNAs are extensively structured. Further, folding of structured RNAs involves well-defined pathways and it remains unclear how DExD/H-box proteins act in the context of specific folding pathways and to what structural features, if any, these proteins are targeted to act.

1.12 *In vitro* chaperone activities of non-specific RNA binding proteins

An "RNA chaperone" is a protein that induces structural changes in the RNA and must be dispensable following the action[8]. Although DExD/H-box proteins are primary candidates for this category, such activities are not restricted solely to DExD/H-box

proteins. Prior to the discovery of DExD/H-box proteins, various single stranded RNA binding proteins have been shown to possess RNA remodeling activities. One class, the hnRNP proteins, coat nascent mRNAs formed during transcription and are thought to either promote or restrict formation of structure. These proteins have been extensively characterized, and many have been found to bind several types of RNAs, independent of sequence (Ref[172] and references therein). The nucleocapsid protein (NC) of the HIV virus has been shown to possess similar activities. Studies showed that NC enhances catalysis of hammerhead ribozyme by acting as a chaperone *in vitro*. The enhanced catalysis was the result of two actions: 1) NC increased the turnover (k_{cat}) by dissociation of the product strands, and 2) NC increased specificity (k_{cat}/K_m) by increasing annealing rate of the substrate strands[173,174]. Other proteins have been shown to act in an analogous manner including the calf thymus protein, UP1[175], the *Escherichia coli* proteins StpA[176-178], Hfq[179] and the ribosomal small subunit protein, S12[180].

1.13 DExD/H-box proteins as general RNA chaperones: CYT-19 and Mss116p

The aforementioned studies with non-specific proteins suggest the possibility that these proteins may function as RNA chaperones *in vivo*. A direct demonstration that a protein functions as an RNA chaperone *in vivo* came in 2002. The *Neurospora crassa* DEAD-box protein, CYT-19, is the first DExD/H-box protein shown to possess RNA chaperone activity *in vivo*[107]. CYT-19 is encoded by a nuclear gene and is targeted to mitochondria for action on group I introns, which are encoded by the mitochondrial genome. Mutations in the CYT-19 gene produced splicing defects *in vivo* but were

rescued by over-expression of CYT-19 from an exogenous plasmid *in vivo* at the non-permissive temperature of 25 °C. Additionally, the same splicing defects were overcome by growing the organisms at the permissive temperature of 37 °C, suggesting that the necessary activation energy for splicing was provided by CYT-19[107]. *In vitro* splicing assays performed in the presence of CYT-19 showed acceleration of splicing and produced a greater amount of spliced products, presumably by resolution of kinetic traps by CYT-19[107]. Finally, footprinting analysis with a well-characterized, misfolded structured RNA from *Tetrahymena* strongly suggests that CYT-19 induces structural changes in RNA as shown by clear changes in the protection/enhancement pattern[107]. Later studies also found that the structural changes can be maintained after complete elimination of CYT-19 action, suggesting that it acts as a true RNA chaperone[73,181,182].

CYT-19 and a related protein, Mss116p, can act on a number of different related RNAs. The first evidence came from the finding that mutations in nuclear MSS116 gene produced splicing defects in RNA transcripts of mitochondrial cytochrome b (cob) and cytochrome c oxidase subunit I (cox1) ORFs[183]. In a different study, over-expression of Mss116p resulted in a marked increase of group II bI1 intron splicing, suggesting an important role in splicing of this group II intron[184]. More recently, Mss116p was found to produce splicing defects in both group I and group II introns in yeast[159]. The CYT-19 mutation also produced splicing defects in several group I introns[107]. Importantly, CYT-19 has been found to rescue, at least partially, splicing for both group I and group II introns as well. Together, the data suggest that both CYT-19 and Mss116p act promiscuously on a variety of group I and group II introns and that some DExD/H-box

proteins can function as general RNA chaperones by recognizing global features of RNAs.

Originally, CYT-19 action was thought to require another protein, CYT-18. This model was based on the earlier finding that three introns that required CYT-19 for splicing (mtLSU, *ND1*-I1 and *cob*-I2) also required CYT-18. Additional introns were tested that were independent of CYT-18 and did not require CYT-19, suggesting that CYT-18 could act as a targeting factor for CYT-19 action. However, later studies revealed that CYT-19 could act on other group I and even on group II introns independent of CYT-18, indicating that the interaction with RNA is more general and a protein cofactor is not a requirement for CYT-19 action[159,181].

Although CYT-19 and Mss116p are the first examples of DExD/H-box proteins that act on RNAs by recognizing general features of RNAs, many DExD/H-box act in the context of defined processes and may require additional co-factors. Specificity of DExD/H-box proteins toward RNA or RNP substrates has not been thoroughly investigated. There are very few cases where specificity has been demonstrated using *in vitro* studies. Examples include *E. coli* DbpA[111-114,118] and to a lesser extent, Prp5 in yeast[121]. The lack of observed specificity may be due to a lack of interacting partners or it could be a common feature of DExD/H-box proteins as these proteins are usually involved in multiple cellular processes and encounter different RNAs or RNPs[64]. The finding that some proteins act as general chaperones is exciting and paves the way to understanding some of the molecular mechanism for targeting and action. Probing the actions of CYT-19 and Mss116p, therefore, is imperative to obtain clues regarding the

mechanisms. The discovery of these proteins also provides an opportunity to investigate the mechanism of how general RNA chaperones function inside cells.

1.14 *Tetrahymena* group I intron as a model structured RNA for studying chaperone function

In order to understand the mechanism of the DExD/H-box RNA chaperone CYT-19 and related proteins on a structured RNA, a well characterized RNA is desired. The *Tetrahymena* group I intron was the first RNA shown to possess catalytic activity. In 1982, Cech and colleagues demonstrated that a ribosomal RNA, transcribed using purified polymerase, exhibited self-splicing activity in the absence of any proteins[185]. This was the first demonstration of such activity and coined the term "ribozyme". Later, the self-splicing construct was modified to cleave an oligonucleotide added in *trans* and act as a true RNA enzyme[186,187]. This development led to an explosion of studies and to *Tetrahymena* group I ribozyme becoming, structurally and functionally, one of the most well-characterized ribozymes.

1.14.1 Structural and biochemical characterization of *Tetrahymena* group I ribozyme

The catalytic nature of ribozymes is due to their ability to adopt specific tertiary structures. The *Tetrahymena* group I ribozyme has been a prototype for studying RNA structure and folding. The overall structure of the ribozyme consists of a catalytic core, highly conserved among group I introns, and the variable peripheral elements that surround the core and increase the overall structural stability. The core elements are P4-

P6 and P3-P9 and the peripheral elements are P5abc, P2 and P9.1 and 9.2 (Figure 1.5). The core and peripheral elements are held together by five long-range tertiary contacts. The first crystal structure of *Tetrahymena* ribozyme was that of P4-P6 domain determined by Cate. *et. al.*[188]. This 2.8 Å crystal structure showed how the two helices interact by forming tertiary contacts through the A-rich bulge and the GAAA tetraloop-receptor, which lead to tight packing[188]. Golden *et. al*. solved the first crystal structure of the core, lacking the peripheral elements such as P1, P2-2.1 and P9-9.1[189]. This structure showed how the domains are packed together by the formation of a base triple that bridges the two domains together[190]. This structure also demonstrated for the first time that the active site of ribozymes can be pre-organized, similar to protein enzymes, to perform catalysis. More recently, a 3.8 Å crystal structure was obtained from the Cech lab that gave a refined view of the active site organization in terms of the positioning of the guanosine nucleophile and the catalytic metal ions[191].

Enzymatic characterization of the *Tetrahymena* group I ribozyme began almost immediately following the discovery of its catalytic activity. Many features of the ribozyme were similar to that of some protein enzymes, such as the existence of a pre-organized active site[189], use of binding energy[192] and metal ion for catalysis[193-195]. Certain other features were unique to the ribozymes. Because the catalyst is an RNA molecule, structural metal ions are required for folding. The ions counteract the highly negative phosphodiester backbone of RNA and allow the RNA to fold into specific structures. Catalysis can occur only if the ribozyme folds to the correct conformation, which includes a binding pocket for the nucleophile guanosine and a docking site for the

substrate RNA. Upon precise positioning of the residues, guanosine attacks and cleaves the substrate RNA strand[196].

1.14.2 The folding pathway of *Tetrahymena* ribozyme

The folding of RNAs is in general hierarchical with the formation of stable secondary elements preceding the formation of tertiary elements. Folding of the *Tetrahymena* ribozyme is an example of hierarchical folding, with early and late events in the pathway[197]. The hallmark of the ribozyme is that it folds to a specific and long-lived misfolded conformation late in the folding pathway (discussed in 1.14.3). Pioneering studies by Zarrinkar and Williamson[198] and Sclavi *et. al.*[199] provided details on the nature of the early folding events. Using hybridization of a DNA oligonucleotide and specific cleavage of the resulting RNA-DNA duplex, Zarrinkar and Williamson[198] demonstrated that formation of P4-P6 occurs on the seconds timescale and is followed by the formation of P3-P7 on the minute timescale. Using time resolved hydroxyl radical footprinting, Sclavi *et. al.*[199] obtained a more quantitative measure of these events, showing that the P4-P6 domain forms early with a rate constant of 2-3 s^{-1}, whereas the regions that are close to P3-P7 (J5/4) and the regions that mediate tertiary contacts with P3-P9 (P2, P2.1 and P9.1) are protected much more slowly at 0.2 to 0.4 s^{-1}. Finally, folding of P3-P7-P9 occurred at an even slower pace of about 0.02 to 0.06 s^{-1}. Using a series of selection for mutations in RNA that accelerate the rate of folding, Treiber *et. al*. found five mutants that had single nucleotide substitutions, all of which were concentrated in the P4-P6 region[200]. These results were surprising considering that the slow folding steps were

shown to occur in the P3-P7 region[199]. The results suggested that native interactions of P4-P6 to the core region stabilized a kinetic trap during folding of RNA. Indeed, addition of urea, which partially unfolds the ribozyme, accelerated folding suggesting the existence of a kinetic trap in the folding of wild type ribozyme[200].

Folding landscapes with multiple folding intermediates, as previously proposed for protein folding, have been suggested for RNA folding. RNA folding proceeds through multiple intermediates and folding pathways on a three dimensional free energy scheme[201]. Using a previously characterized oligo-nucleotide hybridization assay, Rook *et. al.* probed mutants that were faster in folding and identified more folding traps, demonstrating that the folding pathway is rugged with multiple folding intermediates[201,202].

1.14.3 The long-lived misfolded ribozyme

Catalytic activity gives a precise read-out for the formation of the active native state, differentiating it from near-native and other intermediates. Russell and Herschlag demonstrated the presence of a long-lived misfolded species in the folding of the *Tetrahymena* ribozyme at 25 °C[203]. This misfolded species forms late and behaves as a single species during its re-folding to the native state [204].

The long-lived misfolded ribozyme under various solution conditions has been carefully investigated. An alternative base-pairing termed Alt P3 was proposed to occur in the misfolded ribozyme[205]. This was based on chemical modification interference and mutational analysis. The mutations of residues that weakened or stabilized the interaction

led to a decrease or increase, respectively, in the accumulation of this conformer. A later study, however, questioned the existence of Alt P3 in the long-lived misfolded species[206]. Hydroxyl radical footprinting, fluorescence measurements and activity assays did not reveal the presence of Alt P3 but only the native P3[206]. The results of these experiments suggested a model where formation of Alt P3 leads to the accumulation of misfolded species (Figure 1.6); however, this alternative base-pairing is resolved before the formation of misfolded species. Acceleration of re-folding by an increase in temperature or urea addition and strong inhibition of re-folding upon increase in magnesium concentration suggested the existence of a substantial energetic barrier to re-fold to the native species from the misfolded species. Combining experimental data with computer modeling, a misfolded structure differing only in topology from the native species was proposed[206] (Figure 1.6). Importantly, the study emphasizes that, despite the high surface similarity between native and misfolded species, re-folding to the native state can require global unfolding.

1.14.4 *Tetrahymena* ribozyme: A good model system for probing CYT-19 action

The *Tetrahymena* group I ribozyme is a good model system. Its folding pathway has been well studied and thus the properties of the intermediates are well known. In addition, the long-lived misfolded species can be populated and it behaves as a single species (or a family or related ensemble) and refolds slowly to the native state with a rate constant that can be modulated by varying solution conditions. Thus CYT-19's action on native and misfolded species can be effectively separated for comparison, which might provide

insight into how misfolded RNAs are targeted for action. The ability to form a large fraction of misfolded species provides an opportunity to understand how misfolded species are dealt with in nature.

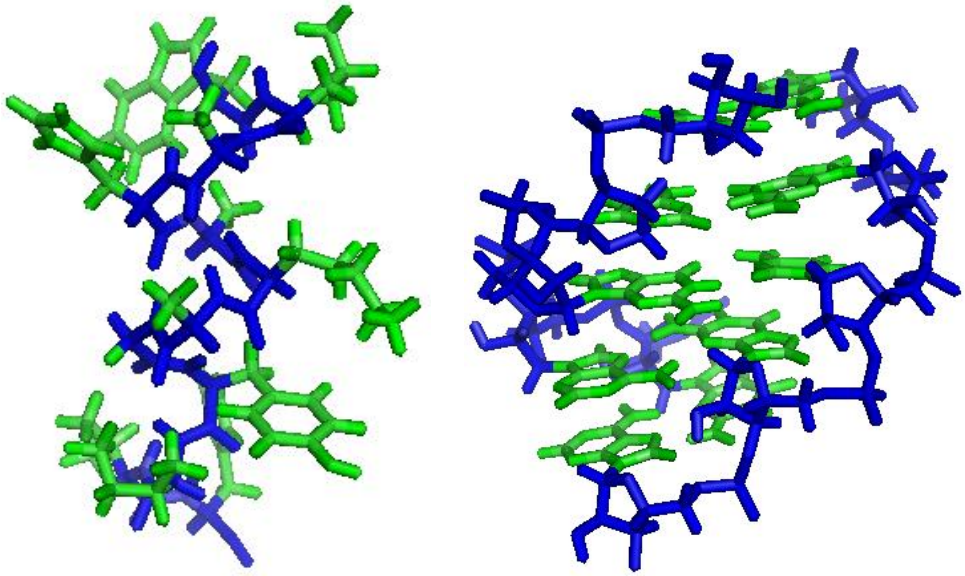


Figure 1.1: Structural elements differ in proteins and RNAs

On the left is the portion of a protein helix with polypeptide backbone (in blue) and amino acid side chains on the exterior (green), that are free to interact with other amino acid residues. On the right is the section of a short RNA helix with the phosphodiester backbone (blue) and bases (green) forming base pairs and are on the interior (Figures made using Pymol and atomic coordinates of 2dx3 and 1yfv).

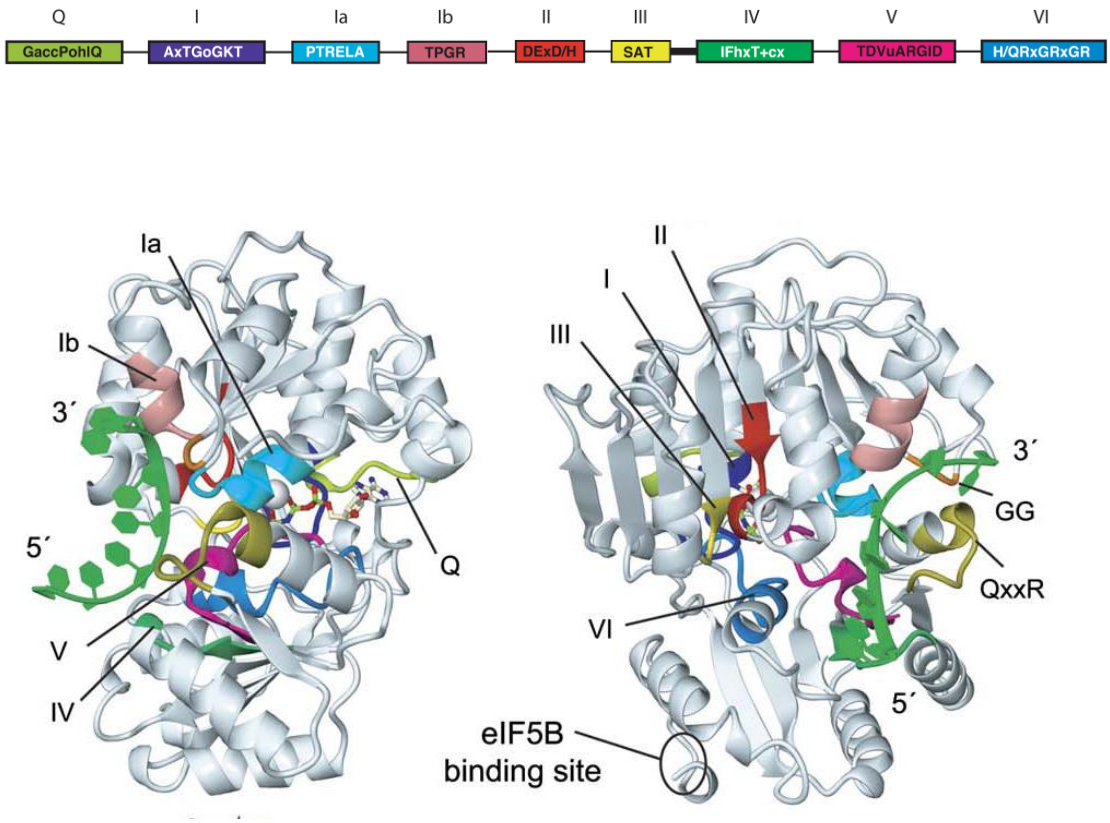


Figure 1.2: Structure of VASA, a *Drosophila* DExD/H-box protein

The structure of a DExD/H-box protein from *Drosophila*, VASA. The picture on top shows the different conserved motifs (N terminus to C terminus) that are common to many DExD/H-box proteins. The Q motif is only present in DEAD-box proteins. There are two RecA like domains connected together by a linker that is indicated by the thick black line. The picture below shows front and side views of VASA, with the different motifs color coded the same as the motifs picture above. (Adapted from Sengoku *et. al*.[99])

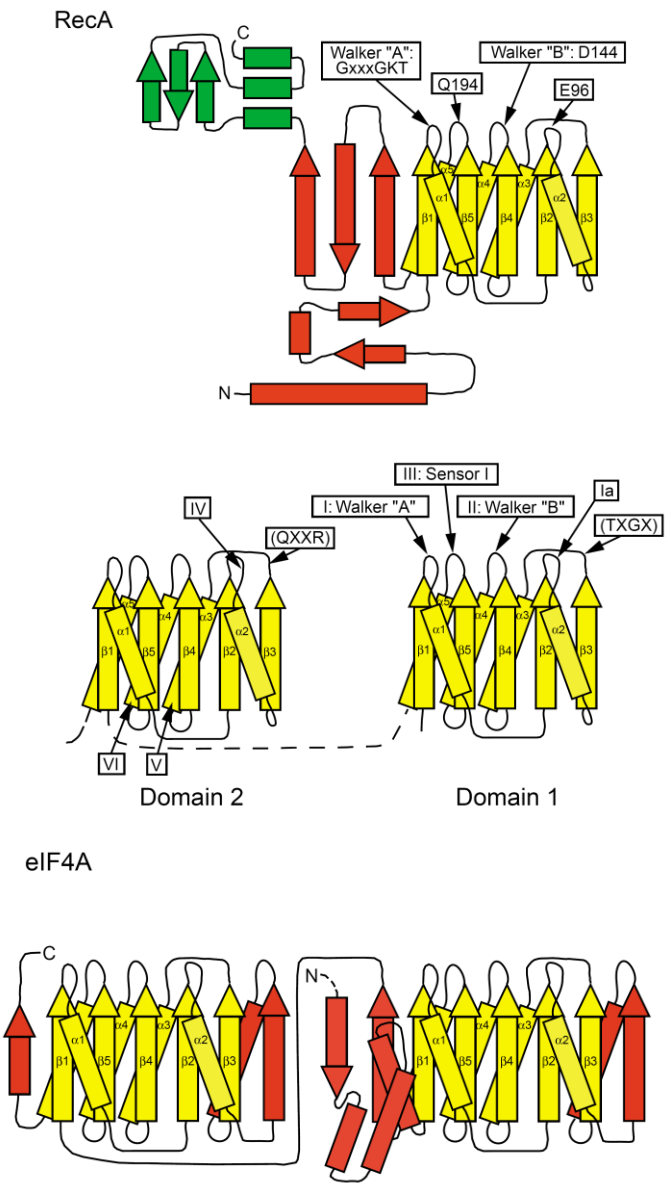


Figure 1.3: DExD/H-box proteins have two RecA-like domains

The top figure is that of RecA protein which contains the large ATP binding domain (in yellow) present in virtually all DExD/H-box proteins. Shown in red is the secondary structure not common to other proteins and in green is the structure specific to RecA protein. The middle structure shows two RecA-like domains that are present in all DExD/H-box proteins. The bottom figure shows the secondary structure of DEAD-box protein, eIF4A[96].

DExD/H-box protein	Organism	Cellular process	Proposed substrates or co-factors	Proposed role of specific substrates or co-factors
eIF4A	Eukaryotes	Translation	eIF4B[152,153]	Enhancement of helicase activity
			eIF4H[153], eIF4G[115,116]	Stimulation of activity
Prp16	Yeast	Pre-mRNA splicing	U2 and U6 snRNA[120]	Facilitate second trans-esterification step
PrP5	Yeast	Pre-mRNA splicing	U2 snRNA[121]	Recruit U2 snRNP to branch point
DbpA	*E.coli*	Ribosome assembly	Specific fragment of 23S rRNA[111-114]	Stimulation of activity
Sub2p	Yeast	Nuclear export of RNAs	Yra1[144]	mRNA export factor
UAP56	Mammals	Nuclear export of RNAs	Aly[143]	mRNA export factor
HCV NS3	Virus	Viral genome replication	*E.coli* SSB[122]	Enhancement of processivity

Table 1.1: Proposed RNA substrates and co-factors for DExD/H-box proteins

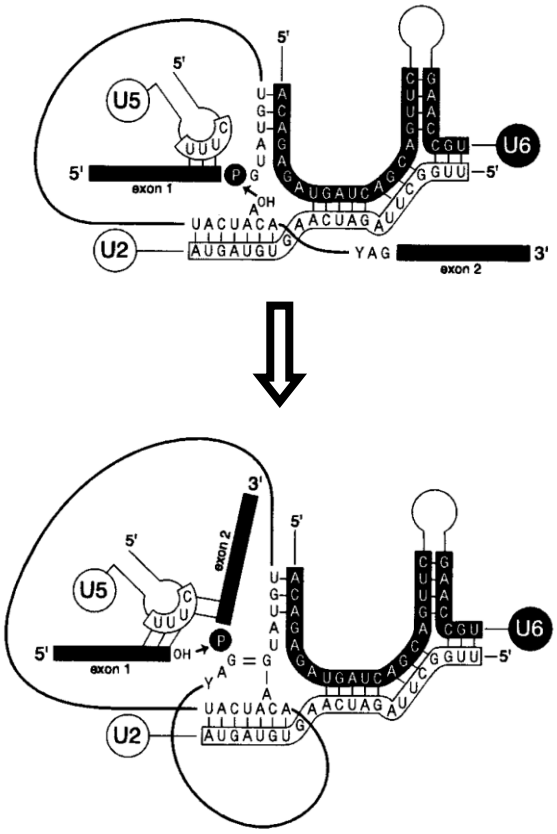


Figure 1.4: An example for structural rearrangement during pre-mRNA splicing

Large scale structural rearrangement involves formation and disruption of base pairs. Shown here is the rearrangement following the first step and before the second step of eukaryotic pre-mRNA splicing. (Adapted from Nagai, K[207].)

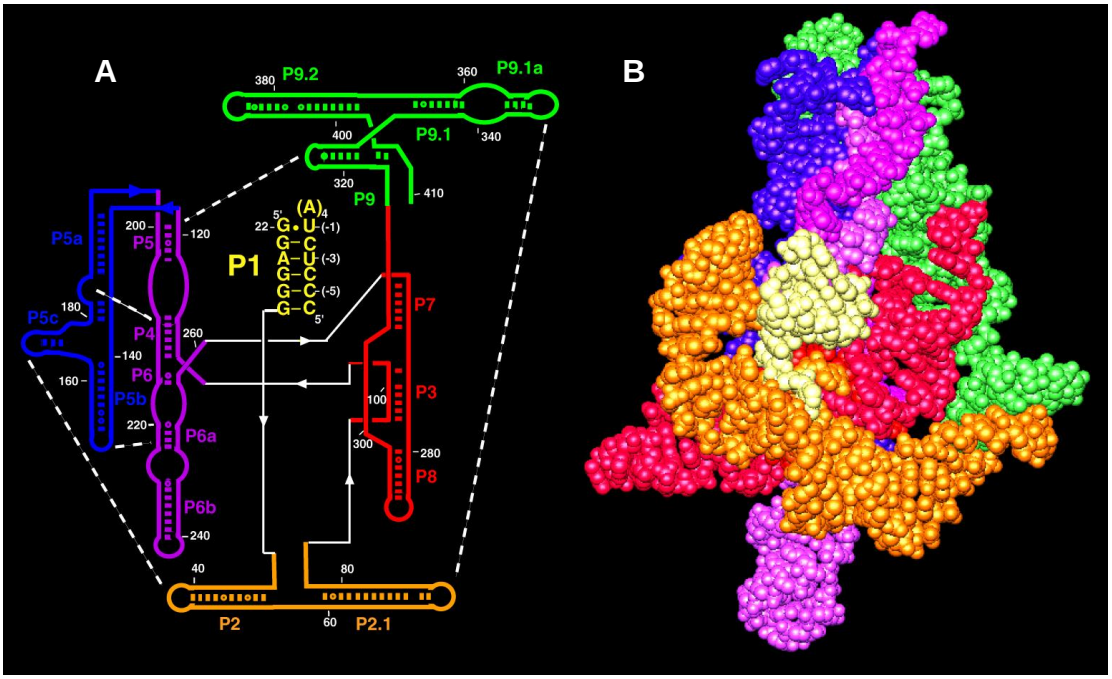


Figure 1.5: Structure of *Tetrahymena* group I ribozyme

(A) Schematic representation of the different elements of *Tetrahymena* group I ribozyme. The core elements are P4-P6 (purple), and P3-P9 (red) domains, which contain the active site that bind the P1 duplex (yellow). The peripheral elements are P9.1-9.2 (green), P2-2.2 (orange) and the P5abc sub-domain (blue). Tertiary interactions are denoted by dotted white lines. (B) Three dimensional structure of *Tetrahymena* ribozyme is color coded similar to (A).

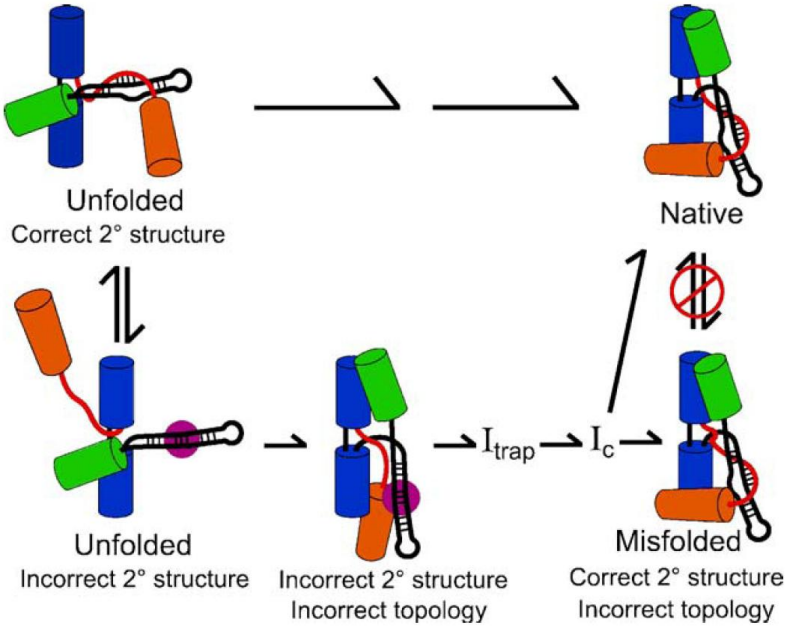


Figure 1.6: Model for the formation of the long-lived misfolded species, a topological isomer of native species in *Tetrahymena* ribozyme

Folding of *Tetrahymena* ribozyme from unfolded state with pre-formed secondary structure. Domains P4-P6 (blue), P2 (orange) and P9 (green) are shown as cylinders and P3-P8 (black) and 5' portion of P3 (red) are shown as strands. Alt P3 (purple disk), which biases folding to the long-lived misfolded species is also shown. The red "stop" symbol indicates that native and misfolded do not interconvert but require extensive unfolding to form their topological isomer (Adapted from Russell, R *et. al.*[206]).

Chapter 2: The DEAD-box proteins give ATP-dependent acceleration of folding transitions

2.1 Introduction

DExD/H-box proteins are extensively involved in RNA mediated processes and many function in the context of defined cellular processes. Although their specific substrates remain to be defined for most DExD/H-box proteins, recent evidence suggests that a subset of DExD/H-box proteins can act on a variety of group I introns found *in vivo*[107,159] and are even active in the folding of group II introns[159,181], suggesting that they have a more general action, recognizing broad features of RNA structure. These are the *Neurospora crassa* DExD/H-box protein, CYT-19, and the related yeast protein, Mss116p.

In an attempt to understand the mechanism of general RNA chaperones on a well defined misfolded RNA, pre-steady state kinetics and equilibrium binding measurements were performed using CYT-19 and the *Tetrahymena* group I ribozyme as a substrate. The catalytic activity of the ribozyme allows one to unambiguously distinguish between the native conformer and all other conformers, including misfolded ones. CYT-19 is found to accelerate the re-folding of the long-lived misfolded species to the native state and is dispensable following this re-folding, suggesting that it functions as a true RNA chaperone. In addition to demonstrating the chaperone action on the well-characterized *Tetrahymena* ribozyme, strong evidence is also provided for chaperone action of CYT-19 on one of its natural substrate, the *Neurospora* mitochondrial large subunit (mtLSU)

intron ribozyme. Results from these experiments provide novel biochemical insight into the molecular mechanism of DEAD-box protein CYT-19. Surprisingly, another DEAD-box protein, Ded1, which primarily acts on substrates unrelated to group I introns *in vivo*[154], was also found to accelerate re-folding of the misfolded *Tetrahymena* group I ribozyme. Together, these studies provide the first empirical evidence for general chaperone action on a defined misfolded state in RNA.

2.2 Materials and Methods

2.2.1 Preparation of ribozymes

L-21 version of *Tetrahymena* mtLSU (pUC18) (Figure 2.2) and a version of *Neurospora* mtLSU (pUC19) group I ribozyme (Figure 2.3) were prepared from *in vitro* run off transcription of *Sca*I enzyme cut cloning vector[206]. Transcription reactions were performed with 1 × Transcription buffer (40 mM Tris.Cl, pH 8.0, 25 mM $MgCl_2$, 2 mM spermidine, 0.01% Triton X-100), 40 mM DTT, NTPs (1 mM each of ATP, GTP, UTP and CTP), 125 µL of approx. 1 mg/mL T7 RNA polymerase, and 5 µg/mL cut vector to a final volume of 5 mL and incubated at 37 °C for 4 hours. Following incubation, 0.5 mL of 0.5M EDTA was added to stop the reaction. 0.6 mL of 3 M sodium acetate and 2M acetic acid were added and mixed. 18 mL of ethanol (100%) was added, mixed and incubated overnight at -20 °C. Following the incubation, the mixture was centrifuged at 12,000 rpm for 30 minutes and the supernatant was discarded. The RNA pellet was allowed to air dry and resuspended in 1 mL of RNase free water. The RNA was purified using RNeasy midi kit (Qiagen).

2.2.2 Purification of CYT-19

2.2.2.1 Materials:

The following is the list of chemicals (with their makers in parenthesis) used in purification of CYT-19. PEI (J.T.Baker), Tris base (J.T.Baker), IPTG (EMD chemicals), EDTA (J.T.Baker), KCl (J.T.Baker), DTT (Fisher Biotech), potassium phosphate (J.T.Baker). The equipment used in purification include a centrifuge (Beckman, Avanti J-30I) and a sonicator (Branson sonifier, VWR).

2.2.2.2 pTwin system of purification of CYT-19:

Plasmid (pTwin2, NEB) encoding CYT-19 protein contains the 524 amino acid ORF lacking the mitochondrial targeting sequence which is fused to the N and C terminal mini inteins containing a chitin binding domain. pTwin-CYT-19 vector was transformed into HMS174 (DE3) strain *Escherichia coli* cells (containing a carbenicillin resistant marker). The cells were grown in LB media at 25 °C with 50 μg/mL carbenicillin to an $O.D._{600}$ value of about 0.5 and then induced with 1 mM IPTG for 16 to 20 hours at 25 °C. The cells were harvested by centrifugation at 6000 rpm for 10 minutes at 4 °C (Beckman Avanti J-30I centrifuge) and the pellets were washed with 100 mL of 150 mM NaCl. Freeze thaw cycles were performed three times with 15 mL of lysis buffer for 2 pellets obtained from a 2 liter culture. The lysis buffer contained 1 × buffer A (25 mM Tris.Cl pH 7.5, 1 mM EDTA pH 7.5; 10% glycerol) and 24 mM KCl. Following freeze thaw, the volume was made up to 200 mL with a final concentration of 10% glycerol, 50 mM Tris.Cl pH 7.5 and 400 mM KCl. The mixture was sonicated at setting 5 (Branson

sonifier, VWR) in the cold room. The lysate was then centrifuged at 10,000 rpm for 15 minutes and the supernatant was transferred to a fresh tube. PEI precipitation (which frees bound proteins from RNA) was performed 5 times. Each time 0.4 mL of PEI was added slowly using a pipet for about 2 minutes (a total of 2 mL) and mixed thoroughly by swirling followed by centrifugation at 10,000 rpm for 15 minutes at 4 °C. The supernatant was transferred to a new tube before the addition of PEI. Following the final PEI precipitation, the lysate was again centrifuged and the supernatant was loaded on to a pre-equilibrated chitin bead column. The column was prepared by adding 8.5 mL chitin beads in 20% ethanol (NEB). This was equilibrated at room temperature with buffer A and loaded with the supernatant from PEI precipitation at about 1 mL/minute. The column was washed with buffer B2 (20 mM Tris.Cl pH 7.0, 500 mM KCl, 1 mM EDTA). After washing was complete, the column was stopped and left overnight at 4 °C. CYT-19 was eluted by thiol induced cleavage using buffer B3 (20 mM Tris.Cl, pH 8.5, 500 mM KCl, 1 mM EDTA and 40 mM DTT). Following elution, the peak fractions were pooled and dialyzed against 500 mL of CYT-19 storage buffer (20 mM Tris.Cl, pH 8.5, 500 mM KCl, 1 mM EDTA, 1 mM DTT and 50% glycerol) at 4 °C for 12 hours using 10 kDa cutoff dialysis tubing (snakeskin® Pierce, Rockford, IL). Following dialysis, aliquots were made, snap frozen using liquid nitrogen and stored at -80 °C.

CYT-19 used in this study was also purified by a slightly modified version of the above protocol[73] and a second method described in Ref[161].

2.2.3 5’ labeling of RNA oligonucleotide substrates and *Tetrahymena* ribozyme[186]

Oligos shown in Figure 2.4 were deprotected using the protocol supplied by the manufacturer (Dharmacon, Inc.) and dried in a speedvac. For the *Tetrahymena* ribozyme, a dephosphorylation reaction was performed using 2 units of shrimp alkaline phosphatase (Promega, Madison, WI) added to 1μL of 50 μM ribozyme, which was incubated at 37 °C for 1 hour followed by incubation at 65 °C for 15 minutes to inactivate the phosphatase. A kinase reaction was then performed using 2’-OH deprotected RNA oligo and 5’-dephosphorylated ribozyme. 1μL each of T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP label (PerkinElmer Life Science) were added and incubated for 1 hour at 37 °C. The reactions were stopped using 20 mM EDTA-formamide dye solution and the labeled oligos and ribozymes were separated by 20% and 8% native polyacrylamide gel electrophoresis (PAGE) respectively. The labeled RNAs were cut out of the gel and eluted overnight in TE (10 mM Tris.Cl, pH 8.0 and 1 mM EDTA) and stored at -20 °C until use.

2.2.4 Kinetic and thermodynamic assays

2.2.4.1 Materials:

The following chemicals were used in the assays. MOPS buffer pH 7.0, $MgCl_2$, $Mg(CH_3COO)_2$, ATP (Sigma), guanosine (Sigma), and proteinase K (Fisher Bioreagents). The equipment used in performing the assays were a temperature controlled water bath (VWR), filter binding apparatus (VWR), Nitrocellulose filter membranes

(Protran), DEAE (Schleicher and Schull), phosphorimager (Stormscan, Molecular Dynamics), Software for data analysis (Kaleidagraph), and standard PAGE apparatus.

2.2.4.2 Buffers used in the study:

All reactions (catalytic and binding assays) were performed in 50 mM Na-MOPS, pH 7.0. For diluting labeled RNA for use in trace amounts, TE (10 mM Tris.Cl, 1 mM EDTA, pH 8.0), which is also the RNA storage buffer, was used. For diluting CYT-19, CYT-19 storage buffer (20 mM Tris.Cl, pH 8.5, 500 mM KCl, 1 mM EDTA, 1 mM DTT and 50% glycerol) was used. All reactions used in the study contained the same 10× dilution of this CYT-19 storage buffer, irrespective of CYT-19 concentration. PAGE running buffer was 1 × TBE (100 mM Tris.Cl, 83 mM Boric acid and 1 mM EDTA).

2.2.4.3 Ribozyme activity assay to follow re-folding of misfolded state to native state:

Re-folding of ribozyme was followed by measuring the fraction of native ribozyme over time (Figure 2.5A, B, C, D). A population of misfolded ribozyme was generated by incubation with 10 mM Mg^{2+} for 10 minutes at 25 °C. This time was sufficient to allow complete conversion of unfolded ribozymes to predominantly misfolded species and a small fraction native ribozyme, but not sufficient for any significant conversion of misfolded species to the native state under these conditions. The magnesium concentration was then decreased to the desired concentration (standard condition is 5 mM) and CYT-19 was added to initiate re-folding. The standard solution conditions of the re-folding reactions were 50 mM Na-MOPS (pH 7.0), 5 mM Mg^{2+}, 50

mM KCl, 2 mM ATP-Mg^{2+} and 5% glycerol. Aliquots from the re-folding reaction were added to a re-folding quench containing 50 mM $MgCl_2$ and 500 μM guanosine without changing the Na-MOPS concentration. Quenching inhibits re-folding of the ribozyme substantially[204] such that it essentially stops the reaction (See Figure A1 in appendix). The fraction of the native ribozyme at each time was determined by adding trace ^{32}P-labeled substrate (S) and performing a cleavage reaction for 1 minute (Figure 2.5B). This time is sufficient for binding of labeled substrate to both native and misfolded ribozyme species (approx. k_{on} = 10^8 $M^{-1}min^{-1}$ for both and at concentration of 30 nM ribozyme, calculated rate constant for binding = 3 min^{-1} and binding is rate-limiting for substrate cleavage; the measured cleavage rate constant was also found to be 3 min^{-1} – see A.1) but not sufficient for the dissociation of substrate from the misfolded ribozyme (0.02 min^{-1}) and rebinding to native ribozyme. Thus the fraction of substrate cleaved in the burst (Figure 2.5B) reflects the fraction of native ribozyme at different folding times (Figure 2.5D). Labeled substrate was separated from the shorter labeled product by using 20% denaturing PAGE. The data was analyzed using a phosphorimager (GE Healthcare, Fairfield, CT). Kinetic analysis was performed using Kaleidagraph.

2.2.4.4 Equilibrium binding assay to measure affinity of CYT-19 to ribozyme:

Equilibrium binding of wild-type CYT-19 and a C-terminally truncated CYT-19 (Δ578-626) to misfolded or native ribozymes was performed using a double nitrocellulose filter binding assay[208]. Predominantly native or predominantly misfolded ribozymes were generated using ^{32}P-labeled ribozyme (<2 nM) and incubated with CYT-

19 for 15 minutes at 25 °C (50 mM Na-MOPS, pH 7.0, 5 mM Mg^{2+}) to allow equilibrium. Pulse-chase dissociation experiment established that 15 minutes was sufficient for dissociation of nearly all CYT-19 from the ribozyme suggesting that this is sufficient length of time to achieve equilibrium (data not shown). Protein-bound RNA was separated from free RNA by applying the mixture to a filter holder fitted with nitrocellulose membrane and a diethylaminoethyl (DEAE) membrane beneath the nitrocellulose membrane. Membranes were washed with 2 mL of 50 mM Na-MOPS buffer and dried along with the retained ^{32}P-label. Both membranes were exposed to a phosphor imager screen and quantitated using a phosphorimager to allow determination of the fraction of ribozyme bound to protein.

2.3 Results

2.3.1 Chaperone action of CYT-19 on the *Tetrahymena* group I ribozyme

Many of the known catalytic activities of DExD/H-box proteins come from studies characterizing their action on model RNA substrates[72,79,102,104-106,136,152,153,165,166,169,209]. For example, it has been demonstrated that Ded1 can catalyze both strand displacement and annealing of complementary RNA strands[106]. Studies on model RNA substrates have been phenomenally useful to understand the capabilities of these proteins. However, although many structured RNAs are made up of duplexes, they are also extensively structured, having motifs such as loops, tertiary contacts, and metal binding motifs, [210-212] all of which can influence folding and local/global stabilities of the structure. In this context, at least two reasons can be envisioned as to why studies of

DExD/H box proteins with structured RNAs are imperative. First, it is plausible that the actions of DExD/H-box proteins are not limited to mere duplex unwinding and their capability extends beyond this action. Second, DExD/H-box proteins *in vivo* are known to act in specific processes, suggesting that they are targeted to act on specific RNA or RNP substrates and it is essential to understand how DExD/H-box proteins achieve this targeting.

In the following research, for the first time, a structural transition from a well-defined misfolded species of *Tetrahymena* ribozyme to native ribozyme was probed in the presence of the DExD/H-box protein CYT-19. In addition, the chaperone action of CYT-19 was also demonstrated on one of its cognate introns with which CYT-19 interacts *in vivo*. Both *Tetrahymena* and *Neurospora* ribozymes used in the study (Figure 2.2 and Figure 2.3) are engineered versions of group I introns, which are self-splicing RNAs found in organellar genomes of plant and fungal mitochondria and chloroplasts[213]. All group I introns catalyze a 2 step trans-esterification reaction as shown in Figure 2.1. In the first step, an exogenous guanosine (G) attacks and cleaves the 5' splice site. This results in the covalent attachment of the 3'-OH of G at the 5' end of the intron. In the second step, a conformational change brings the 3' splice site to the active site. The 3' end of the 5'exon attacks and cleaves the 3' splice site resulting in spliced exons and a free intron. An engineered version of the intron in which the first 21 nucleotides and the last 5 nucleotides are deleted is called the L-21*Sca*I ribozyme (Figure 2.2), and this version can cleave a substrate that is complementary to the internal guide sequence (IGS)[213].

The re-folding of the ribozyme was followed using catalytic activity. The catalytic activity gives an accurate read-out for the formation of the native state, thus unambiguously distinguishing the native ribozyme from all other related but misfolded and unfolded species. The re-folding in the absence of CYT-19 is slow (Figure 2.6A), on the time-scale of hours, under the standard solution conditions (5 mM Mg^{2+}). Upon addition of CYT-19, the reaction showed acceleration in re-folding and this acceleration required the presence of ATP, as leaving out ATP did not significantly increase the re-folding rate over the basal re-folding rate. This result suggests that CYT-19 can mediate an ATP-dependent conformational transition and resolve misfolded structures of a complex structured RNA. The observed rate constant for re-folding increased linearly up to at least 500 nM, giving an efficiency or k_{cat}/K_m value of 9.5×10^4 $M^{-1}min^{-1}$ (Figure 2.6B). The lack of saturation suggests that CYT-19 binds relatively weakly with a lower limit of 500 nM for the K_d for functional binding of CYT-19. An alternative explanation for the weak binding could be rate-limiting binding such that 9.5×10^4 $M^{-1}min^{-1}$ represents bimolecular association rate constant. However, previous results demonstrate that CYT-19 unwinds a duplex (P1 duplex of *Tetrahymena* ribozyme) with an efficiency of about 4.1×10^7 $M^{-1}min^{-1}$([Ref.161]), imposing a lower limit of the same value on the association rate constant. Thus it is likely that 500 nM represents a lower limit for functional binding of CYT-19. This result implies that CYT-19 does not bind tightly to the intron substrate and this is presumably important for its action as a general RNA chaperone of group I introns. The result is also consistent with previous findings that CYT-19 binds relatively non-specifically to group I introns[107]. In the re-folding assay,

CYT-19 is completely inactivated before determination of the fraction of native ribozyme (Figure A2 in Appendix), indicating that CYT-19 chaperones folding of the ribozyme to the native state rather than give formation of a species that requires its continued presence. Additionally, the final fraction of native ribozyme either in the presence of CYT-19 or in its absence was the same (Figure 2.6A), suggesting that CYT-19 does not substantially shift the distribution of species away from their values at equilibrium. Together all of these data suggest that CYT-19 functions as an ATP-dependent enzyme to accelerate formation of native structure from a misfolded structure and is dispensable following its action, thus acting as a true RNA chaperone[8].

The action of CYT-19 strongly indicates that it induces structural modifications in the misfolded ribozyme and reduces the activation energy for the transition between misfolded and native states. In order to probe these conformational transformations mediated by CYT-19, it is essential to determine the rate-limiting steps during the process. At the standard solution conditions of 5 mM Mg^{2+}, no saturation behavior was observed up until 500 nM (Figure 2.6B), suggesting weak binding under these conditions. One way to increase binding is to increase CYT-19 concentration. An alternative approach is to reduce the magnesium concentration. Prior characterization of the ribozyme in the absence of CYT-19 demonstrated that at lower magnesium concentration the re-folding becomes faster[203]. In addition, lowering magnesium salt concentration presumably increases the efficiency of CYT-19 because of increased electrostatic interaction with the highly basic residues of CYT-19. Experiments were therefore performed at lower magnesium concentration (2 mM) to potentially saturate binding in an

attempt to identify the rate-limiting first-order step(s). The plot of observed rate constant against the concentration of CYT-19 shows saturation behavior at 2 mM Mg^{2+} (Figure 2.7) with a k_{cat} of about 0.58 ± 0.02 min^{-1}. This indicates that a rate-limited step(s) does exist during CYT-19 mediated conversion of misfolded to native state. The simplest explanation is that this slow step involves a conformational change in the RNA. It remains to be determined whether the rate limiting step is unfolding accelerated by CYT-19 or the subsequent re-folding of the ribozyme to the native state.

The lack of saturation under standard solution conditions (5 mM Mg^{2+}) suggests weak binding. Therefore, it was desired to directly determine the binding affinity of CYT-19 to the ribozyme to corroborate this finding. Equilibrium binding experiments demonstrate that CYT-19 binds the wild type ribozyme with an affinity of 30 nM (Figure 2.8), a value much lower than the predicted lower limit for K_d for functional binding to ribozyme. This suggests that, although misfolded ribozymes are apparently bound at lower concentrations of CYT-19, it takes substantially higher concentrations of CYT-19 to facilitate re-folding of misfolded to native ribozyme (Figure 2.6B). Equilibrium binding experiments were also performed to determine whether CYT-19 preferentially binds the misfolded ribozyme relative to native ribozyme. Such preference would indicate how misfolded ribozymes are targeted by CYT-19, but it would not be clear how such differential binding could be elicited given that the native and misfolded are structurally similar species. The results are not surprising, therefore, that CYT-19 binds the native and misfolded ribozymes with similar affinity (Figure 2.9). However, there

could be differences with respect to functional binding such that the action on misfolded ribozyme is greater than that on the native species.

CYT-19, like other proteins of DExD/H-box family, have both amino and carboxyl terminal extensions. Although these extensions have been proposed to play a role in targeting, similar to that of the internal extensions of DNA helicases[64], it was unclear how CYT-19, which has to recognize multiple group I introns, can achieve specificity. Comparative binding studies of wild type and a C-terminal truncated CYT-19 were performed as part of a broader study to investigate the significance of the C-terminus. The C-terminus of CYT-19 was shown to be involved in general recognition of structured RNA[161]. In this study, the C-terminal truncated CYT-19 (Δ578-626) was much less efficient compared to wild type CYT-19 in unwinding a P1 duplex when the duplex was attached to *Tetrahymena* ribozyme than when it was free in solution[161]. In order to directly probe whether this difference in efficiency arose as a result of difference in binding affinities between wild type and truncated CYT-19, equilibrium binding experiments were performed.

The Δ578-626 protein bound the ribozyme with a K_d of 200 nM, about 7-fold weaker than the wild type CYT-19 (Figure 2.8). This difference in binding affinity is similar to the difference in efficiency of unwinding of P1 duplex that is attached to the ribozyme[161]. Thus the C-terminus domain potentially plays an important role in general recognition of structured RNA by CYT-19.

2.3.2 Chaperone action of CYT-19 on *Neurospora* group I ribozyme

The *Tetrahymena* group I ribozyme is one of the most highly characterized RNA in terms of structure and folding pathway (discussed in chapter 1). Thus it acts as a well-defined substrate to probe the properties and actions of the DEAD-box protein CYT-19. The activity of CYT-19 on *Tetrahymena* ribozyme suggests a non-specific action. However, it remains to be determined whether there is specific action of CYT-19 on one or more of its cognate substrates from *Neurospora* because of co-evolution. In addition, CYT-18 may play a role in targeting of CYT-19 to specific substrates. To explore the aforementioned possible modes of chaperone action, the activity of CYT-19 on a cognate group I intron from *Neurospora crassa*, the mitochondrial large ribosomal subunit (mtLSU) ribozyme, was investigated.

The *Neurospora* mtLSU ribozyme gave little or no observable cleavage of the oligonucleotide substrate in *trans* in the absence of CYT-18, consistent with the previous findings that CYT-18 gives functional binding to the ribozyme[214-219] (Figure 2.10). Although much needs to be investigated about the ribozyme's structure(s) and folding pathway(s), data presented here provides strong evidence that the *Neurospora* mtLSU ribozyme misfolds and that CYT-19 can facilitate formation of native species thus acting as chaperone. *In vitro*, the substrate cleavage reaction at 25 °C (after incubation with CYT-18 for 30 minutes at 37 ºC) did not go to completion and it showed a 20% burst of product formation followed by a slow phase of product accumulation (Figure 2.10). This most likely reflects the presence of non-native species among native species, with the amplitude of the burst corresponding to the fraction of native ribozyme.

An alternative interpretation of the 20% burst in substrate cleavage is that it represents a rapid formation of equilibrium between substrate cleavage and re-ligation. In this model, substrate cleavage does not go to completion because there is concurrent re-ligation of the products. However, this model is less likely because the burst amplitudes of substrate cleavage reactions depend on their prior incubation times and conditions (Figure 2.13 and data not shown). In a purely dynamic equilibrium model, the burst amplitudes would be expected to stay the same regardless of prior incubation times. In addition, two experiments strongly argue against the equilibrium model. First, a pulse-chase experiment performed suggests that the second phase involves dissociation of substrate from non-native species and rebinding to a species that can cleave the substrate, suggesting that the latter species is native (Figure 2.11A). Second, a dilution experiment in which the concentration of guanosine was diluted from 200 μM to 10 μM, the latter being less than that of the K_d value previously determined for other group I introns[2,220,221], did not result in decrease in the fraction product (Figure 2.11B), arguing against the equilibrium model. Additional evidence for misfolding comes from experiments with CYT-19, which show larger bursts following CYT-19 action and its inactivation as described later in this section.

In Figure 2.11A, it is possible that the 20% burst is not an accurate reflection of the fraction of native ribozyme because of differences in rate constants for binding between native and misfolded species. For example, there might be 5% native ribozyme that bind with a rate constant 4-fold faster than that of the misfolded species, which would also give a 20% burst of product formation. To more accurately measure the

fraction of native ribozyme, pre-steady state multiple turnover substrate cleavage reactions were performed with a slight excess of substrate over the ribozyme concentration (Figure 2.11C). This experiment was performed under the same solution conditions as that of a single turnover reaction (data not shown) similar to the one shown in Figure 2.11A, except that the reaction was transferred to 4 °C prior to addition of substrate. The reaction at low temperature presumably makes product dissociation rate-limiting. Thus, even if there are differences in the binding rate constants between native and misfolded species, they would not influence the estimation of fraction native species from the magnitude of burst amplitude. Two independent reactions with a slight excess of labeled substrate concentrations (400 nM and 250 nM) over the ribozyme concentration (200 nM) gave burst amplitudes of 7% and 17% respectively. Thus the actual fraction of native ribozyme is likely to be about 14 to 21% in contrast to 29% estimated from the single turnover reaction (Figure 2.11A). The higher fraction of native species in the single turnover reaction is presumably because S binds the native ribozyme about 1.4- to 2-fold faster than it binds the misfolded ribozyme(s), overestimating the fraction native ribozyme in the single turnover reaction.

Magnesium induced folding of mtLSU ribozyme and the effect of CYT-18 was monitored by activity assay at 25 °C (Figure 2.12A). For simplicity, the fraction of ribozyme species that become native and non-native species upon CYT-18 addition will be called native and misfolded species respectively but both conformers could have non-native contacts in the absence of CYT-18. At 2 mM and 20 mM Mg^{2+}, the maximum fraction of native species obtained in the absence of CYT-18 was 32% at equilibrium and

longer incubations did not affect this value, although experimental uncertainty makes it slightly variable. The equilibrium value suggests that, under these conditions, neither the native nor misfolded mtLSU ribozyme are strongly favored over each other. The results are consistent with the previous findings that the RNA remains largely devoid of tertiary structure, even at 25 mM Mg^{2+}, in the absence of CYT-18, however the results here demonstrates that a small fraction of the ribozyme folds to a species that rapidly forms native ribozyme upon CYT-18 addition. The rate constant for folding at 20 mM magnesium concentration was faster (0.05 min^{-1}) than at 2 mM magnesium (0.003 min^{-1}), presumably because of formation of different intermediates at different magnesium concentrations in the absence of CYT-18. On the other hand, addition of CYT-18 and concurrent dilution to 3 mM Mg^{2+}, to a reaction pre-equilibrated with the maximum fraction of native ribozyme obtained at 25 °C and 10 mM Mg^{2+}, gave a rapid decrease in the fraction of native ribozyme from 0.34 to about 0.15 with a rate constant of 1.8 min^{-1}, whereas addition of CYT-18 dilution buffer, 1× Na-MOPS only slightly affected the equilibrium presumably because of change in Mg^{2+} concentration from 10 mM to 3 mM (Figure 2.12C). Moreover, addition of CYT-18 at higher magnesium (50 mM) also affected the equilibrium value, albeit at a slower rate (Figure 2.12B). CYT-19 did not seem to affect this equilibrium in the presence of CYT-18 but decreases the rate constant for approach to this apparent equilibrium value, although absence of any CYT-19 effect cannot be ruled out because of uncertainty in the data. These results suggest a surprising scenario in which CYT-18 binds the non-native species of *Neurospora* mtLSU ribozyme more favorably than it binds the native species under these solution conditions.

In order to determine whether CYT-19 acts as a chaperone of mtLSU ribozyme, CYT-19 was incubated with mtLSU ribozyme, in the absence of CYT-18, at low magnesium where the re-folding of misfolded to native species is slow. Under these conditions (50 mM KCl, 5% glycerol, 3 mM Mg^{2+}), CYT-19 was found to transiently give formation of a species that rapidly became native ribozyme upon CYT-18 addition. Figure 2.13 shows the increase in the fraction of native ribozyme from about 17% to 40%, strongly supporting the notion that CYT-19 can convert non-native species of *Neurospora* mtLSU ribozyme to native species. However, this increase is followed by a slow decrease in the fraction of native ribozyme. The slow decrease is presumably because of the inability of CYT-19 to maintain its action. These data are consistent with a model in which CYT-19 acts on the misfolded ribozyme to set up a steady state, where there is interconversion between native and misfolded ribozymes with misfolded species being only slightly favored less than at equilibrium. The higher fraction of native ribozyme is not maintained presumably because of CYT-19 death, inhibition by accumulating ADP, or ATP depletion. One of the predictions of this model is that a second addition of CYT-19, replenishing ATP, or relieving ADP inhibition would increase the fraction of native ribozyme again. This was indeed observed experimentally. Figure 2.13B shows a second transient increase in fraction native ribozyme upon addition of 500 nM CYT-19 along with compensatory magnesium to maintain its concentration at 3 mM and addition of fresh Mg^{2+}·ATP at 2 mM. This second increase had higher amplitude presumably because of higher CYT-19 and ATP concentration in the reaction.

In addition to evidence for chaperone activity, data presented here and in Ref[73], suggest that CYT-19 can accelerate dissociation of substrate bound to *Neurospora* ribozyme. Using catalytic activity, it was found that the substrate cleavage reaction in the presence of CYT-19 apparently goes to completion. However, CYT-19 was unable to allow completion of substrate cleavage when excess unlabeled substrate was added prior to CYT-19 addition (Figure 2.14). This suggests that the excess unlabeled substrate blocks the dissociation of substrate and that the increased fraction native upon addition of CYT-19 is not acceleration of re-folding from misfolded to native species. These results demonstrate that CYT-19 accelerates dissociation of substrate from species that is/are presumably misfolded and allows rebinding to a fraction of native ribzoyme in the population. Additional more direct evidence for the acceleration of substrate release from *Neurospora* ribozyme by CYT-19 was demonstrated using gel shift assay[73]. Thus, one action of CYT-19, acceleration of P1 duplex in both *Neurospora* and *Tetrahymena* ribozyme, appears to be similar.

2.3.3 An unrelated DExD/H-box protein can function as a chaperone of group I ribozyme

A highly conserved core in all DExD/H-box proteins suggests a common mechanism of action amongst these proteins, although many are involved in specific cellular processes. *In vivo*, group I introns are the substrates of DEAD-box protein CYT-19. It is possible that other DEAD-box proteins, which function in different processes, can act as a chaperone of group I introns. To probe this possibility, and to determine a

potential general mechanism of action, Ded1, a related DEAD-box protein involved in translation initiation in yeast[165], was tested.

Figure 2.15 shows plots comparing re-folding of the ribozyme in the presence and absence of either CYT-19 or Ded1. Although the solution conditions are slightly different, the very similar rate constants in the absence of proteins suggest that these slight differences in solution conditions do not affect the re-folding behavior of the ribozyme and hence, any difference arising could be attributed to the difference in mechanism of action of the proteins. On the other hand, any difference in efficiency of action in the presence of proteins would not necessarily rule out similar mechanism of action. Both the proteins showed similar rate constants, within error, in the presence of either CYT-19 or Ded1, suggesting a similar general mechanism of action.

2.4 Discussion

DExD/H box proteins are some of the most diverse and abundant proteins found in nature, yet so little is known about how they act on structured RNAs or what strategies they use. Here, using the highly characterized *Tetrahymena* group I ribozyme, the action of a general RNA chaperone, CYT-19, was investigated. CYT-19 was found to interact with a kinetically stable misfolded form of the ribozyme and accelerate its folding to the native functional state. Prior studies indicated that this misfolded state resembles the native state on the surface but has topological differences within the core[206] and it requires high activation energy to fold to the active conformer[206]. Under standard solution conditions, CYT-19 action was strongly dependent on the presence of ATP, suggesting

that CYT-19 uses the energy of ATP binding and hydrolysis to overcome a kinetic barrier. The efficiency of CYT-19 mediated re-folding was low compared to efficiency of unwinding of a 6 bp duplex[73] under similar solution conditions, most likely because partial unfolding of a highly structured RNA requires more energy than unwinding of a simple duplex.

Magnesium ions are essential for folding and stabilization of RNA structures. Previous studies showed that the re-folding of L-21*Sca*I ribozyme is slowed at higher magnesium concentration[204] and increased at lower magnesium concentration, suggesting that magnesium ions have to be displaced during re-folding of the ribozyme. Higher concentrations of magnesium salt also have an inhibitory effect on the efficiency of CYT-19 action. This is because CYT-19 is a highly basic protein and the positively charged residues are presumably counteracted by the anions derived from the divalent salt. Electrostatic interactions have been shown to be important for the action of many basic single-stranded binding proteins that can cause RNA structural rearrangements. Such interactions presumably play a role in function of CYT-19 as well, facilitating conformational changes by stabilizing the negative charges on the phosphodiester backbone of RNA thereby replacing the role of divalent cations.

The K_d for CYT-19 to ribozyme was much lower than the apparent $K_{1/2}$ for re-folding suggesting weaker functional binding than that detected in equilibrium binding experiments. This could also reflect a requirement of multimerisation or action of multiple CYT-19 molecules for re-folding. However, the former seems less likely because sedimentation equilibrium experiments suggest that CYT-19 remains as a

monomer under a wide range of solution conditions (data not shown). The effect of CYT-19 can be saturated at higher CYT-19 and lower magnesium concentrations, giving a k_{cat} of 0.6 min^{-1}. This suggests that a rate-limiting step(s) exists during modulation of misfolded ribozyme structure by CYT-19 to allow folding to the native state. Presumably, this step involves multiple cycles of partial unfolding of the misfolded state to fold to the native state, perhaps even through the same intermediates, such as I_{trap}, that exist in the pathway in the absence of CYT-19. Further studies need to be performed to validate this model.

The actions of DExD/H-box proteins on structured RNAs are in general, difficult to study *in vitro* because of the lack of defined substrates or interacting partners. The *Tetrahymena* group I ribozyme is a great system for studying the action of CYT-19 because it folds to a defined misfolded species (or a family of rapidly interconverting species)[203] and therefore behaves as a single species in kinetic studies. This is important because it allows one to interpret the parameters directly from kinetic studies rather than from the more complicated modeling of multiple species. In fact, the structural changes of RNA induced by CYT-19 observed in footprinting studies, which was key to providing strong support for chaperone action, was demonstrated using this misfolded ribozyme presumably because of its homogenous nature.

The folding of *Neurospora* mtLSU ribozyme to the native species is quite unfavorable even under high magnesium (20 mM), consistent with the previous footprinting data that mtLSU intron remains largely unfolded in the absence of CYT-18 even up to 25 mM Mg^{2+} (Ref.[214]). However, the results obtained here show that at least a

small fraction is able to fold to the native species or form a conformer that rapidly becomes native species upon addition of CYT-18. It is intriguing that the folding of this species is enhanced at higher magnesium concentration. Perhaps here, the magnesium ions play a role in stabilizing distinct structural intermediates, which fold with different rate constants upon binding to CYT-18.

Non-native conformation(s) of *Neurospora* is favored over the native species at equilibrium under certain solution conditions. Much needs to be investigated about the type and number of misfolded species, however, CYT-19 can give accumulation of a ribozyme species which is predisposed to fold to the native state upon CYT-18 addition. The accumulation of native ribozyme is transient suggesting that CYT-19 action is not maintained either because it forms an unproductive complex following its action or because of unfavorable solution conditions, such as increased ADP concentration, that can be inhibitory to the action of CYT-19. Additional evidence for CYT-19 inactivation comes from the finding that the transient action can be regenerated by a second addition of CYT-19. It has to be emphasized that these results support a hypothesis, where under certain conditions, when misfolded ribozymes are favored at equilibrium, the fraction of native ribozymes are increased due to the constant presence of active CYT-19. This in turn suggests that CYT-19 action is either due to 1) repeated cycles of structure disruption followed by dissociation or 2) CYT-19 binding in the ground state to native ribozyme. The latter model, however, is inconsistent with CYT-19's low affinity of binding. The action on misfolded ribozyme suggests a crucial finding, that is, CYT-19's ability to apparently create a steady-state distribution that differs from equilibrium. It is

plausible that CYT-19 shifts the equilibrium distribution by setting up a dynamic steady state between the different populations of ribozymes. Such an effect has been demonstrated in case of Ded1 on model duplex substrates[209] and perhaps could also be demonstrated for *Tetrahymena* ribozyme where under some conditions, CYT-19 action on the stable native ribozyme leads to an increase in the population of the thermodynamically less favored species.

A second action of CYT-19 on *Neurospora* mtLSU ribozyme, which is distinct from the one described above, was also demonstrated. Previously, CYT-19 has been shown to accelerate dissociation of substrate from the misfolded *Tetrahymena* ribozyme. Here, catalytic assays provide strong evidence that CYT-19 accelerates dissociation of substrate from the P1 duplex of *Neurospora* ribozyme. The efficiency of duplex unwinding was comparable to that of *Tetrahymena* ribozyme, suggesting that CYT-19 is not specifically targeted to structural features of either ribozyme, consistent with its role as a general chaperone.

The mtLSU ribozyme is one of the three group I introns in *Neurospora* that requires CYT-18 for its function[218]. CYT-18 induces tertiary structure formation upon binding to its cognate group I introns including the mtLSU intron[214]. Footprinting studies and *in vitro* splicing assays have indeed demonstrated that formation of tertiary structure and enhancement of splicing requires the continued presence of CYT-18, suggesting that the protein provides functional binding to the mtLSU intron[214,218,222]. In addition, a recent X-ray crystal structure shows CYT-18 in complex with native conformation of a non-cognate RNA[217]. It was therefore surprising to find that, in case of a cognate RNA, an

apparently misfolded conformation was stabilized relative to the native-like conformation by CYT-18. The paradoxical behavior requires further investigation. Perhaps, this is a peculiar feature of the ribozyme version and is due to the manner in which the ribozyme was generated by *in vitro* transcription, hence may not be biological.

Initial comparative studies of CYT-19 and Ded1 suggest that these proteins may have similar mechanism of action. Given that CYT-19 and other DExD/H-box proteins share the same conserved core which is involved in RNA/ATP binding and hydrolysis, the present results are not surprising. The C-terminal region of CYT-19 was previously demonstrated to recognize general features of the RNA for efficient unwinding of duplexes. A similar mode of action could be envisioned for Ded1, which could explain the similar efficiency of action on *Tetrahymena* ribozyme. On the other hand, CYT-19 is a chaperone of group I introns but Ded1 is involved in translation initiation in yeast and has no known function in folding of group I introns. This raises the question as to how specificity is achieved *in vivo* for action of Ded1 on its substrate. Perhaps, one way to achieve specificity is to have additional recognition motifs in its N- or C-terminal regions. There are a few known examples for such an active specificity among DExD/H-box proteins where the protein uses these extensions to potentially make additional interactions with its substrates and thereby increase binding affinity. On the other hand, the specificity could be achieved in a more passive manner where some of these proteins are confined in space, for example, by having barriers such as nuclear envelope or in time, by expression of specific DExD/H-box proteins during different stages of gene function.

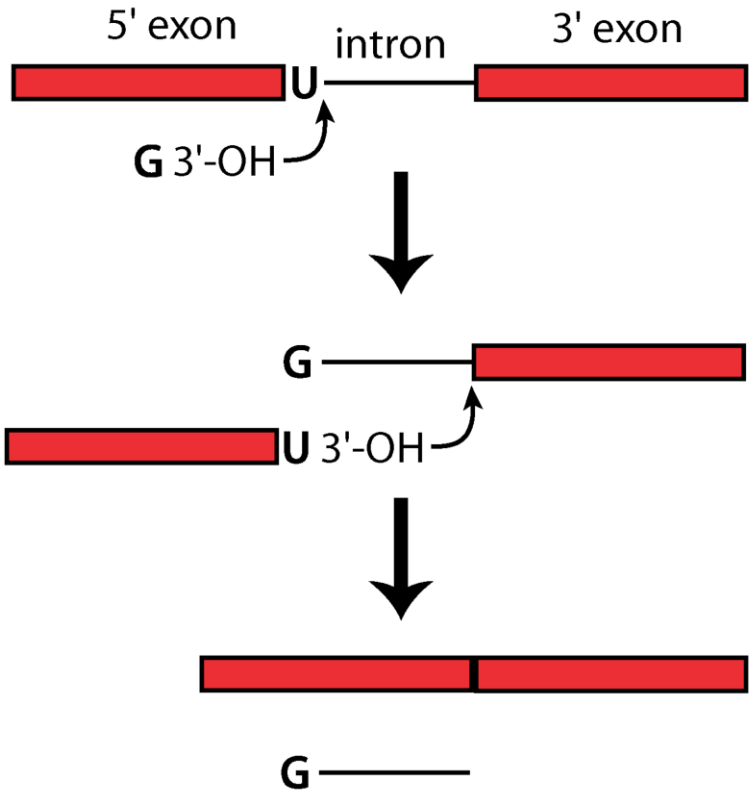


Figure 2.1: Overview of group I intron splicing

A simplified 2-step mechanism for group I intron. In the first step, a nucleophile, guanosine, attacks and cleaves the 5' splice site. The resulting 5' exon attacks and cleaves the 3' splice site, splicing out the intron and ligating the exons together. (Adapted from Tanner, K[213])

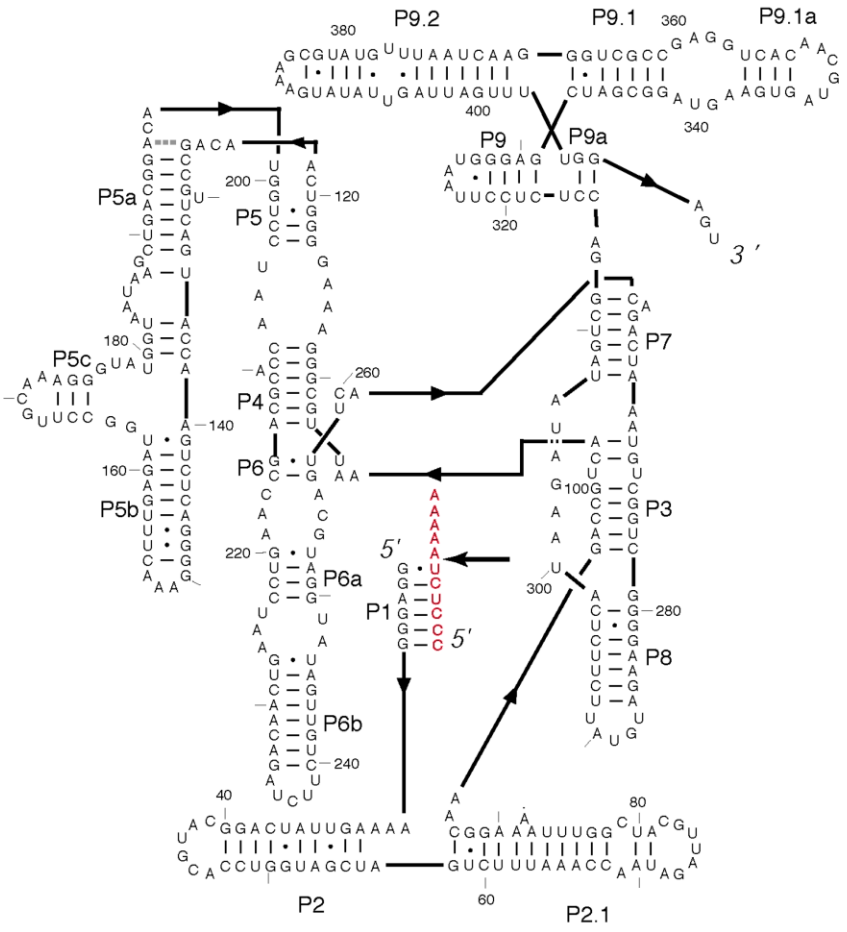


Figure 2.2: Secondary structure of *Tetrahymena thermophila* mtLSU ribozyme

Secondary structure of L-21*Sca*I ribozyme with core (P4-P6 and P3-P8) & peripheral elements (P5abc, P2-P2.1, and P9) are shown. P1 duplex is formed when substrate (red) binds the IGS region by base-pairing. The cleavage site on the substrate is indicated by black arrow; cleavage results in 5'CCCUCU3' and 5'AAAAA3' products.

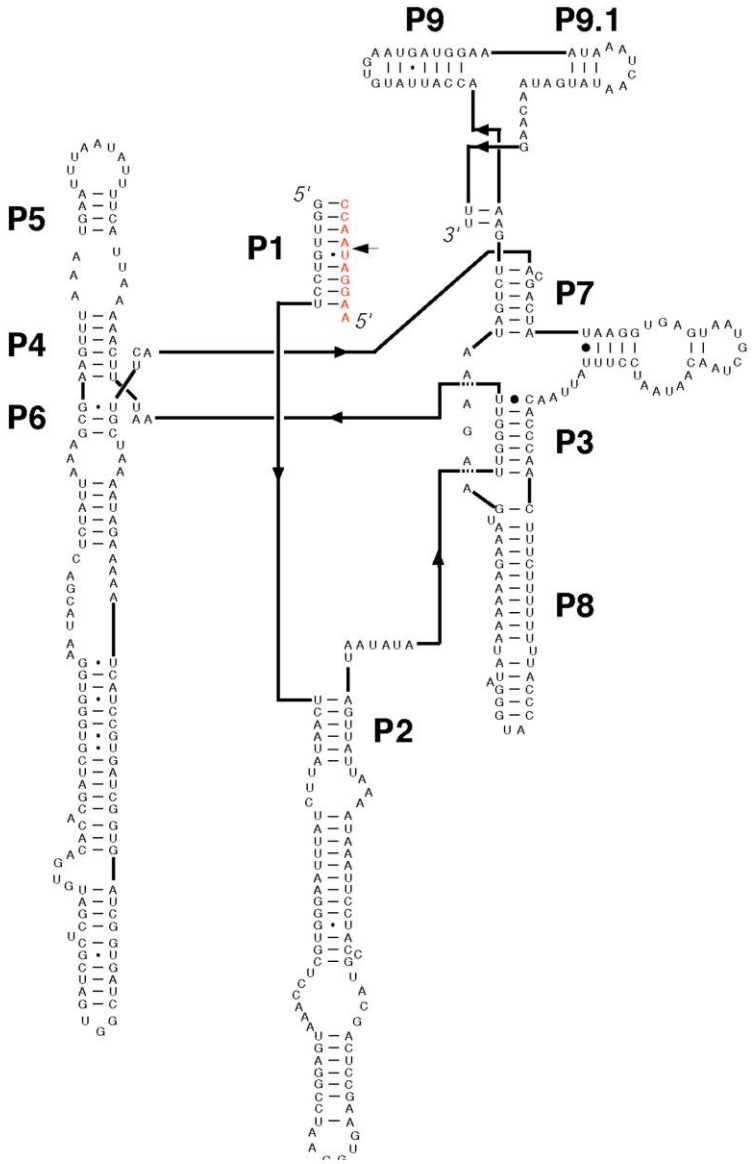


Figure 2.3: Secondary structure of *Neurospora crassa* mtLSU ribozyme
Neurospora crassa mtLSU ribozyme shows core (P4-P6 and P3-P8) and peripheral elements (P2, P9, and extension on P6). The substrate (red) binding to internal guide sequence (IGS) results in the formation of P1 duplex. The substrate cleavage (indicated by arrow) results in the formation of a long 5' product (5'AAGGAU) and a shorter 3' product (5'AACC).

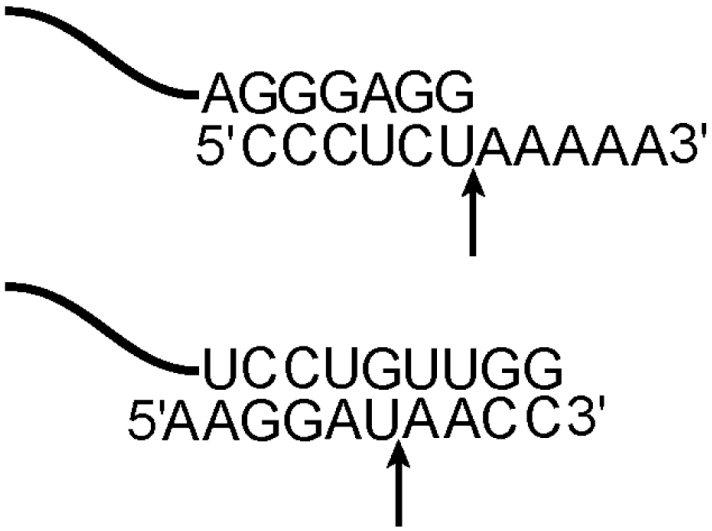


Figure 2.4: Sequence of oligonucleotide substrates that base pair with internal guide sequence (IGS) of ribozyme to form P1 duplex

P1 duplex forms upon binding of specific substrates, indicated by 5' and 3' at the ends, to *Tetrahymena* mtLSU (above) and *Neurospora* mtLSU (below) group I intron ribozymes by base pairing with their respective IGS to form the P1 duplex. The arrow indicates the site of cleavage which is 3' to a G.U wobble pair in both ribozymes.

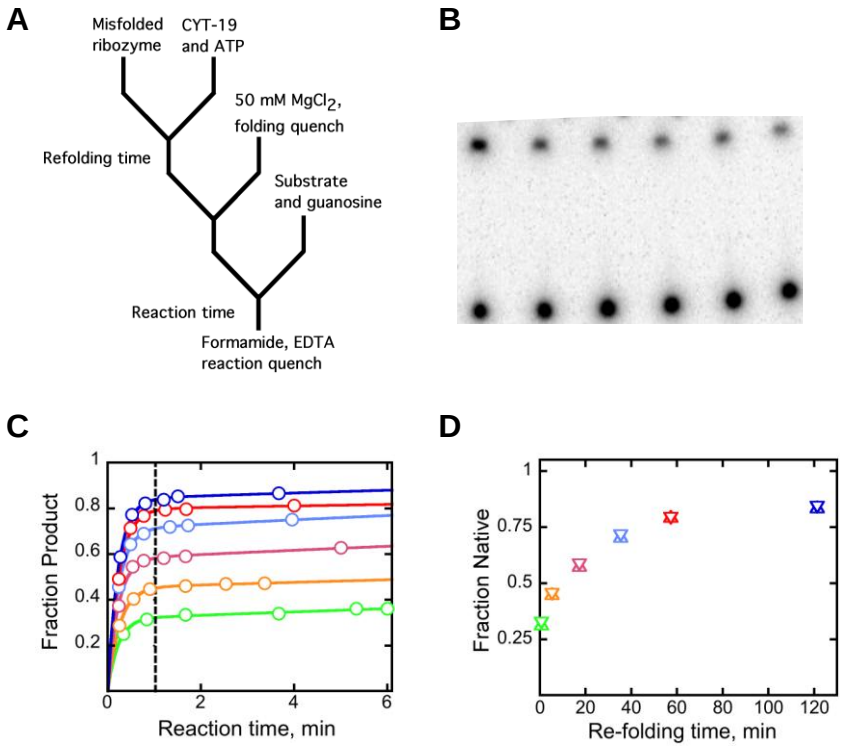


Figure 2.5: Assay to monitor re-folding of the *Tetrahymena* group I misfolded ribozyme

(A) Scheme of the ribozyme re-folding reaction. Predominantly misfolded ribozyme (about 90%) was formed by incubating *Tetrahymena* ribozyme at 25 °C for 10 minutes at 10 mM Mg^{2+}; CYT-19 and Mg^{2+}-ATP were added and solution conditions adjusted to 5 mM Mg^{2+}, 1 × Na-MOPS, 50 mM KCl, 5% glycerol and 2 mM ATP-Mg^{2+} (final concentration of ribozyme = 100 to 200 nM). Over time (re-folding time), aliquots were quenched in high $MgCl_2$ folding quench. After incubation for 5 minutes, oligonucleotide substrate and guanosine were added to aliquots and a second incubation was performed. The cleavage reactions were monitored by taking a second set of aliquots, which were quenched over time (reaction time) with formamide and EDTA. **(B)** Labeled substrate and product are separated on a 20% denaturing gel containing 8M Urea and quantitated. **(C)** Substrate cleavage reactions at varying re-folding time – 1.33 min (green), 6 min (orange), 18 min (magenta), 36 min (light blue), 58 min (red) and 122 min (dark blue). The fraction cleaved at 1 minute for the aliquots quenched at different folding times is indicated by the dotted line **(D)** The amplitude of substrate cleavage reaction (△) and the fraction of substrate cleaved at 1 minute (▽) are plotted against the folding time (same color coding as that of **(C)**). [Note: The ribozyme reactions shown here are for the P5a mutant (see chapter 3), whose re-folding in the absence of CYT-19 is much faster than that of the L-21*Sca*I ribozyme without any difference in substrate cleavage rate constants]

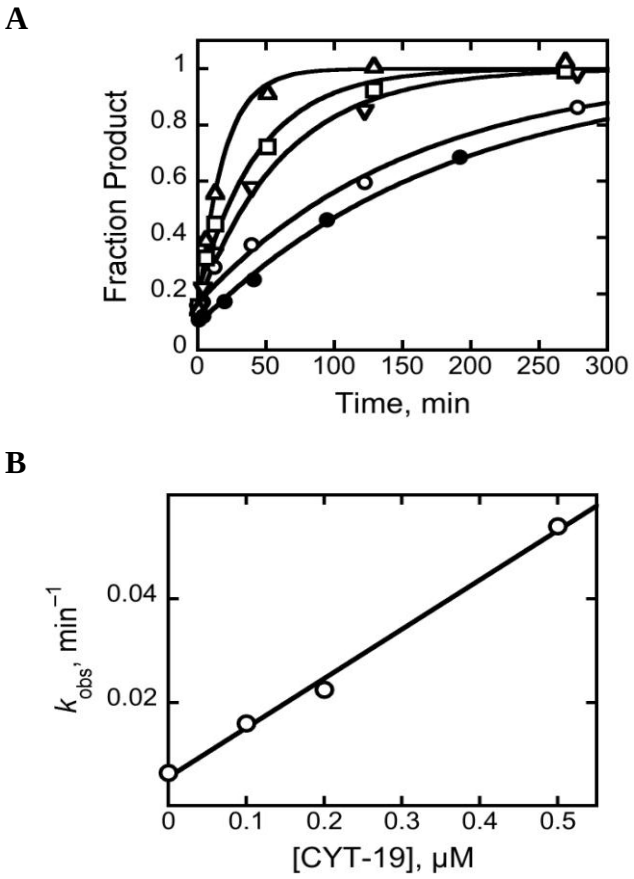


Figure 2.6: CYT-19 accelerates re-folding of the *Tetrahymena* group I misfolded ribozyme

(A) Fraction native was plotted as a function of time (minutes) to give rate constants for re-folding of misfolded to native species in the presence of CYT-19 storage buffer (○), 100 nM (▽), 200 nM (□) and 500 nM (△) CYT-19. All reactions had a final concentration of 100 nM ribozyme and 2 mM mg^{2+}·ATP. An equivalent reaction in the presence of 500 nM CYT-19 but without 2 mM ATP was also performed (●). Rate constants for re-folding were 0.0064 (○), 0.016 (▽), 0.0225 (□), 0.054 (△) and 0.0054 (●). All data points are normalized to the average end points of the three reactions in the presence of CYT-19 and ATP. Curve fitting was performed by forcing the end points of all the reactions to 1 and floating the starting point. **(B)** Rate constants obtained from (A) were plotted against CYT-19 concentration and the data fit to a straight line. The slope, which represents k_{cat}/K_m, gave a value of $9.5 \times 10^4 M^{-1}min^{-1}$.

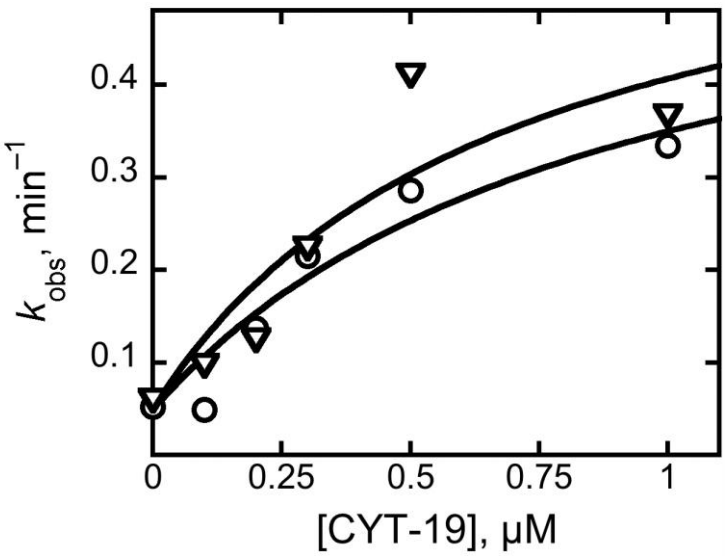


Figure 2.7: Efficiency of re-folding of L-21*Sca*I ribozyme by CYT-19 at 2 mM Mg^{2+}

Efficiency of CYT-19 action for re-folding can be saturated at 2 mM Mg^{2+}. Re-folding experiments analogous to that shown in Figure 2.6 were performed but at 2 mM instead of 5 mM Mg^{2+}. Reactions were at 100 nM ribozyme and at varying CYT-19 concentrations as shown. The rate constants (k_{obs}) obtained from two independent sets of experiments, 1 (○) and 2 (▽) were plotted against CYT-19 concentration. Data from either sets were fit to Michaelis-Menten equation (k_{cat}*[S_0]/ K_m+[S_0])+0.05, where k_{cat} represents the maximal rate constant, K_m signifies Michaelis constant, which is the concentration that gives ½ maximal rate constant, S_0 is the free CYT-19 concentration approximated to initial concentration and '+0.05' represents an offset in the equation which represents the rate constant for re-folding in the absence of CYT-19. k_{cat} and K_m obtained from experimental set 1 were 0.56 min^{-1} and 900 nM respectively and that from set 2 were 0.6min^{-1} and 687 nM. The efficiency of re-folding calculated from the independent measurements of k_{cat}/K_m were 8.7 × 10^5 $M^{-1}min^{-1}$ and 6.3 × 10^5 $M^{-1}min^{-1}$ for the two sets.

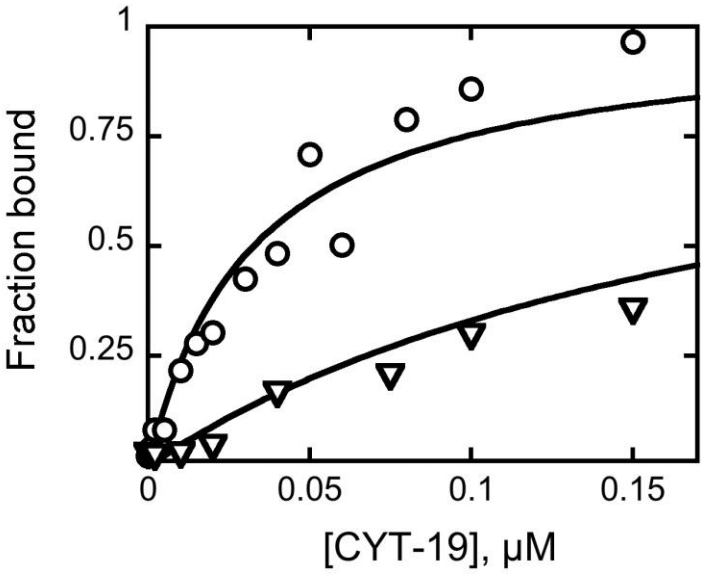


Figure 2.8: The C-terminal truncated CYT-19 binds *Tetrahymena* group I ribozyme less tightly compared to wild-type CYT-19

Equilibrium binding assays were performed using nitro-cellulose/DEAE double filter binding approach (experimental details in methods). The fraction of labeled ribozyme bound by wild type (○) or C-terminal truncated CYT-19, Δ 578-626 (▽), were plotted against CYT-19 concentration and the data were fit to hyperbolic function, fraction bound = $[S_o]/ [S_o] + K_d$, where $[S_o]$ represents concentration of unbound CYT-19, approximated to total CYT-19 concentration. The fits gave a K_d value of 30 nM for the wild type and 200 nM for the C-terminal truncated CYT-19, a difference of 6.7-fold weaker binding.

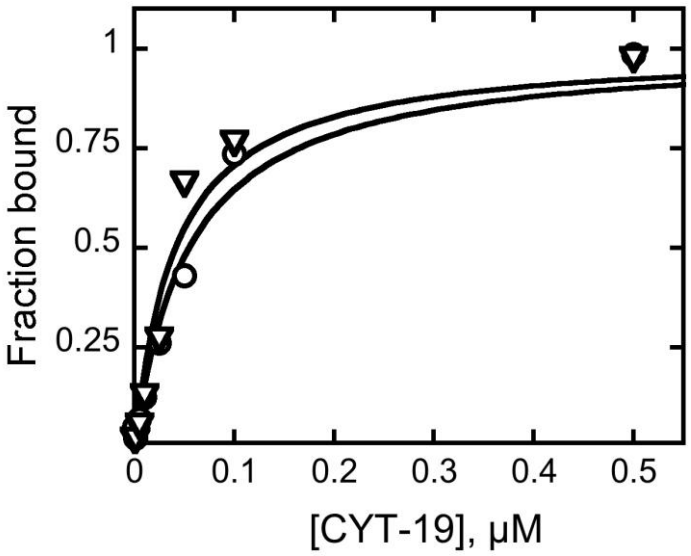


Figure 2.9: Binding affinity of wild type CYT-19 to predominantly native or predominantly misfolded ribozymes

Filter binding experiments were performed at 25 ºC. ^{32}P-labeled ribozymes were either allowed to misfold (○) (25 ºC for 10 minutes at 10 mM Mg^{2+}) or allowed to fold to native species (▽) (50ºC for 30 minutes at 10 mM Mg^{2+}) followed by incubation of either ribzoymes at 25 ºC for 15 minutes at 5 mM Mg^{2+} with CYT-19 at concentrations shown before applying to the double membrane filters. Other solution conditions were 50 mM Na-MOPS and pH 7.0. The data of fractions of labeled ribozymes bound at equilibrium for each concentration of CYT-19 were fit to the hyperbolic binding equation, $[S_o]/K_d+[S_o]$, where $[S_o]$ is the free CYT-19 concentration, approximated to initial concentration and K_d is the equilibrium dissociation constant. The affinity of CYT-19 to native ribozyme (K_d) was 54 nM, the same within error as that of CYT-19's affinity to misfolded ribozyme (K_d = 41 nM). Both these values were same within error as that of the affinity of CYT-19 to misfolded ribozyme measured in a separate experiment (Figure 2.8).

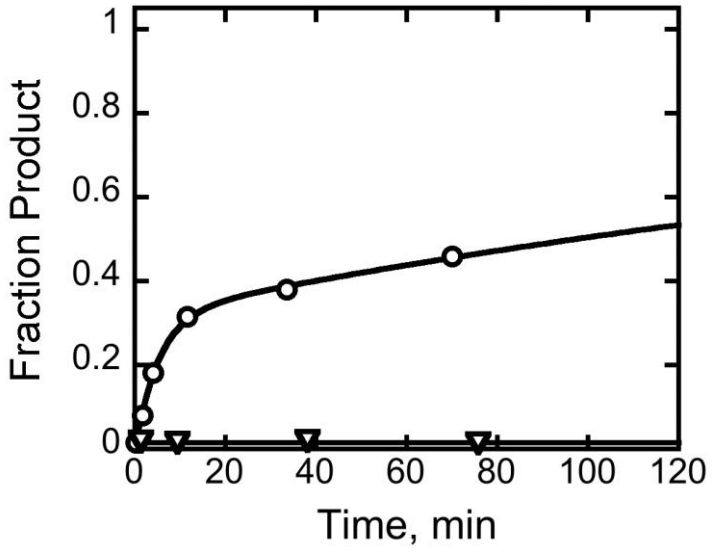


Figure 2.10: ***Neurospora*** **mtLSU group I ribozyme cleavage requires CYT-18**

Neurospora Crassa mtLSU group I ribozyme (200 nM) was incubated with (○) and without (▽) 400 nM CYT-18, 500 μM guanosine and 5.5 mM Mg^{2+} at 37 °C for 30 minutes. Reactions were transferred to 25 °C and trace substrate added to initiate the cleavage reaction. The final concentration of Mg^{2+} was 5 mM. The ribozyme cleavage occurs only in the presence of CYT-18 (○) and leaving out CYT-18 does not produce any appreciable amount of cleavage (▽). The reaction produces a burst of product with a rate constant of 0.19 min^{-1}, followed by a slow phase of product accumulation (0.0039 min^{-1}).

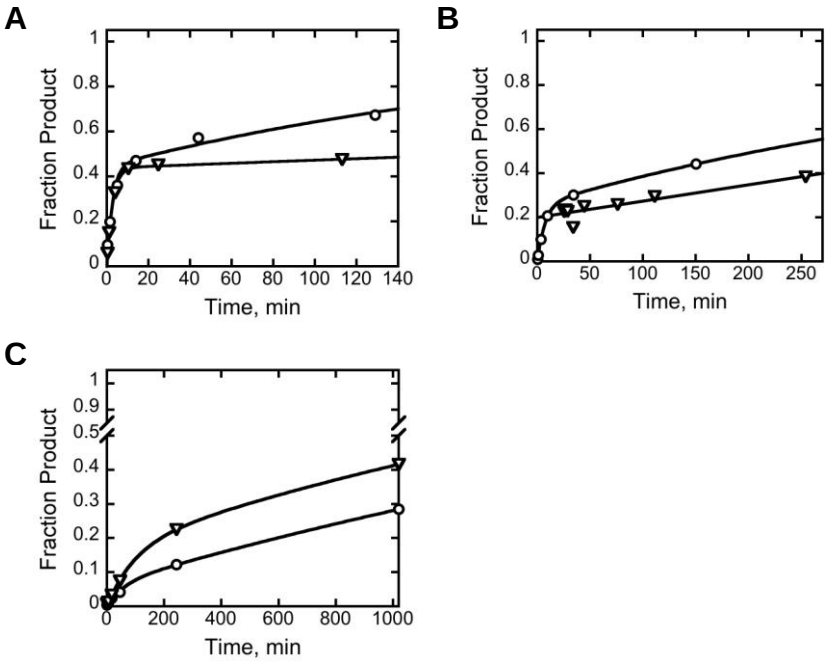


Figure 2.11: ***Neurospora*** **mtLSU group I ribozyme misfolds**

(A) *Neurospora* mtLSU group I ribozyme cleavage reaction demonstrates evidence for the presence of native and alternative conformation(s). The ribozyme was incubated with CYT-18 (1:2 ratio) at 37 ºC for 30 minutes before transferring to 25 ºC. Trace labeled substrate was added and incubated for 10 minutes before 500 μM G (final concentration), with (▽) and without (○) excess unlabeled substrate (chase) was added to initiate the cleavage reaction. The final Mg^{2+} concentration was 5 mM. The reaction shows a biphasic behavior with a 7-fold difference in rate constants for the slow phase in the absence (○, 0.004 min^{-1}) and presence (▽, 0.0006 min^{-1}) of excess unlabeled chase (16.9 μM over 200 nM ribozyme/400 nM CYT-18 concentration). **(B)** Does the burst of fraction native reflect an equilibrium between product formation and re-ligation? Diluting the substrate cleavage reaction 20 fold into 1 × Na-MOPS buffer containing 5 mM Mg^{2+} after completion of the burst shows that the products did not re-ligate to give substrate. Substrate cleavage reactions with ribozyme and CYT-18 concentrations as in (A). The reactions with (▽) and without (○) dilution of guanosine (200 μM to 10 μM) following the completion of the burst phase (25 minutes) are shown. **(C)** Pre-steady state multiple turnover reaction at 4 °C to estimate the true fraction of native ribozyme. 200 nM ribozyme was pre-incubated with CYT-18 as in **(A)** and transferred to 4 °C before the addition of 400 nM (○) or 250 nM (▽) labeled ‘S’. The burst amplitudes obtained were 7% (○) and 17% (▽) respectively giving an estimated fraction of native ribozyme as 14 to 21%.

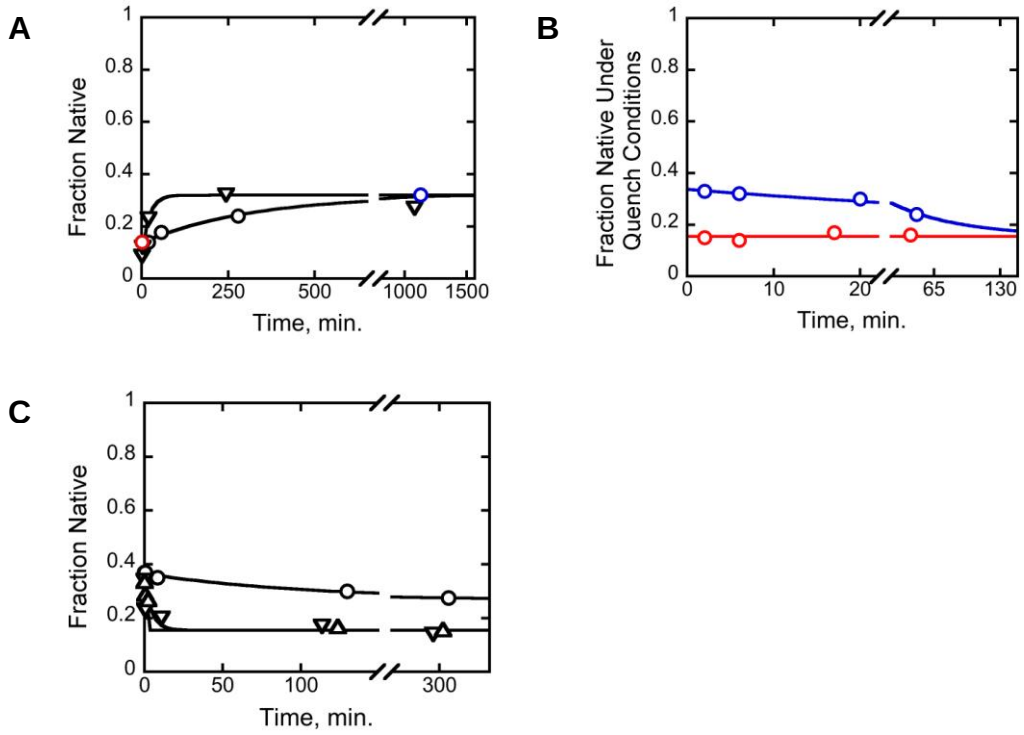


Figure 2.12: CYT-18 affects the equilibrium of *Neurospora* mtLSU ribozyme

(A) *Neurospora* mtLSU ribozyme was heated at 92 °C for 2 minutes and rapid cooled on ice before use. The folding reaction was initiated by adding magnesium to 2 mM (○) or 20 mM Mg^{2+} (▽). Aliquots were quenched in 50 mM $MgCl_2$ containing 600 μM guanosine and 400 nM CYT-18 (200 nM ribozyme) followed by addition of labeled S and performing complete cleavage reactions. The burst amplitudes were plotted against re-folding time and the first (red, 1 min) and last time points (blue, 1129 min) for the reaction at 2 mM are indicated. The rate constant for formation of native state (from unfolded or non-native state) is 0.003 min^{-1} at 2 mM and 0.05 min^{-1} at 20 mM Mg^{2+}. **(B)** Fraction native monitored over time in the folding quench containing 400 nM CYT-18 at early (red, 1 min) and late (blue, 1129 min) re-folding times corresponding to the points in plot (A). **(C)** Effects of CYT-18 and CYT-19 on native ribozyme. 200 nM ribozyme was heated and rapidly cooled as in (A). 1.33 μM ribozyme was incubated with 10 mM Mg^{2+} and 6.66 mM Mg^{2+}-ATP. A '0' time point was taken after incubation for 80 to 92 minutes before the addition of 1× Na-MOPS (○), 500 nM CYT-18 (▽), or 500 nM CYT-18 and 400 nM CYT-19 (△). Aliquots were quenched in high $MgCl_2$ (50 mM) along with 600 nM guanosine (and 500 nM CYT-18 for reaction with 1× Na-MOPS). The fraction native ribozyme was determined as in (A). Addition of Na-MOPS shifted the equilibrium slightly from 0.37 to 0.27 toward misfolded species with a rate constant of 0.01 min^{-1}. Addition of CYT-18 also shifted the equilibrium but from 0.34 to about 0.155 with a rate constant of 1.8 min^{-1}. Addition of CYT-19 slowed this transition (0.2 min^{-1}).

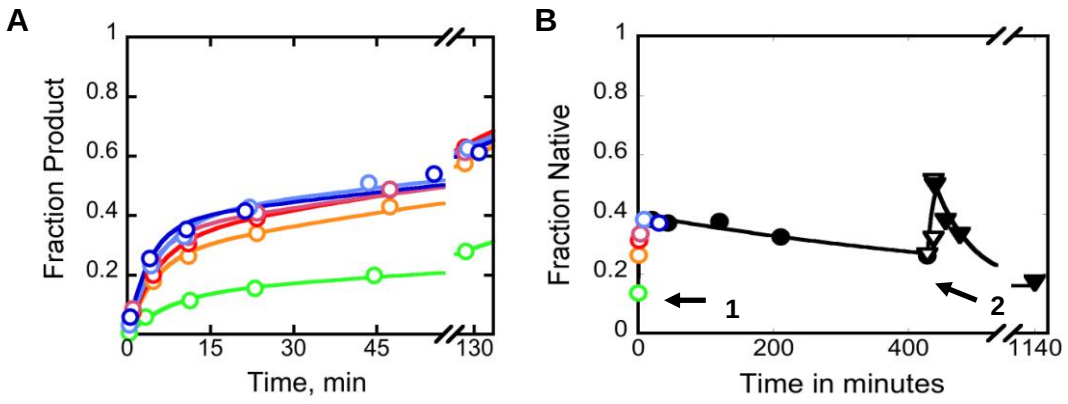


Figure 2.13: CYT-19 gives increase in fraction native species of *Neurospora* group I ribozyme

(A) Substrate cleavage reactions of *Neurospora* mtLSU ribozyme (pre-incubated by heating to 92 °C for 2 minutes and cooled for 5 minutes) after incubation with 3 mM Mg^{2+} 400 nM CYT-19 (indicated by an arrow and the number 1 in **B**) and ATP-Mg^{2+} for 0.33 (orange), 1 (red), 3 (magenta), 8.5 (cyan) and 31 (blue) minutes. (Light green), No CYT-19. Following incubation, the magnesium concentration was adjusted to 50 mM $MgCl_2$ along with final concentrations of 500 nM guanosine and 550 nM CYT-18 (about 2 × higher than ribozyme concentration). Fraction of substrate cleaved were plotted against reaction time and the data are fit to double exponential model – $(A_1 \times (1-\exp^{(-k_1 \times time)}) + (1-A1) \times (1-\exp^{(-k_2 \times time)})$, where A1 represents amplitude of the fast phase and k_1 and k_2 represent rate constants for the fast and slow phases respectively. **(B)** The amplitudes from (A) were plotted against the folding time with the same color code as in **(A)**. Solution conditions were 3 mM Mg^{2+}, 1 × Na-MOPS, 2 mM Mg^{2+}.ATP, 50 mM KCl and 5% glycerol. Aliquots from the reaction were diluted three fold into folding quench containing final concentrations of 500 nM guanosine, 550 nM CYT-18 (about 2 × higher than ribozyme concentration) and 50 mM $MgCl_2$. Complete substrate cleavage reactions **(A)** were performed to determine the burst amplitudes that represents fraction native. The fraction of native ribozyme increases rapidly at a rate constant of 1.3 min^{-1} (○) upon addition of 400 nM CYT-19 (first addition), followed by a slow decrease (●, k_{obs} = 1.3 × 10^3 min^{-1}). The second addition of CYT-19 (indicated by an arrow and the number 2), increases the fraction of native ribozyme (▽), followed by a rapid decrease in fraction native (▼, k_{obs} = 0.018 min^{-1}). The last data point at about 1140 minutes presumably reflects the equilibrium for native ribozyme.

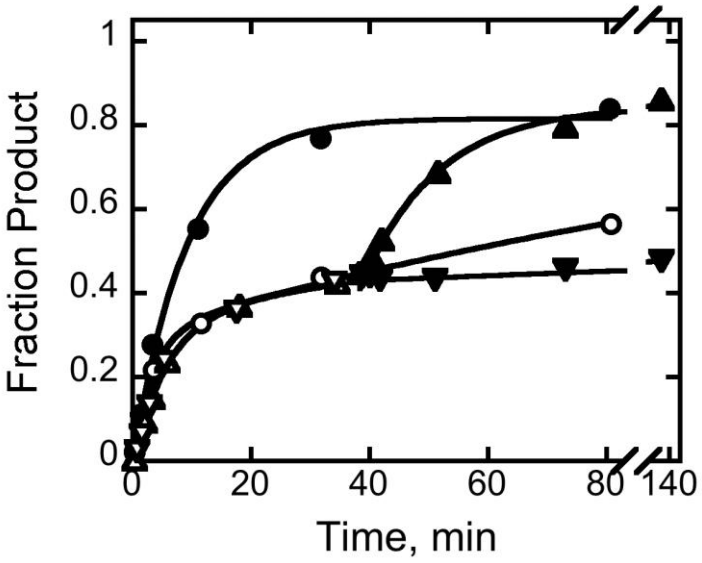


Figure 2.14: CYT-19 acts on *Neurospora* mtLSU group I ribozyme and accelerates substrate dissociation from the ribozyme

Neurospora ribozyme (200 nM) was incubated with CYT-18 (400 nM) at 37 °C for 30 minutes and substrate cleavage reactions were performed with 500 nM CYT-19 (●) or without CYT-19 (○). The substrate cleavage reaction in the absence of CYT-19 (○) showed biphasic kinetics with a fast rate constant of 0.29 min^{-1} and an amplitude of 0.3, followed by a slow phase which has a rate constant of 0.0059 min^{-1}. The substrate cleavage reaction in the presence of CYT-19, gave maximum amplitude of about 0.8 (substantially higher than the reaction in the absence of CYT-19) and with an apparent single exponential (●). Substrate cleavage reactions were performed, in which the rapid first phase was allowed to complete (△, ▽) before the addition of either CYT-19 alone (▲) or addition of excess unlabeled substrate (1.3 μM over 160 nM ribozyme) followed by the addition of CYT-19 (400 nM, ▼). The reaction went to the higher endpoint of 0.8 in the presence of CYT-19 (▲). Addition of excess unlabeled substrate chase obliterated this activity and further slowed the rate constant for the slow phase (▼ versus ○).

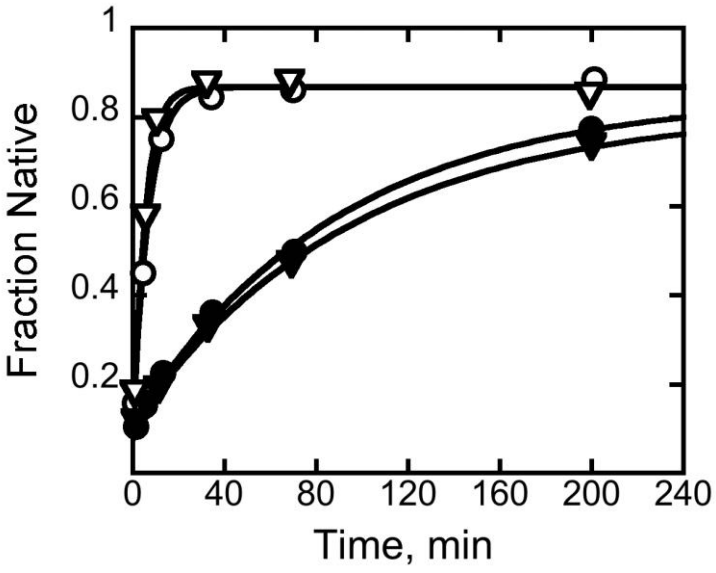


Figure 2.15: Comparison of chaperone activity of CYT-19 and Ded1

Tetrahymena L-21*Sca*I ribozyme was allowed to misfold (25 °C for 10 minutes at 10 mM Mg^{2+}); adjusted to 5 mM Mg^{2+} and allowed to refold either in the presence of 500 nM CYT-19 (○) or in the presence of 500 nM Ded1 (▽). Control reactions in the absence of proteins are also shown (filled symbols). The rate constant for re-folding in the presence of either CYT-19 storage buffer (final concentration of 50 mM KGlu, 5% glycerol) (●) or Ded1 storage buffer (final concentration of 30 mM NaCl, 1% glycerol) (▼); common solution conditions are 5 mM $Mg(OAc)_2$; 50 mM Na-MOPS pH 7.0, 25 ºC. The same rate constant of 0.11 min^{-1} for both CYT-19 and Ded1 mediated reactions suggests that these slightly different buffer conditions do not affect the rate constants of re-folding. The rate constants for re-folding in the presence of 500 nM CYT-19 (○) was 0.15 min^{-1}, the same, within error, as that of re-folding in the presence of 500 nM Ded1 (▽) (0.20 min^{-1}).

Chapter 3: Kinetic redistribution of native and misfolded RNAs by the DEAD-box chaperone protein, CYT-19*

3.1 Introduction

All organisms encode a host of RNAs that must fold into functional structures and undergo extensive conformational transitions as they mediate essential cellular processes such as pre-mRNA splicing and translation. Essentially all processes that are mediated by structured RNAs also require DExD/H-box proteins, which use cycles of ATP binding and hydrolysis to accelerate RNA conformational changes[11,167]. Although they are related to DNA helicases[223], and some viral DExD/H-box proteins possess at least modest canonical RNA or DNA helicase activity[102,104,224], many DExD/H-box proteins display very low activity in conventional helicase assays and are poorly processive, consistent with roles in disrupting local structural elements rather than unwinding long helices[63,167]. There has been much progress using short model duplex RNAs and RNA-protein complexes to elucidate the basic capabilities of DExD/H-box proteins[72,152,165,209]. However, relatively little is known at the molecular level about how DExD/H-box proteins interact with structured RNAs to mediate conformational changes, and what determines which RNAs, and which of their conformations, are targeted for action.

* Significant portions of this chapter has been previously published in *Nature* 449, 1044-1088

Whereas many DExD/H-box proteins are thought to function in the context of a defined RNA or RNA-protein complexes, some proteins of the major subfamily, DEAD-box, function as general RNA chaperones by interacting less discriminately with structured RNAs to promote their folding[107,159]. This latter group includes the *Neurospora crassa* CYT-19 protein, which is required for efficient splicing of several mitochondrial group I introns and can facilitate folding of a diverse set of group I and group II introns *in vitro* or when expressed in *Saccharomyces cerevisiae*[107,159,181].

These systems are valuable experimentally because the relatively simple groups I RNAs are likely to provide insight into the mechanisms of DExD/H-box proteins in more complex systems. A particularly attractive candidate for detailed mechanistic studies of CYT-19 is the ribozyme derived from a group I intron of *Tetrahymena thermophila*, because its *in vitro* folding has been extensively characterized[16,198-200,225]. Further, because it folds preferentially to a long-lived misfolded conformation which then slowly re-folds to the native structure[201,203,204,226,227], it is possible to generate populations of either predominantly native or predominantly misfolded ribozyme.

3.2 Results

3.2.1 Redistribution of native and misfolded ribozyme

Using a quantitative assay for ribozyme catalytic activity, it was shown previously that CYT-19 interacts with the misfolded ribozyme, giving ATP-dependent re-folding to the catalytically active, native state[73,161]. Because the misfolded ribozyme is extensively structured, including all five long-range native tertiary contacts, and must unfold

substantially to reach the native state[206], CYT-19 apparently accelerates this folding reaction by promoting partial unfolding of the misfolded ribozyme.

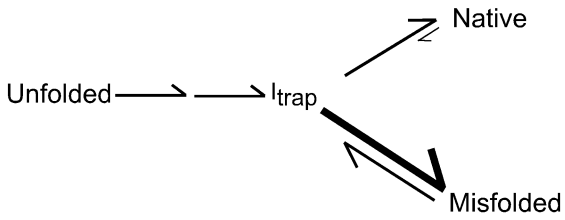
Here, in order to probe the mechanism of CYT-19 action and to explore whether it recognizes structural features that are specific to the misfolded conformer or whether it can also mediate unfolding of the native ribozyme, CYT-19 along with ATP was added to a preparation of ribozyme that was pre-folded to the native state (Figure 3.1A). At various times thereafter, CYT-19 was inactivated by adding proteinase K and increasing Mg^{2+} concentration. The fraction of the ribozyme present in the native conformation was then determined, without interference from CYT-19 (Figure A3 in Appendix), by measuring the fraction of added oligonucleotide substrate ($CCCUCUA_5$) that was rapidly cleaved by the ribozyme[73]. Prior work showed that the substrate binds the native and misfolded species with similar rate constants but is only cleaved by the native ribozyme[204], so that a burst of product is obtained, with the amplitude indicating the fraction of native ribozyme (Figure 3.1B). This burst is followed by a slower phase of product accumulation, which reflects dissociation of the substrate from the misfolded ribozyme and subsequent binding and cleavage by the native ribozyme[73,204]. As expected, no net unfolding was observed under conditions shown previously to give complete accumulation of native ribozyme (5 mM Mg^{2+}, Figure 3.1C). However, under less stabilizing conditions (1 mM Mg^{2+}), the fraction of native ribozyme decreased upon addition of high CYT-19 concentrations, indicating that CYT-19 can also unfold the native ribozyme (Figure 3.1B, 3.1C and 3.2A; a subsequent slow increase reflects time-dependent inactivation of CYT-19 and inhibition by accumulating ADP, data not shown).

The reaction was dependent on ATP, as its omission or replacement with ADP or AMP-PNP gave only low levels of residual activity (Figure 3.1C, 3.2A and 3.2B and data not shown; higher concentrations of CYT-19, not shown, gave significant ATP-independent activity, presumably reflecting passive 'strand capture'[228]). The ribozyme ultimately reached an apparent steady state between the native and alternative conformers, which was dependent on CYT-19 concentration (Figure 3.2A). The observed rate constant for steady-state formation increased modestly with CYT-19 concentration, giving an efficiency of $< 10^5$ M^{-1} min^{-1}. The same steady-state distribution was obtained whether starting from a population of predominantly native or misfolded ribozyme (Figure 3.3), indicating that the entire ribozyme population is subject to the action of CYT-19.

Additional insight into the action of CYT-19 on the native ribozyme and the nature of the steady-state redistribution came from the finding that the decrease in native ribozyme was accompanied by formation of the previously-characterized misfolded conformer. Upon inactivation of CYT-19 by proteinase K, the native ribozyme accumulated with the same rate constant as that for re-folding from the long-lived misfolded species and gave the same Mg^{2+} and urea concentration dependences (Figure 3.4A, 3.4B and 3.4C). The very similar rate constants for re-folding reactions of the misfolded ribozyme and reactions of the intermediates that are formed from CYT-19 action under a wide range of solution conditions shown indicate that a large population of the misfolded conformer was either formed during the CYT-19 reaction or immediately upon its inactivation. These observations suggested a simple model in which CYT-19

partially unfolds both the native and misfolded ribozyme species, giving intermediates that subsequently fold along the same pathway that predominates in the absence of CYT-19. This pathway includes a late commitment point for folding from a trapped intermediate (I_{trap}) to the native and misfolded species, with preferential partitioning to the misfolded species[200,204] (**Scheme 1**).

Scheme 1



This kinetic preference could allow the misfolded species to accumulate despite its lower stability than the native species.

3.2.2 Unfolding efficiency depends on RNA stability

3.2.2.1 Stability of the ribozyme structure determines the efficiency of unfolding by CYT-19

To explore the redistribution model (section 3.2.1) further, features of the ribozyme were investigated to potentially identify those that affect the efficiency of unfolding by CYT-19. Although CYT-19 is able to act upon the native ribozyme, the efficiency was estimated to be at least 50-fold lower than for the misfolded ribozyme (Figure 3.2A and ref.[73]). This difference raised the possibility that CYT-19 derives

specificity from structural features of the misfolded species, but an alternative possibility was that the native ribozyme is unfolded less efficiently because it is more stable[204,229]. Therefore, two ribozyme variants with decreased native stability were examined (Figure 3.5). First, it was found that the native species of a ribozyme variant with a disrupted tertiary contact between the P4 helix and an A-rich internal loop in the peripheral helix P5a[230,231] was unfolded efficiently by CYT-19 even at 5 mM Mg^{2+} (Figure 3.6A and 3.7), conditions that do not give detectable unfolding of the wild-type ribozyme. Again, the reaction was strongly ATP-dependent (Figure 3.9A and 3.9C) and, for a given CYT-19 concentration, the same steady state was reached when starting from populations of predominantly native or misfolded ribozyme (Figure 3.6B).

To test the model that CYT-19 produces intermediates that subsequently fold along the pathway that predominates in its absence, the unfolding reactions were compared to kinetic simulations (curves in Figure 3.6B). The data were well-described using the experimental value for CYT-19-mediated unfolding of the native species and values for subsequent folding that were determined in the absence of CYT-19 (Figure 3.10 and 3.11). Thus, CYT-19 apparently gives partial unfolding of the native and misfolded species and then allows the unfolded intermediates to fold again without significant interference in this latter process.

3.2.2.2 Quantitative correlation between degree of destabilization and efficiency of unfolding

To probe whether a quantitative relationship exists between the degree of destabilization and the efficiency of action by CYT-19, a ribozyme variant, with the entire P5abc peripheral element deleted, was examined ($E^{\Delta P5abc}$, Figure 3.5)[232]. This ribozyme populates the native and misfolded species nearly equally at equilibrium,[229] leading to the prediction that the two species would be unfolded with comparable efficiency. CYT-19 readily acted on the native ribozyme to decrease the fraction native (1.4×10^6 M^{-1} min^{-1}, Figure 3.8A and 3.8B) and again, the reaction was ATP dependent (Figure 3.9B and 3.9C). However, when starting from a population of misfolded ribozyme, net re-folding was not observed; instead, a steady state was reached with the fraction of native ribozyme approximately equal to the small fraction that avoids the misfolded species and folds to the native state directly (Figure 3.8A). This behavior supports the model because, if the native and misfolded species are rapidly unfolded with nearly equal efficiencies, even at modest CYT-19 concentrations the relative populations will be determined by the kinetic partitioning that occurs between them during folding. Nevertheless, the lack of native ribozyme accumulation prevented determination of the efficiency of CYT-19 for unfolding the misfolded ribozyme.

This limitation was circumvented by including a small excess of the group I intron-binding protein CYT-18 (ref.[233]), which binds the native $E^{\Delta P5abc}$ ribozyme tightly[216] but does not bind stably to the misfolded ribozyme under these conditions (A. Chadee and R.R., unpublished results), such that it traps any newly-formed native ribozyme. Net re-

folding of misfolded $E^{\Delta P5abc}$ ribozyme to the native state was indeed observed. After correcting for the bias of the ribozyme to misfold again after CYT-19 action[234], this measurement gave an efficiency for unfolding of the misfolded ribozyme of 3.1×10^6 M^{-1} min^{-1} (Figure 3.8B). This value is ~2-fold larger than that for unfolding of the native $E^{\Delta P5abc}$ ribozyme, similar to the equilibrium value of 1.4 (ref.[229]). Thus, the relative stabilities of the native and misfolded species appear to play a central role in determining their efficiencies of unfolding by CYT-19, although the possibility that the region of the ribozyme near P5abc is the preferred site of action by CYT-19 and it is the local stability of this region, rather than the global RNA stability, that governs CYT-19 efficiency cannot be excluded. Again, kinetic simulation of the approaches to a steady-state mixture of native and misfolded ribozyme (Figure 3.8A) gave good agreement with the data using the experimental values for CYT-19-mediated unfolding of the native and misfolded species and prior experimental values for folding of the $E^{\Delta P5abc}$ ribozyme in the absence of CYT-19 (ref.[234])

3.3 Model for kinetic redistribution by CYT-19

All of the results are consistent with the model shown in Figure 3.12. CYT-19 unfolds both the native and misfolded species, with efficiencies that depend on their relative stabilities, and then allows folding to proceed along the same pathway that dominates in its absence[204]. This generates an ATP-dependent steady state in which the misfolded species, and presumably earlier folding intermediates, are populated to much larger extents than at equilibrium because their formation is favored kinetically during

folding. It is intriguing that CYT-19 does not appear to affect subsequent folding of the ribozyme even though folding proceeds through intermediates that are, by definition, less stable than the native and misfolded species. Most likely, CYT-19 transiently unfolds these intermediates also, but the unfolding is not observed because the same intermediates are readily re-formed. Physical studies, probably at the single molecule level, will be necessary to explore the intermediates that result directly from CYT-19 action.

It was proposed prior to the discovery of RNA chaperones that proteins might be required to accelerate formation of the most stable conformations of RNA by facilitating transitions that require unfolding[8,175], and these general ideas were made more specific a decade later in a model in which CYT-19 actively unfolds misfolded group I RNAs, allowing them multiple chances to fold properly[235]. Whereas it was assumed for simplicity that the native species would be impervious to further chaperone action and would therefore accumulate, here the findings suggest that CYT-19 can have the opposite effect, decreasing the fraction of native ribozyme even though it is the most stable species. The mechanism of CYT-19 action is presumably the same as proposed earlier, but because it acts on the native as well as misfolded species it changes the distribution from equilibrium to kinetic control, allowing increased population of intermediates that can form rapidly, even if they are less stable.

3.4 Implications

These results suggest that, *in vivo*, DExD/H-box proteins can act with sufficient breadth and efficiency to allow structured RNAs to populate a wider range of conformations than would be present at equilibrium. This redistribution of intermediates is analogous to an effect demonstrated for DEAD-box proteins using model RNA duplexes of varying stability[209], and it is reasonable to imagine that depletion of the native state, rather than accumulation, is a general issue for RNA[58]. Indeed, co-expression of an unrelated RNA chaperone protein was shown to decrease *in vivo* self-splicing of several mutant group I RNAs with reduced thermostability by unfolding their native structures[236]. The underlying reason for this behavior is probably that RNA chaperones are unable to distinguish unambiguously between native and misfolded RNAs. Although an analogous challenge exists for protein chaperones, they can achieve a strong bias for interacting productively with misfolded species by recognizing exposed hydrophobic surfaces (ref. [237]and references therein). In contrast, native and misfolded RNAs can be highly similar in global structure and perhaps identical on their surfaces[230].

Thus, RNAs that are required to populate one native structure may face selective pressure to minimize the extent to which their native structures are disrupted by DExD/H-box chaperones. Presumably, one important strategy is to 'hyper-stabilize' the native structures relative to alternative structures, beyond the level necessary simply to ensure their accumulation at equilibrium. Such an effect may contribute to the observation that the native state of the wild-type *Tetrahymena* ribozyme is ~6 kcal/mol more stable than the misfolded conformation under standard *in vitro* conditions[229], as this

large energy gap results in the native species being unfolded by CYT-19 so infrequently compared to the misfolded species that the kinetic preference for misfolding is overcome, and CYT-19 action gives accumulation of the native state[73]. For many RNAs, proteins that associate to form functional complexes also presumably contribute to this energy gap.

There may be additional strategies available to structured RNAs to minimize action of DExD/H-box proteins on their native states. CYT-19 is strongly biased to unwind a duplex when it is unable to form tertiary contacts with the body of the *Tetrahymena* ribozyme[73] suggesting that tight packing of a native structured RNA would direct CYT-19 to interact more efficiently with extended or loosely-packed misfolded structures. The quality control protein Ro selectively binds single-stranded ends of structured RNAs[238], so native RNAs can presumably evade its action by protecting their ends. A further strategy, available for certain RNAs, is for the native structure to rapidly undergo an irreversible process, such as self-splicing, allowing mass action to drive the equilibrium toward the native form even if it is efficiently unfolded by chaperones.

Last, the ability of DExD/H-box proteins to increase the relative populations of less stable RNA structures is also likely to present opportunities for a broad range of RNAs whose functions require formation of multiple structures. A striking example is the spliceosome, which relies on several DExD/H-box proteins to facilitate conformational changes during its reaction cycle[170,239]. Some of these transitions are likely to be thermodynamically unfavorable, as exemplified by the required separation of the extensively-base-paired U4 and U6 snRNAs, and the action of DExD/H-box proteins

may be necessary to prevent the more stable complex from dominating the steady-state population. MicroRNAs, which regulate gene expression by forming base-paired complexes with mRNA targets, may also use the action of chaperones to increase the sampling of alternative complexes, allowing a broader spectrum of physiological targets than would be expected from the relative stabilities of the target complexes[240]. This action of DExD/H-box proteins may also assist in the evolution of structured RNAs by allowing them to sample alternative structures, some of which could fortuitously possess beneficial activities[241] and would then be subject to further selection for stability and activity.

3.5 Yeast Mss116p, a related protein of CYT-19, can also give ATP dependent unfolding of native *Tetrahymena* ribozyme

Neurospora protein, CYT-19 has a related DEAD-box protein in yeast - Mss116p, which has 52% sequence similarity to CYT-19 in the core ATPase domain[159]. Both CYT-19 and Mss116p are encoded in the nucleus[107,183] and function in similar processes, that is to resolve kinetic traps during folding of mitochondrial group I and group II introns[107,159,181,242]. It is reasonable to imagine that they share a similar mechanism of action to facilitate folding of these RNAs. Indeed overexpression of CYT-19 in *MSS116*-deleted mutant of yeast partially rescues the splicing defects suggesting that they are interchangeable and act non-specifically on their substrates[159].

The role of CYT-19 and Mss116p in splicing of group II introns has been investigated previously[159,181,242]. *In vitro* splicing of many group II introns can be

achieved under high monovalent or divalent ion concentrations but can be very slow, presumably because of kinetically trapped forms, that it is not observed under near-physiological ion concentrations. In the case of yeast aI5γ group II intron, it was shown previously that both CYT-19[181] and Mss116p[242] can accelerate folding of this intron most likely by disrupting non-native structures using RNA unwinding activity, thereby acting as chaperones under such conditions. The acceleration was strictly dependent on the presence of ATP suggesting that resolution of such structures require energy from binding and hydrolysis of ATP. Nevertheless, ATP-independent effects have been observed at high concentrations of Mss116p in group II intron splicing[242].

In contrast to the actions described above for Mss116p on group II intron splicing, recent evidence suggested a model in which Mss116p functioned not by resolving kinetic traps but by binding and stabilizing on-pathway intermediates in yeast aI5γ group II intron[243]. Solem *et. al*. reported that Mss116p can facilitate splicing of aI5γ intron without the unwinding of duplexes. This was based on the finding that a motif III mutant (amino acids SAT replaced by AAA), which was assumed to completely uncouple ATP hydrolysis from unwinding activity, was highly defective in unwinding of 12 bp duplexes whereas showed only two-fold reduction in formation of spliced products.

The experiments described above and their conclusions are quite intriguing and contrary to the role of DExD/H-box proteins as ATP-dependent RNA unwinding helicases. If this is true, then it is possible that such an action would also be significant for remodeling RNA by DExD/H-box proteins *in vivo*. However, there is contrasting evidence that Mss116p facilitates group II intron, including aI5γ, splicing through its

helicase activity. First, on a more general level, all DExD/H-box proteins have strong structural homology to that of DNA helicases suggesting duplex unwinding activity as the primary mode of action and indeed many DExD/H-box proteins tested *in vitro* have been found to possess helicase activity[65]. Second, the efficiency of unwinding of RNA duplexes by Mss116p was strongly dependent on the duplex length with increasing duplex length decreasing the efficiency of unwinding. The 12 bp duplex used in solem *et. al.* study is not typically found in structured RNAs and the inability to unwind such long duplexes may be consistent with its physiological role of unwinding short duplexes in a non-processive manner. Last and more importantly, SAT/AAA mutant may not completely unlink ATPase activity from RNA unwinding activity. Indeed, there is evidence for residual *in vivo* splicing activity of SAT/AAA mutant[244]. It is imperative to rule out the incorrect mechanism and therefore experiments were performed as part of a larger study to investigate the mechanism of Mss116p on aI5γ splicing. The assays developed here, unfolding of native *Tetrahymena* ribozyme, proved to be instrumental because this group I ribozyme is a simple yet highly structured and well studied RNA that it is likely to provide mechanistic insight into the more complex and less well studied group II aI5γ intron.

To investigate, whether duplex unwinding is completely abolished by SAT/AAA mutant and compare its efficiency to wild type Mss116p, Del Campo *et. al*. used a shorter 6 base pair duplex which are typically found in RNAs[228]. CYT-19 mediated unwinding of the isolated P1 duplex (that forms following binding of the oligonucleotide substrate in the context of the ribozyme) had been shown previously to be much faster when it is

attached to the ribozyme[73]. Wild type Mss116p showed similar results in unwinding assays but SAT/AAA mutant also showed ATP-dependent activity, albeit, with much less efficiency (7-fold less when P1 was free in solution and 15-fold when attached to the ribozyme)[228]. More importantly, splicing of group II intron aI5γ by SAT/AAA mutant protein was affected to about the same extent relative to wild type (8 to 28 fold). These data indeed suggest that SAT/AAA mutant can have residual helicase activity and that the mutation does not completely uncouple ATP binding and splicing activity. They are also consistent with the results of Solem *et. al.* without the need to invoke a model in which Mss116p stabilized on-pathway intermediates.

Here, in order to compare the action of wild type Mss116p and SAT/AAA on the unfolding of a structured RNA, *Tetrahymena* ribozyme unfolding assays were performed as described for CYT-19 (chapter 3). The results are summarized in Figure 3.13. Both Mss116p and SAT/AAA proteins were active giving ATP-dependent unfolding of the ribozyme and thus a decrease in fraction of native species. However, the efficiency of unfolding of SAT/AAA mutant was about 15-fold lower than that of wild-type consistent with results from splicing assays[228]. Further, both proteins showed comparable ATP-independent activity ($<0.8 \times 10^5$ $M^{-1}min^{-1}$, data not shown) at much higher concentration presumably due to passive strand capture[228]. Taken together, all of these results strongly suggest that ATP-dependent helicase action is the predominant mode of action of DExD/H-box proteins Mss116p and related protein, CYT-19 and that they function primarily by resolution of kinetic traps during folding and splicing of group I and group II introns including aI5γ.

3.6 Summary

DExD/H-box proteins are ubiquitously involved in RNA-mediated processes and use ATP to accelerate RNA conformational changes. However, their mechanisms of action, and what determines which RNA species are targeted, are not well understood. Here, the DExD/H-box protein CYT-19, a general RNA chaperone was shown to mediate ATP-dependent unfolding of both the native and a long-lived misfolded conformation of a group I catalytic RNA with efficiencies that depend on the stabilities of the RNA species but not on specific structural features. CYT-19 then allows the RNA to re-fold, changing the distribution from equilibrium to kinetic control. Because misfolding is favored kinetically, conditions that allow unfolding of the native RNA give large increases in the population of misfolded species. The results suggest that DExD/H-box proteins act with sufficient breadth and efficiency to allow structured RNAs to populate a wider range of conformations than would be present at equilibrium. Thus, RNAs may face selective pressure to stabilize their active conformations relative to inactive ones to avoid significant redistribution by DExD/H-box proteins. Conversely, RNAs whose functions depend on forming multiple conformations may rely on DExD/H-box proteins to increase the populations of less stable conformations, thereby increasing their overall efficiencies.

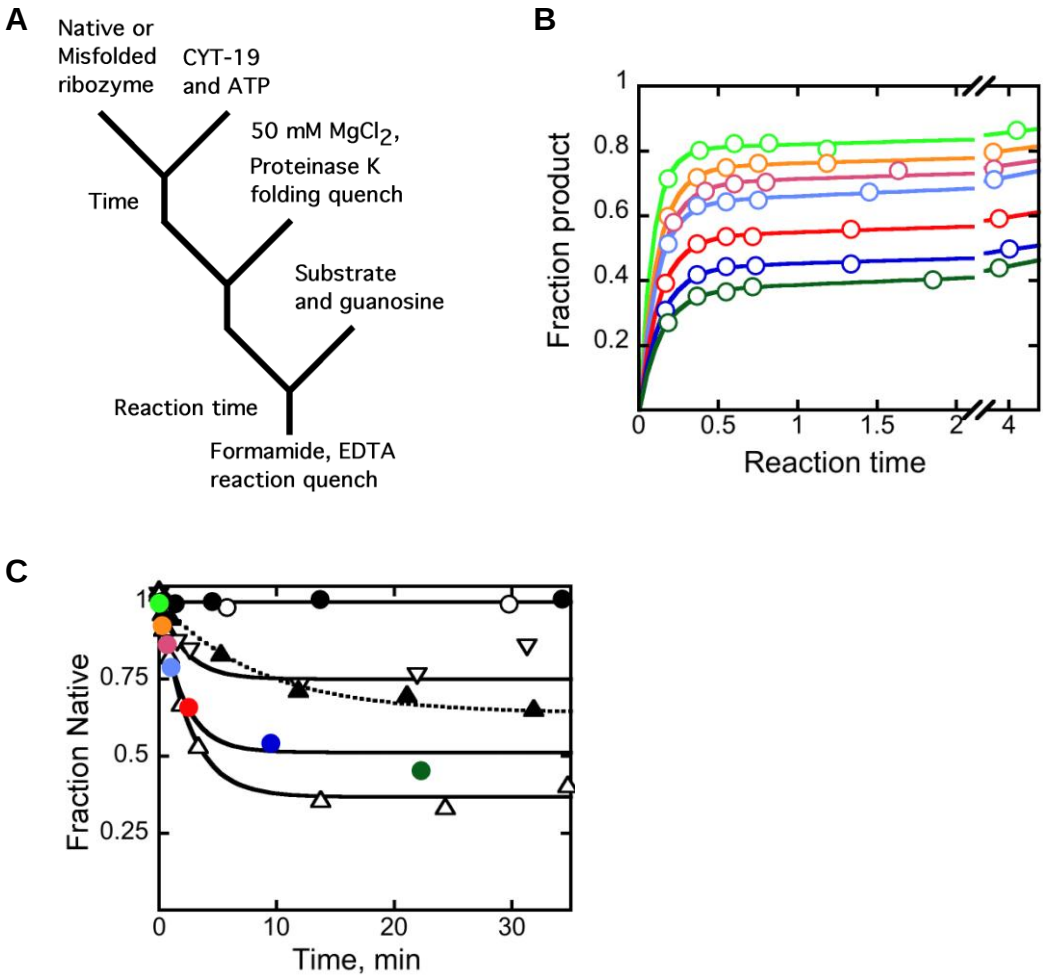
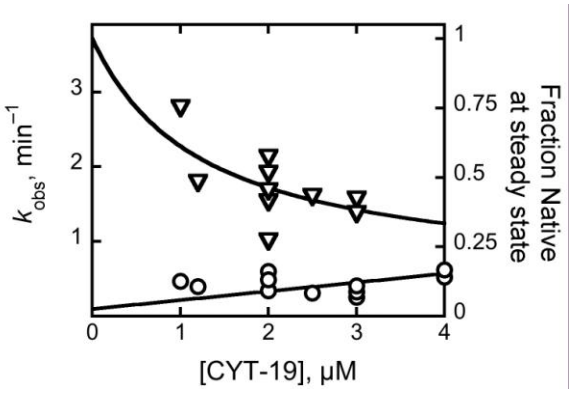


Figure 3.1: Unfolding of Native *Tetrahymena* ribozyme

(A) Reaction scheme. **(B)** Substrate cleavage after incubation with 1 mM Mg^{2+}, 2 µM CYT-19 and 2 mM Mg^{2+}·ATP for 0.25 (orange), 0.67 (red), 1 (cyan), 2.5 (magenta), 9.5 (blue), or 22 min (dark green). (Light green), no CYT-19. Fraction of substrate cleaved plotted against reaction time and the data are fit to double exponential model – $(A_1 \times (1-\exp^{(-k_1 \times time)}) + (1-A1) \times (1-\exp^{(-k_2 \times time)})$, where A1 represents amplitude of the fast phase and k_{obs1} and k_{obs2} represent rate constant for the fast and slow phases respectively. **(C)** Amplitudes obtained from 1B are plotted as fraction native against time. Native ribozyme unfolding (1 mM Mg^{2+}). CYT-19 was 1 µM (▽), 2 µM (solid colored circles), or 3 µM without (○) or with 2 mM Mg^{2+}•ATP (△). Colored circles show burst amplitudes from corresponding curves (panel b). ○, no CYT-19; ●, 2 µM CYT-19, 2 mM Mg^{2+}·ATP, 5 mM Mg^{2+}.

A



B

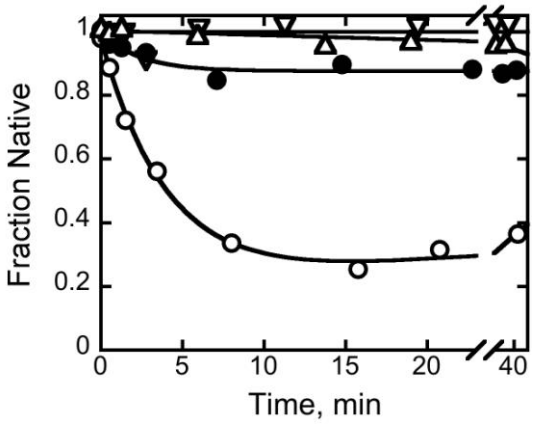


Figure 3.2: Unfolding of native ribozyme by CYT-19 is moderately efficient and requires ATP binding and hydrolysis

(A) Plot of rate constant (circles) and steady state value (triangles) against CYT-19 concentration. The k_{obs} data are fit to a straight line giving a modest k_{cat}/K_m value of less than $10^5 M^{-1}min^{-1}$. **(B)** Unfolding of native ribozyme by CYT-19 requires ATP. Unfolding reactions were performed by 2 μM CYT-19 with 2 mM ATP.Mg^{2+} (○); No ATP (●), 2 mM ADP.Mg^{2+} (▽); and 2 mM AMP.PNP.Mg^{2+} (△).

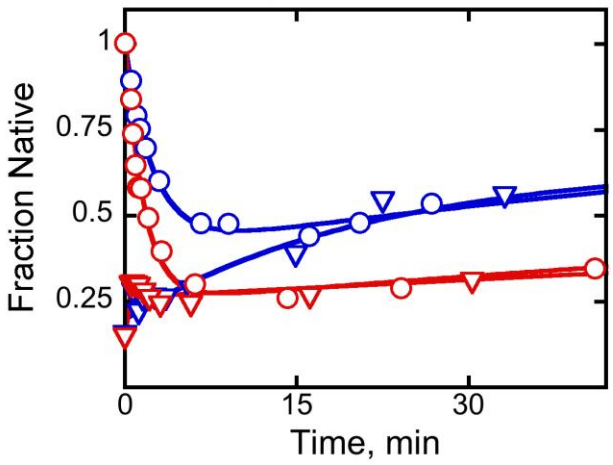


Figure 3.3: CYT-19 sets up a steady state of unfolding

Approach to steady state from native (○) or misfolded (▽) ribozyme with 1.2 μM (blue) or 2 μM (red) CYT-19. Native ribozyme was formed by incubation in 10 mM Mg^{2+} at 50 °C for 30 minutes and allowed to cool to reaction temperature (25 °C); Predominantly misfolded ribozyme (about 90%) was formed by incubating the ribozyme in 10 mM Mg^{2+} at 25 °C for 10 minutes to trap the misfolded species

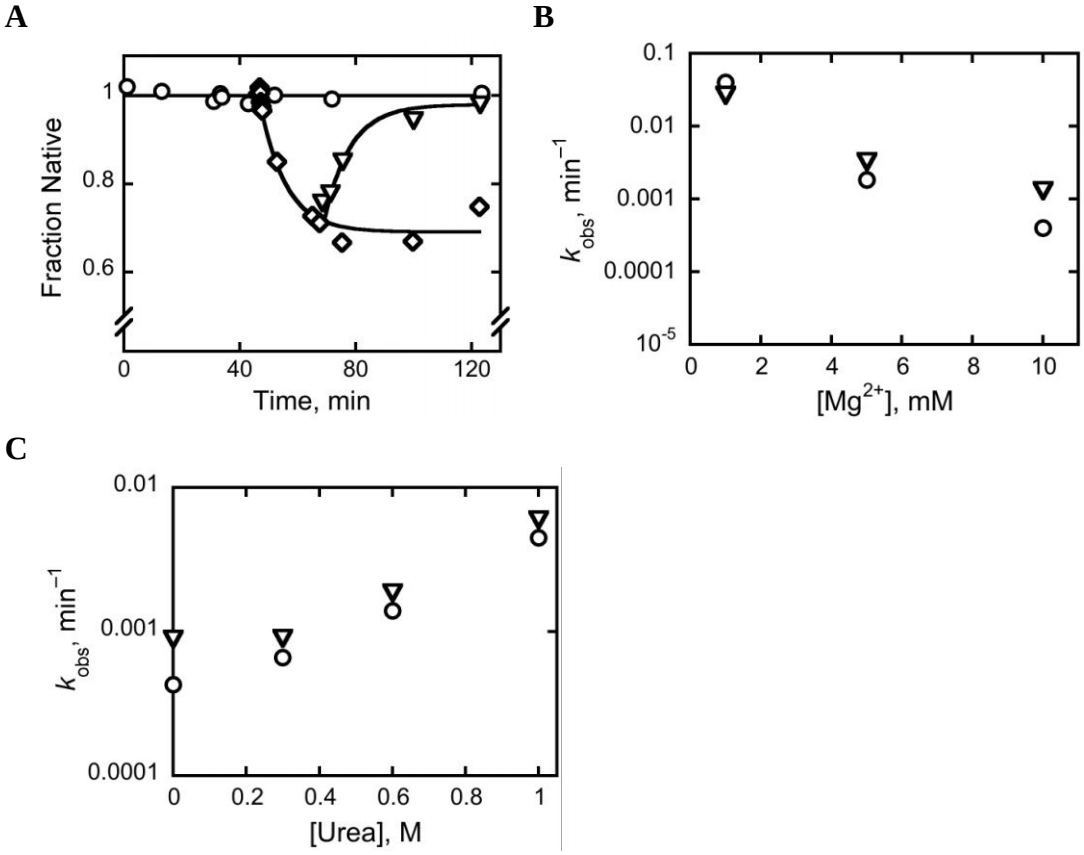


Figure 3.4: Native ribozyme gets converted to the long-lived misfolded state during CYT-19 action or upon its inactivation

(A) Re-folding to the native state (▽) after unfolding by CYT-19 (◇) and inactivation by proteinase K. ○, no CYT-19. **(B)** and **(C)** Magnesium and Urea dependence on re-folding. Reactions were performed in the absence of CYT-19 (○) or following CYT-19 action (▽) in the presence of varying concentrations of Mg^{2+} (panel a) or urea (panel b). The wild-type ribozyme was misfolded under standard conditions (25 °C, 10 mM Mg^{2+}, 50 mM Na-MOPS, pH 7.0 for 10 min, ○) or first allowed to fold to the native state (50 °C, 10 mM Mg^{2+}, 50 mM Na-MOPS, pH 7.0 for 30 min, ▽). CYT-19 (2 μM) was then added to the native ribozyme with 2 mM ATP-Mg^{2+}, and reactions were incubated at 25 °C for 15–30 min to allow formation of a steady-state mixture of the native state and intermediates. CYT-19 was then inactivated by addition of 1 mg/ml proteinase K (and incubated for an additional 2 min for reactions to which urea would subsequently be added). Reactions with and without CYT-19 were adjusted to the conditions shown and native state formation was monitored by ribozyme activity. The rate constant under standard conditions (25 °C, pH 7.0, 10 mM Mg^{2+}, no urea) is shown from independent determinations in each panel, performed side-by-side with the set of reactions shown.

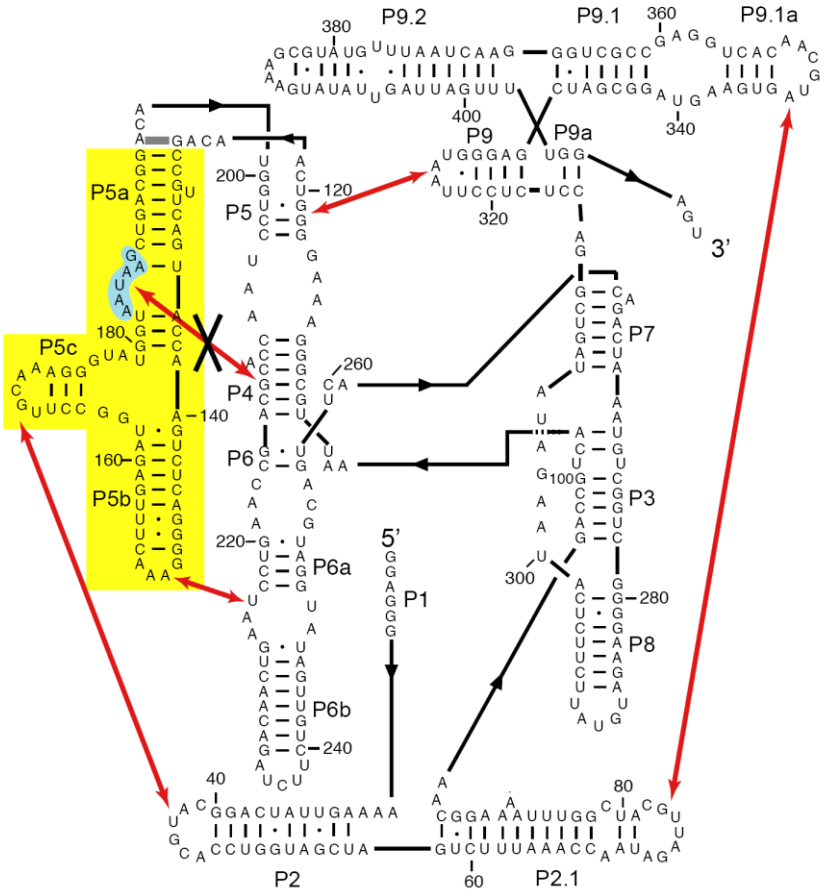


Figure 3.5: Secondary structure, long-range tertiary contacts, and mutations of the *Tetrahymena* ribozyme

The five long-range tertiary contacts are indicated with red arrows. In the P5abc deletion variant ($E^{\Delta P5abc}$), the region shown in yellow is deleted and nucleotides 126 and 196 are directly connected (thick gray line above yellow region). In the P5a mutant, nucleotides 183-188 (shaded cyan) are each changed to uridine[230]. This mutation disrupts the tertiary contact indicated by the black 'X'.

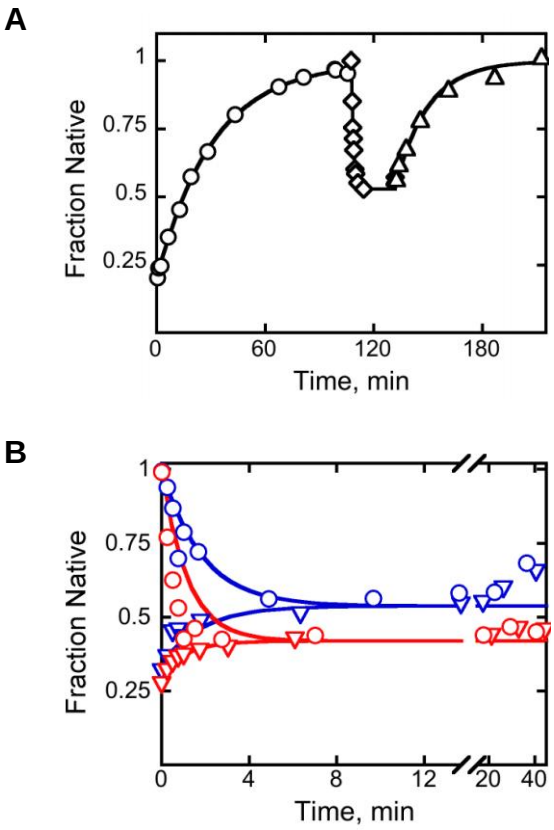


Figure 3.6: CYT-19 mediated unfolding of destabilized ribozyme variant, P5a

(A) Unfolding of destabilized ribozyme variant, P5a at 5 mM Mg^{2+} with 0.8 µM CYT-19 (◇). Upon CYT-19 inactivation, native ribozyme accumulated (△, 0.056 min^{-1}) with the same rate constant within error as for re-folding of the misfolded ribozyme (○, 0.032 min^{-1}). (B) Approach to steady state for native (○) or misfolded (▽) P5a variant with 0.5 µM (blue) or 1 µM (red) CYT-19. Curves depict kinetic simulations using experimentally-derived values (Figure 11A), not fits to the data.

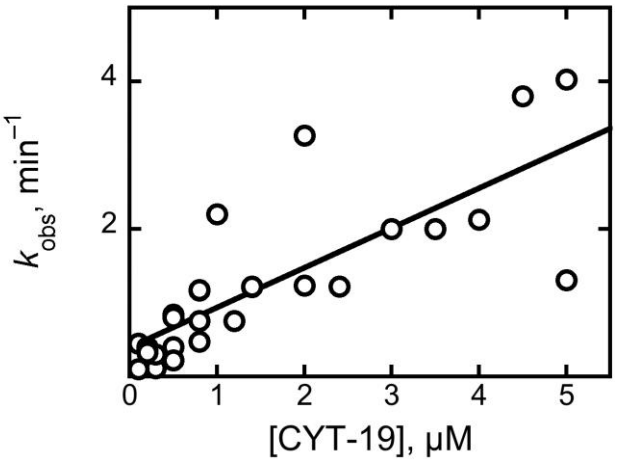


Figure 3.7: CYT-19 concentration dependence for unfolding the native P5a variant ribozyme

A linear fit to the data (solid line) gives an efficiency of $5 \pm 1 \times 10^5$ M^{-1} min^{-1}. (25 °C, pH 7.0, and 5 mM Mg^{2+}). k_{obs} was plotted against CYT-19 concentration and the data was fit to a straight line equation (k_{obs} = (k_{cat}/K_m ×[CYT-19]) + Y intercept), where k_{cat}/K_m represents slope of the line, equal to $5 \pm 1 \times 10^5$ $M^{-1}min^{-1}$. The Y intercept potentially represents the rate-limiting step during CYT-19 mediated unfolding of P5a mutant.

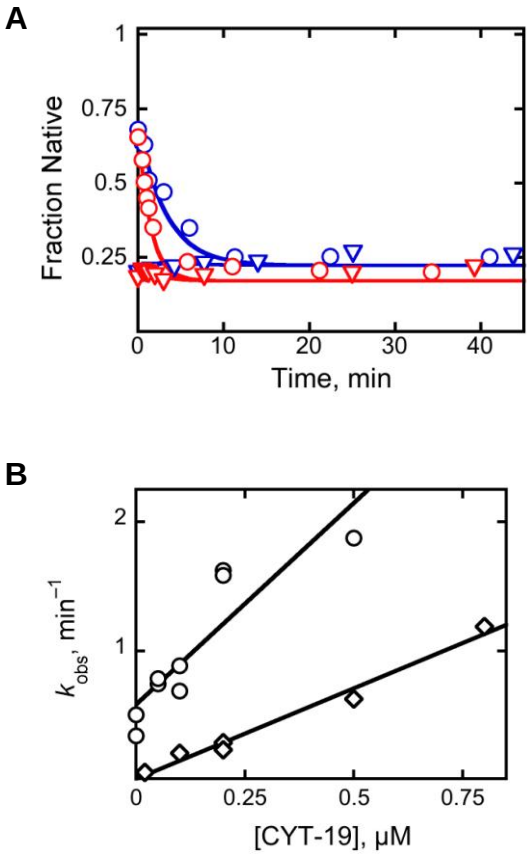


Figure 3.8: CYT-19 mediated unfolding of destabilized ribozyme variant, $E^{\Delta P5abc}$

(A) Approach to steady state for native (○) or misfolded (▽) $E^{\Delta P5abc}$ ribozyme with 0.5 μM (blue) or 1.2 μM (red) CYT-19. Curves depict kinetic simulations using experimentally-derived values (Figure 11B and ref.[234]), not fits to the data **(B)** Unfolding of native (◇) and misfolded (○) $E^{\Delta P5abc}$ ribozyme. All reactions were 25 °C, pH 7.0, 5 mM Mg^{2+}.

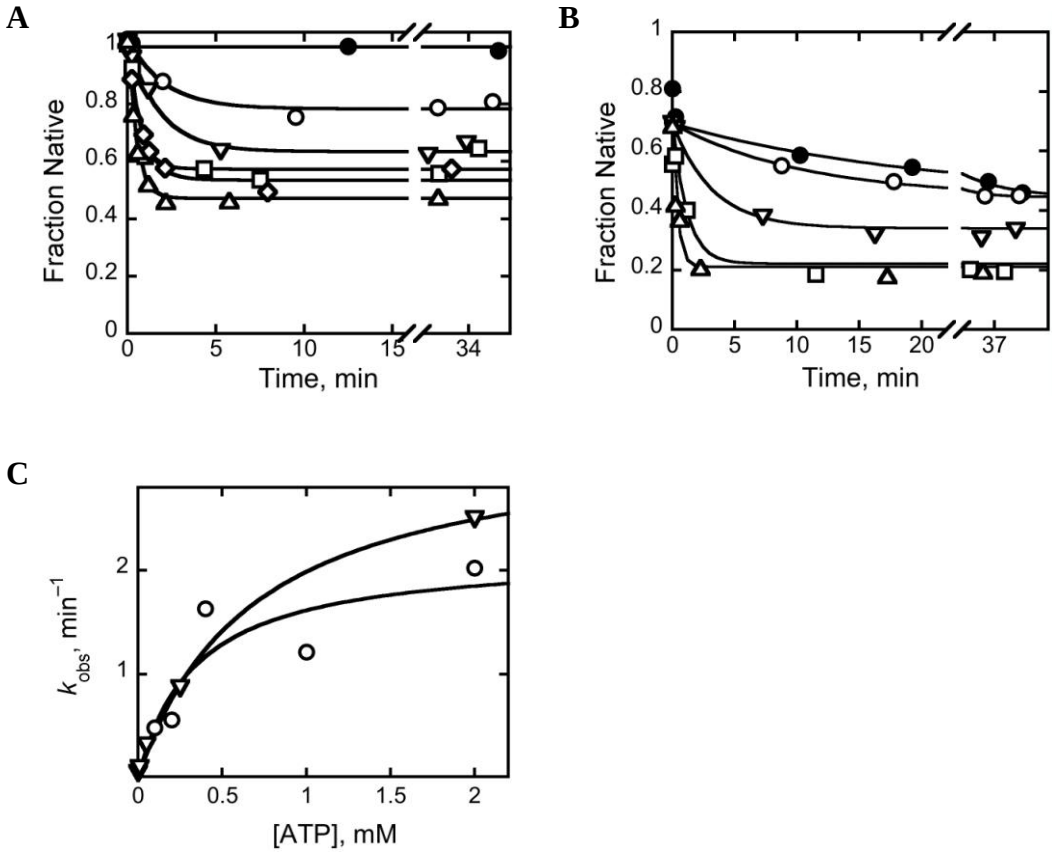


Figure 3.9: ATP concentration dependence for CYT-19 mediated unfolding of ribozyme variants

(**A)** Native P5a variant ribozyme was incubated in 5 mM Mg^{2+} with 1 μM CYT-19 and no ATP-Mg^{2+} (●) or ATP-Mg^{2+} concentrations of 100 μM (○), 200 μM (▽), 400 μM (□), 1 mM (◇), or 2 mM (△). **(B)** P5abc-deleted ribozyme ($E^{\Delta P5abc}$) was folded to an equilibrium mixture of the native and misfolded conformers (10 mM Mg^{2+}) and then transferred to conditions of 5 mM Mg^{2+}, 0.5 μM CYT-19 and no ATP-Mg^{2+} (●) or ATP-Mg^{2+} concentrations of 10 μM (○), 50 μM (▽), 250 μM (□), or 2 mM (△). The slow decrease in the fraction of native ribozyme observed in the absence of CYT-19 reflects adjustment of the equilibrium of native and misfolded ribozyme upon the decrease in Mg^{2+} concentration from 10 mM to 5 mM (ref.[229]). **(C)** Dependence of rate constant on ATP concentration. Hyperbolic fits to the data gave $K_{1/2}$ values of 0.3 ± 0.2 μM for the P5a variant (○) and 0.7 ± 0.1 μM for $E^{\Delta P5abc}$ (▽).

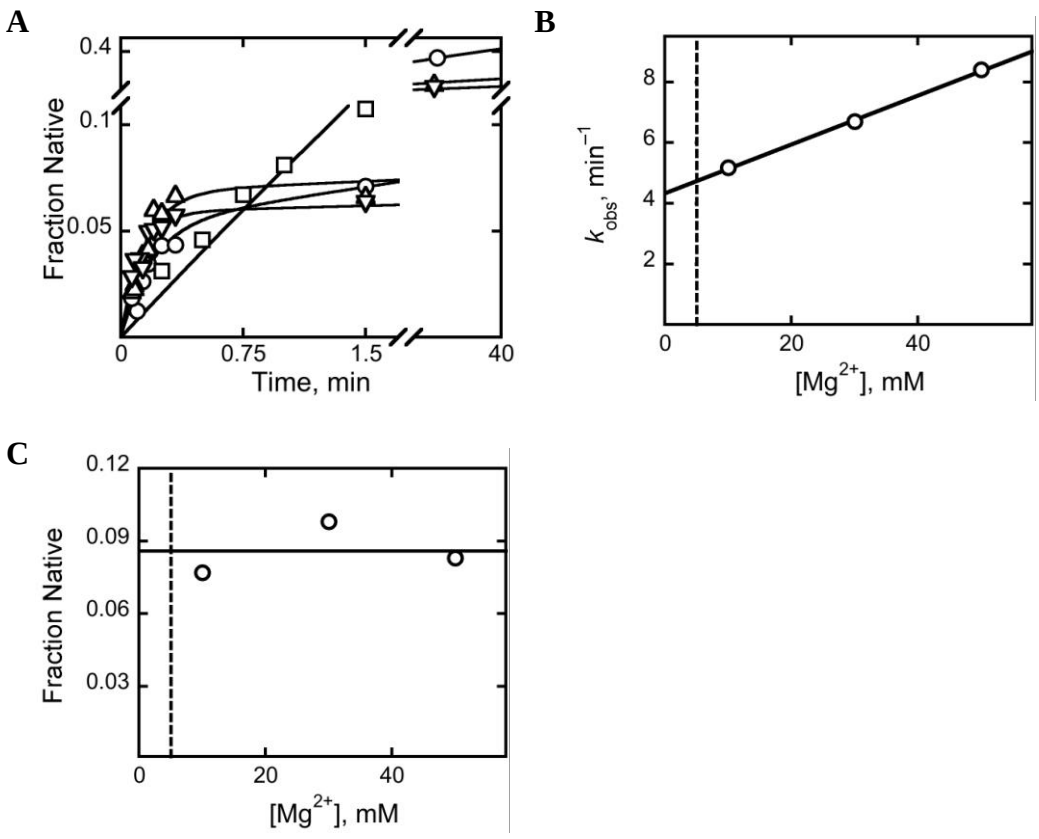


Figure 3.10: Folding to the native and misfolded species by the P5a ribozyme variant

Progress curves of native ribozyme formation, as measured by substrate cleavage (see Methods). Folding was performed in the presence of 5 mM (□), 10 mM (○), 30 mM (△), or 50 mM (▽) Mg^{2+}. **b,** Dependence of rate constant for formation of native ribozyme on Mg^{2+} concentration, obtained from the progress curves in panel a. This rate constant reflects formation of the native state by the small fraction of the ribozyme that avoids the long-lived misfolded species (folding from I_{trap} to the native species). **c,** Fraction of the ribozyme that folds to the native state without forming the misfolded species. No dependence of this amplitude on Mg^{2+} concentration was observed, analogous to prior observations for the wild-type ribozyme[203]. In panels b and c, dashed lines are shown at 5 mM Mg^{2+}, the conditions of the CYT-19 unfolding experiments, and the values obtained from extrapolation of the data to 5 mM Mg^{2+} were used in simulations (see Figure 3.11, Figure 3.6B and Figure 3.8A). It was not possible to determine these values directly at 5 mM Mg^{2+} because re-folding of the misfolded ribozyme is fast enough under these conditions to obscure the burst of native ribozyme that reflects folding directly from I_{trap} (shown by the squares in panel a, where a discrete burst of product formation was not observed).

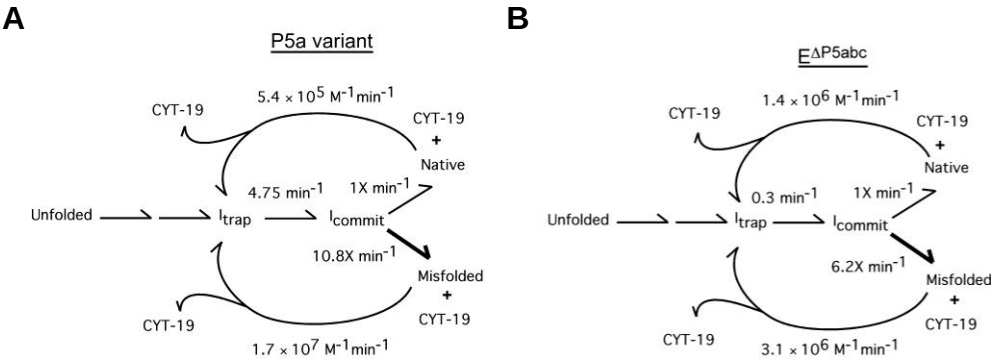


Figure 3.11: Schemes and kinetic parameters used for simulation of CYT-19-mediated approach to steady state of the native and misfolded ribozyme species

(A) Simulation of CYT-19-mediated unfolding of the P5a variant. **(B)** Simulation of CYT-19-mediated-unfolding of the P5abc deletion variant ($E^{\Delta P5abc}$).
For each scheme, values for CYT-19-mediated unfolding of the native ribozyme were obtained from experiments described herein. Values for subsequent folding of the ribozymes were determined here for the P5a variant (Figure 3.10) or previously for the $E^{\Delta P5abc}$ variant[234]. These schemes are essentially equivalent to the one shown in Figure 3.12, but the versions shown here allow the observed folding parameters to translate more directly into individual rate constants. For example, experiments establish a rate constant for the rate-limiting step, which in this scheme represents folding from I_{trap} as shown. The experiments also reveal the fraction that folds to the native state without forming the misfolded intermediate, which is reflected in the partitioning between pathways to the native and misfolded species. This partitioning is shown as occurring from a late intermediate, as determined previously[204,245], and is depicted as ratios of 'X', where $X \geq 5$ min^{-1}. (Provided that X is at least 5 min^{-1}, its value does not affect the simulations.) Additional experiments (not shown) indicated that, consistent with the model shown, CYT-19 does not affect the rate constant for folding of the ribozyme (≤ 500 nM CYT-19; higher concentrations cannot be used because they increase the rate of re-folding of misfolded ribozyme such that the initial folding cannot be distinguished from re-folding). For CYT-19-mediated unfolding of the misfolded ribozyme, the value for the $E^{\Delta P5abc}$ ribozyme was obtained directly from experiment (Figure 3.8B) and the value for the P5a variant was allowed to vary in the simulation. We found that with the other parameters constrained by their experimental values, a value within 2-fold of that shown, 1.7×10^7 M^{-1} min^{-1}, was required to give a reasonable description of the data. In order for the simulations to accurately model what would be observed in experiments, it was necessary to include in the simulation the process of inactivating CYT-19 and allowing subsequent folding to occur. Thus, the simulated curves in Figure 3.6B and Figure 3.8A show the fraction of ribozyme represented by ([Native] + $[I_{trap}]$ × the fraction that folds to native from I_{trap} (0.085 for the P5a variant and 0.14 for $E^{\Delta P5abc}$). Simulations were performed using Berkeley Madonna and Kinetic Explorer (Kintek, Austin TX).

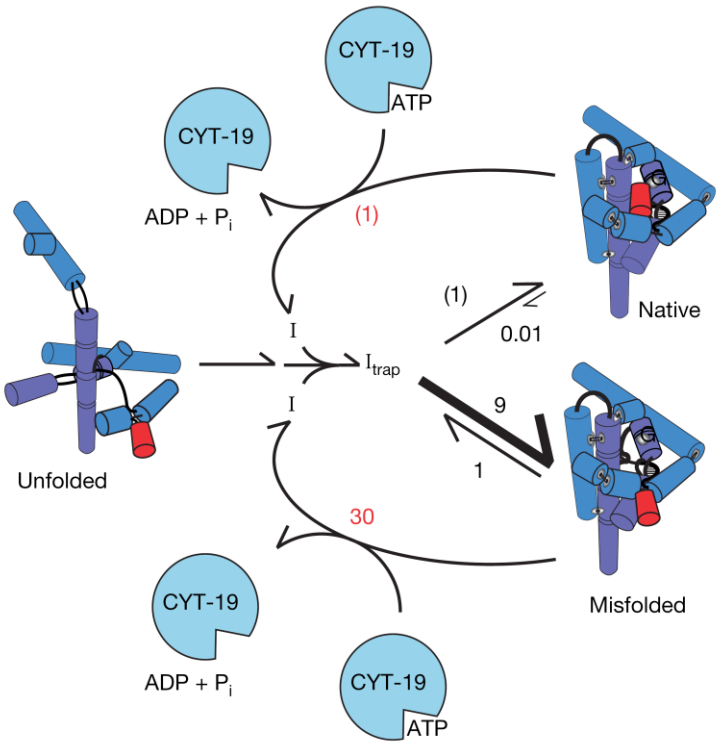


Figure 3.12: Model for chaperone activity

CYT-19 generates ribozyme intermediates (I) and then allows them to fold again, thus increasing the population of less stable intermediates that are kinetically favored. Values are for the P5a variant, normalized by native ribozyme unfolding (red; (1) indicates 5×10^5 M^{-1} min^{-1}, Figure 3.7), or 'direct' native state formation (black; (1) indicates 0.4 min^{-1}, Figure 3.10). The wild-type ribozyme behaves similarly at low Mg^{2+} concentration, but at higher Mg^{2+} concentration the native species is sufficiently stable that CYT-19 unfolds it poorly and therefore can accelerate its formation from the kinetically-trapped misfolded species.

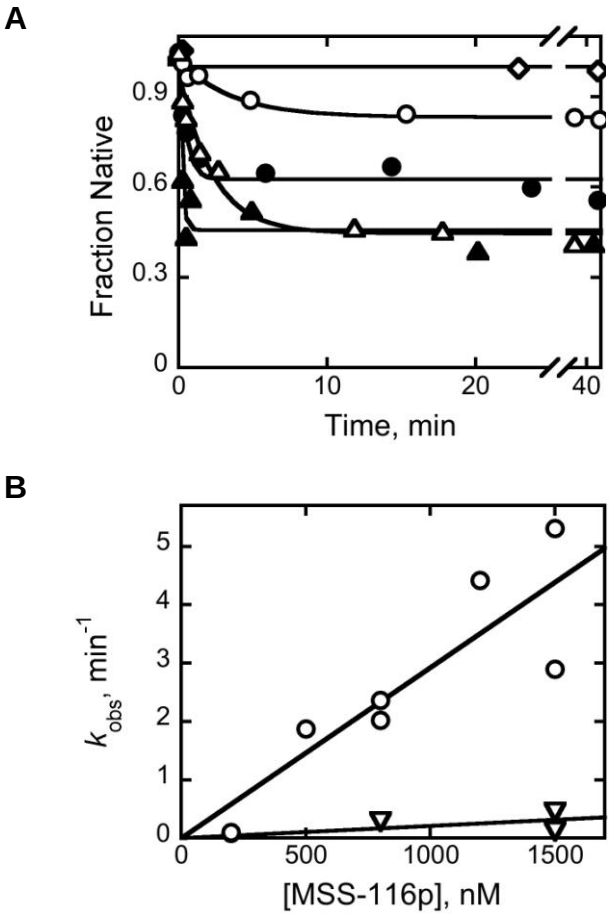


Figure 3.13: Efficient unfolding of ribozyme by DEAD-box protein, Mss116p compared to less efficient unfolding by SAT/AAA mutant

(A) P5a ribozyme was incubated at 50 °C for 30 minutes with 10 mM Mg^{2+} to allow formation of native ribozyme before adjusting the magnesium concentration to reaction conditions (5 mM). Native ribozyme unfolding of P5a variant by Mss116p and SAT/AAA mutant was performed at either 800 nM (circles) or 1500 nM (triangles). The reactions were quenched by adding $MgCl_2$ to 50 mM followed by addition of substrate to determine the fraction of native ribozyme. Progress curves of unfolding with Mss116p WT (filled symbols) or SAT/AAA (open symbols) is shown; No protein control with just storage buffer added (diamonds) is also shown. **(B)** Comparison of efficiency of WT Mss116p (circles) versus SAT/AAA mutant (triangles). k_{cat}/K_m values for WT Mss116p is 3×10^6 $M^{-1}min^{-1}$ and SAT/AAA (triangles) is 0.2×10^6 $M^{-1}min^{-1}$.

Chapter 4: The role of global stability versus local stability of RNA for efficient action by CYT-19

4.1 Introduction

To form functional structures, RNAs have to fold and stabilize their native states relative to both unfolded and misfolded states. The strategies that RNAs employ to stabilize structures include formation of networks of hydrogen bonded interactions, induction of physical constraints due to structures such as pseudoknots and binding of site-specific metal ions[246,247]. The factors that affect the stability of *Tetrahymena* ribozyme have been studied extensively[229,248-250] and suggest involvement of both localized and cooperative interactions. The core of the *Tetrahymena* ribozyme is highly conserved and is stabilized by peripheral elements that surround the core. The peripheral elements are in turn held by long-range tertiary contacts between them and the core. One such contact, the A-rich bulge interaction (Figure 4.1), was shown to be equally stable in the isolated P4-P6 sub-domain and when it is part of the whole ribozyme, suggesting that its effect is local; whereas, the loop 5b tetraloop-receptor interaction was found to be more stable in the whole ribozyme than as isolated P4-P6, suggesting a cooperative type of interaction[251]. Within the P4-P6 sub-domain, base stacking, hydrogen bonding and Mg^{2+} binding call contribute to a favorable enthalpy change upon tertiary structure formation[249].

Magnesium ions play important roles in stabilizing RNA structure. Diffuse magnesium ions counteract the negative charge on the phosphodiester backbone whereas

site-bound magnesium coordinates interactions with the tightly folded RNA[252]. In the *Tetrahymena* ribozyme, phosphorotioate substitutions of oxygen in the backbone of peripheral element, P5abc, drastically affected the entire ribozyme structure[250]. This indicates that backbone interactions to the known metal ions are weakened in the mutants, strongly suggesting that local, site-bound magnesium ions are crucial for the global stability of the whole ribozyme.

RNA stability may play an important role during structural modification processes involving DExD/H-box proteins. The efficiency of unwinding simple RNA duplexes by some DExD/H-box proteins has been shown to depend on duplex length with longer and more stable duplexes unwound with much lower efficiency compared to shorter ones[106,152,253]. Structured RNAs, containing RNA duplexes, may also depend on their stability to modulate the efficiency of action of DExD/H-box proteins upon them. Indeed, the stability of the *Tetrahymena* group I ribozyme was demonstrated to play an important role in determining the efficiency of its unfolding by CYT-19 (Ref[182] and Chapter 3). In addition, the degree of stability seems to quantitatively correlate with the efficiency of unfolding at least in the case of $E^{\Delta P5abc}$ version of the ribozyme[182] (See Chapter 3). Here, to explore this correlation further, each of the five long-range tertiary contacts was disrupted. The equilibrium values for native species relative to misfolded species for a mutant (P5a in Figure 4.1), in which an A-minor tertiary contact has been disrupted, was shown to strongly correlate with the unfolding efficiency of CYT-19 on the native and misfolded species.

Thermodynamic analyses were performed to obtain an estimate of the degree of destabilization of the different mutants. The efficiencies of re-folding of misfolded state by CYT-19, which is an indirect measure of the unfolding efficiencies, were determined for the mutants. Initial studies show that, although global destabilization leads to an increase in the efficiency of re-folding by CYT-19, there is not an absolute correlation between the degree of destabilization and re-folding efficiency. The results suggest that local stabilities of certain interactions could play a role in efficient unfolding by CYT-19.

4.2 Results and Discussion

4.2.1 Construction of mutants:

There are five long-range tertiary contacts in the *Tetrahymena* ribozyme (Figure 4.1), each of which has been disrupted in separate mutants. Single stranded loop sequences that form tertiary contacts were changed to the canonical "UUCG" tetraloops in four out of the six mutants. The rationale behind this approach was to disrupt the tertiary contacts without affecting the stability of the secondary structure[230]. Plasmids for all tertiary contact mutants (except L9) (Figure 4.1) were made by *in vitro* transcription as described for wild type ribozyme in section 2.2.1 of chapter 2. The L9 mutation was introduced in the L-21*Sca*I plasmid using Quikchange site-directed mutagenesis kit (Stratagene, La Jolla, CA) and transcribed using *in vitro* transcription.

4.2.2 Equilibrium for native P5a ribozyme relative to misfolded ribozyme corresponds to difference in efficiency of their unfolding by CYT-19:

Previous data suggests that the efficiency of unfolding of native conformer of P5a mutant by CYT-19 is much higher compared to that of native wild type ribozyme[182] (Chapter 3). In addition, the native and the misfolded species of $E^{\Delta P5abc}$ ribozyme which have comparable stability are unfolded with equal efficiency, suggesting that stability might be a key factor in determining efficiency of unfolding by CYT-19. In order to determine whether there is quantitative correlation between CYT-19 mediated unfolding efficiency for P5a mutant and its stability, the stability of the native ribozyme relative to the misfolded species was determined by measuring their equilibrium values. The ribozyme was allowed to fold to equilibrium under one solution condition, and the equilibrium was perturbed by changing the solution condition to that of the CYT-19 reaction (Figure 4.2). During re-equilibration, the fraction native shifted from 0.91 to a new equilibrium value of 0.86, giving a calculated equilibrium value of native relative to misfolded as 10. This value is similar to the 30-fold difference in efficiency between the experimentally determined value for unfolding of native species (5×10^5 $M^{-1}min^{-1}$) and the predicted value from simulation for misfolded ribozyme (1.7×10^7 $M^{-1}min^{-1}$) (see Chapter 3). These data bolster the previous finding that the efficiency of unfolding inversely correlates with the global stability of the ribozyme structure[182].

4.2.3 Preliminary evidence for correlation between local stability of ribozyme and the efficiency of its unfolding by CYT-19:

The native RNAs of both the P5a and $E^{\Delta P5abc}$ mutants are unfolded efficiently by CYT-19. Since the tertiary contact P5a (Figure 4.1) is disrupted in both mutants it could be that the region close to P5a tertiary contact is important for efficient unfolding by CYT-19. However, an alternative possibility is that disruption of specific tertiary contacts destabilizes the RNAs globally and that it is purely global stability that matters for efficient unfolding by CYT-19. In order to probe whether global stability plays a crucial role or whether local stability is important, specific tertiary contacts were disrupted, and their effects on global stability of native state were compared to their efficiency of unfolding by CYT-19. If there is a strict correlation between the degree of destabilization and the efficiency of unfolding, this would strongly suggest that global stability of the ribozyme is the principal determinant for CYT-19 mediated unfolding. On the other hand, if certain tertiary contact mutants showed increased efficiency of unfolding by CYT-19 without any correlation with the degree of global destabilization, this would suggest that CYT-19 mediated unfolding involves disruption of specific tertiary contacts such that it followed a specific pathway.

Folding of *Tetrahymena* ribozyme to the native state at equilibrium was dependent on the concentration of magnesium at a low background monovalent ion concentration. Ribozyme mutants were made in which each of the five tertiary contacts was disrupted (Figure 4.1). The magnesium concentration required for folding to the native state was determined using an activity assay (Ref[234] and Figure 4.3A). Briefly, the

assay involves incubating the ribozyme at different concentrations of magnesium to allow equilibration between the native and unfolded (or other intermediate states). This is followed by addition of high magnesium to a concentration of 50 mM which allows rapid conversion of all the unfolded and partial folded intermediates to fold to mostly (about 90%) the misfolded species and the rest (about 10%) to fold to native ribozyme. A substrate cleavage assay was then performed to determine the fraction of native versus misfolded ribozyme at each concentration of magnesium.

Magnesium binding to folded RNA is highly cooperative[254], and thus, Hill plot analysis of magnesium binding can provide information on the stability of folded RNA. $K_{1/2}$ for magnesium obtained from such experiments can give an idea as to how stable the native RNA is relative to other intermediates. Although quantitative analysis of such data to obtain free energy relationships between two states is possible if the two states are defined, the analyses become complicated if there are multiple states[255]. Thus, direct comparison of free energy changes between the wild type and the mutants becomes difficult because the higher magnesium concentration required for folding may also allow accumulation of a different ensemble of unfolded species for mutants compared to wild type. Nonetheless, magnesium $K_{1/2}$ provides an approximation of degree of destabilization of native species of mutant ribozyme compared to native ribozyme of wild type.

The magnesium $K_{1/2}$ data for the different mutants are shown in Figure 4.3B and summarized in Table 4.1. The efficiency for re-folding of misfolded species to native species of these mutants by CYT-19 was also obtained by measuring rate constants at

different concentrations of CYT-19 and plotting them against CYT-19 concentration (Figure 4.3C). These results are also summarized in Table 4.1.

All the mutants require increased Mg^{2+} for folding as indicated by higher $K_{1/2}$ values compared to wild type. This suggests that local disruption of any tertiary contact destabilizes the ribozyme globally. $K_{1/2}$s for mutants vary, suggesting that the degree of destabilization is different for different mutants. Consistent with previous observations[231], the P5a mutant was, indeed, destabilizing for the native state; however, it is noteworthy to mention that the P5a mutant was the most destabilizing of all the mutants tested (observed from the $K_{1/2}$ values).

In the absence of direct data for CYT-19 mediated unfolding, it is reasonable to assume that re-folding efficiency reflects efficiency of unfolding because re-folding to the native state probably involves multiple cycles of unfolding of the misfolded ribozyme[182]. A comparison of the efficiency of re-folding by CYT-19 for all of the different mutants shows that, indeed, the most destabilizing of the tertiary mutants, the P5a, was refolded with the greatest efficiency by CYT-19. This contact was also disrupted in another mutant, $E^{\Delta P5abc}$, whose stability was shown to directly correlate with the degree of unfolding by CYT-19[182]. Perhaps, this contact is crucial for efficient unfolding by CYT-19, presumably by a pathway that involves its disruption, facilitating large scale rearrangement of the ribozyme. CYT-19 mediated efficiency (k_{cat}/K_m) for re-folding of the misfolded species for all other mutants was higher compared to wild type, indicating that stability, indeed, determines efficient CYT-19 action and is also consistent with previous results[182]. In addition, the mutant with the most global destabilization (P5a)

showed the most efficient re-folding of its misfolded state by CYT-19. On the other hand, there is no strict correlation between the $K_{1/2}$ and the efficiency of re-folding of misfolded structures among the different mutants. These data suggest that the stability of local structures such as the region encompassing the P5a contact might be important.

There are limitations to the experiments described above. First, the re-folding efficiency is not a direct estimate of unfolding efficiency because multiple cycles of unfolding of misfolded state, based on the kinetic partitioning, would have to be taken into account to estimate the unfolding efficiency of native state. Second, the $K_{1/2}$ values obtained here reflect the stability of the native state of mutants relative to all other unfolded states and intermediates and not exclusively the misfolded state. Therefore, even if re-folding is a good measure of the unfolding of the misfolded ribozyme, it is necessary to estimate the stability of the misfolded species of various mutants for direct comparison with unfolding efficiencies.

The limitations stated above can be overcome by directly measuring the efficiency of unfolding of native ribozymes for the different mutants under the same solution conditions. Direct comparisons using the $K_{1/2}$ can then be made between the efficiency of unfolding of native ribozyme and the degree of destabilization of native species for the various mutants.

4.3 Summary and prospectus

Many DExD/H- box proteins act at discrete steps during RNA mediated processes suggesting that they act on specific RNA substrates. However, some proteins of the

DEAD-box family, namely CYT-19 and Mss116p, act more generally on group I and group II intron substrates[107,159,181], and it is unclear how their non-specific action would allow distinguishing native from misfolded RNAs. Using the catalytic property of *Tetrahymena* group I ribozyme, it was demonstrated that the efficiency of unfolding of the ribozyme primarily depends on the stability of the *Tetrahymena* ribozyme[182]. In order to further probe the effect of stability on the action of CYT-19 and to determine whether local stability or global stability plays a crucial role, several tertiary mutants were constructed. CYT-19's action was compared between mutant ribozymes and wild type. The native conformers of mutants were found to be globally destabilized by disruption of local structures. Initial results suggest that disrupting any local structure results in ribozymes that are destabilized to varying degrees. Results also suggest that CYT-19 action is more efficient at re-folding misfolded ribozymes for all mutants compared to wild type. In addition, there is no absolute correlation between the degree of global destabilization and efficiency of re-folding, suggesting that certain local structures might play a role. Efficiency of unfolding must be directly monitored to corroborate these findings.

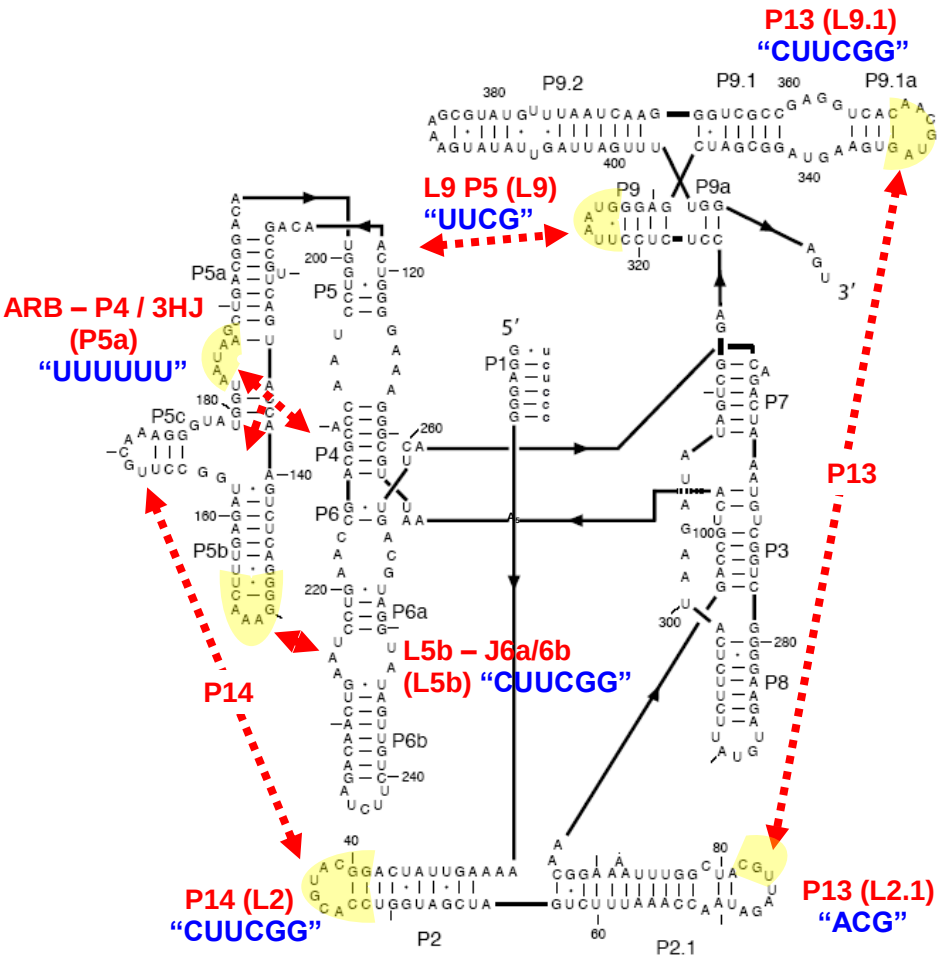


Figure 4.1: Tertiary contact disruption mutants of *Tetrahymena* group I ribozyme

Secondary structure of *Tetrahymena* group I ribozyme showing the five tertiary contacts (red broken arrows). Single contact disruption mutants were made at the residues highlighted in yellow. New residues that were introduced are shown in blue and the names of tertiary contacts are shown in red (the abbreviated names used in text, also in red, are in parenthesis). The five mutants are P5a, L2, L9, L5b, L9.1 and L2.1 (the last two mutants disrupt the same tertiary contact, P13)

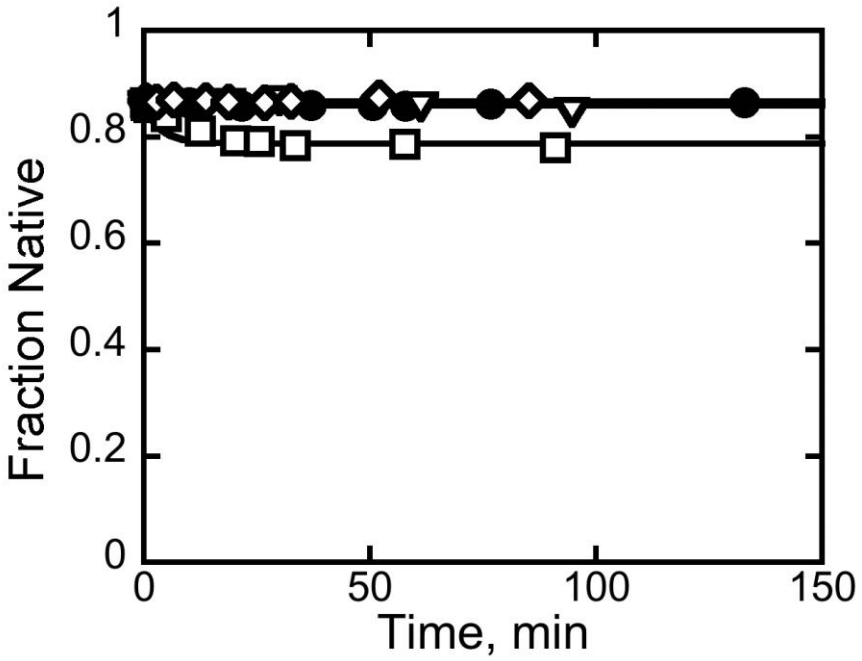


Figure 4.2: Equilibrium for native state of P5a mutant relative to misfolded species under the same solution conditions as CYT-19-mediated unfolding is about 10

P5a ribozyme was incubated at 50 °C for 30 minutes to make native ribozyme and cooled to 25 °C before use. The ribozyme at 2 μM with 50 mM Mg^{2+} was diluted to either 200 nM ribozyme and 5 mM Mg^{2+} (▽) or to 200 nM ribozyme, 5 mM Mg^{2+} and 50 mM KCl (□). Control reactions in which the ribozyme alone was diluted from 2μM to 200 nM, maintaining the 50 mM Mg^{2+} (◇) or when no dilution was performed (2 μM ribozyme and 50 mM Mg^{2+}(●)) showed no decrease in fraction of native ribozyme, indicating 50 mM KCl concentration affects equilibrium of native ribozyme of P5a mutant

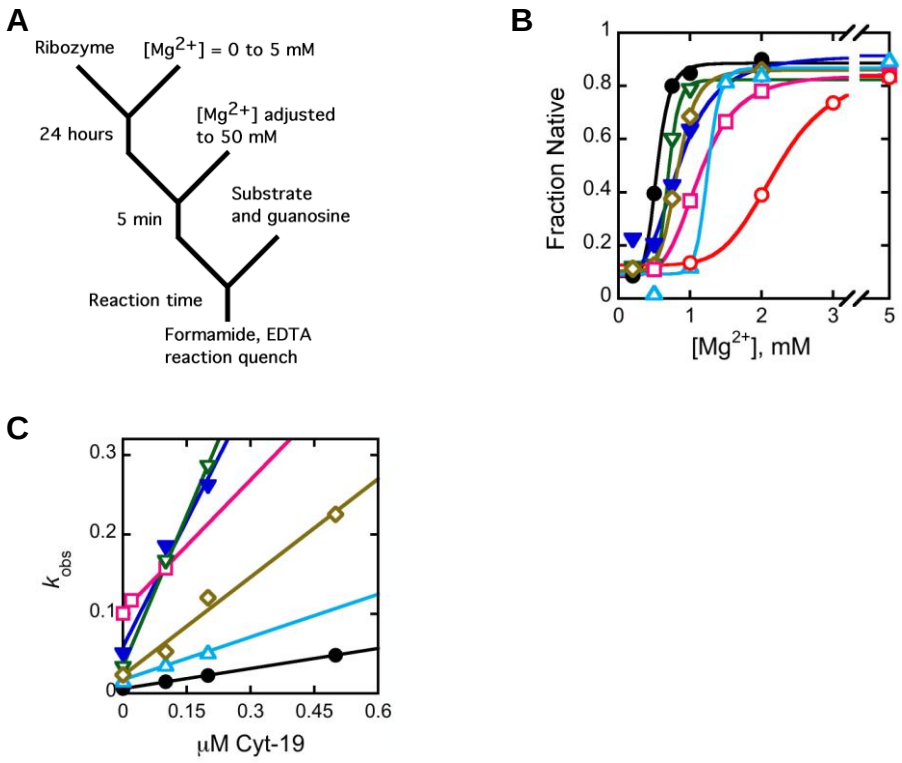


Figure 4.3: Tertiary contact disruption mutants require higher magnesium to fold to native state and all mutants are efficiently re-folded by CYT-19

(A) Scheme for measuring fraction of the ribozyme that fold to native state as a function of magnesium concentration at equilibrium. The wild type and mutant ribozymes were incubated at different magnesium concentrations for 24 hours to allow equilibrium. Magnesium was adjusted to 50 mM in all reactions. The reactions were incubated for 5 minutes following this adjustment to allow complete folding to either native or misfolded ribozymes. Substrate and guanosine were added and complete time course reactions were performed for substrate cleavage. At each magnesium concentration the fraction of native ribozyme relative to misfolded species are plotted against each magnesium concentration shown in **(B)**. Conditions were 50 mM Na.MOPS and 25 °C. **(B)** The fraction of native ribozyme at equilibrium plotted against Mg^{2+} concentration for wild type (●) and L2.1 (▼), L9.1 (▽), L9 (◇), L2 (□), P5a (○), L5b (△) mutants. The data were fit to the Hill equation – $([Mg^{2+}]^n)/(K_{1/2}{}^n + [Mg^{2+}]^n) + C$, where $K_{1/2}$ represents magnesium concentration required to give ½ of native ribozyme, n represents Hill coefficient and C represents offset which is the fraction of the ribozyme that directly folds to native from unfolded or partially unfolded species giving a minimum fraction native species (~10%) even at 0 mM Mg^{2+} concentration **(C)** Efficiency of re-folding for wild type and mutants at different concentration of CYT-19. Colors and symbols the same as in **(B)** Data were fit to straight line and the slopes (k_{cat}/K_m) obtained are summarized in Table 4.1.

Ribozyme	Average $K_{1/2}$ mM	Efficiency of re-folding by CYT-19 (at 5 mM Mg^{++}) $M^{-1}min^{-1}$
Wild type	0.54	9×10^4 (Ref.[73])
L9P5	0.61	4.1×10^5
P13 (L9.1)	0.73	1.1×10^6
P13 (L2.1)	0.84	1.1×10^6
P14	1.17	5.5×10^5
L5bTL	1.2	1.8×10^5
P5a	2.3	6.5×10^5 #
$E^{\Delta P5abc}$	2.4 [234]	4.4×10^5

 # at 7 mM Mg^{2+}; Results from simulation predicts a value of 1.4×10^6 $M^{-1}min^{-1}$ at 5 mM Mg^{2+}

Table 4.1: Comparison of efficiency of CYT-19 mediated re-folding reactions and Mg^{2+} $K_{1/2}$ values for various tertiary contact disruption mutants

Chapter 5: Chaperone activities of nucleocapsid protein (NC) on the *Tetrahymena* group I ribozyme

5.1 Introduction

Non-specific RNA binding proteins have long been known to possess RNA chaperone activities *in vitro*[103,172,174,175,180,256,257]. They interact with RNAs with little or no sequence specificity and disrupt base pairs to allow formation of alternative, more favorable pairings. Many single stranded RNA binding proteins, including hnRNP, stpA, S12, HIV nucleocapsid (NC) protein and Hfq, contain RNA binding motifs such as the RGG box and the zinc knuckle domain, and they modify RNA structures both locally and globally[258]. The mode of interaction is thought to be largely electrostatic between the protein's positively charged basic residues and the negatively charged phosphodiester backbone of RNA[259].

The first demonstrations of chaperone activity by single-stranded binding proteins came from the pioneering studies of Karpel, Fresco and colleagues[103,175,260]. Small structured RNAs, such as the $tRNA^{Leu}$, were shown to misfold *in vitro*[260], and in one study, UP1, a nucleic acid binding protein, was demonstrated to induce helix destabilization in $tRNA^{Leu}$ and *E. coli* 5S RNA, presumably by binding tightly to exposed single stranded regions[175]. hnRNP proteins also bind single stranded RNAs, coating the nascent mRNA[172]. These proteins provide the strongest evidence for non-specific RNA binding[261,262], which probably is a requirement because they have to bind to several different mRNAs made by the cell. Catalytic activity allows estimation of fraction of

native folded structures versus inactive or unfolded species in a mixture. This allows one to unambiguously distinguish correctly folded from incorrect structures. Using an activity assay, Coetzee *et. al*. demonstrated that the *Eshcerichia coli* ribosomal protein, S12, can promote splicing of inactive ribozymes[180]. *In vivo*, overexpression of non-specific RNA binding protein, StpA, has been shown to resolve a folding trap in T4 phage td group I intron splicing construct[176-178,256]. StpA can also facilitate strand annealing and exchange activities on model substrates *in vitro*[10,178] and may have such activities *in vivo*. Although these proteins were not shown directly to function as chaperones *in vivo*, quantitative comparisons of their *in vitro* chaperone properties with those of DExD/H-box proteins may be useful for understanding how DExD/H-box proteins transduce energy from ATP to facilitate RNA rearrangements.

HIV nucleocapsid protein is a non-specific RNA binding protein that was shown to possess chaperone activity based on its enhancement of substrate strand annealing and product strand dissociation activities of the hammerhead ribozyme[173,174]. NC is a small, 55-amino acid protein containing two zinc-finger motifs (Figure 5.1) and is involved in various aspects of retrovirus maturation[263,264]. One of the steps involves utilization of the 3'end of a cellular tRNA as a primer for making viral DNAs. In this context, NC-mediated strand exchange reactions might be especially significant *in vivo* because they are crucial for replication of the HIV virus genome[257,265-267]. The mechanism of tRNA primer annealing to primer binding site (PBS), including negative and positive strand transfers by NC, have been well investigated[268,269]. In addition, structural modifications in tRNAs were shown to be induced by NC mediated non-specific binding[270] and structure

disruption[271]. For several of these functions, stoichiometric binding is essential, and NC binds to tRNAs in a 1:1 or a 5:1 ratio, depending upon the function[271,272]. Recently, single molecule studies showed that a recognition element TAR exists in two distinct conformations and that NC shifts the equilibrium population of one to another[273,274] presumably by disrupting base-pairs and allowing formation of more stable conformers.

Here, in order to probe whether the non-ATP dependent non-specific RNA binding protein NC can facilitate re-folding of a highly structured RNA and to compare mechanistic features of a DEAD-box protein, CYT-19, with NC, ribozyme re-folding kinetics were investigated by performing reactions with NC under identical solution conditions to that of CYT-19. NC was found to accelerate re-folding of the long-lived misfolded conformation of *Tetrahymena* group I ribozyme, suggesting that it can act as a chaperone of a structured RNA. Additional results also indicate that, although the two classes of chaperones act, in *in vitro* assays, in a generally similar manner, there might be key differences in their action.

5.2 Results

NC acts as a chaperone of the misfolded *Tetrahymena* group I ribozyme. Re-folding assays were performed similarly to the one shown in **scheme 1** (Chapter 2) (NC instead of CYT-19), and NC was found to accelerate re-folding of the long-lived misfolded species to the native state in an ATP-independent manner (Figure 5.2). The efficiency of re-folding, under similar solution condition, was similar (within error) to that of CYT-19; 1.6×10^5 $M^{-1}min^{-1}$ for NC (Figure 5.3A, B) versus 9×10^4 $M^{-1}min^{-1}$ for

CYT-19 (Separate assays to monitor the efficiency of partial unfolding of native ribozyme also gave comparable efficiencies for NC and CYT-19). This suggests that NC acts as a chaperone and can resolve misfolded conformations in structured RNA to facilitate re-folding. However, in contrast to CYT-19, there is an effect of ribozyme concentration on the rate constant for re-folding, even though NC is in excess. The rate constant for re-folding increases with increasing NC concentration but is sharply dependent on NC concentration after an initial non-dependent phase (Figure 5.3B). This effect is similar to the ATP-independent effect seen for unfolding of native wild type ribozyme by CYT-19 at higher concentrations (data not shown). The concentration of NC, at which the linear dependence on rate constant begins increases with increases in ribozyme concentration (Figure 5.4). This suggests a stoichiometric effect with respect to ribozyme concentration rather than upward curvature reflecting a requirement for NC multimerization. In this experiment, two concentrations of ribozyme were used – 150 nM and 500 nM. With 150 nM ribozyme concentration, the rate constant increases linearly starting at about 500 nM NC suggesting a possibility of approximately 3 potential tight binding sites in the RNA bound by NC. At 500 nM ribozyme concentration, uncertainty in the data suggests a range of stoichiometry of 2:1 to 8:1 NC to ribozyme (Figure 5.4). Additional evidence for the stoichiometric effect comes from experiments probing unwinding of the P1 duplex of *Tetrahymena* ribozyme in which the acceleration of re-folding by NC was dependent on the ribozyme concentration (Pilar Tijerna and Rick Russell - data not shown). More data needs to be obtained in order to evaluate both the dependence of rate constants and the stoichiometry.

Prior studies have demonstrated the requirement of a threshold concentration of about 1 NC molecule for every 7 nucleotides in RNA or DNA for chaperone action [263]. This stoichiometry would most simply predict about 60 NC (for the 413 nucleotide ribozyme). This value is much more than the 2-8 NC implied by the data, but it is possible that the number of binding sites is much lower for a highly structured RNA such as the *Tetrahymena* group I ribozyme.

5.3 **Summary and prospectus**

Cells employ two types of proteins in various cellular processes involving RNA - ATPases such as DExD/H-box proteins and non-ATP dependent simple basic proteins. Although many non-specific RNA binding proteins possess RNA chaperone activities *in vitro*, it remains to be seen whether these proteins actually function as RNA chaperone *in vivo*. Comparative studies of DExD/H-box proteins and non-specific RNA binding proteins on structured RNAs will potentially give mechanistic insights and help provide clues to their evolution. With this goal, activities of NC protein were compared with a bona fide DExD/H-box chaperone, CYT-19. On a general level, both proteins seem to perform well and mediate re-folding of a misfolded structured RNA to its native state with equal efficiency. However, detailed probing suggests an important difference: NC protein shows a stoichiometric effect, that is, the efficiency of re-folding is much higher at higher concentration than at lower concentration with a distinct cutoff point, which in turn, is dependent on the concentration of ribozyme. The significance of this "threshold" effect awaits further investigation.

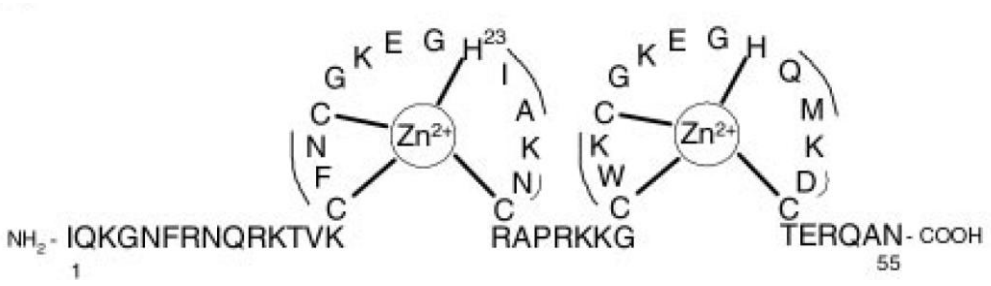


Figure 5.1: Structure of the 55-amino acid HIV NC protein

HIV Nucleocapsid (NC) protein, pNL4-3 isolate used in the study, has two zinc finger motifs, with the two site-bound zinc ions specifically coordinating interactions between conserved residues. The variable amino acids in each domain are indicated by curved lines. Figure adapted from ref[264].

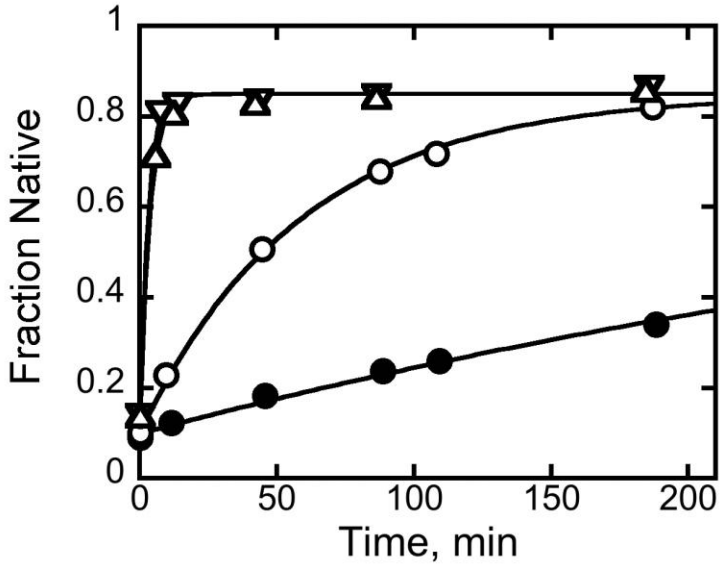


Figure 5.2: Non-ATP dependent NC protein accelerates re-folding of misfolded ribozyme of *Tetrahymena* to the native state.

Misfolded WT ribozyme was formed by incubating the unfolded ribozyme with 10 mM Mg^{2+} at 25 °C for 10 minutes before adjusting the concentration to that of the reaction (5 mM); Final ribozyme concentration in the reaction is 100 nM. Re-folding of misfolded L-21*ScaI* ribozyme in the presence (○,△,▽) or absence (●) of NC. Rate constants for re-folding were 0.002 min^{-1} in the absence of NC (●), 0.017min^{-1} in the presence of 500 nM NC (○), and 0.28min^{-1} in the presence of 2 μM NC (▽). Addition of 2 mM Mg^{2+}-ATP (△) gave a rate constant of 0.35 min^{-1}, the same within error as that of the reaction in the absence of ATP (▽). Reactions were performed under similar solution conditions as that of CYT-19 mediated re-folding (1 × Na-MOPS, 50 mM KCl, 5% glycerol).

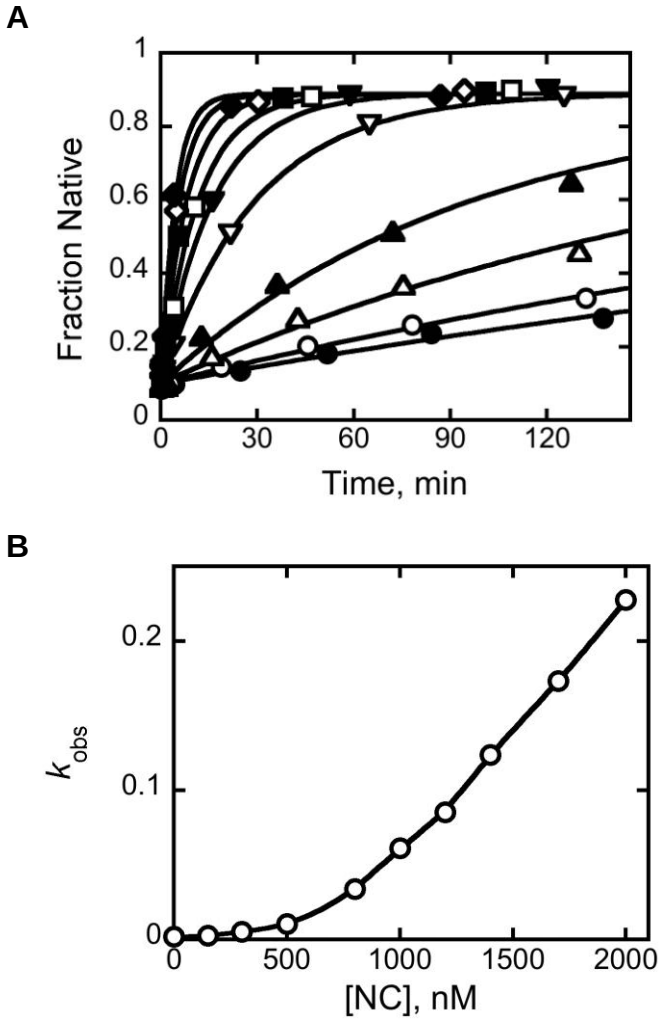


Figure 5.3: Plot of concentration dependence on re-folding shows upward curvature at higher concentrations of NC

(A) Misfolded L-21*Sca*I ribozyme was formed by incubating in 10 mM Mg^{2+} at 25 °C for 10 min before adjusting the magnesium concentration to reaction conditions (5 mM). Ribozyme concentration in the reaction is 100 nM. Re-folding reactions at different NC concentrations gave rate constants that increased with increasing NC concentrations – 0 nM, 0.002 min^{-1} (●), 150 nM, 0.0028 min^{-1} (○), 300 nM, 0.005 min^{-1} (△), 500 nM, 0.01 min^{-1} (▲), 800 nM, 0.034 min^{-1} (▽), 1000 nM, 0.061 min^{-1} (▼), 1200 nM, 0.085 min^{-1} (□), 1400 nM, 0.123 min^{-1} (■), 1700 nM, 0.17 min^{-1} (◇), 2000 nM, 0.23 min^{-1} (◆). **(B)** The dependence of rate constant (obtained from fits in Figure 5.3A) on NC concentration gave a sharp upward curvature with the slope representing k_{cat}/K_m (efficiency) of 1.6×10^5 $M^{-1}min^{-1}$. Solution conditions were 5 mM Mg^{2+}, 50 mM KCl, and 5% glycerol.

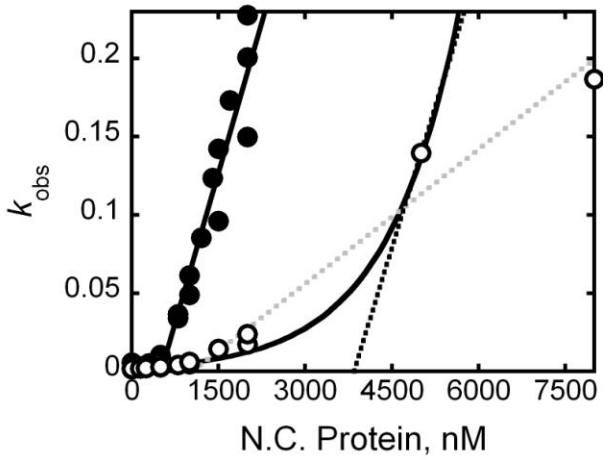
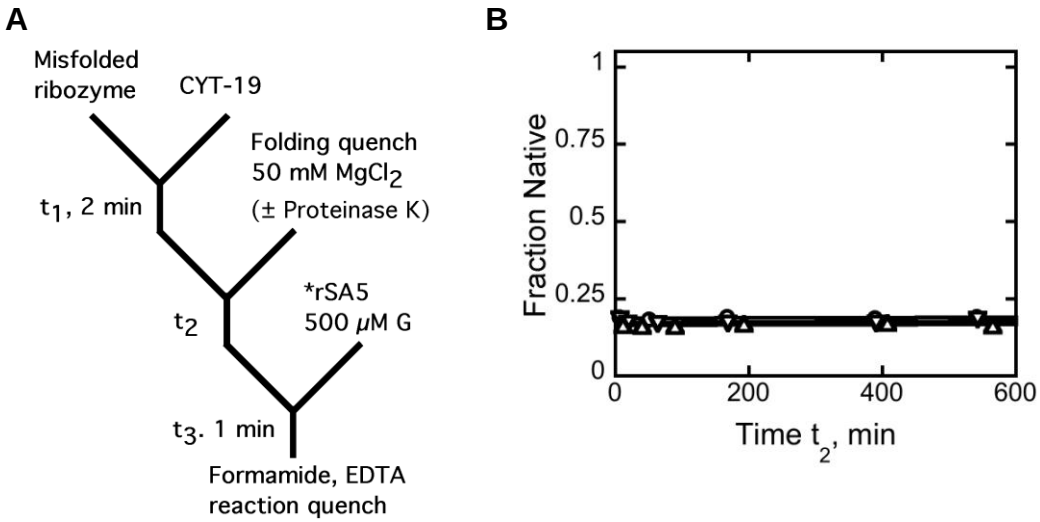


Figure 5.4: Plot of concentration dependence of NC on re-folding shows stoichiometric effect of ribozyme concentration

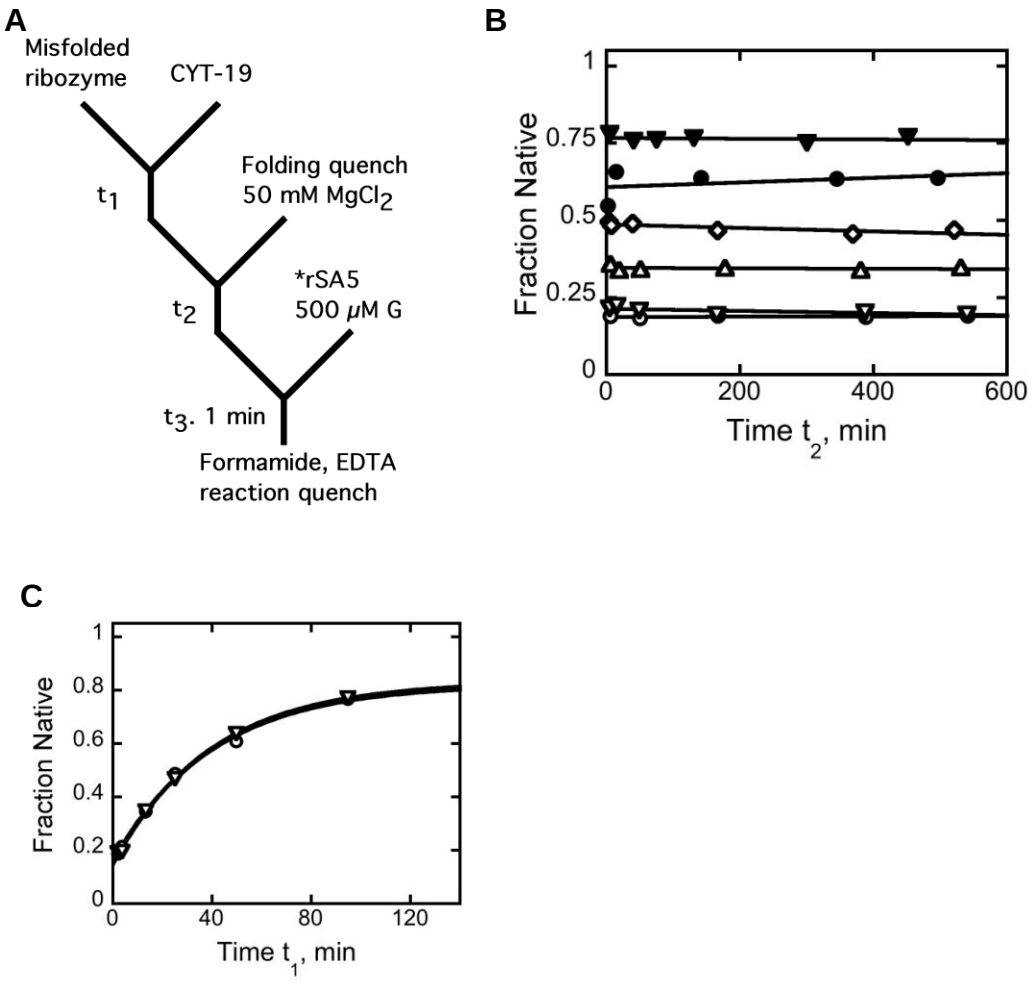
Rate constants plotted against NC concentration. Reactions were performed at two different ribozyme concentrations – 150 nM (●) and 500 nM (○). The steep linear increase of rate constants at 150 nM ribozyme concentration (●) extrapolates back to the x-intercept at about 500 nM NC. A possible fit at 500 nM ribozyme concentration (○) extrapolates to about 3880 nM NC (black dotted line). However, it is also possible that the slope of the linear increase in rate constant is different (gray dotted line) at 500 nM ribozyme concentration. At 150 nM ribozyme concentration, the stoichiometry of NC to ribozyme is about 3:1 ratio. At 500 nM ribozyme, uncertainty in the data gives an estimated range for NC to ribozyme stoichiometry of 2:1 to 8:1.

Appendix



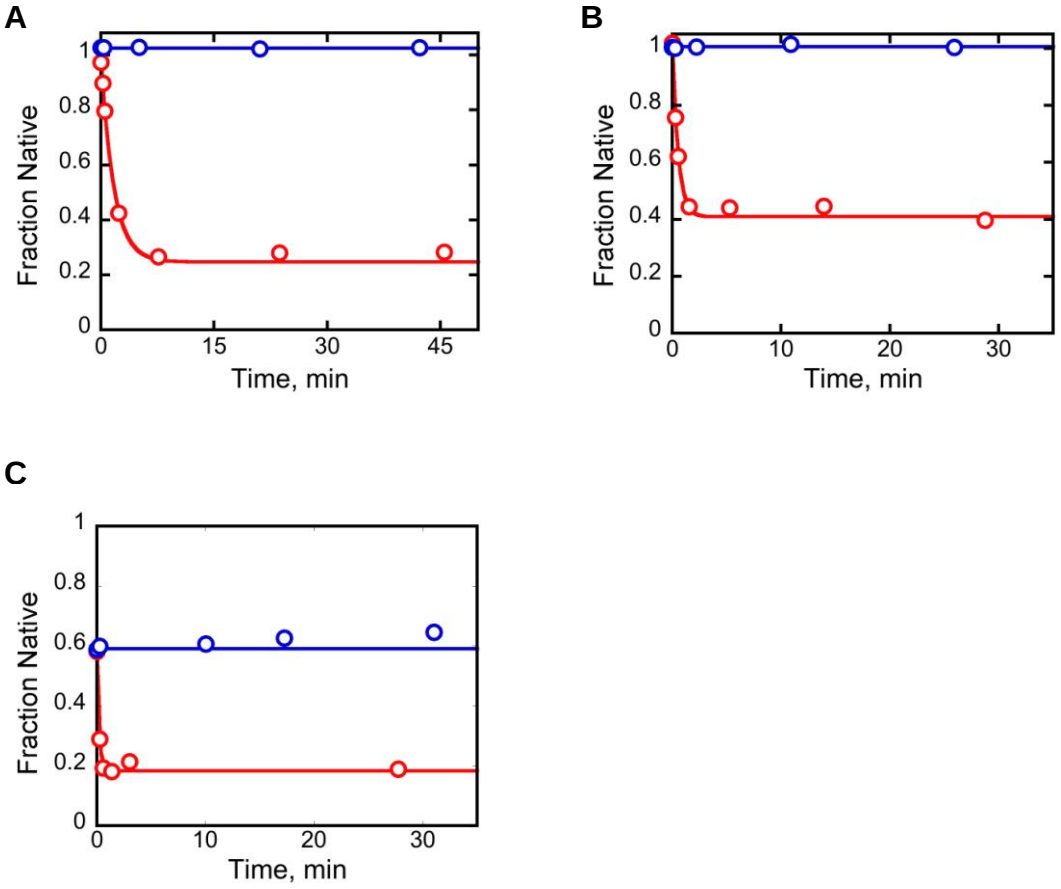
A1: Control reactions demonstrating that CYT-19 is inactivated under folding quench conditions

(A) The reaction scheme. CYT-19 (500 nM) was added to the misfolded ribozyme in the presence of ATP for 2 min (t_1), at which time further re-folding was blocked by addition of $MgCl_2$ to a final concentration of 50 mM. The quenched solutions were then incubated at 25 °C for various times, t_2, before substrate cleavage reactions were performed to determine the fraction of native ribozyme. **(B)** Fraction native in the folding quench that contains 50 mM $MgCl_2$ (○) or 50 mM $MgCl_2$ and 1 mg/mL Proteinase K (▽) plotted against t_2. A control reaction in which CYT-19 was replaced by storage buffer (△) is also shown. The fraction native ribozyme remains the same in the folding quench even for long incubations greater than 500 minutes suggesting that CYT-19 was completely eliminated in the folding quench and that the quench stops re-folding.



A2: Control reactions demonstrating that the fraction native ribozyme does not decrease in the folding quench.

(A) Reaction scheme. This experiment was identical to A1 except that the incubation time with CYT-19 was also varied. Incubation time, t_1, was 2 min (○), 3.67 min (▽), 13.17 min (△), 25 min (◊), 49.67 min (●), or 94.67 min (▼). (B) The fraction of native ribozyme at various times, t_2, in the re-folding quench is plotted. The fraction of native ribozyme obtained from panel B were plotted against t_1 to give a progress curve for re-folding of the misfolded ribozyme. The same curve obtained whether the Y-intercepts were plotted (○) or the final time points greater than 400 minutes were plotted (▽) against folding time, t_1, clearly demonstrating that little or no re-folding occurs in the folding quench and that the folding quench is effective at any concentration of CYT-19 used.



A3: Control reactions to ensure effectiveness of the folding quench in CYT-19 mediated unfolding reactions.

Wild type ribozyme **(A),** P5a variant **(B),** or P5abc-delted ribozyme ($E^{\Delta P5abc}$) **(C)** were pre-folded to give predominantly native ribozyme and then transferred to folding quench conditions (1 mg/mL Proteinase K for (A) and (B), 50 mM $MgCl_2$ for all panels, blue curves) or to the folding reaction conditions (red curves). CYT-19 (4 μM, 1 μM and 1.2 μM respectively for (A), (B) and (C)) was added and reactions performed. The fraction of native ribozyme does not decrease upon CYT-19 addition under quench conditions (blue curves) for any ribozyme indicating CYT-19 is inactivated under the quench conditions. Analagous experiments in which the folding quench did not include proteinase K for P5a variant and $E^{\Delta P5abc}$ showed similar results as shown here suggesting CYT-19 mediated unfolding was blocked even in the absence of proteinase K.

List of abbreviations and short forms used in thesis

Alt P3 – P stands for paired region; Alt P3 is the proposed alternative pairing of P3 region of *Tetrahymena* group I ribozyme, in which J8/7 strand is paired.
ATP – Adenosine triphosphate
ADP – Adenosine diphosphate
AMPPNP – Analog of ATP: 5'-adenylyl-beta, gamma-imidodiphosphate
CYT – Associated with cytochrome function in mitochondria
DEAD-box or DEAH-box – The ATP binding motif in the DExD/H-box proteins, representing the amino acids aspartic acid (D), glutamic acid (E), alanine (A) and either aspartic acid (D) or histidine (H) from N-terminus to C-terminus.
DTT - Dithiothreitol
$E^{\Delta P5abc}$ – Ribozyme mutant of *Tetrahymena* group I intron which lacks the P5abc domain
EDTA – ethylenediaminetetraacetic acid
eIF – Eukaryotic initiation factor
EJC – Exon junction complex
G – Guanosine
IPTG – Isopropyl β-D-1-thiogalactopyranoside
LB media - Luria-Bertani media (1.0% tryptone, 0.5% yeast extract, and 0.5% NaCl)
L-21*Sca*I – Wild type ribozyme, containing 413 nucleotides and lacking the first 21 nucleotides, derived from self-splicing *Tetrahymena* group I intron
MOPS – 3-(N-morpholino) propanesulfonic acid
mtLSU – Mitochondrial Large subunit
mRNA – Messenger RNA
NPH – II – Nucleotide phosphohydrolase II
O.D. – Optical density
P5abc – P stands for paired region; 5abc stand for subdomains a, b and c in P5 element of group I ribozyme
PAGE – Polyacrylamide gel electrophoresis
PEI – Polyethyleneimine
Prp – Precursor RNA processing
RNP – Ribonucleoprotein
rpm – Revolutions per minute
rRNA – Ribosomal RNA
***Sca*I** - Restriction enzyme
snoRNA – small nucleolar RNA
snRNA – small nuclear RNA
tRNA – transfer RNA
TBE buffer – 100 mM Tris, 83 mM boric acid and 1 mM EDTA
TE buffer – 10 mM Tris and 1 mM EDTA
TRIS – Trishydroxymethylaminomethane
Triton-X 100 – Surfactant
WT – Wild type

References:

1. Noller, H. F. RNA structure: reading the ribosome. *Science* **309**, 1508-14 (2005).
2. Herschlag, D. & Cech, T. R. Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site. *Biochemistry* **29**, 10159-71 (1990).
3. Nakano, S., Chadalavada, D. M. & Bevilacqua, P. C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493-7 (2000).
4. Perrotta, A. T., Shih, I. & Been, M. D. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, 123-6 (1999).
5. Treiber, D. K. & Williamson, J. R. Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* **11**, 309-14 (2001).
6. Uhlenbeck, O. C. Keeping RNA happy. *RNA* **1**, 4-6 (1995).
7. Russell, R. RNA misfolding and the action of chaperones. *Front Biosci* **13**, 1-20 (2008).
8. Herschlag, D. RNA chaperones and the RNA folding problem. *J Biol Chem* **270**, 20871-4 (1995).
9. Woodson, S. A. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci* **57**, 796-808 (2000).
10. Rajkowitsch, L., Semrad, K., Mayer, O. & Schroeder, R. Assays for the RNA chaperone activity of proteins. *Biochem Soc Trans* **33**, 450-6 (2005).
11. Tanner, N. K. & Linder, P. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* **8**, 251-62 (2001).
12. Jankowsky, E. & Bowers, H. Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res* **34**, 4181-8 (2006).
13. Silverman, E., Edwalds-Gilbert, G. & Lin, R. J. DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene* **312**, 1-16 (2003).
14. Sigler, P. B. An analysis of the structure of tRNA. *Annu Rev Biophys Bioeng* **4**, 477-527 (1975).
15. Levinthal, C. Are there pathways for protein folding? *Extrait du journal de chimie physique* **65**, 44-45 (1968).
16. Russell, R., Millett, I. S., Doniach, S. & Herschlag, D. Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* **7**, 367-70 (2000).
17. Das, R. et al. The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the Tetrahymena ribozyme. *J Mol Biol* **332**, 311-9 (2003).
18. Swisher, J. F., Su, L. J., Brenowitz, M., Anderson, V. E. & Pyle, A. M. Productive folding to the native state by a group II intron ribozyme. *J Mol Biol* **315**, 297-310 (2002).
19. Fang, X. W., Thiyagarajan, P., Sosnick, T. R. & Pan, T. The rate-limiting step in the folding of a large ribozyme without kinetic traps. *Proc Natl Acad Sci U S A* **99**, 8518-23 (2002).

20. Fang, X. W., Pan, T. & Sosnick, T. R. Mg2+-dependent folding of a large ribozyme without kinetic traps. *Nat Struct Biol* **6**, 1091-5 (1999).
21. Buchmueller, K. L., Webb, A. E., Richardson, D. A. & Weeks, K. M. A collapsed non-native RNA folding state. *Nat Struct Biol* **7**, 362-6 (2000).
22. Wu, M. & Tinoco, I., Jr. RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci U S A* **95**, 11555-60 (1998).
23. Treiber, D. K. & Williamson, J. R. Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* **9**, 339-45 (1999).
24. Fresco, J. R., Adams, A., Ascione, R., Henley, D. & Lindahl, T. Tertiary structure in transfer ribonucleic acids. *Cold Spring Harb Symp Quant Biol* **31**, 527-37 (1966).
25. Gartland, W. J. & Sueoka, N. Two interconvertible forms of tryptophanyl sRNA in E. coli. *Proc Natl Acad Sci U S A* **55**, 948-56 (1966).
26. Cole, P. E. & Crothers, D. M. Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (Escherichia coli). *Biochemistry* **11**, 4368-74 (1972).
27. Cole, P. E., Yang, S. K. & Crothers, D. M. Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry* **11**, 4358-68 (1972).
28. Aubert, M., Scott, J. F., Reynier, M. & Monier, R. Rearrangement of the conformation of Escherichia coli 5S RNA. *Proc Natl Acad Sci U S A* **61**, 292-9 (1968).
29. Weidner, H. & Crothers, D. M. Pathway-dependent refolding of E. coli 5S RNA. *Nucleic Acids Res* **4**, 3401-14 (1977).
30. Weidner, H., Yuan, R. & Crothers, D. M. Does 5S RNA function by a switch between two secondary structures? *Nature* **266**, 193-4 (1977).
31. Breaker, R. R. Complex riboswitches. *Science* **319**, 1795-7 (2008).
32. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8**, R239 (2007).
33. Barrick, J. E. & Breaker, R. R. The power of riboswitches. *Sci Am* **296**, 50-7 (2007).
34. Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**, 487-517 (2005).
35. Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751-8 (1981).
36. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**, 701-7 (2003).
37. Altman, S. & Guerrier-Takada, C. M1 RNA, the RNA subunit of Escherichia coli ribonuclease P, can undergo a pH-sensitive conformational change. *Biochemistry* **25**, 1205-8 (1986).
38. Fedor, M. J. & Uhlenbeck, O. C. Substrate sequence effects on "hammerhead" RNA catalytic efficiency. *Proc Natl Acad Sci U S A* **87**, 1668-72 (1990).

39. Walstrum, S. A. & Uhlenbeck, O. C. The self-splicing RNA of Tetrahymena is trapped in a less active conformation by gel purification. *Biochemistry* **29**, 10573-6 (1990).
40. Esteban, J. A., Walter, N. G., Kotzorek, G., Heckman, J. E. & Burke, J. M. Structural basis for heterogeneous kinetics: reengineering the hairpin ribozyme. *Proc Natl Acad Sci U S A* **95**, 6091-6 (1998).
41. Esteban, J. A., Banerjee, A. R. & Burke, J. M. Kinetic mechanism of the hairpin ribozyme. Identification and characterization of two nonexchangeable conformations. *J Biol Chem* **272**, 13629-39 (1997).
42. Zhuang, X. et al. Correlating structural dynamics and function in single ribozyme molecules. *Science* **296**, 1473-6 (2002).
43. Held, W. A. & Nomura, M. Rate determining step in the reconstitution of Escherichia coli 30S ribosomal subunits. *Biochemistry* **12**, 3273-81 (1973).
44. Hogan, J. J. & Noller, H. F. Altered topography of 16S RNA in the inactive form of Escherichia coli 30S ribosomal subunits. *Biochemistry* **17**, 587-93 (1978).
45. Talkington, M. W., Siuzdak, G. & Williamson, J. R. An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628-32 (2005).
46. Gluick, T. C., Gerstner, R. B. & Draper, D. E. Effects of Mg2+, K+, and H+ on an equilibrium between alternative conformations of an RNA pseudoknot. *J Mol Biol* **270**, 451-63 (1997).
47. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35**, 161-75 (2006).
48. Donahue, C. P. & Fedor, M. J. Kinetics of hairpin ribozyme cleavage in yeast. *RNA* **3**, 961-73 (1997).
49. Yadava, R. S., Choi, A. J., Lebruska, L. L. & Fedor, M. J. Hairpin ribozymes with four-way helical junctions mediate intracellular RNA ligation. *J Mol Biol* **309**, 893-902 (2001).
50. Donahue, C. P., Yadava, R. S., Nesbitt, S. M. & Fedor, M. J. The kinetic mechanism of the hairpin ribozyme in vivo: influence of RNA helix stability on intracellular cleavage kinetics. *J Mol Biol* **295**, 693-707 (2000).
51. Brehm, S. L. & Cech, T. R. Fate of an intervening sequence ribonucleic acid: excision and cyclization of the Tetrahymena ribosomal ribonucleic acid intervening sequence in vivo. *Biochemistry* **22**, 2390-7 (1983).
52. Bass, B. L. & Cech, T. R. Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: implications for biological catalysis by RNA. *Nature* **308**, 820-6 (1984).
53. Nikolcheva, T. & Woodson, S. A. Facilitation of group I splicing in vivo: misfolding of the Tetrahymena IVS and the role of ribosomal RNA exons. *J Mol Biol* **292**, 557-67 (1999).
54. Koduvayur, S. P. & Woodson, S. A. Intracellular folding of the Tetrahymena group I intron depends on exon sequence and promoter choice. *RNA* **10**, 1526-32 (2004).
55. Antal, M., Boros, E., Solymosy, F. & Kiss, T. Analysis of the structure of human telomerase RNA in vivo. *Nucleic Acids Res* **30**, 912-20 (2002).

56. Pan, T., Artsimovitch, I., Fang, X. W., Landick, R. & Sosnick, T. R. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci U S A* **96**, 9545-50 (1999).
57. Wong, T., Sosnick, T. R. & Pan, T. Mechanistic insights on the folding of a large ribozyme during transcription. *Biochemistry* **44**, 7535-42 (2005).
58. Mahen, E. M., Harger, J. W., Calderon, E. M. & Fedor, M. J. Kinetics and thermodynamics make different contributions to RNA folding in vitro and in yeast. *Mol Cell* **19**, 27-37 (2005).
59. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* **5**, 908-19 (2004).
60. Dennis, P. P., Russell, A. G. & Moniz De Sa, M. Formation of the 5' end pseudoknot in small subunit ribosomal RNA: involvement of U3-like sequences. *RNA* **3**, 337-43 (1997).
61. Steitz, J. A. & Tycowski, K. T. Small RNA chaperones for ribosome biogenesis. *Science* **270**, 1626-7 (1995).
62. Hughes, J. M. Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J Mol Biol* **259**, 645-54 (1996).
63. Linder, P. Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res* **34**, 4168-80 (2006).
64. Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. The DEAD-box protein family of RNA helicases. *Gene* **367**, 17-37 (2006).
65. Rocak, S. & Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* **5**, 232-41 (2004).
66. Hall, M. C. & Matson, S. W. Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* **34**, 867-77 (1999).
67. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17**, 4713-30 (1989).
68. Linder, P. et al. Birth of the D-E-A-D box. *Nature* **337**, 121-2 (1989).
69. Linder, P. & Daugeron, M. C. Are DEAD-box proteins becoming respectable helicases? *Nat Struct Biol* **7**, 97-9 (2000).
70. Delagoutte, E. & von Hippel, P. H. Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: Structures and properties of isolated helicases. *Q Rev Biophys* **35**, 431-78 (2002).
71. Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annu Rev Biochem* **69**, 651-97 (2000).
72. Yang, Q. & Jankowsky, E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* **13**, 981-6 (2006).
73. Tijerina, P., Bhaskaran, H. & Russell, R. 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**, 16698-703 (2006).

74. Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci* **25**, 156-65 (2000).
75. Pause, A. & Sonenberg, N. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J* **11**, 2643-54 (1992).
76. Rozen, F., Pelletier, J., Trachsel, H. & Sonenberg, N. A lysine substitution in the ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. *Mol Cell Biol* **9**, 4061-3 (1989).
77. Blum, S. et al. ATP hydrolysis by initiation factor 4A is required for translation initiation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **89**, 7664-8 (1992).
78. Story, R. M., Li, H. & Abelson, J. N. Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proc Natl Acad Sci U S A* **98**, 1465-70 (2001).
79. Rogers, G. W., Jr., Komar, A. A. & Merrick, W. C. eIF4A: the godfather of the DEAD box helicases. *Prog Nucleic Acid Res Mol Biol* **72**, 307-31 (2002).
80. Iost, I., Dreyfus, M. & Linder, P. Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. *J Biol Chem* **274**, 17677-83 (1999).
81. Yao, N. et al. Structure of the hepatitis C virus RNA helicase domain. *Nat Struct Biol* **4**, 463-7 (1997).
82. Schwer, B. & Meszaros, T. RNA helicase dynamics in pre-mRNA splicing. *EMBO J* **19**, 6582-91 (2000).
83. Cheng, Z., Coller, J., Parker, R. & Song, H. Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA* **11**, 1258-70 (2005).
84. Pause, A., Methot, N. & Sonenberg, N. 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**, 6789-98 (1993).
85. Tanner, N. K., Cordin, O., Banroques, J., Doere, M. & Linder, P. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* **11**, 127-38 (2003).
86. Cordin, O., Tanner, N. K., Doere, M., Linder, P. & Banroques, J. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* **23**, 2478-87 (2004).
87. Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. Crystal structure of a DExx box DNA helicase. *Nature* **384**, 379-83 (1996).
88. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**, 635-47 (1997).
89. Moore, K. J. & Lohman, T. M. Helicase-catalyzed DNA unwinding: energy coupling by DNA motor proteins. *Biophys J* **68**, 180S-184S; discussion 184S-185S (1995).
90. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75-84 (1999).

91. Kim, J. L. et al. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89-100 (1998).
92. Benz, J., Trachsel, H. & Baumann, U. Crystal structure of the ATPase domain of translation initiation factor 4A from Saccharomyces cerevisiae--the prototype of the DEAD box protein family. *Structure* **7**, 671-9 (1999).
93. Carmel, A. B. & Matthews, B. W. Crystal structure of the BstDEAD N-terminal domain: a novel DEAD protein from Bacillus stearothermophilus. *RNA* **10**, 66-74 (2004).
94. Caruthers, J. M., Johnson, E. R. & McKay, D. B. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A* **97**, 13080-5 (2000).
95. Shi, H., Cordin, O., Minder, C. M., Linder, P. & Xu, R. M. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc Natl Acad Sci U S A* **101**, 17628-33 (2004).
96. Caruthers, J. M. & McKay, D. B. Helicase structure and mechanism. *Curr Opin Struct Biol* **12**, 123-33 (2002).
97. Ye, J., Osborne, A. R., Groll, M. & Rapoport, T. A. RecA-like motor ATPases--lessons from structures. *Biochim Biophys Acta* **1659**, 1-18 (2004).
98. Linder, P. & Lasko, P. Bent out of shape: RNA unwinding by the DEAD-box helicase Vasa. *Cell* **125**, 219-21 (2006).
99. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* **125**, 287-300 (2006).
100. Bleichert, F. & Baserga, S. J. The long unwinding road of RNA helicases. *Mol Cell* **27**, 339-52 (2007).
101. De la Cruz, J., Kressler, D. & Linder, P. Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. *Trends Biochem Sci* **24**, 192-8 (1999).
102. Jankowsky, E., Gross, C. H., Shuman, S. & Pyle, A. M. The DExH protein NPH-II is a processive and directional motor for unwinding RNA. *Nature* **403**, 447-51 (2000).
103. Karpel, R. L. et al. Acceleration of RNA renaturation by nucleic acid unwinding proteins. *Brookhaven Symp Biol*, 165-74 (1975).
104. Pang, P. S., Jankowsky, E., Planet, P. J. & Pyle, A. M. The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *Embo J* **21**, 1168-76 (2002).
105. Ray, B. K. et al. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem* **260**, 7651-8 (1985).
106. Yang, Q. & Jankowsky, E. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**, 13591-601 (2005).

107. Mohr, S., Stryker, J. M. & Lambowitz, A. M. A DEAD-box protein functions as an ATP-dependent RNA chaperone in group I intron splicing. *Cell* **109**, 769-79 (2002).
108. Grifo, J. A., Abramson, R. D., Satler, C. A. & Merrick, W. C. RNA-stimulated ATPase activity of eukaryotic initiation factors. *J Biol Chem* **259**, 8648-54 (1984).
109. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* **37**, 2180-93 (1998).
110. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry* **37**, 2194-206 (1998).
111. Fuller-Pace, F. V., Nicol, S. M., Reid, A. D. & Lane, D. P. DbpA: a DEAD box protein specifically activated by 23s rRNA. *Embo J* **12**, 3619-26 (1993).
112. Nicol, S. M. & Fuller-Pace, F. V. The "DEAD box" protein DbpA interacts specifically with the peptidyltransferase center in 23S rRNA. *Proc Natl Acad Sci U S A* **92**, 11681-5 (1995).
113. Pugh, G. E., Nicol, S. M. & Fuller-Pace, F. V. Interaction of the Escherichia coli DEAD box protein DbpA with 23 S ribosomal RNA. *J Mol Biol* **292**, 771-8 (1999).
114. Tsu, C. A., Kossen, K. & Uhlenbeck, O. C. The Escherichia coli DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA. *RNA* **7**, 702-9 (2001).
115. Korneeva, N. L., First, E. A., Benoit, C. A. & Rhoads, R. E. Interaction between the NH2-terminal domain of eIF4A and the central domain of eIF4G modulates RNA-stimulated ATPase activity. *J Biol Chem* **280**, 1872-81 (2005).
116. Oberer, M., Marintchev, A. & Wagner, G. Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev* **19**, 2212-23 (2005).
117. Linder, P. Yeast RNA helicases of the DEAD-box family involved in translation initiation. *Biol Cell* **95**, 157-67 (2003).
118. Kossen, K., Karginov, F. V. & Uhlenbeck, O. C. The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA. *J Mol Biol* **324**, 625-36 (2002).
119. Wang, S. et al. 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**, 959-67 (2006).
120. Umen, J. G. & Guthrie, C. The second catalytic step of pre-mRNA splicing. *RNA* **1**, 869-85 (1995).
121. O'Day, C. L., Dalbadie-McFarland, G. & Abelson, J. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem* **271**, 33261-7 (1996).
122. Rajagopal, V. & Patel, S. S. Single strand binding proteins increase the processivity of DNA unwinding by the hepatitis C virus helicase. *J Mol Biol* **376**, 69-79 (2008).

123. Tai, C. L., Chi, W. K., Chen, D. S. & Hwang, L. H. The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J Virol* **70**, 8477-84 (1996).
124. Fuller-Pace, F. V. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res* **34**, 4206-15 (2006).
125. Nakajima, T. et al. RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* **90**, 1107-12 (1997).
126. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**, 5-14 (2003).
127. Stevens, S. W. et al. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* **9**, 31-44 (2002).
128. Nilsen, T. W. The spliceosome: no assembly required? *Mol Cell* **9**, 8-9 (2002).
129. Azubel, M., Wolf, S. G., Sperling, J. & Sperling, R. Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol Cell* **15**, 833-9 (2004).
130. Malca, H., Shomron, N. & Ast, G. The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol Cell Biol* **23**, 3442-55 (2003).
131. Brow, D. A. Allosteric cascade of spliceosome activation. *Annu Rev Genet* **36**, 333-60 (2002).
132. Staley, J. P. & Guthrie, C. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* **3**, 55-64 (1999).
133. Kim, S. H. & Lin, R. J. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol* **16**, 6810-9 (1996).
134. Schwer, B. & Guthrie, C. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *Embo J* **11**, 5033-9 (1992).
135. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494-9 (1991).
136. Wagner, J. D., Jankowsky, E., Company, M., Pyle, A. M. & Abelson, J. N. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *Embo J* **17**, 2926-37 (1998).
137. Company, M., Arenas, J. & Abelson, J. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349**, 487-93 (1991).
138. Martin, A., Schneider, S. & Schwer, B. Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J Biol Chem* **277**, 17743-50 (2002).
139. Arenas, J. E. & Abelson, J. N. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A* **94**, 11798-802 (1997).
140. Linder, P. & Stutz, F. mRNA export: travelling with DEAD box proteins. *Curr Biol* **11**, R961-3 (2001).

141. Zhang, M. & Green, M. R. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev* **15**, 30-5 (2001).
142. Gatfield, D. et al. The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. *Curr Biol* **11**, 1716-21 (2001).
143. Luo, M. L. et al. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644-7 (2001).
144. Strasser, K. & Hurt, E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**, 648-52 (2001).
145. Jensen, T. H., Boulay, J., Rosbash, M. & Libri, D. The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr Biol* **11**, 1711-5 (2001).
146. Tseng, S. S. et al. Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *Embo J* **17**, 2651-62 (1998).
147. Schmitt, C. et al. Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *Embo J* **18**, 4332-47 (1999).
148. Williamson, J. R. After the ribosome structures: how are the subunits assembled? *RNA* **9**, 165-7 (2003).
149. Kressler, D., Linder, P. & de La Cruz, J. Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. *Mol Cell Biol* **19**, 7897-912 (1999).
150. Luking, A., Stahl, U. & Schmidt, U. The protein family of RNA helicases. *Crit Rev Biochem Mol Biol* **33**, 259-96 (1998).
151. Charollais, J., Pflieger, D., Vinh, J., Dreyfus, M. & Iost, I. The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in Escherichia coli. *Mol Microbiol* **48**, 1253-65 (2003).
152. Rogers, G. W., Jr., Richter, N. J. & Merrick, W. C. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J Biol Chem* **274**, 12236-44 (1999).
153. Rogers, G. W., Jr., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J Biol Chem* **276**, 30914-22 (2001).
154. Chuang, R. Y., Weaver, P. L., Liu, Z. & Chang, T. H. Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science* **275**, 1468-71 (1997).
155. Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *Embo J* **22**, 1370-80 (2003).
156. Py, B., Higgins, C. F., Krisch, H. M. & Carpousis, A. J. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **381**, 169-72 (1996).
157. Liou, G. G., Chang, H. Y., Lin, C. S. & Lin-Chao, S. DEAD box RhlB RNA helicase physically associates with exoribonuclease PNPase to degrade double-stranded RNA independent of the degradosome-assembling region of RNase E. *J Biol Chem* **277**, 41157-62 (2002).

158. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713-25 (2006).
159. Huang, H. R. et al. 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**, 163-8 (2005).
160. Schmidt, U., Lehmann, K. & Stahl, U. A novel mitochondrial DEAD box protein (Mrh4) required for maintenance of mtDNA in Saccharomyces cerevisiae. *FEMS Yeast Res* **2**, 267-76 (2002).
161. Grohman, J. K. et al. 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**, 3013-22 (2007).
162. Leeds, N. B., Small, E. C., Hiley, S. L., Hughes, T. R. & Staley, J. P. The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis. *Mol Cell Biol* **26**, 513-22 (2006).
163. Lebaron, S. et al. The splicing ATPase prp43p is a component of multiple preribosomal particles. *Mol Cell Biol* **25**, 9269-82 (2005).
164. Combs, D. J., Nagel, R. J., Ares, M., Jr. & Stevens, S. W. Prp43p is a DEAH-box spliceosome disassembly factor essential for ribosome biogenesis. *Mol Cell Biol* **26**, 523-34 (2006).
165. Fairman, M. E. et al. Protein displacement by DExH/D "RNA helicases" without duplex unwinding. *Science* **304**, 730-4 (2004).
166. Rogers, G. W., Jr., Lima, W. F. & Merrick, W. C. Further characterization of the helicase activity of eIF4A. Substrate specificity. *J Biol Chem* **276**, 12598-608 (2001).
167. Jankowsky, E. & Fairman, M. E. RNA helicases--one fold for many functions. *Curr Opin Struct Biol* **17**, 316-24 (2007).
168. Rossler, O. G., Straka, A. & Stahl, H. Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res* **29**, 2088-96 (2001).
169. Jankowsky, E., Gross, C. H., Shuman, S. & Pyle, A. M. Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. *Science* **291**, 121-5 (2001).
170. Schwer, B. A new twist on RNA helicases: DExH/D box proteins as RNPases. *Nat Struct Biol* **8**, 113-6 (2001).
171. Von Hippel, P. H. Helicases become mechanistically simpler and functionally more complex. *Nat Struct Mol Biol* **11**, 494-6 (2004).
172. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321 (1993).
173. Tsuchihashi, Z., Khosla, M. & Herschlag, D. Protein enhancement of hammerhead ribozyme catalysis. *Science* **262**, 99-102 (1993).
174. Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R. L. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *Embo J* **13**, 2913-24 (1994).

175. Karpel, R. L., Miller, N. S. & Fresco, J. R. Mechanistic studies of ribonucleic acid renaturation by a helix-destabilizing protein. *Biochemistry* **21**, 2102-8 (1982).
176. Mayer, O., Waldsich, C., Grossberger, R. & Schroeder, R. Folding of the td pre-RNA with the help of the RNA chaperone StpA. *Biochem Soc Trans* **30**, 1175-80 (2002).
177. Waldsich, C., Grossberger, R. & Schroeder, R. RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo. *Genes Dev* **16**, 2300-12 (2002).
178. Zhang, A., Derbyshire, V., Salvo, J. L. & Belfort, M. Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *RNA* **1**, 783-93 (1995).
179. Mikulecky, P. J. et al. Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol* **11**, 1206-14 (2004).
180. Coetzee, T., Herschlag, D. & Belfort, M. Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev* **8**, 1575-88 (1994).
181. Mohr, S., Matsuura, M., Perlman, P. S. & Lambowitz, A. M. 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**, 3569-74 (2006).
182. Bhaskaran, H. & Russell, R. Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* **449**, 1014-8 (2007).
183. Seraphin, B., Simon, M., Boulet, A. & Faye, G. Mitochondrial splicing requires a protein from a novel helicase family. *Nature* **337**, 84-7 (1989).
184. Niemer, I., Schmelzer, C. & Borner, G. V. Overexpression of DEAD box protein pMSS116 promotes ATP-dependent splicing of a yeast group II intron in vitro. *Nucleic Acids Res* **23**, 2966-72 (1995).
185. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-57 (1982).
186. Zaug, A. J., Been, M. D. & Cech, T. R. The Tetrahymena ribozyme acts like an RNA restriction endonuclease. *Nature* **324**, 429-33 (1986).
187. Zaug, A. J. & Cech, T. R. The intervening sequence RNA of Tetrahymena is an enzyme. *Science* **231**, 470-5 (1986).
188. Cate, J. H. et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-85 (1996).
189. Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* **282**, 259-64 (1998).
190. Tanner, M. A. & Cech, T. R. Joining the two domains of a group I ribozyme to form the catalytic core. *Science* **275**, 847-9 (1997).
191. Guo, F., Gooding, A. R. & Cech, T. R. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* **16**, 351-62 (2004).
192. Narlikar, G. J., Khosla, M., Usman, N. & Herschlag, D. Quantitating tertiary binding energies of 2' OH groups on the P1 duplex of the Tetrahymena ribozyme: intrinsic binding energy in an RNA enzyme. *Biochemistry* **36**, 2465-77 (1997).

193. Shan, S., Yoshida, A., Sun, S., Piccirilli, J. A. & Herschlag, D. Three metal ions at the active site of the Tetrahymena group I ribozyme. *Proc Natl Acad Sci U S A* **96**, 12299-304 (1999).
194. Shan, S., Kravchuk, A. V., Piccirilli, J. A. & Herschlag, D. Defining the catalytic metal ion interactions in the Tetrahymena ribozyme reaction. *Biochemistry* **40**, 5161-71 (2001).
195. Stahley, M. R. & Strobel, S. A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587-90 (2005).
196. Cech, T. R., Herschlag, D., Piccirilli, J. A. & Pyle, A. M. RNA catalysis by a group I ribozyme. Developing a model for transition state stabilization. *J Biol Chem* **267**, 17479-82 (1992).
197. Batey, R. T. & Doudna, J. A. The parallel universe of RNA folding. *Nat Struct Biol* **5**, 337-40 (1998).
198. Zarrinkar, P. P. & Williamson, J. R. Kinetic intermediates in RNA folding. *Science* **265**, 918-24 (1994).
199. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. & Woodson, S. A. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* **279**, 1940-3 (1998).
200. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. Kinetic intermediates trapped by native interactions in RNA folding. *Science* **279**, 1943-6 (1998).
201. Russell, R. et al. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci U S A* **99**, 155-60 (2002).
202. Rook, M. S., Treiber, D. K. & Williamson, J. R. Fast folding mutants of the Tetrahymena group I ribozyme reveal a rugged folding energy landscape. *J Mol Biol* **281**, 609-20 (1998).
203. Russell, R. & Herschlag, D. New pathways in folding of the Tetrahymena group I RNA enzyme. *J Mol Biol* **291**, 1155-67 (1999).
204. Russell, R. & Herschlag, D. Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway. *J Mol Biol* **308**, 839-51 (2001).
205. Pan, J. & Woodson, S. A. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J Mol Biol* **280**, 597-609 (1998).
206. Russell, R. et al. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* **363**, 531-44 (2006).
207. Nagai, K. et al. Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture. *Biochem Soc Trans* **29**, 15-26 (2001).
208. Wong, I., Chao, K. L., Bujalowski, W. & Lohman, T. M. DNA-induced dimerization of the Escherichia coli rep helicase. Allosteric effects of single-stranded and duplex DNA. *J Biol Chem* **267**, 7596-610 (1992).
209. Yang, Q., Fairman, M. E. & Jankowsky, E. DEAD-box-protein-assisted RNA structure conversion towards and against thermodynamic equilibrium values. *J Mol Biol* **368**, 1087-100 (2007).

210. Batey, R. T., Rambo, R. P. & Doudna, J. A. Tertiary Motifs in RNA Structure and Folding. *Angew Chem Int Ed Engl* **38**, 2326-2343 (1999).
211. Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys* **38**, 221-43 (2005).
212. Moore, P. B. Structural motifs in RNA. *Annu Rev Biochem* **68**, 287-300 (1999).
213. Tanner, N. K. Ribozymes: the characteristics and properties of catalytic RNAs. *FEMS Microbiol Rev* **23**, 257-75 (1999).
214. Caprara, M. G., Mohr, G. & Lambowitz, A. M. A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J Mol Biol* **257**, 512-31 (1996).
215. Guo, Q. & Lambowitz, A. M. A tyrosyl-tRNA synthetase binds specifically to the group I intron catalytic core. *Genes Dev* **6**, 1357-72 (1992).
216. Mohr, G., Caprara, M. G., Guo, Q. & Lambowitz, A. M. A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the Tetrahymena ribozyme. *Nature* **370**, 147-50 (1994).
217. Paukstelis, P. J., Chen, J. H., Chase, E., Lambowitz, A. M. & Golden, B. L. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* **451**, 94-7 (2008).
218. Wallweber, G. J., Mohr, S., Rennard, R., Caprara, M. G. & Lambowitz, A. M. Characterization of Neurospora mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3' splice site for an intron ORF. *RNA* **3**, 114-31 (1997).
219. Lambowitz, A. M., Caprara, M.G., Zimmerly, S., Perlman, P.S. in *The RNA world* (ed. Gesteland, R. F., Cech, T.R., Atkins, J.F.) 451-485 (CSHL press, 1999).
220. Karbstein, K. & Herschlag, D. Extraordinarily slow binding of guanosine to the Tetrahymena group I ribozyme: implications for RNA preorganization and function. *Proc Natl Acad Sci U S A* **100**, 2300-5 (2003).
221. McConnell, T. S., Cech, T. R. & Herschlag, D. Guanosine binding to the Tetrahymena ribozyme: thermodynamic coupling with oligonucleotide binding. *Proc Natl Acad Sci U S A* **90**, 8362-6 (1993).
222. Guo, Q. B., Akins, R. A., Garriga, G. & Lambowitz, A. M. Structural analysis of the Neurospora mitochondrial large rRNA intron and construction of a mini-intron that shows protein-dependent splicing. *J Biol Chem* **266**, 1809-19 (1991).
223. Gorbalenya, A. E. & Koonin, E. V. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* **3**, 419-429 (1993).
224. Shuman, S. Vaccinia virus RNA helicase: an essential enzyme related to the DE-H family of RNA-dependent NTPases. *Proc Natl Acad Sci U S A* **89**, 10935-9 (1992).
225. Latham, J. A. & Cech, T. R. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276-82 (1989).
226. Pan, J., Deras, M. L. & Woodson, S. A. Fast folding of a ribozyme by stabilizing core interactions: evidence for multiple folding pathways in RNA. *J Mol Biol* **296**, 133-44 (2000).

227. Treiber, D. K. & Williamson, J. R. Concerted kinetic folding of a multidomain ribozyme with a disrupted loop-receptor interaction. *J Mol Biol* **305**, 11-21 (2001).
228. Del Campo, M. et al. Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol Cell* **28**, 159-166 (2007).
229. Johnson, T. H., Tijerina, P., Chadee, A. B., Herschlag, D. & Russell, R. Structural specificity conferred by a group I RNA peripheral element. *Proc Natl Acad Sci U S A* **102**, 10176-81 (2005).
230. Russell, R. et al. The Paradoxical Behavior of a Highly Structured Misfolded Intermediate in RNA Folding. *J Mol Biol* (2006).
231. Battle, D. J. & Doudna, J. A. Specificity of RNA-RNA helix recognition. *Proc Natl Acad Sci U S A* **99**, 11676-81 (2002).
232. Joyce, G. F., van der Horst, G. & Inoue, T. Catalytic activity is retained in the Tetrahymena group I intron despite removal of the large extension of element P5. *Nucleic Acids Res* **17**, 7879-89 (1989).
233. Lambowitz, A. M., caprara, M. G., Zimmerly, S. & Perlman, P. S. in *The RNA world* (eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 451-485 (Cold Spring Harbor Laboratory Press, New York, 1999).
234. Russell, R., Tijerina, P., Chadee, A. B. & Bhaskaran, H. Deletion of the P5abc peripheral element accelerates early and late folding steps of the Tetrahymena group I ribozyme. *Biochemistry* **46**, 4951-61 (2007).
235. Thirumalai, D. & Hyeon, C. RNA and protein folding: common themes and variations. *Biochemistry* **44**, 4957-70 (2005).
236. Grossberger, R. et al. Influence of RNA structural stability on the RNA chaperone activity of the Escherichia coli protein StpA. *Nucleic Acids Res* **33**, 2280-9 (2005).
237. Lin, Z. & Rye, H. S. GroEL-mediated protein folding: making the impossible, possible. *Crit Rev Biochem Mol Biol* **41**, 211-39 (2006).
238. Stein, A. J., Fuchs, G., Fu, C., Wolin, S. L. & Reinisch, K. M. Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. *Cell* **121**, 529-39 (2005).
239. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-26 (1998).
240. Rajewsky, N. microRNA target predictions in animals. *Nat Genet* **38 Suppl**, S8-13 (2006).
241. Schultes, E. A. & Bartel, D. P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448-52 (2000).
242. Halls, C. et al. 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**, 835-55 (2007).
243. Solem, A., Zingler, N. & Pyle, A. M. A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol Cell* **24**, 611-7 (2006).

244. Huang, H.-R. & University of Texas Southwestern Medical Center at Dallas. Graduate School of Biomedical Sciences. xix, 214 leaves (2004).
245. Zhuang, X. et al. A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048-51 (2000).
246. Draper, D. E. Strategies for RNA folding. *Trends Biochem Sci* **21**, 145-9 (1996).
247. Ferre-D'Amare, A. R. & Doudna, J. A. RNA folds: insights from recent crystal structures. *Annu Rev Biophys Biomol Struct* **28**, 57-73 (1999).
248. Engelhardt, M. A., Doherty, E. A., Knitt, D. S., Doudna, J. A. & Herschlag, D. The P5abc peripheral element facilitates preorganization of the tetrahymena group I ribozyme for catalysis. *Biochemistry* **39**, 2639-51 (2000).
249. Szewczak, A. A., Podell, E. R., Bevilacqua, P. C. & Cech, T. R. Thermodynamic stability of the P4-P6 domain RNA tertiary structure measured by temperature gradient gel electrophoresis. *Biochemistry* **37**, 11162-70 (1998).
250. Cate, J. H., Hanna, R. L. & Doudna, J. A. A magnesium ion core at the heart of a ribozyme domain. *Nat Struct Biol* **4**, 553-8 (1997).
251. Ralston, C. Y., He, Q., Brenowitz, M. & Chance, M. R. Stability and cooperativity of individual tertiary contacts in RNA revealed through chemical denaturation. *Nat Struct Biol* **7**, 371-4 (2000).
252. Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335-43 (2004).
253. Levin, M. K., Wang, Y. H. & Patel, S. S. The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity. *J Biol Chem* **279**, 26005-12 (2004).
254. Celander, D. W. & Cech, T. R. Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251**, 401-7 (1991).
255. Das, R., Travers, K. J., Bai, Y. & Herschlag, D. Determining the Mg2+ stoichiometry for folding an RNA metal ion core. *J Am Chem Soc* **127**, 8272-3 (2005).
256. Clodi, E., Semrad, K. & Schroeder, R. Assaying RNA chaperone activity in vivo using a novel RNA folding trap. *EMBO J* **18**, 3776-82 (1999).
257. Egele, C. et al. HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. *J Mol Biol* **342**, 453-66 (2004).
258. Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615-21 (1994).
259. Auweter, S. D., Oberstrass, F. C. & Allain, F. H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* **34**, 4943-59 (2006).
260. Karpel, R. L., Miller, N. S., Lesk, A. M. & Fresco, J. R. Stabilization of the native tertiary stucture of yeast tRNALeu3 by cationic metal complexes. *J Mol Biol* **97**, 519-32 (1975).
261. Wilk, H. E., Angeli, G. & Schafer, K. P. In vitro reconstitution of 35S ribonucleoprotein complexes. *Biochemistry* **22**, 4592-600 (1983).

262. Thomas, J. O., Glowacka, S. K. & Szer, W. Structure of complexes between a major protein of heterogeneous nuclear ribonucleoprotein particles and polyribonucleotides. *J Mol Biol* **171**, 439-55 (1983).
263. Rein, A., Henderson, L. E. & Levin, J. G. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **23**, 297-301 (1998).
264. Williams, M. C., Gorelick, R. J. & Musier-Forsyth, K. Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc Natl Acad Sci U S A* **99**, 8614-9 (2002).
265. Urbaneja, M. A., Wu, M., Casas-Finet, J. R. & Karpel, R. L. HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. *J Mol Biol* **318**, 749-64 (2002).
266. Wu, T., Heilman-Miller, S. L. & Levin, J. G. Effects of nucleic acid local structure and magnesium ions on minus-strand transfer mediated by the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. *Nucleic Acids Res* **35**, 3974-87 (2007).
267. Cruceanu, M. et al. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **34**, 593-605 (2006).
268. Dib-Hajj, F., Khan, R. & Giedroc, D. P. Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity. *Protein Sci* **2**, 231-43 (1993).
269. Zeng, Y. et al. Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer. *Proc Natl Acad Sci U S A* **104**, 12651-6 (2007).
270. Mely, Y. et al. Binding of the HIV-1 nucleocapsid protein to the primer tRNA(3Lys), in vitro, is essentially not specific. *J Biol Chem* **270**, 1650-6 (1995).
271. Khan, R., Chang, H. O., Kaluarachchi, K. & Giedroc, D. P. Interaction of retroviral nucleocapsid proteins with transfer RNAPhe: a lead ribozyme and 1H NMR study. *Nucleic Acids Res* **24**, 3568-75 (1996).
272. Khan, R. & Giedroc, D. P. Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA. *J Biol Chem* **267**, 6689-95 (1992).
273. Liu, H. W. et al. Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys J* **89**, 3470-9 (2005).
274. Cosa, G. et al. Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. *Biophys J* **87**, 2759-67 (2004).

Vita

Hari Bhaskaran was born in Chennai, India to parents that have strong art backgrounds. He attended N.S.N Matriculation higher secondary school and obtained his B.V.Sc degree from Madras veterinary college. He then completed a masters degree in poultry microbiology at the University of Arkansas, Fayetteville followed by a doctoral degree in molecular biology at the University of Texas at Austin. Throughout his education, his parents provided him with their love and support.

Address: c/o Rick Russell, MBB2.148 A2400 University of Texas, Austin, TX 78712

This dissertation was typed by Hari Bhaskaran.