The Dissertation Committee for Cynthia Pan certifies that this is the approved version of the following dissertation:

INVESTIGATIONS INTO THE MECHANISM FOR RNA STRUCTURAL REMODELING BY DEAD-BOX HELICASE PROTEINS

Committee:

Rick Russell, Supervisor

Alan Lambowitz

Karen S. Browning

Scott W. Stevens

Christian P. Whitman

INVESTIGATIONS INTO THE MECHANISM FOR RNA STRUCTURAL REMODELING BY DEAD-BOX HELICASE PROTEINS

by

Cynthia Pan, B.A.

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

August 2014

Dedication

To my parents and sisters.

Acknowledgements

I would like to express my gratitude to all of the people who have supported me throughout my time in graduate school. First of all, I would like to thank my advisor, Rick Russell, for his advice and mentoring, and especially for his incredible patience with me. I thank Dr. Alan Lambowitz for allowing me to use materials and equipment from his laboratory. I would also like to thank both Dr. Lambowitz and Dr. Mark Del Campo for the generous gift of purified Ded1 for my studies. My gratitude to Dr. Lambowitz, Dr. Karen Browning, Dr. Scott Stevens and Dr. Christian Whitman for serving on my committee and for their comments and suggestions.

I would also like to thank my labmates, both former and current, for creating a stimulating working environment. Pilar Tijerina was always very helpful, and taught me a great deal, especially when I first entered the lab. Former members Cindy Chen, Hari Bhaskaran, Amanda Chadee, Yaqi Wan, and David Mitchell made me feel welcome and were a great source of knowledge both inside and outside of the lab. Thank you to current members Luke Ward and Brant Gracia for your energy and humor. Special thanks to former members Jeffrey Potratz and Inga Jarmoskaite for interesting discussions of scientific and not-so-scientific topics. Also, I am especially grateful to Dr. Brian Cannon for his time and patience in teaching me single molecule fluorescence microscopy. His availability and willingness to discuss all of my questions allowed me to learn so much.

Thank you also to the friends I have met in Austin. You have become my second family. Lastly, I would like to express my gratitude and love to all of my family back home who have encouraged and supported me during my long graduate career.

INVESTIGATIONS INTO THE MECHANISM FOR RNA STRUCTURAL REMODELING BY DEAD-BOX HELICASE PROTEINS

Cynthia Pan, Ph.D.

The University of Texas at Austin, 2014

Supervisor: Rick Russell

Structured RNAs and RNA-protein complexes (RNPs) are involved in many essential biological processes and the specific conformations of these RNAs are crucial to their various functions. However, *in vitro* studies have found that RNA has propensity for misfolding into inactive species that often consist of extensive secondary and tertiary interactions, which can be locally and globally stabilizing, resulting in long-lived non-native conformers. DEAD-box helicases are one class of proteins that have been found to accelerate folding and rearrangements of highly structured RNAs. While these proteins have been shown to use ATP to unwind short RNA helices, it is not known how they disrupt the tertiary interactions that often stabilize both native and misfolded RNA conformations.

We used single molecule fluorescence to probe the mechanism by which DEAD-box proteins facilitate global unfolding of a structured RNA. DEAD-box protein CYT-19, a mitochondrial protein from *Neurospora crassa*, was found to destabilize a specific tertiary interaction with the *Tetrahymena* group I intron ribozyme using a helix capture mechanism. The protein molecule binds to a helix within the structured RNA only after

the helix spontaneously loses its tertiary contacts, and then uses ATP to unwind the helix, liberating the product strands. Ded1, a multi-functional DEAD-box protein found in *Saccharomyces cerevisiae*, gives analogous results with small but reproducible differences that may reflect its *in vivo* roles. The requirement for spontaneous dynamics likely targets DEAD-box proteins toward less stable RNA structures, which are likely to experience greater dynamic fluctuations, and provides a satisfying explanation for previous correlations between RNA stability and CYT-19 unfolding efficiency. Biologically, the ability to sense RNA stability probably biases DEAD-box proteins to act preferentially on misfolded structures and thereby to promote native folding while minimizing spurious interactions with stable, natively-folded RNAs. In addition, this straightforward mechanism for RNA remodeling does not require any specific structural environment of the helicase core and is likely to be relevant for DEAD-box proteins that promote RNA rearrangements of RNP complexes including the spliceosome and ribosome.

Table of Contents

List of Tables

List of Figures

Chapter 1: DEAD-box proteins and RNA remodeling*

1.1 IMPORTANCE OF RNAS *IN VIVO*

Within the central dogma of molecular biology, RNAs serve as an intermediary messenger molecule, transcribed from DNA in the nucleus of the cell to be translated into proteins in the cytoplasm. RNAs are also ubiquitous in the various pathways that make up the central dogma, including pre-mRNA and transfer RNA (tRNA) processing, protein synthesis as well as many aspects of gene regulation[1–5]. Many structural and catalytic components of the cellular machinery required for these processes consist of non-coding RNAs (ncRNAs). For example, the ribosome, the large ribonucleoprotein (RNP) complex that performs protein translation, is largely made up of structured RNAs, with RNAs performing the required catalytic reactions for protein production[1,6,7]. RNAs also play critical roles in pathways beyond the central dogma, such as intracellular protein trafficking[8,9] and the maintenance of chromosome ends[10,11]. Indeed, recent findings of pervasive transcription of prokaryotic and eukaryotic genomes have established that ncRNAs greatly exceed mRNAs[12–15], and large, complex RNAs continue to be discovered[16,17].

* Previously published as Pan C, Russell R (2010) Roles of DEAD-box proteins in RNA and RNP Folding. RNA Biol 7: 667–676. Pan and Russell co-wrote the review paper.

1.2 SECONDARY AND TERTIARY STRUCTURE OF FUNCTIONAL RNAS

Many cellular RNAs have functions that require folding into highly specific three-dimensional conformations that can vary greatly in size and structure. As mentioned above, structured RNAs are known to make up essential components of larger RNP complexes, such as the ribosome or spliceosome, that involve formation of particular secondary and/or tertiary structures or in some cases, cycling between multiple conformations in order to assemble and function. Riboswitches, a comparatively smaller and simpler type of functional RNA involved in gene regulation, also adopt different and specific conformations depending on conditions such as ligand binding[18]. Additionally, self-cleaving group I and II introns must fold correctly to highly structured, native conformations in order for excision from precursor RNAs to proceed. Since formation of the native species can easily be monitored by following the splicing reactions for these RNAs[19,20], group I and II introns, as well as ribozymes derived from these RNAs that include a 5' splice site that mimics the first step in the splicing reaction but lack the adjacent exons, have commonly been used as model systems in RNA folding studies[21,22].

1.2.1 RNA has a propensity to misfold

Short helices are pervasive in structured RNAs, so much so that they typically encompass most of the nucleotides in the RNA and constitute the basic unit of RNA secondary structure. Even helices of modest length (five or six base pairs) can form stably in the absence of higher-order, enforcing structure. This stability depends principally on local interactions, hydrogen bonding, and base stacking, such that helical

stability can be predicted by considering only the composition of each base pair and its immediate neighbors[23]. Secondary structure can form extremely fast during folding, such that RNA folding may be considered as a hierarchical process in which secondary structure forms first, and then tertiary connections are established between pre-formed secondary structure elements[24]. This mechanism generates a folding landscape in which sequential intermediates have progressively greater stability **(Figure 1.1, left).** These folding models have been generated largely from *in vitro* experiments monitoring folding of full-length RNA, but the basic features are also likely to be relevant *in vivo*, where folding of local helices can occur during transcription before longer-range tertiary partners are synthesized[25–27]. This behavior is in striking contrast to a common feature of protein folding. Although it is by no means a universal property, many small proteins and domains fold without accumulating intermediates. Because its local secondary structure and tertiary contacts tend to be unstable, protein folding does not become "downhill" until these locally unstable structural elements coalesce into a tertiary fold that is stabilized by global cooperativity[28,29] **(Figure 1.1, right).**

With only four standard RNA bases, there is potential for formation of short alternative helices in even the smallest RNA sequences. Thus, non-native secondary structures are likely to form at least transiently for most RNAs. While an interaction of one or two base pairs would be expected to dissociate spontaneously without causing problems for folding, incorrect RNA helices of as few as five or six base pairs can create substantial blocks for subsequent rearrangement to the native state. Early work demonstrated that several different tRNAs can adopt two different conformations, only

one of which is active for aminoacylation[30,31], and further work on diverse tRNAs indicated that the inactive conformations can have substantial differences in secondary structure from the native "cloverleaf" arrangement[32,33]. An alternative secondary structure was also suggested for 5S rRNA[34], and more recently, misfolded conformations with secondary structure changes have been identified for several RNAs, including group I self-splicing introns[35–39] and the hepatitis delta virus (HDV) ribozyme[27,40] (**Figure 1.2**).

In addition to forming the full set of native base pairs, structured RNAs must also form tertiary interactions that bring together the secondary elements into a precise native arrangement. Some tertiary contacts are formed by base-pairing interactions involving loops, termed pseudoknots, and these have the same thermodynamic properties as secondary structure. Other types of tertiary interactions are formed between motifs in terminal or internal loops and specific contacts within helical segments[41–44]. As with secondary structures, RNA tertiary contacts are frequently modular and can have substantial independent stability. Indeed, a tetraloop-receptor interaction has been removed from its natural context and used in extensive quantitative studies on the kinetics and energetics of tertiary-contact formation[45,46].

With stable, modular tertiary contacts, it is perhaps not surprising that RNA can misfold at the level of tertiary structure as well as secondary structure. Extensive work has been done on the folding and misfolding of group I introns, particularly one from *Tetrahymena thermophila* and its shortened "ribozyme" form **(Figure 1.3).** Upon addition of Mg^{2+} *in vitro*, this ribozyme folds through a complex landscape along pathways that lead to the native state as well as a long-lived misfolded conformation that

is similar in overall size and shape[47–56]. Although formation of a non-native helix in the ribozyme core, termed alt-P3, biases folding of the ribozyme towards the misfolded conformation[56,57], chemical footprinting and fluorescence experiments have shown that this misfolded RNA has all of the native secondary structures and all five native long-range tertiary contacts[58,59] (**arrows in Figure 1.3**). This result indicates that the non-native alt-P3 is resolved to the correct base-pairing during folding to the misfolded conformation[56,58]. Further, in spite of its predominantly native structure, extensive unfolding of the misfolded ribozyme, which includes not only unwinding of the P3 helix but also disruption of all five tertiary contacts, is required for refolding to the correct, functional state[56,58]. Thus, native tertiary contacts can stabilize misfolded conformations by "locking in" non-native structure elsewhere[60]. In addition, there is growing evidence that a topological difference within the core of the ribozyme is the source of misfolding for this ribozyme and modeling studies have proposed a specific non-native strand crossing that in order to be resolved would necessitate the type of global unfolding described above[59,58].

Other group I introns have also been shown to misfold, broadening the conclusion that RNA misfolding is common. A small group I intron from the bacterium *Azoarcus* forms a long-lived misfolded conformation that is globally compact and highly structured[39], which is consistent with observations from previous work showing that this RNA rapidly collapses into one or more intermediates that strongly resemble the native state[61–65]. Like for the *Tetrahymena* ribozyme, transient unfolding is required for this misfolded species to refold to its more thermodynamically stable native conformation[39].

A *Candida albicans* intron from the same subgroup as the *Tetrahymena* intron adopts a misfolded conformation that is linked to its P3 pseudoknot and can be avoided by pre-incubation in monovalent ions, analogous to results for the *Tetrahymena* ribozyme[26,66]. In contrast, an intron from the mitochondria of *Saccharomyces cerevisiae* misfolds to a conformation that includes an extensive non-native secondary structure and drastic changes in tertiary structure relative to the functional structure[38,67]. Therefore, while misfolding for large RNAs may be prevalent, the details of the misfolded structures can vary substantially between related RNAs.

1.2.2 Functional remodeling of structured RNAs

Beyond folding to the native state, conformational transitions involving secondary structure changes are also required for the reaction cycles of many cellular RNAs and RNP complexes. RNA secondary structure changes underlie the regulatory roles of some groups of riboswitch RNAs[18,68]. Analogously, miRNAs, siRNAs and small regulatory RNAs in bacteria rely on a functional cycle of base pair formation and dissociation from target mRNAs[5,69]. In addition, the spliceosome, the large, highly dynamic RNP complex involved in pre-mRNA splicing, undergoes a series of extensive secondary structure rearrangements involving changes in base pairing of the snRNA components of the spliceosome with each other and the pre-mRNA[2,70,71].

This requirement for RNA structural remodeling creates a fundamental challenge *in vivo* since local RNA structure can be independently stable, and thus, the energy barriers between alternative conformations, misfolded species as well as intermediates

along the RNA folding pathway, are often large enough to block interconversion on the biological timescale[72]. Consideration of these basic properties led to the prediction that RNA chaperone proteins would be necessary to interact transiently with structured RNAs and RNPs *in vivo* to accelerate transitions between alternative conformations or to resolve misfolded species[72]. Since then, RNA chaperone activities have been demonstrated for diverse sets of proteins (reviewed in refs. 22,73–75).

1.3 DEAD-BOX PROTEINS ACT AS CHAPERONES OF RNA FOLDING

Several groups of proteins have been shown to possess RNA chaperone activity using a systematic assay in which overexpression of a chaperone protein rescues self-splicing of a T4 intron that is otherwise trapped by non-native secondary structure[76,77]. Proteins that have demonstrated RNA chaperone activity for group I introns, using this and other approaches, include the HIV protein NC, the *E. coli* histone-like protein StpA, and several of the ribosomal proteins[78–80]. These proteins are small and quite positively-charged, and they are thought to function in RNA folding by binding strongly to single-stranded segments, thereby stabilizing intermediates required for exchange of alternative structures or contacts[74,75].

One critical group of proteins shown to facilitate RNA structural changes are the DEAD-box helicase proteins[39,81,82]. These proteins make up the largest subfamily within the superfamily 2 (SF2) RNA helicases, which include DEAH-box and Ski-like proteins among others, and are present in all three branches of life[83]. At least one DEAD-box or related protein is required for essentially all RNA-mediated processes[84,85]. These proteins

have diverse roles and biochemical activities, including protein displacement[86] and strand annealing[87,88] but many of them undergo a cycle of ATP-dependent conformational changes that is thought to facilitate folding and remodeling transitions of RNAs and RNPs[89]. Indeed, conformational changes in the spliceosome have proven to be fertile ground for exploring specific roles of DEAD-box proteins and related SF2 helicases[90,91]. RNA secondary structure rearrangements have been identified for the DEAD-box protein Prp28[92–94], the DEAH-box proteins Prp16[95] and Prp22[96], and the Ski2-like helicase Brr2[97]. In particular, Prp22 has been shown directly to disrupt base pairs between the newly-spliced mRNA and U5 snRNP, allowing release of the spliced product from the spliceosome[96]. Proper splicing dictates that the conformational transitions proceed in a defined order, and it has been shown that certain mutations in the helicases, typically leading to cold sensitivity, allow accumulation and isolation of relatively homogenous intermediates stalled just upstream of the conformational transition mediated by the mutated protein[98–100]. These results identify the importance of RNA helicase proteins in promoting the structural rearrangements necessary for many essential cellular processes to proceed.

1.3.1 Structural characteristics of DEAD-box proteins

DEAD-box proteins share with other superfamily 1 and 2 helicases a conserved core of two RecA-like domains connected by a flexible linker[83–85,101]. At least eleven conserved motifs lie near the interface of these two domains[101–107] **(Figure 1.4)**. ATP binds on one side of the interface, forming interactions with several motifs. The

triphosphate moiety interacts with motif I (specifically the P-loop), motif II including the eponymous D-E-A-D sequence, and motif VI. In addition, motif V is thought to contact the ribose of ATP[101] and the adenine base forms specific interactions with the Q-motif, which is diverged in most other SF2 helicases and gives rise to a strong preference in DEAD-box proteins for ATP over other nucleoside triphosphates[108,109]. Motif III lies near the γ-phosphoryl group of ATP and is suggested to assist in positioning it for hydrolysis[110]. On the opposite side of the domain interface, motifs Ia, GG and Ib from domain 1, along with motifs IV, QxxR and V from domain 2, mediate binding of a segment of single-stranded RNA (ssRNA)[101].

A conserved feature of DEAD-box proteins is that the binding sites for ssRNA and adenosine nucleotide interact energetically to give coupled binding of the two ligands. Recent structural studies have found that the isolated domain 2 of the *Saccharomyces cerevisiae* DEAD-box helicase Mss116p stably binds dsRNA while ATP primarily forms contacts with domain 1[111]. Using ATP and non-hydrolyzable analogs, Mss116p was shown to undergo a conformational change upon binding both RNA and ATP, coordinated by a network of interactions between several motifs, which "closes" the two domains of the helicase core[102]. This closed conformation gives tight binding of nucleotide and ssRNA[110,112–115] and the extensive contacts formed between the two domains are thought to result in displacement of one of the strands of bound dsRNA[102,111], resulting in local secondary structure disruption (further discussion below in **Section 1.3.2**). Upon hydrolysis of ATP, release of inorganic phosphate (P_i) is a key step[116]. Prior to release of P_i, nucleotide and ssRNA binding are positively coupled, and the RNA is

bound at least as tightly with ADP and P_i as with ATP[117]. In contrast, upon release of P_i, a conformational change occurs that results in much weaker binding of RNA in the ADP-bound state due to anti-cooperativity[112]. This cycle of regulated ssRNA binding affinity is thought to be critical for most or all of the activities of DEAD-box proteins, including RNA unwinding activity.

Many DEAD-box proteins have ancillary sequences in addition to the core helicase domains, and these regions are suggested, and in some cases shown, to form tethering interactions with RNA substrates[118–123]. DEAD-box proteins can be appended with domains that recognize specific RNA structures and direct the proteins to these structures. The archetypal protein of this group is the bacterial DbpA/YxiN protein, which has an ancillary C-terminal domain that adopts an RNA recognition motif fold (RRM) and binds specifically to a structure within the large ribosomal RNA, allowing the protein to function in large subunit assembly[120,122,124,125]. In keeping with their versatile nature, there is strong evidence that other DEAD-box proteins rely on a tethering interaction formed by the helicase core of a second monomer in a transiently-formed dimer or multimer[121,126].

Some ancillary regions of DEAD-box proteins are highly basic and predicted to be unstructured in solution, most likely contacting RNA non-specifically[118,119,123,127]. This interaction has been proposed to enhance binding of the chaperone protein to RNA substrate. Indeed, a truncation mutant of *Neurospora crassa* DEAD-box protein CYT-19, which lacks its positively charged C-terminal region (C-tail), was found to have weaker binding to RNA than the wildtype protein and consequently, have a decreased overall

efficiency for helix unwinding and refolding of structured RNAs[118]. Furthermore, using small angle X-ray scattering (SAXS), the basic C-tail of Mss116p, as well as that of CYT-19, was shown to interact with structures adjacent to the ssRNA regions bound by the helicase core and does so regardless of the polarity of the extension[123]. The core domains, however, can only bind ssRNA in one orientation and SAXS experiments with RNA-DNA hybrid substrates, where a rigid DNA duplex extends from either the 3' or 5' end of a ssRNA region (substrates 1 and 2 shown in **Figure 1.5, top**), showed that the C-tail is flexible enough to interact with the dsDNA segment in spite of extensive changes in its positioning relative to the core domains[123] (**Figure 1.5**). Together, these results support the proposed model of DEAD-box protein chaperone activity where the protein is non-specifically tethered to the structured RNA through its flexible C-tail while its helicase core performs local structure disruption[118,123].

1.3.2 DEAD-box proteins are not conventional helicases

Although their unwinding activity is referred to as helicase activity, the mechanisms of DEAD-box proteins are substantially different from conventional helicases in ways that make them poorly equipped to unwind long, continuous helices but very well-suited to physiological roles in manipulating structured RNAs and RNPs. First, whereas conventional helicases unwind helices by translocating along one of the strands and typically have processivity measured in the range of hundreds to thousands of base pairs[128,129], DEAD-box proteins have very low processivity, displaying little or no unwinding of helices longer than about 25-40 base pairs[88,130,131] (**Figure 1.6**). Increased

duplex stability, without increased length, also severely inhibits DEAD-box proteins and as a result, they are thought to use an unwinding mechanism that does not involve translocation[126,131–133].

From structural analyses, we know that DEAD-box proteins lack the additional domains present in many processive helicases that form interactions with the bases of the bound ssRNA and are involved in directional translocation[134–138]. In contrast, DEAD-box proteins contact ssRNA only along its sugar-phosphate backbone[102–107]. These proteins also lack a β-hairpin that is present in conventional helicases and is thought to facilitate duplex disruption by acting as a "pin" or "wedge" to separate the two strands during translocation[83,135]. Thus, DEAD-box proteins are apparently not set up structurally for efficient and directional translocation along ssRNA. Instead of moving along the duplex, DEAD-box proteins are though to initiate unwinding internally and unwind short duplexes using a single cycle of ATP-dependent changes in RNA affinity[126,133,139]. Recent crystal structures of DEAD-box protein Mss116 suggest a reasonable mechanism for the internal initiation of strand separation by DEAD-box proteins. When the two domains of the protein core are in the closed conformation, bound ssRNA includes one or two sharp bends (termed a "crimp"), adopting a conformation that would prevent base-pairing with a complementary strand[102–107,111]. This structural feature also provides a good explanation for observations that local strand separation can precede hydrolysis of ATP, with subsequent hydrolysis and product release being coupled to release of the DEAD-box protein from the tightly bound strand of short duplexes[111,126,133].

A second key difference between DEAD-box proteins and conventional helicases concerns the role of RNA structures that flank the helix. Most conventional helicases strongly prefer a single-stranded extension, which must have a defined polarity because it serves as a starting point for directional translocation and unwinding[128,129] **(Figure 1.6A)**. In contrast, some DEAD-box proteins can unwind blunt-ended helices just as efficiently as those with extensions[132,140–143], again suggesting the unwinding is initiated by a direct interaction with a duplex. Although some DEAD-box proteins are activated by extensions, there is no requirement for a defined polarity (reviewed in ref. 139). The RNA extension also is not required to be single-stranded, since adjacent double-stranded regions[118,144] and even a structured RNA "extension" has been shown to enhance unwinding[127]. Furthermore, this increased unwinding efficiency is preserved even when the extension is separated from the duplex through a streptavidin linker[145]. Together, these results indicate that the extensions do not serve as a starting point for translocation but suggest instead that they provide an additional point of contact to tether the helicase core where it can unwind nearby helices[89] **(Figure 1.6B)**.

These characteristics of DEAD-box protein structure and activity suggest an unwinding mechanism in which they are localized by specific or non-specific tethering interactions by their ancillary domains, disrupt nearby helices in an inherently non-specific manner, and then release the single-stranded RNA strands upon ATP hydrolysis and release of products. Although this mechanism is very different from conventional helicases and would be clearly ineffective for tasks involving unwinding of long, continuous duplexes, it would allow DEAD-box proteins to efficiently promote

conformational transitions of RNAs and RNP complexes, which are typically made up of short RNA helical segments.

1.3.3 Other activities of DEAD-box proteins

Detailed biochemical work has shown that DEAD-box proteins have activities extending beyond the local disruption of RNA duplexes. For example, the DEAD-box protein Ded1p was shown to promote release of the exon junction complex of proteins (EJC) from an RNA oligonucleotide in an ATP-dependent manner and to release the U1 snRNP from RNA in a reaction that requires disruption of both RNA-protein and RNA-RNA contacts[86,146]. DDX42, a human DEAD-box protein, can displace the single-stranded binding protein T4gp32 from ssRNA[147]. Although these activities do not necessarily reflect the *in vivo* functions of these proteins, they establish that remodeling RNA-protein complexes is within the catalytic repertoire of DEAD-box proteins. The ability to remodel complexes by displacing proteins most likely arises from the basic RNA binding properties of the helicase core, that is, tight binding to a segment of ssRNA sequesters it away from its former protein partner.

There is strong genetic evidence and emerging biochemical evidence that these activities are used *in vivo* to promote folding and rearrangements of structured RNAs, or to remodel or dissemble RNPs. In the spliceosome, Prp5p is thought to displace the protein Cus2p from the branchpoint to allow association of the U2 snRNP[148], and Prp28p is thought to remove U1C from the RNA helix formed between U1 snRNA and the 5´-splice site in addition to disrupting the helix[92,93]. The DEAD-box protein Dbp5p (DDX21

in humans), which shuttles between the nucleus and cytosol and functions in nuclear export of mRNAs, has been the subject of intense recent scrutiny[149–153]. Genetic and biochemical evidence suggests that Dbp5p facilitates dissociation from newly-exported mRNA of at least two proteins, Mex67 and Nab2, which bind mRNA in the nucleus and promote export[149,150] **(Figure 1.7A)**. The timing of this displacement activity appears to be directed by Gle1, which is located on the cytosolic face of the nuclear pore complex (NPC) and stimulates the ATPase activity of Dbp5p[153,154]. Thus, a model has emerged in which Dbp5p travels with mRNA and its associated proteins through the NPC and then interacts with Gle1, which promotes dissociation of the proteins, blocking re-entry of the mRNA into the nucleus and conferring directionality on the export process.

DEAD-box proteins have also been shown to promote duplex formation, or annealing, of complementary oligonucleotides present at low concentrations *in vitro*[87,88,147,155] **(Figure 1.7B)**. Ded1p and Mss116p accelerate RNA annealing by several orders of magnitude, such that the rate constant for duplex formation approaches the diffusion limit for a bimolecular reaction[87,155]. The annealing activities of both proteins are enhanced by their positively charged C-terminal regions[87,119,155], which, as previously discussed, is a conserved feature of many DEAD-box proteins that function in manipulation of RNA and RNP structures[87,88,101,118,119,130].

This strand annealing activity may be important for the functions of DEAD-box proteins as RNA chaperones. It is intuitive that the formation of RNA secondary structure is an intramolecular version of duplex formation between two ssRNAs, and it is possible that annealing activity accelerates formation of intramolecular RNA secondary or tertiary

structure **(Figure 1.7B)**. This role is difficult to establish experimentally, however, because a folding step that gives accumulation of structure may nevertheless be rate-limited by a local structure disruption, which then allows the larger-scale movements that lead to the new structure. More generally, because intramolecular RNA segments are held in proximity by their covalent connections, it is not clear that the process of bringing these regions together to form a contact is rate-limiting for folding, even for individual steps, so acceleration of annealing may not typically lead to accelerated folding. This remains an interesting question for the future.

1.4 CYT-19 AND THE *TETRAHYMENA* RIBOZYME: A MODEL SYSTEM FOR STUDYING RNA STRUCTURAL REMODELING BY DEAD-BOX PROTEINS

DEAD-box proteins are known to participate in folding of complex, structured RNAs that include group I and group II introns[39,81,82] as well as facilitate conformational changes between or following the splicing steps[156]. The *Neurospora crassa* mitochondrial protein CYT-19 was the first DEAD-box protein found to have general chaperone activity[81]. CYT-19 is essential *in vivo* for efficient splicing of several mitochondrial group I introns of different structural sub-groups and to participate in other RNA processing reactions[81,82,157]. Further, co-expression of CYT-19 rescued splicing defects in a strain lacking the *S. cerevisiae* ortholog Mss116p[82], which is a DEAD-box protein required *in vivo* for efficient splicing of all nine mitochondrial group I introns and all four group II introns[82,158], underscoring the role of CYT-19 as a general chaperone and the breadth of RNA structures on which CYT-19 can act productively.

Any RNA chaperone protein that functions nonspecifically faces a fundamental challenge. How can the native, functional RNAs be differentiated from misfolded species or intermediates? The same conceptual challenge is faced by protein chaperones, and an important part of their general mechanism is that they bind preferentially to exposed hydrophobic sequences, which are typically buried in functional, folded proteins. However, RNA chaperones probably cannot use an analogous approach because misfolded RNAs can strongly resemble their native counterparts, being compact and monomeric, and differing principally in local secondary structure or the local arrangement of helices[55,58,60].

Biochemical work *in vitro* has probed the mechanisms by which DEAD-box proteins target non-native RNA conformations and promote correct folding of structured RNAs. In particular, group I introns have served as a valuable model system to explore protein-facilitated RNA folding, dating back to the mid 1990s when it was shown that several *E. coli* ribosomal proteins have the ability to accelerate splicing of phage T4 introns by participating in folding of the introns[159]. Ribozyme versions of group I introns have subsequently been used to study RNA folding and the role of chaperone proteins since correct folding to the native state can be directly monitored via their catalytic activity[20,22,139].

1.4.1 CYT-19 targets less stable RNA structure

CYT-19 has been shown to accelerate refolding of the inactive, misfolded conformer of the *Tetrahymena* group I intron ribozyme to its native state[127]. Upon

reaching the native conformation, the ribozyme is fully active, and no further assistance from CYT-19 is needed. Thus, CYT-19 acts as a true chaperone in this process, interacting transiently with the RNA to promote correct folding[81,127]. CYT-19 chaperone activity was also demonstrated for group II introns in reverse branching and reverse splicing reactions, with CYT-19 being proteolyzed prior to initiation of the reaction by substrate addition[157].

Additional studies using the *Tetrahymena* ribozyme have provided further insight into the mechanism by which CYT-19 facilitates native folding of structured RNA. Because refolding of the misfolded ribozyme to its native form requires extensive transient unfolding[58,59], it was inferred that CYT-19 promotes refolding by accelerating partial unfolding of the RNA. Further work supported and extended this model by demonstrating that CYT-19 and Mss116p are inherently non-specific and can efficiently unfold the native as well as the misfolded species if the two structures have comparable global stability[160,161]. This was directly established by monitoring unfolding and refolding of a mutant version of the ribozyme for which the large natural energy gap between the misfolded and native conformations is decreased[160]. After promoting unfolding, CYT-19 appears to allow subsequent folding along the same pathways and rates as in its absence. This activity leads to a distribution of native and misfolded conformers in which the relative populations depend on both kinetics and thermodynamics[160] **(Figure 1.8)**. For the wild-type ribozyme, this activity leads to rapid accumulation of the native state, whereas for the destabilized mutants, the same activity favors the misfolded state because the reduced energy gap is not sufficient to overcome a substantial kinetic bias for

misfolding[160]. An analogous redistribution was observed for model duplexes by the *S. cerevisiae* DEAD-box protein Ded1p[162]. Thus, DEAD-box proteins can differentiate between native and misfolded or alternate RNA conformations by their relative stability.

Together, these results lead to a model in which DEAD-box proteins target less stable RNA structures and facilitate global refolding of these RNAs by disrupting base pairs locally[116,126,133]. This mechanism bears a strong resemblance to one used by protein chaperones, particularly the Hsp70 group. Akin to DEAD-box proteins, Hsp70 proteins bind short polypeptides in an extended conformation, preventing the bound segment from forming other interactions[163,164]. Also in analogy to DEAD-box proteins, the peptide binding affinity of Hsp70 proteins is coupled to ATP binding and hydrolysis, and the ATPase cycle is regulated in turn by proteins that accelerate or inhibit specific steps[165]. The resulting cycles of regulated peptide binding and release are used by Hsp70 proteins to promote protein folding, to translocate proteins across membranes, and to disrupt protein aggregates alone or in combination with additional chaperone proteins[166–170]. For DEAD-box proteins, multiple rounds of helix binding, unwinding and strand release by one or more protein molecules could potentially lead to the wider scale disruption of structured RNAs.

However, structured RNAs are typically stabilized by tertiary interactions between secondary elements and global unfolding of these RNAs is expected to involve disruption of these tertiary contacts. Indeed, as previously mentioned, the misfolded *Tetrahymena* ribozyme not only has all of the native tertiary contacts but refolding to its native conformation requires disruption of these interactions[58]. While DEAD-box

proteins have been shown to unwind short RNA helices, their interaction with and effect on tertiary contacts had not been elucidated and it is unknown whether these chaperone proteins can directly promote the disruption of tertiary interactions required for some refolding processes.

1.4.2 How do DEAD-box proteins disrupt RNA tertiary structure?

The secondary and tertiary structure changes of the P1 helix within the *Tetrahymena* ribozyme has been extensively studied as local folding transitions in the context of a globally structured RNA[171]. This helix docks into tertiary contacts with the ribozyme core, principally through hydrogen bonds between 2´-OH groups within the helix and nucleotides within the core[173–179]. While CYT-19 has been shown to accelerate strand separation of the P1 helix[127], whether the DEAD-box protein also destabilizes the P1 docking interaction has not been established.

Previous studies have shown that the P1 helix is predominantly docked in the native ribozyme but mostly undocked for the misfolded conformation, which also has greatly reduced catalytic activity[58,180]. Furthermore, CYT-19 can unwind the P1 helix more efficiently from the misfolded than the native conformation, and additional studies have found that unwinding activity by CYT-19 was greatly reduced when the tertiary interactions between the P1 helix and the ribozyme core were strengthened[127]. The extent of CYT-19 inhibition tracks with the docking stability of P1[127], suggesting that CYT-19 is not involved in disruption of RNA tertiary structure. Although it is unknown whether the sensitivity to tertiary contact stability is general for other DEAD-box chaperones and

other RNAs, disruption of loosely-associated structural elements may predispose these chaperone proteins to destabilize non-native regions. This bias could be an important strategy for targeting DEAD-box proteins to act on misfolded RNAs, which are less likely than their native counterparts to pack their structural elements into a globular structure[67]. However, the physical basis for this proposed model for DEAD-box protein chaperone activity had not been established and is the subject of the current work.

Since destabilization of RNA tertiary structure is expected to be necessary for at least some RNA folding or remodeling processes *in vivo*, determination of the mechanism by which DEAD-box proteins can facilitate these conformational transitions is an important part of establishing how DEAD-box proteins act as general chaperones in the cell. Single molecule methods have proven useful for studying both intrinsic and protein-facilitated RNA folding. We were able to directly monitor the formation and disruption of the P1 docking interaction by CYT-19 as a model system for tertiary structure remodeling by DEAD-box proteins. In addition, secondary structure disruption of the P1 helix could also be directly observed, allowing us to examine the multistep process by which DEAD-box proteins can facilitate both the local structure changes required for some RNPs and structured RNAs to function, and the extensive unfolding and refolding necessary for resolution of inactive misfolded RNA species. We propose a mechanism for DEAD-box protein chaperone activity that provides a physical model for previous observations that these proteins can act nonspecifically yet preferentially on misfolded and non-native RNA conformations.

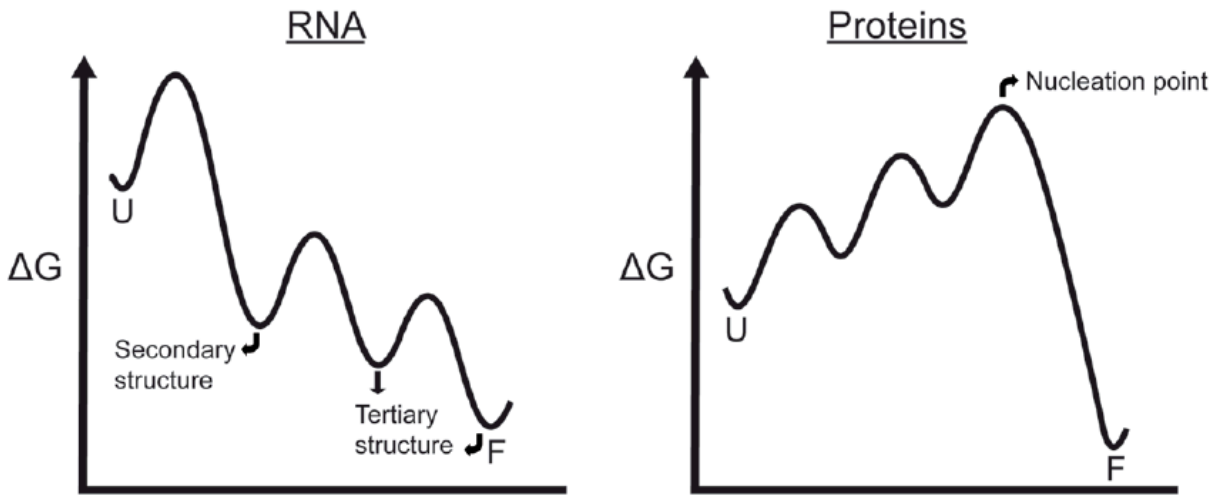

Figure 1.1: Free energy diagrams comparing RNA and protein folding

Folding of RNA typically proceeds through intermediates of increasing stability, with secondary structure forming first from unfolded structures (U), followed by steps of tertiary contact formation to give a fully folded conformation (F). Many small proteins, on the other hand, do not accumulate intermediates because partially-ordered conformations are less stable than unfolded forms even under conditions that favor global folding.

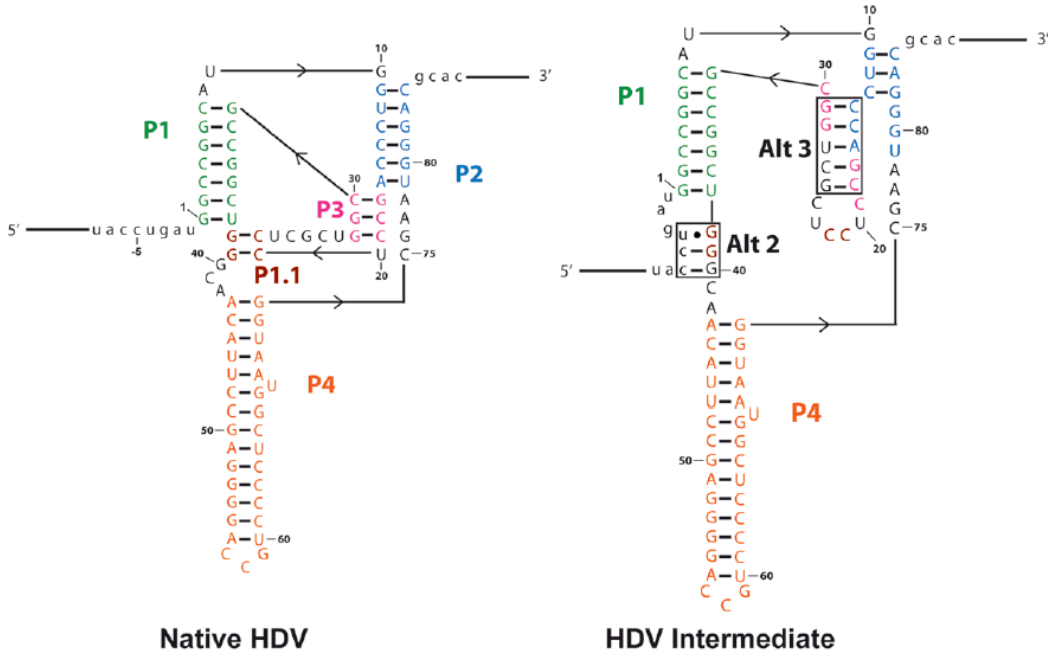

Figure 1.2: Secondary structures of the native hepatitis delta virus (HDV) and an intermediate with non-native base pairings

Nucleotides are color-coded according to their native fold. Alternate base pairings in the HDV intermediate are boxed and labeled.

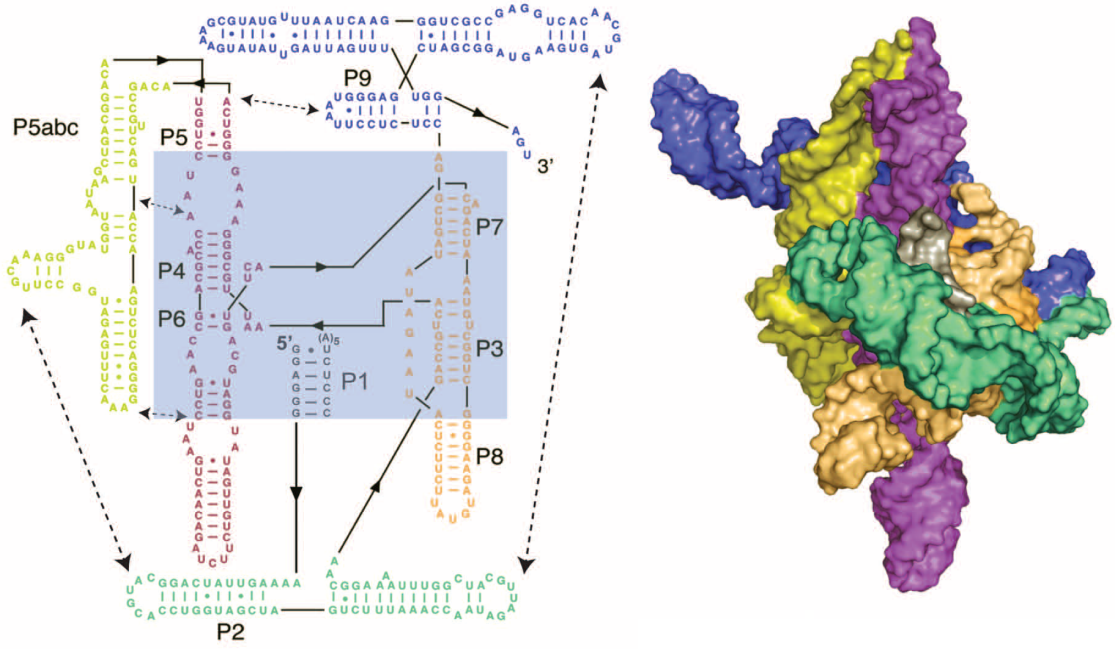

Figure 1.3: Secondary and tertiary structure of the *Tetrahymena* group I intron

The "shortened" ribozyme version is shown, in which the exons are removed and the ribozyme forms a core helix, termed P1, with an oligonucleotide that mimics the 5'-splice site. The ribozyme core is highlighted in blue in the secondary structure. Regions of the secondary structure are color-coded according to structural domains, and tertiary contacts are indicated with dashed lines. At the right is the tertiary structure model of the intron[181] with domains in the same colors as in the secondary structure diagram.

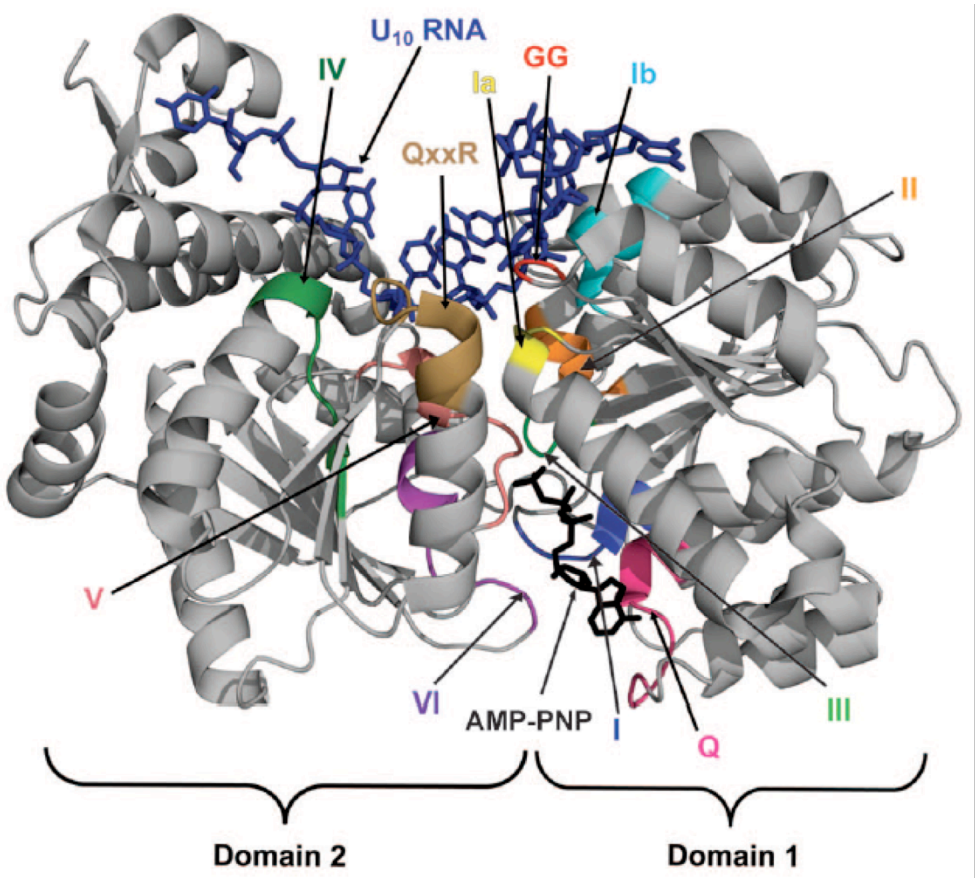

Figure 1.4: Crystal structure of DEAD-box protein Mss116p with bound AMPPNP and U_{10} ssRNA

AMPPNP is shown in black (bottom of the domain interface) and the ssRNA is shown in dark blue (top of the domain interface). The conserved motifs are highlighted and labeled. Domain 1 is on the right. This structure was originally published in Del Campo and Lambowitz (2009).

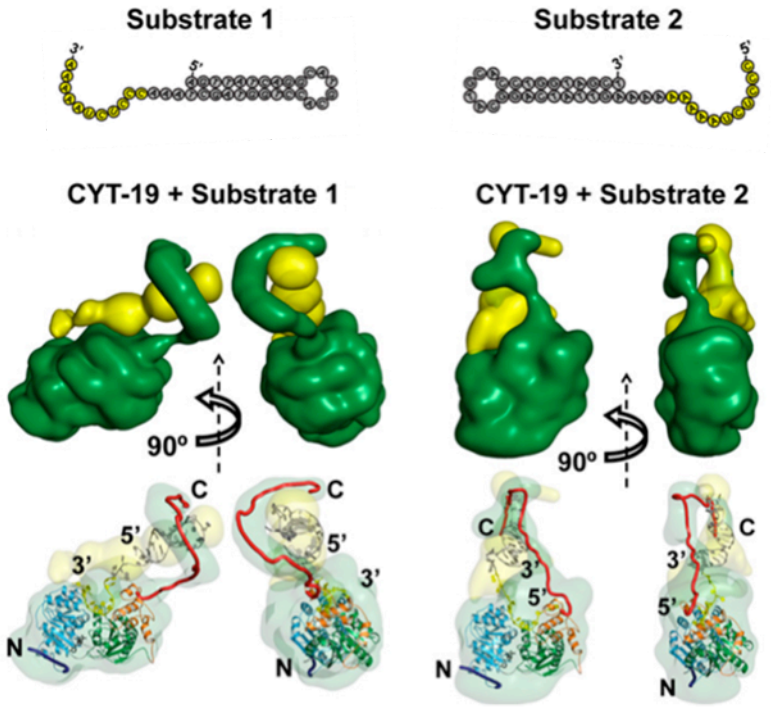

Figure 1.5: The flexible C-tail of DEAD-box protein CYT-19 can interact with adjacent structures regardless of their polarity and orientation relative to the helicase core

The secondary structures of the two RNA-DNA substrates used are shown with the ssRNA region in yellow and dsDNA segment shown in grey. SAXS analysis showing CYT-19 (green) in complex with the indicated substrates (yellow) with atomic models for protein and nucleic acid manually placed inside the corresponding SAXS envelope (bottom). These structures were originally published in Mallam, et al (2011).

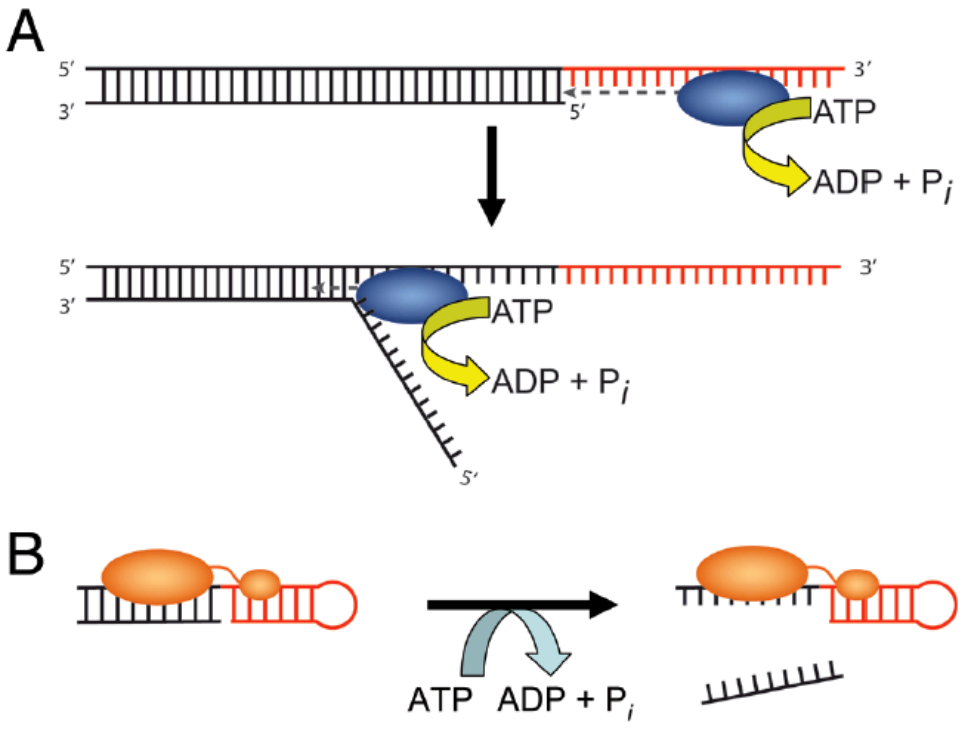

Figure 1.6: Models for duplex unwinding by conventional helicases and DEAD-box proteins

(A) Conventional helicases load onto single-stranded overhangs of a defined polarity (red) and then translocate into the duplex region (black), displacing the complementary strand. **(B)** DEAD-box proteins unwind short duplexes in an ATP-dependent but non-processive manner by initiating unwinding internally, within a duplex region (black). Some DEAD-box proteins form additional interactions with adjacent single-stranded or double-stranded segments (red), which are suggested to tether the core in position to disrupt nearby helices.

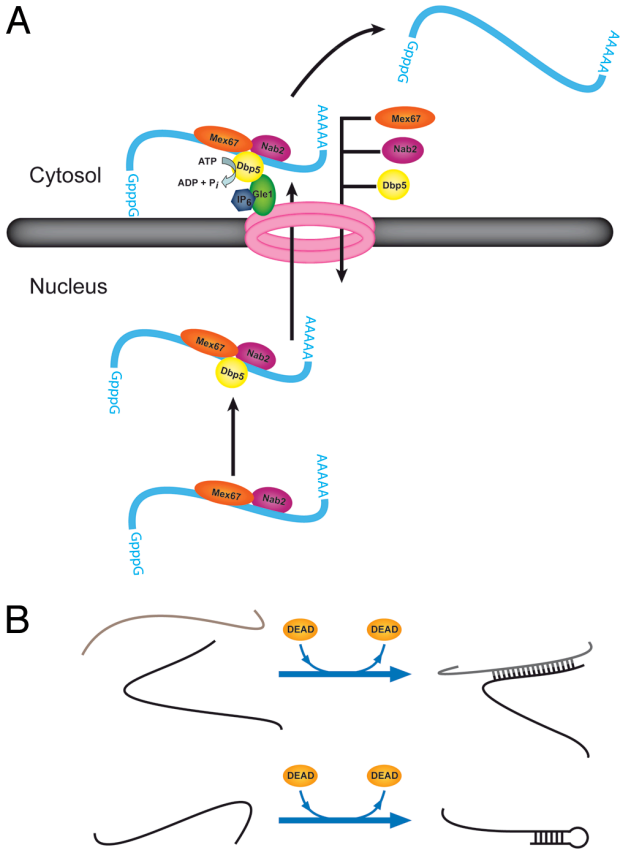

Figure 1.7: Additional activities of DEAD-box proteins in promoting RNA rearrangements and RNP remodeling

(A) Dbp5p facilitates transport of mRNA from the nucleus to the cytosol. The role of Dbp5p is thought to include removal of the nuclear export factor Mex67p and the nuclear RNA-binding protein Nab2p. Upon protein removal, the mRNA is blocked from re-entering the nucleus. **(B)** Strand annealing capabilities of DEAD-box proteins. DEAD-box proteins have been shown to accelerate intermolecular duplex formation by two strands of RNA in solution without a requirement for ATP (top). It is possible that this activity is important for intramolecular RNA folding by facilitating formation of local or long-range secondary structure (bottom).

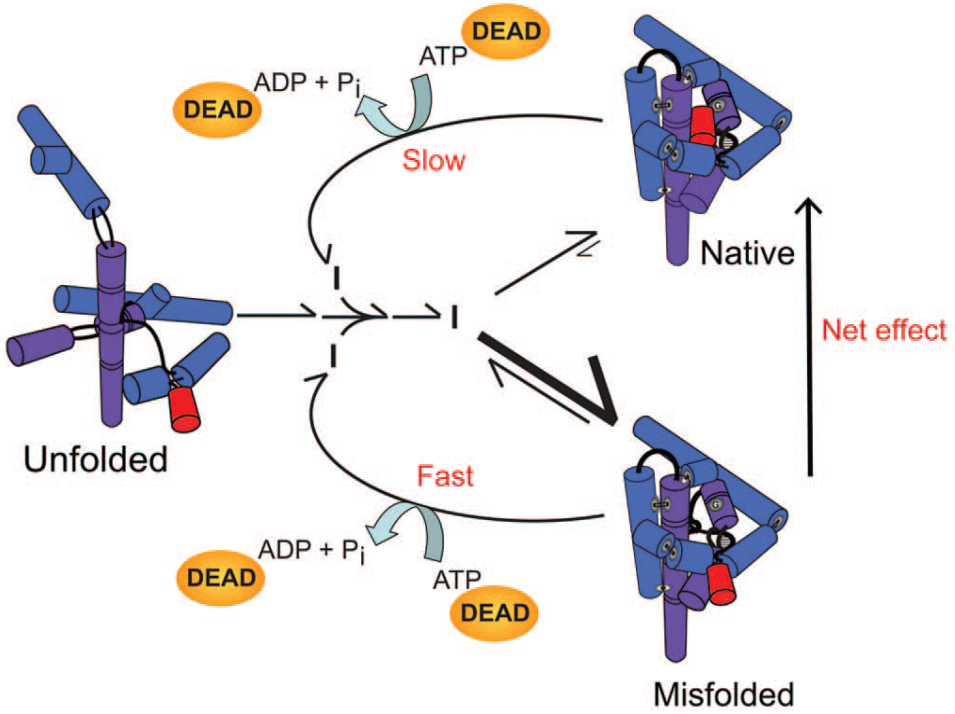

Figure 1.8: RNA chaperone activity of DEAD-box proteins by non-specific disruption of local structure

The RNA, shown in cartoon form as helical cylinders, represents the Tetrahymena group I intron ribozyme. The ribozyme folds primarily to the long-lived misfolded conformation (bottom; larger rate constants are shown with longer, thicker arrows), which therefore accumulates to high levels despite being less stable than the native state (top). DEAD-box proteins facilitate local structure disruptions of the native and misfolded RNA species in an ATP-dependent manner, without recognizing any structural features of the misfolded RNA. The RNA then folds again without any further influence of the DEAD-box protein. Despite the lack of specific recognition by the DEAD-box protein, the misfolded RNA is disrupted with greater efficiency than the native ribozyme because of its lower stability, and this difference allows the non-specific activity of DEAD-box proteins to accelerate net refolding to a population of predominantly native ribozyme (indicated by arrow at right). Modified from Bhaskaran and Russell (2007) with permission.

Chapter 2: DEAD-box helicase proteins disrupt RNA tertiary structure through helix capture

2.1 Introduction

Structured RNAs are involved in many essential biological processes such as pre-mRNA splicing, regulation of gene expression, protein synthesis and maintenance of chromosome ends[1,2,11,18,182]. These functions require the RNAs to fold into specific structures and for some, to transition between functional conformations. Because RNA structure is inherently stable, even at the local level, resolution of misfolded RNAs or rearrangements of structured RNAs can be slow on the biological timescale. These properties suggest a general requirement for RNA folding chaperones *in vivo*[72], and diverse proteins have been shown to possess ATP-dependent or ATP-independent RNA chaperone activity[22,75].

DEAD-box proteins are superfamily 2 RNA helicases that can function as RNA chaperones to promote the formation and remodeling of functional RNAs and RNPs[183,184] and are linked to essentially all RNA metabolic processes in all three branches of life[184–186]. They use a conserved helicase core of two RecA-like domains to perform a broad range of activities including protein displacement from RNA[86], RNA structure formation[87,187] and their hallmark activity, ATP-dependent unwinding of short RNA helices[127,184,188], including those within structured RNAs[127]. However, in addition to the helical segments that constitute RNA secondary structure, structured RNAs typically

include tertiary contacts that must be disrupted during many remodeling processes[39,54,58,60]. Although it has been proposed that regulated binding to ssRNA might be sufficient to accelerate disruption of tertiary contacts[133,139], such disruptions have not been demonstrated for any DEAD-box protein, leaving the mechanisms of these RNA remodeling reactions unclear.

CYT-19, a DEAD-box protein found in the mitochondria of *Neurospora crassa*, functions as a general RNA chaperone[81], facilitating correct folding of diverse group I intron RNAs by accelerating unfolding of misfolded intermediates[39,127,160]. Here, we probe how CYT-19 promotes such unfolding by monitoring changes in the secondary and tertiary structure of the P1 helix within the *Tetrahymena* group I intron ribozyme, which has been extensively studied as an isolated tertiary folding event within a globally folded RNA[171,189]. CYT-19 can unwind the P1 helix, and previous results have shown that the unwinding efficiency depends on the docking stability of the P1 helix, suggesting that unwinding requires loss of the tertiary contacts prior to or during secondary structure disruption[127]. However, it was unclear how CYT-19 accomplished the RNA tertiary unfolding and whether it resulted from a known or a novel activity.

To dissect this multi-step remodeling process, we used single molecule Fo¨rster resonance energy transfer (smFRET) to observe CYT-19 action on both secondary and tertiary structure of the 11-bp P1 helix (**Figures 2.1** and **2.2A**). Formation and disruption of this tertiary contact as well as helix unwinding were directly monitored, allowing us to independently resolve and quantify the effects of CYT-19 on the secondary and tertiary folding transitions of the P1 helix. Thus, we generated a detailed view of the process by

which a DEAD-box protein can disrupt tertiary and secondary contacts within a globular RNA structure.

2.2 MATERIALS AND METHODS

2.2.1 Protein purification

CYT-19 was purified as previously described[160]. Ded1 was a gift from the Lambowitz lab and purified by M. Del Campo.

2.2.2 Ribozyme preparation

For ensemble experiments, the L-21/ScaI form of the *Tetrahymena thermophila* group I ribozyme was prepared by *in vitro* transcription (>4 hours at 37 °C with 25 mM $MgCl_2$)[127]. For single molecule experiments, L-21/T2, a form of the ribozyme that is extended at the 3´-end with the tail sequence ACCAAAAUCAACCUAAAACUUACACA, was prepared under the same conditions[180]. L-16/ScaI, a version of the ribozyme with a 5´-extension of GGUUU (resulting in an 11-bp P1 helix), and L-16/T2, which includes both the 5´- and 3´-extensions, were transcribed *in vitro* at 30 °C for 30 min with 4 mM $MgCl_2$ to minimize self-cleavage[190]. All RNAs were then purified with RNeasy columns (Qiagen) and stored in TE buffer at -20 °C.

RNA oligonucleotides with and without dye labels were purchased from IDT (Coralville, IA) and stored in TE buffer at -20 °C. For ensemble experiments, substrate oligonucleotides were 5´ end-labeled with [γ^{32}-P]ATP (PerkinElmer) using T4

polynucleotide kinase (New England Biolabs). See **Table 2.1** for sequences of all oligonucleotides used.

2.2.3 Ensemble unwinding experiments

Benchtop and rapid quench-flow experiments monitoring the unwinding activity of CYT-19 or Ded1 were performed at 25 °C in 50 mM Na-MOPS (pH 7.0), 10 mM $MgCl_2$, 50 mM KCl, 2 mM ATP-Mg^{2+} (ATP mixed with an equal amount of $MgCl_2$) and 5% glycerol as previously described[127]. Ribozymes were pre-folded to the native state in 50 mM Na-MOPS (pH 7.0) and 10 mM $MgCl_2$ for 30 min at 50 °C[127,180,190]. Alternatively, the misfolded ribozyme was generated by incubation in 50 mM Na-MOPS (pH 7.0) and 10 mM $MgCl_2$ for 5 min at 25 °C[58,127]. Trace radiolabeled substrate was incubated with pre-folded native or misfolded ribozyme for 5 min at 25 °C. Unwinding reactions were initiated by adding CYT-19 or Ded1 and at least 25-fold excess unlabeled substrate and quenched to a solution of 33 mM $MgCl_2$ and 1 mg/ml Proteinase K. Bound and unbound substrates were separated on a 20% native gel at 4 °C and quantified using a PhosphorImager and ImageQuant (GE Healthcare). Data were analyzed using Kaleidagraph (Synergy Software).

2.2.4 Total internal reflection fluorescence (TIRF) microscope

A diode-pumped solid-state green laser (532 nm; CrystaLaser GCL-100-M) and a red laser (637 nm; Coherent, maximum power 50 mW) were directed by dichroic mirrors through a prism at an angle that allows total internal reflection at the surface of the

sample channel, which was constructed from a glass cover slip adhered to a quartz slide with double-sided tape. The surfaces of both the cover slip and slide were passivated with a mixture of mPEG and biotin-PEG, allowing for ribozyme immobilization while preventing protein adsorption to the slide surface (see below). Images were collected using a 60X water-immersion Olympus UPlanApo objective (numerical aperture, 1.2), filtered through a 550 nm long-pass filter (Chroma Technology) to remove scattered excitation light, separated into "green" and "red" images using dichroic mirrors, and focused onto the two halves of a microchannel plate intensified charge-multiplying charge-coupled device (CCD) (I-Penta*MAX*, Princeton Instruments, Roper Scientific, Inc).

2.2.5 Slide preparation for single molecule experiments

Quartz slides were incubated in 10% Alconox solution overnight. After a brief sonication (1 min) and thorough rinsing with ePure water, slides were incubated in a solution of NoChromix (Godax Laboratories) and sulfuric acid for at least 24 hours. They were then cleaned by successive 20-min sonications in acetone, ethanol and 1 M KOH with thorough rinsing with ePure water between each step. Slides were then pyrolyzed before a 10-min sonication in methanol in preparation for aminosilanization. Glass coverslips were sonicated in 1 M KOH (20 min) and then in methanol (10 min) before aminosilanization.

Cleaned slides and coverslips were incubated for 10 min in a methanol solution of 1% aminosilane (United Chemical Technologies) and 5% acetic acid (Fisher Scientific),

sonicated for one min and then incubated for another 10 min in the same solution before being rinsed with methanol and then ePure water, and dried with nitrogen. Coverslips and slides were then incubated for at least three hours in a 40:1 solution of methoxy-poly(ethylene glycol)-succinimidyl valerate (5000 kD mPEG-SVA, Laysan Bio) and biotin-poly(ethylene glycol)-succinimidyl valerate (5000 kD biotin-PEG-SVA, Laysan Bio) made in 100 mM sodium bicarbonate buffer (pH 8.5). Coverslips and slides were then rinsed thoroughly with ePure water, dried with nitrogen and then stored in the dark at -20 °C.

2.2.6 Single molecule fluorescence data acquisition

The ribozyme was annealed to biotinylated, Cy5-labeled tether (≥10:1 molar ratio of ribozyme to tether) in 50 mM Na-MOPS (pH 7.0) with 200 mM NaCl by heating at 95 °C for 1 min before cooling at 0.1 °C/s to 50 °C. The ribozyme was then folded to its native conformation by adding $MgCl_2$ to a final concentration of 10 mM and incubating the solution at 50 °C for 30 min. Cy3-labeled substrate oligonucleotides were then added to the pre-folded ribozyme at approximately 7-fold excess and incubated for 5 min at 25 °C in ribozyme buffer (50 mM MOPS, pH 7.0, 10 mM $MgCl_2$). The ribozyme-substrate-tether complex was then diluted to 10-25 pM in ribozyme buffer and immobilized onto PEG slides via a biotin-streptavidin linkage.

To measure P1 docking and unwinding, various concentrations of CYT-19 or Ded1 protein were diluted in CYT-19 buffer solution (50 mM Na-MOPS, pH 7.0, 10 mM $MgCl_2$, 50 mM KCl, 5% glycerol). For some experiments, ATP or another nucleotide

(see **Table 2.2**) was added to a final concentration of 2 mM. The solution was then flowed through the sample channel along with an oxygen scavenging system (OSS) of 1 mM Trolox ((±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid, Aldrich, >97%), 500 mM glucose, 0.1 mg/ml glucose oxidase, and 0.06 mg/ml catalase. Images of the dye-labeled molecules within the sample channel were collected in 40-ms or 100-ms frames for 10-30 s (fully intensified at ~1000 V). To measure CYT-19 dissociation, slide-immobilized ribozyme was incubated with subsaturating concentrations of CYT-19 along with 2 mM AMP-PNP for >2 min at 25 °C before the sample channel was washed with CYT-19 buffer (and OSS) to remove the protein from solution, preventing CYT-19 from re-binding. Data recordings were then acquired at 2-s frames for 5-10 seconds (to reduce dye photobleaching) every 30 s over a period of 30 min. Data were collected under green laser excitation and then under red laser excitation for colocalization of Cy3 with Cy5. The average signal-to-noise ratio was ~5 with green laser intensity averaging ~15 mW (measured near the laser aperture).

2.2.7 Single molecule data analysis

2.2.7.1 Selection of molecules

For each data acquisition, fluorescent 'spots' of interest within each field of view were identified by mapping pixels from the "red" side to the "green" side based on an affine transformation determined from images using fluorescent beads as fiducial markers. Local background was subtracted from each selected spot for each frame.

Fluorescence intensity traces were first automatically chosen based on intensity thresholds and then inspected visually. Single RNA molecules were selected based on fulfillment of the following criteria: (1) the presence of active Cy5, based on average fluorescence intensity under the red laser or the presence of a high FRET state under green laser, (2) the presence of active Cy3, based on average fluorescence intensity under green laser, (3) stable signal strength, indicating a single molecule. Approximately 65% of automatically selected traces in each field of view were ultimately selected for further analysis. FRET traces for these molecules were calculated from the recorded fluorescence intensities of the Cy3 and Cy5 dyes with 9-pt smoothing applied.

$$\mathrm{FRET} = \frac{(\mathrm{I_{Cy5}} - \beta \times \mathrm{I_{Cy3}})}{(\mathrm{I_{Cy3}} + \mathrm{I_{Cy5}})},$$

where β represents the degree of Cy3 emission that crosses over to the red channel.

2.2.7.2 Determination of P1 docking and undocking kinetics

Transitions between two FRET states (high FRET ~0.9 and low FRET ~0.2), representing docked and undocked states, were observed for these ribozyme molecules as previously described[189,190]. Thresholds were set from 0 to 0.5 for detection of the low FRET state and from 0.7 to 1 for the high FRET state. The docking equilibrium constant (K_{dock}) was determined from the ratio of total time spent in the docked and undocked states as calculated from the Gaussian fit of each peak in the FRET histogram. Time constants for undocked and docked P1 ($\tau_{undocked}$ and τ_{docked}, respectively) were calculated using the dwell times in the low and high FRET states, respectively[189]. The corresponding

rate constants for most molecules under CYT-19 buffer conditions (see above), were well-described by a single-exponential fit. However, taken as an ensemble, heterogeneity in intrinsic docking behavior contributed to a small second phase (**Figure 2.3** and **Table 2.1**).

The lifetime of the docked state, τ_{docked}, was not changed significantly by CYT-19 or Ded1. However, the lifetime in the undocked state became biphasic, with a fraction of events having a larger $\tau_{undocked}$, i.e. a slower than intrinsic observed docking rate constant (k_{obs}). Several parallel processes contribute to k_{obs}: (1) redocking, (2) unwinding of P1, (3) photobleaching of Cy3, and (4) truncation of the time trace. That is,

$$k_{obs} = k_{dock} + k_{unwind} + k_{Cy3\ photobleach} + k_{truncation} \quad .$$

The rate constants for each of these processes in the presence of CYT-19 or Ded1 were determined by multiplying k_{obs} by the relative probability (F) of each "fate" for those events that contribute to the slow phase. These long undocked events were defined as having a minimum lifetime of ~2 s for CYT-19 and ~1 s for Ded1 (these lifetimes are based on the double exponential fit of the cumulative number of undocked events). For example, the docking rate constant for P1 in the presence of either CYT-19 or Ded1 (k_{dock}) was calculated by determining the fraction of these long-lived undocked events that resulted in redocking of P1 (F_{docked}, see below). For our RNA construct, the signals for P1 unwinding and Cy3 photobleaching are the same. Therefore, both processes are included in the calculation of $F_{unwound/PB}$ (see below).

$$F_{docked} = \frac{N_{docked}}{N_{undocked}},$$

$$F_{unwound/PB} = \frac{N_{unwound/PB}}{N_{undocked}},$$

$$F_{truncated} = \frac{N_{truncated}}{N_{undocked}},$$

where $N_{undocked}$ is the total number of long undocked events and N_{docked} is the number of these transition events that dock after a long dwell time (see above) in the undocked state. Likewise, $N_{unwound/PB}$ is the number of P1 unwinding and Cy3 photobleaching events, and $N_{truncated}$ is the number of long undocked states that are limited by the observation time. The rate constant for Cy3 photobleaching ($k_{photobleach}$) was measured independently (0.55 min^{-1}, **Figure 2.4**) and subtracted from k_{unwind}.

$$k_{dock} = F_{docked} \times k_{obs},$$

$$k_{unwind} = (F_{unwound/PB} \times k_{obs}) - k_{photobleach},$$

$$k_{truncation} = F_{truncated} \times k_{obs}.$$

The results for CYT-19-dependent k_{dock} are shown in Tables 2 and 3. The average k_{dock} in the presence of CYT-19 and ATP is 5.2±2.1 min^{-1} (**Table 2.3**). Calculations were performed with an observation time of 10 s. Data for longer (~15-30 s) observation times yielded similar results (within 2-3-fold, not shown). Results for k_{unwind} are shown in **Table 2.3** and reflect unwinding of P1 by CYT-19 from the undocked conformation. The results for Ded1 are shown in **Table 2.4**.

2.2.7.3 P1 unwinding monitored using single molecule fluorescence

P1 unwinding by CYT-19 was measured by single molecule fluorescence by monitoring the intact P1 helices over time. Data were taken over multiple fields of view and the number of Cy3-labeled P1 helices was normalized by the amount of Cy5-tether within the same field of view. Rate constants for unwinding by CYT-19 determined this way are shown in **Table 2.2** and are within 3-5-fold of benchtop results (**Figure 2.5**).

While CYT-19 increased P1 unwinding from the low FRET state, a fraction of unwinding events appeared to occur from the high FRET state (~20% of all molecules). The frequency of these events, however, was not affected significantly by the presence of CYT-19 or Ded1, indicating that this behavior was not a protein-dependent effect and most likely reflected photophysical processes[191–193].

2.2.7.4 Single molecule measurement of CYT-19 dissociation from the ribozyme

The average FRET value of the ribozyme molecules can serve as a sensor for the CYT-19-dependent shift in K_{dock} and therefore, as an indicator of the lifetime of CYT-19 association. Because CYT-19 increases $\tau_{undocked}$, the average FRET value of CYT-19-bound molecules would be closer to that of the undocked state, while molecules without bound CYT-19 would have a higher average FRET that reflects strong intrinsic P1 docking and short undocked lifetimes. After removal of CYT-19 from solution by a washout, the average FRET value for the molecules that were initially in the low FRET state would increase as CYT-19 dissociates from the ribozyme, returning to the intrinsic average FRET value calculated from the docking equilibrium in the absence of CYT-19

(~0.85). After washing out CYT-19, we observed a fast initial increase from the average FRET for undocked P1 (~0.2) with a rate constant close to that expected for re-equilibration of P1 docking in the presence of bound CYT-19 ($k_{obs} + k_{undock} \sim 23$ min^{-1}; **Figure 2.6** shows an initial phase of 30 min^{-1}). In comparison, without CYT-19, docking re-equilibration occurs with a rate constant of ~130 min^{-1} (**Figure 2.6, black**). The subsequent slow phase for the CYT-19 washout data has a rate constant of 0.43 min^{-1} and is thought to reflect dissociation of CYT-19 because it is not present when CYT-19 is not pre-bound. The additional slow phase of 0.01 min^{-1} is most likely due to the slow interconversion of ribozyme populations displaying heterogeneous docking behaviors[190]. This very slow phase was not probed in the absence of CYT-19 because the observation times were too short to detect it. Nevertheless, it probably accounts for the difference between the observed endpoint in the absence of CYT-19 (~0.73) and the expected value based on independent measurements of the docking equilibrium (~0.85).

2.3 RESULTS

To measure secondary and tertiary transitions of the P1 helix, we used a smFRET system that was designed previously[180,189,190]. The ribozyme was extended from its 3' end and annealed to a complementary oligonucleotide that was immobilized on a microsope slide by a biotin/streptavidin linkage for visualization using total internal reflection (TIR) microscopy (see **Section 2.2.4** and **Table 2.1**). Cy3 and Cy5 dyes were positioned such that docked P1 gives efficient energy transfer from Cy3 to Cy5 and a correspondingly high FRET value (~0.9), while undocked P1 gives a greatly reduced FRET value (~0.2;

Figure 2.2A)[180,189,190]. Unwinding of P1 was detected as a loss of Cy3, because the Cy3-labeled strand is released into solution. Cy3 disappearance can also reflect photobleaching, which was measured separately and subtracted (**Figure 2.4** and **Section 2.2.7.2**).

2.3.1 CYT-19 does not accelerate loss of tertiary contacts between the P1 helix and ribozyme core

In the absence of CYT-19, P1 was predominantly docked in most molecules but underwent cycles of undocking and redocking, as observed previously under similar conditions[189,190], with rate constants of 20 min^{-1} and 120 min^{-1}, respectively (**Figure 2.2B, top trace, Figure 2.2C-D, black, Figure 2.3** and **Table 2.2**). Spontaneous unwinding of the P1 helix was not detectable. However, addition of CYT-19 and ATP led to robust unwinding (**Figure 2.2B, second trace, Figure 2.5** and **Table 2.2**). The secondary structure disruption of the P1 helix by CYT-19 occurred primarily from the low FRET state (see **Section 2.2.7.3**). Thus, the CYT-19-mediated remodeling process occurs in two steps, with tertiary undocking preceding helix unwinding. Strikingly, the rate of P1 undocking was not accelerated (**Figure 2.2C**), even with CYT-19 concentrations that approached saturation (see below) and gave substantial increases in the overall unwinding rate (**Figure 2.5** and **Table 2.2**). Thus, CYT-19 apparently 'waits' for spontaneous loss of the tertiary contacts and then interacts with the undocked P1 helix to unwind it.

2.3.2 CYT-19 captures the undocked P1 helix, preventing redocking

While CYT-19 does not actively disrupt the P1 docking contact, we found that it increased the lifetime of the P1 helix in the undocked state. In the presence of CYT-19, a substantial fraction of undocked events had longer than intrinsic lifetimes, resulting in a slow phase with an observed rate constant of 20 min^{-1} (**Figure 2.2D, blue and green**). Other undocked events were followed by rapid redocking with the intrinsic docking rate constant (120 min^{-1}, **Figure 2.2D**), presumably because CYT-19 was not bound or was not positioned to interact with the P1 helix. Supporting a contribution from incomplete protein binding, the fraction of undocking events with long lifetimes increased with CYT-19 concentration (**Table 2.2**), and additional experiments indicated that CYT-19 was approaching saturation at these concentrations but not fully saturated (**Figure 2.5**).

For the long-lived undocked complexes, we observed a competition between alternative fates. For undocked events that were not truncated by the termination of data collection, the P1 helix was either unwound (56% of the time) or it redocked into the ribozyme core (**Figure 2.2B, middle traces,** and **Figure 2.7**). We calculated unwinding and docking rate constants from the lifetime distributions of these complexes and the probabilities of the alternative outcomes, and we found that CYT-19 slows P1 docking by ~20-fold to 5.2±2.1 min^{-1} (**Tables 2.2** and **2.3,** see **Section 2.2.7.2** for details of the analysis). Comparable slowing was observed with modified versions of P1 that dock less strongly. Previous studies have established that 2' methoxy substitutions at positions -1 and -3 of the P1 helix shift the docking equilibrium significantly towards the undocked state[171,189] (**Figure 2.1** and **Tables 2.1** and **2.2**). The modified P1 helix that predominantly

occupies the undocked position (–3m,rSA$_3$C$_2$ with K_{dock} ~0.03, **Table 2.2**) does not show a significant population of slowed docking events in the presence of CYT-19, probably since the undocked lifetimes for this P1 helix are intrinsically long-lived. However, CYT-19 slows P1 docking for the modified helix that is moderately destabilized (–1m,rSA$_3$C$_2$ with K_{dock} ~0.54) by approximately 5-fold without significantly affecting its undocking rate constant, (**Figure 2.8** and **Table 2.2**), indicating that even when the P1 docking interaction is weakened, CYT-19 can not actively facilitate tertiary contact disruption. These results indicate that CYT-19 interferes with P1 docking by binding and capturing the P1 helix after it undocks spontaneously. This 'helix capture' mechanism allows CYT-19 to destabilize tertiary docking of the P1 helix, shifting the equilibrium toward the undocked state, without actively disrupting the tertiary contacts like it does for secondary structure.

2.3.3 ATP is not required for P1 helix capture by CYT-19

To probe the role of ATP in CYT-19-mediated destabilization of P1 tertiary docking, we monitored P1 docking behavior with ATP analogs and in the absence of nucleotide. We found that upon replacing ATP with ATP analog AMP-PNP, ADP, or in the absence of nucleotide, CYT-19 does not unwind the P1 helix significantly but it retains the ability to block tertiary docking (**Figure 2.2B, bottom trace, Figure 2.2E** and **Table 2.2**). With AMP-PNP, the redocking rate is the same within error as with ATP, whereas the rate is modestly increased with ADP or in the absence of nucleotide (~2-3-fold, **Table 2.2**). Overall, the lack of a strong dependence of helix capture on nucleotide

suggests that it does not require closure of the two RecA-like domains in the DEAD-box helicase core, a conformational transition that typically depends on bound nucleotide[102,111]. Thus, helix capture may result primarily from interactions of the RNA helix with domain 2 (see **Discussion**)[111].

2.3.4 Helix unwinding can be limited by the rate of tertiary contact disruption

When CYT-19 interacts with the 11-bp P1 helix, helix unwinding is partially rate limiting for the overall disruption process, as indicated by the substantial fraction of long-lived undocking events that result in P1 redocking rather than unwinding (**Table 2.2**). Most helical segments in structured RNAs are shorter than 11 bp and correspondingly less stable, such that unwinding of these helices may be fast enough that the overall process is fully rate limited by the intrinsic loss of the tertiary contacts. We tested this idea using a ribozyme construct with a shorter P1 helix of 6 bp, which also displayed extended undocked lifetimes in the presence of CYT-19 and AMP-PNP (**Figure 2.9**). This helix was indeed unwound much faster by CYT-19 in the presence of ATP[127], which precluded generating robust statistics with smFRET (data not shown). Therefore, we used rapid quench-flow techniques to measure the maximum rate of unwinding by CYT-19 (k_{max}). At saturating concentrations of CYT-19, the 6-bp P1 helix was unwound with an overall rate constant of ~6 min^{-1}, which is comparable to the intrinsic undocking rate constant for this helix, suggesting rate-limiting undocking (**Figure 2.9C, left** and **2.10A, red**). As expected from our model, CYT-19 unwinding of the P1 helix from the misfolded conformation of the ribozyme, for which the P1 helix is known to be

predominantly undocked[180], is significantly increased (**Figure 2.10A, blue**). In addition, when docking stability was weakened by introducing a 2' methoxy substitution within the P1 helix (–3m,rSA$_5$ and **Table 2.1**), we found that the difference in k_{max} between the native and misfolded ribozyme was essentially abolished, confirming that CYT-19 acts preferentially to unwind the undocked P1 helix rather than recognizing minute structural differences between the native and misfolded conformers (**Figure 2.10B)**. Conversely, when docking is strengthened by forming the P1 helix with the product oligonucleotide[194] (rP, **Table 2.1**), P1 unwinding by CYT-19 for the native ribozyme decreased significantly and the difference in k_{max} between the native and misfolded ribozyme was exacerbated (**Figure 2.10C**). Thus, unwinding of a short helix is indeed rate limited by loss of the tertiary interactions, and this tertiary disruption is not accelerated by CYT-19.

2.3.5 CYT-19 can remain associated with the ribozyme for multiple cycles of helix capture

We next used the CYT-19-dependent destabilization of P1 docking to monitor the lifetime of the DEAD-box protein interaction with the ribozyme, testing whether CYT-19 remains associated with the ribozyme after it releases the P1 helix. As above, we added CYT-19 to immobilized ribozyme, and then we washed CYT-19 out of the sample channel so that its dissociation from the ribozyme would be irreversible. We then monitored the FRET value of ribozyme molecules for which the P1 helix was initially undocked. The average FRET value of these molecules increased over time, ultimately returning to a value that reflected the intrinsic P1 docking equilibrium (see **Section**

2.2.7.4). After an initial rapid increase in FRET with a rate constant of ~30 min^{-1}, reflecting re-equilibration of P1 docking with bound CYT-19, the subsequent phase gave a rate constant of 0.43 min^{-1}. This slow phase was not present in a control reaction lacking CYT-19 (**Figure 2.6, black**). This increase in average FRET value most likely reflects CYT-19 dissociation, which would allow the P1 helix to return to its intrinsic behavior. Previous studies indicated that CYT-19 could associate with structured RNA through its flexible C-tail[118,123], and the current results support and extend this model. In fact, a CYT-19 mutant, which has the C-tail region truncated, does not efficiently capture the P1 helix, suggesting that the C-tail is required for long-lived association between CYT-19 and its RNA substrate (**Figure 2.11**). The continued binding is expected to allow CYT-19 to participate in multiple cycles of helix capture and unwinding, with the helicase core likely remaining poised to capture P1 or other helical elements as they become exposed by transient fluctuations.

2.3.6 The DEAD-box protein Ded1 also uses a helix-capture mechanism

We tested the generality of the helix-capture mechanism by using Ded1, a multi-functional *S. cerevisiae* DEAD-box protein[86,87]. In the presence of ATP or AMP-PNP, we found that Ded1 uses the same basic mechanism to destabilize tertiary docking of the P1 helix. Specifically, Ded1 does not accelerate the loss of tertiary contacts, but slows their subsequent formation (**Figure 2.12** and **Table 2.4**), indicating that like CYT-19, Ded1 captures the P1 helix after spontaneous undocking. There are also some interesting differences. Unlike for CYT-19, the fraction of P1 undocking events that resulted in helix

capture did not depend on Ded1 concentration (**Figure 2.12B, left** and **center,** and **Table 2.4**), suggesting that Ded1 binding is saturating under our experimental conditions. However, Ded1 unwinding activity is not saturated at these concentrations (**Figure 2.13**), consistent with previous suggestions that helix unwinding may be promoted by multiple Ded1 monomers[142,145]. Also, the fraction of captured helices is lower than that for CYT-19 and the rate constant for redocking is approximately 3-fold faster (**Tables 2.2** and **2.4**), suggesting weaker helix binding by Ded1. The shorter-lived complex between Ded1 and RNA helices may reflect that Ded1 acts primarily on cytoplasmic RNAs such as mRNAs[184], which are typically less stable and therefore more dynamic than the group I intron substrates of CYT-19. Finally, Ded1 requires nucleotide, ATP or AMP-PNP, to capture the undocked helix (**Figure 2.12B, right,** and **Table 2.4**), suggesting the involvement of both RecA-like domains in this interaction (see **Discussion**). Despite these differences, Ded1 shares the basic behaviors delineated for CYT-19, capturing the transiently exposed RNA helices and preventing re-formation of tertiary contacts.

2.4 DISCUSSION

While DEAD-box proteins have previously been shown to promote global refolding of highly structured RNAs, which can require extensive disruption of tertiary interactions, it was not known how they disrupt RNA tertiary structure. Here, we used single molecule fluorescence to dissect an RNA unfolding process into discrete steps involving losses of tertiary and secondary structure. Together, our results suggest a straightforward mechanism by which DEAD-box helicase proteins can disrupt RNA

tertiary structure (**Figure 2.14**). Even if the protein is pre-associated with the RNA, the helicase core does not actively disrupt tertiary contacts. Instead, it captures RNA helices that become exposed transiently by spontaneous fluctuations. This helix capture process does not require ATP and may result from RNA binding by just one of the two core domains, as closure of the two domains typically requires a bound nucleotide[102,111]. Supporting this idea, domain 2 of the *S. cerevisiae* DEAD-box protein Mss116 was shown recently to bind double-stranded RNA (dsRNA) in the absence of an adenosine nucleotide[111]. Ultimately, closure of the domains and unwinding of the RNA helix commits the ssRNA product strands to form new contacts, allowing refolding to a functional structure or exchange between structures.

This helix capture process is reminiscent of a mechanism described for some processive helicases, termed 'passive unwinding', in which the helicase does not actively disrupt base pairs but instead captures the nucleotides from the terminal base pair upon spontaneous fraying, preventing the base pair from reforming. Processive unwinding can be achieved by this mechanism if the helicase protein repetitively captures the frayed end of the helix while it tracks directionally along one of the strands[195,196]. As each frayed base pair is successively captured, the loss of base stacking is expected to weaken the adjacent base pair, accelerating its fraying and therefore accelerating unwinding[196]. In a conceptually analogous manner, when a DEAD-box protein captures a helix from a structured RNA, it not only destabilizes tertiary structure by preventing reformation of tertiary contacts by the captured helix, but it also weakens all other contacts that form

cooperatively[197–199]. Thus, despite its passive nature, this helix capture mechanism is expected to accelerate the kinetics of large-scale tertiary unfolding of structured RNAs.

This mechanism for unfolding RNA tertiary structure is likely to be used broadly by DEAD-box proteins that function to promote RNA folding, as it relies on their inherent abilities to bind dsRNA and induce ATP-dependent helix unwinding[111], and does not depend on any specific protein binding site or structural context. Previous work showed that CYT-19 can unfold the *Tetrahymena* ribozyme with an efficiency that depends on the overall stability of the RNA[160], and helix capture provides a physical model for this result. Less stable structures are expected to undergo more frequent dynamic fluctuations, allowing for more frequent capture events and therefore more efficient unfolding. Thus, this mechanism allows DEAD-box proteins to sense RNA stability, leading to preferential action on less stable misfolded intermediates, regardless of specific structural features in the misfolded RNAs, while minimizing activity upon stable, natively folded RNA. Consistent with this view, CYT-19 is activated for ATPase activity to a lower extent by the natively folded wild-type *Tetrahymena* ribozyme than by less stable mutants, suggesting fewer productive interactions with the more stable structure[200]. In fact, when protruding helices of the ribozyme are deleted, the ATPase activity of CYT-19 is decreased and small angle X-ray studies have shown that increased stimulation of the ATPase activity of CYT-19 correlates with a decrease in overall compactness of the ribozyme[200]. A corollary of the model is that groups of cellular RNAs that lack stable tertiary structure, such as mRNAs, are potentially subject to persistent unfolding by DEAD-box proteins. Indeed, recent work has shown that cellular mRNAs

are continually remodeled, such that they are less structured on average than they are under standard *in vitro* conditions[201,202]. Furthermore, this remodeling requires ATP[202], highlighting the roles of RNA helicase proteins as general manipulators of RNA structure *in vivo*.

In addition to DEAD-box proteins that function as general RNA chaperones, the helix capture mechanism may also be important for DEAD-box proteins that function more specifically in processes such as assembly of the ribosome and spliceosome[203–205]. In these processes, capture and unwinding of dynamic helices would be expected to promote conformational transitions, while formation of a stable, folded surface would indicate that an RNA folding or protein assembly step has proceeded correctly. Thus, this helix capture mechanism is likely to be used widely by DEAD-box proteins, ranging from those that function as general chaperones of RNA folding to those that promote specific RNA structural transitions in complex biological processes.

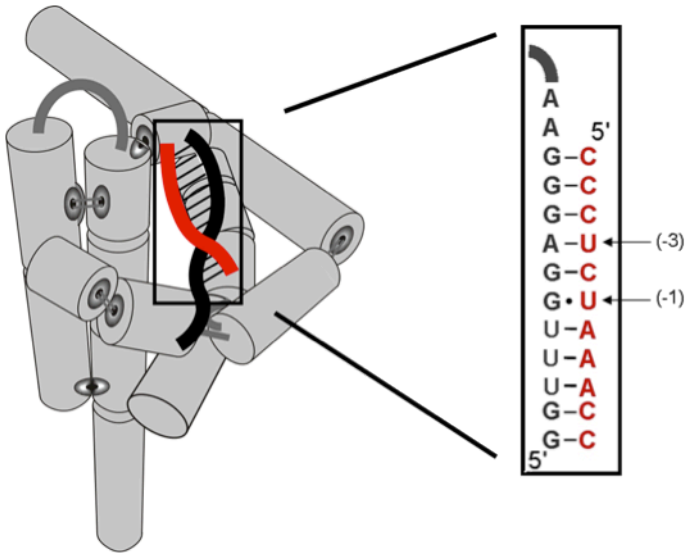

Figure 2.1: Cartoon representation of the *Tetrahymena* group I ribozyme with the P1 helix in the docked position

The secondary structure of the 11-bp P1 helix is shown with positions -1 and -3 indicated by arrows.

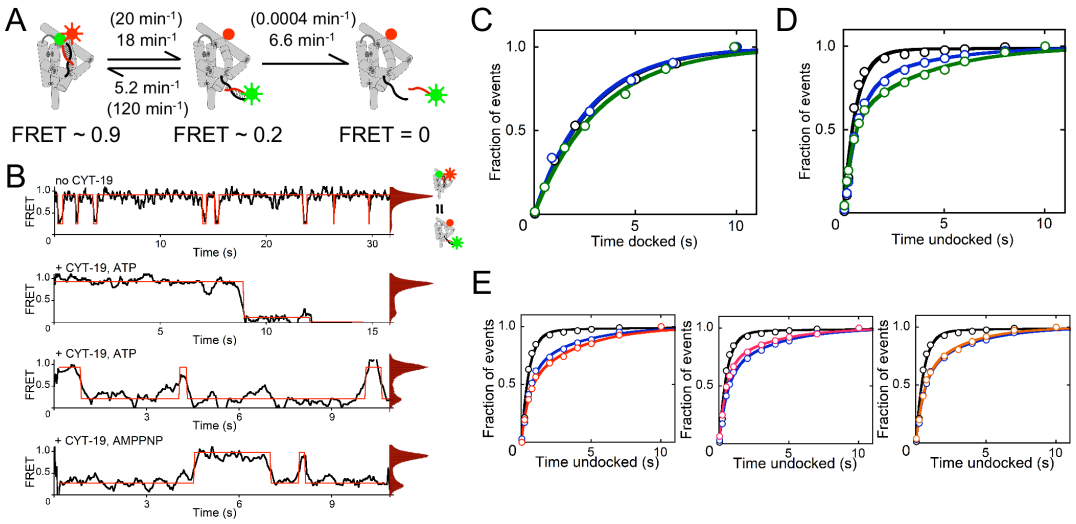

Figure 2.2: CYT-19 destabilizes tertiary docking of the P1 helix into the *Tetrahymena* ribozyme core

(A) Cartoon of the ribozyme showing P1 helix docking, undocking and unwinding rate constants in the presence of CYT-19, with the corresponding rate constants without CYT-19 in parentheses (**Table 2.2**). **(B)** Representative FRET traces and histograms showing reversible docking (transitions shown in red) without CYT-19 (top), with CYT-19 and ATP (middle traces), and with CYT-19 and AMP-PNP (bottom). **(C and D)** Lifetime distributions of the docked (panel C) and undocked (panel D) states without CYT-19 (black) or with 0.5 μM (blue) or 1 μM (green) CYT-19 and 2 mM ATP-Mg^{2+} (**Table 2.2**). **(E)** Lifetime distributions of undocked P1 in the presence of 2 μM CYT-19 with AMP-PNP (red, left plot), without nucleotide (pink, center plot), and with ADP (orange, right plot). In each plot, corresponding data in the absence of CYT-19 and for 2 μM CYT-19 with ATP are shown for comparison in black and blue, respectively.

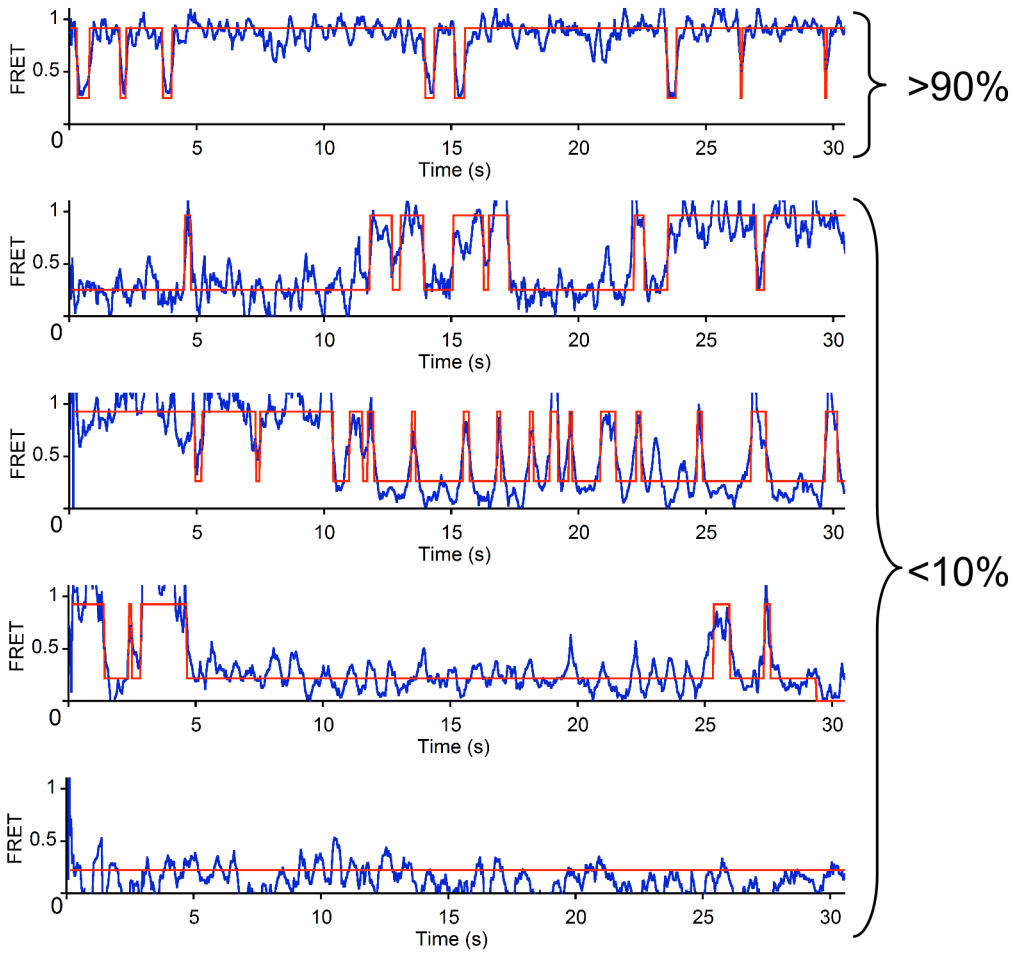

Figure 2.3: Representative FRET traces showing heterogeneous P1 docking behavior in the absence of CYT-19

While most molecules gave behavior as shown in the top FRET trace (>90%), longer undocked dwell times were observed for some molecules (transitions shown in red). Some of these molecules may be misfolded and therefore not support stable docking of P1[180]. In addition, conformational heterogeneity in docking behavior has been previously observed for this ribozyme construct in single molecule experiments[190,206]. As a result of this small population of ribozymes for which the P1 helix does not dock stably (<10%), a minor phase with an increased $\tau_{undocked}$ is observed in the absence of CYT-19 (**Table 2.2**).

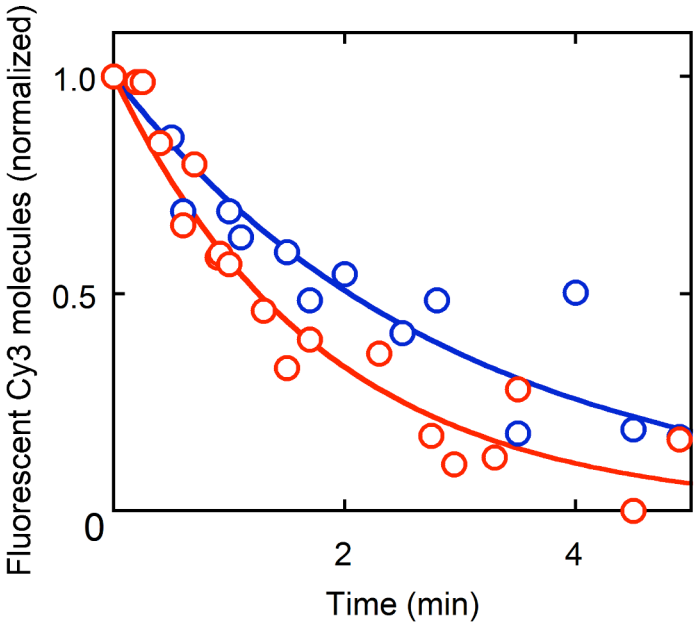

Figure 2.4: Measurement of the rate constant for Cy3 photobleaching

A Cy3-labeled oligonucleotide corresponding to the 'tether' oligonucleotide was immobilized on a PEG-treated slide and excited constantly by the green laser (532 nm) at 15 mW. Photobleaching of Cy3 under our OSS conditions was measured by monitoring the time dependence of Cy3 fluorescence under CYT-19 buffer conditions (see **Section 2.2.6**) in the absence of CYT-19 (blue, 0.34 min^{-1}). Analogous data were collected with 2 μM CYT-19 and AMP-PNP in solution (red, 0.55 min^{-1}) to determine whether these solutes affect photobleaching.

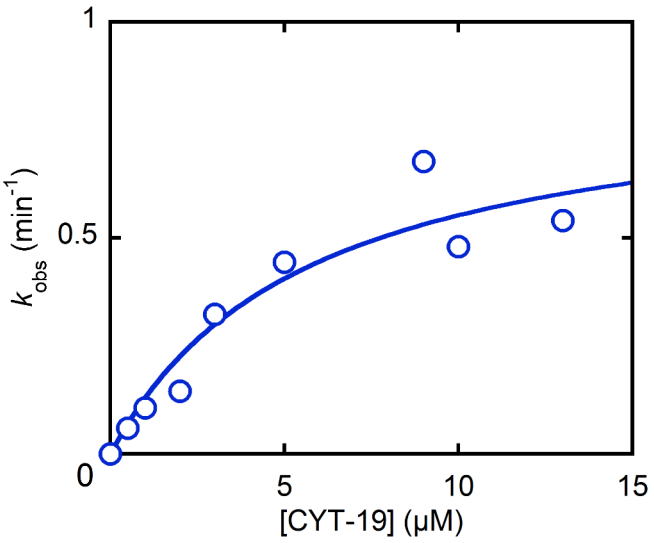

Figure 2.5: Unwinding of the standard 11-bp P1 helix by CYT-19

Observed rate constants for P1 unwinding determined in ensemble measurements are plotted against CYT-19 concentration. The hyperbolic fit gives a second order rate constant of 1.5×10^5 $M^{-1}min^{-1}$ with a maximum unwinding rate constant (k_{max}) of 0.86 min^{-1} and a $K_{1/2}$ value of 5.7 μM. Analogous single molecule measurements, in which the number of remaining substrate molecules was determined over time from multiple fields of view, gave comparable observed rate constants (within 3-5-fold, **Table 2.2**).

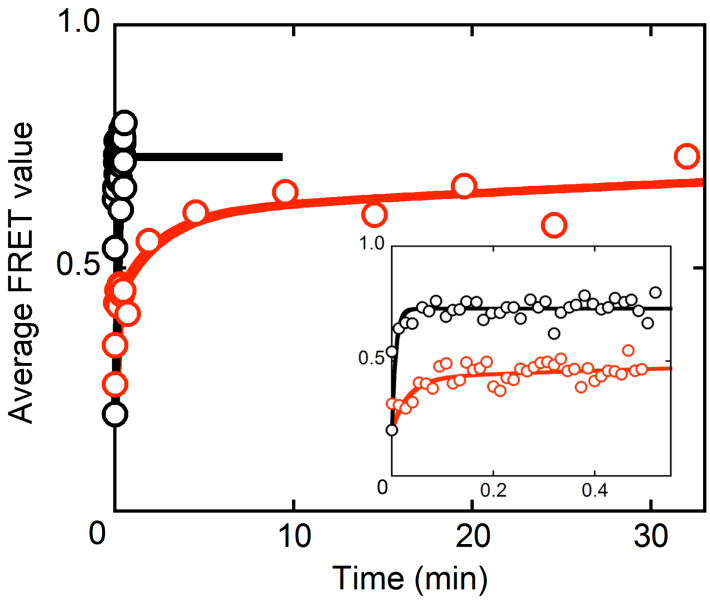

Figure 2.6: CYT-19 dissociation from the ribozyme

After CYT-19 washout, the average FRET value for ribozyme molecules with the 11-bp P1 helix undocked (red, 62 molecules) was fit by three phases with rate constants and relative amplitudes of 30 min^{-1} (0.36), 0.43 min^{-1} (0.29) and 0.01 min^{-1} (0.35). We infer that the rate constant of 0.43 min^{-1} reflects CYT-19 dissociation because this phase was not observed in the absence of CYT-19. The initial fast phase reflects P1 docking re-equilibration with bound CYT-19 and is predicted from the model, while the slowest phase most likely reflects the slow conversion of ribozyme molecules that initially give poor docking or are misfolded (see **Section 2.2.7.4** for further details). In the absence of CYT-19 (black, 64 molecules), re-equilibration of P1 docking gave a single observed phase of 130 min^{-1} (inset).

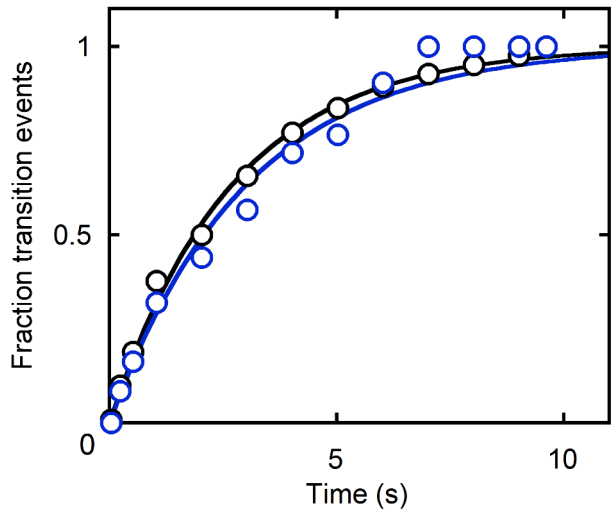

Figure 2.7: From the CYT-19-bound undocked state, the P1 helix can redock into tertiary contacts with the ribozyme core or be unwound by CYT-19

To determine whether these alternative fates arise from a kinetic competition from the same population of undocked molecules or whether they are different populations that are predetermined to undergo one fate or the other, we separately monitored the lifetimes of P1 undocking events that led to redocking or to unwinding. The corresponding rate constants for events that led to redocking (black, 22 min^{-1}) and helix unwinding (blue, 20 min^{-1}) are comparable to each other and to k_{obs} when all of the undocked complexes are considered (~20 min^{-1}; **Figure 2.2D**). Therefore, these results indicate that P1 unwinding and redocking are competing processes that originate from the same initial population of undocked P1.

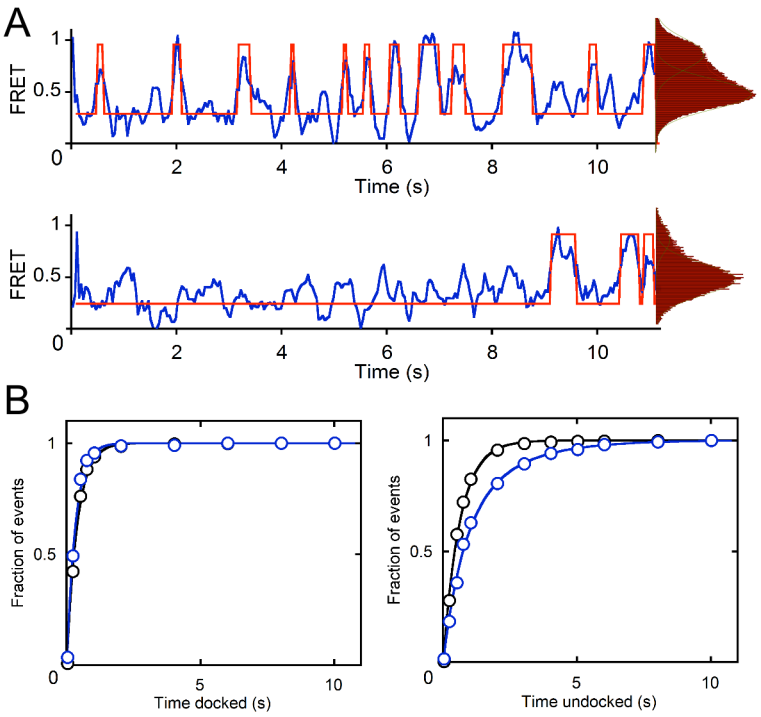

Figure 2.8: Effect of CYT-19 on docking of the 11-bp P1 helix of the *Tetrahymena* ribozyme with $K_{dock} \sim 0.54$

(A) Representative FRET traces (transitions shown in red) and corresponding histograms of the docking equilibrium in the absence of CYT-19 (top) and with 1 μM CYT-19 and 2 mM ATP-Mg^{2+} (bottom) for a P1 helix formed with the oligonucleotide –1m,rSA_3C_2 (see **Table 2.1**). **(B)** Lifetime plots for docked and undocked P1 in the absence of CYT-19 (black) and with 1 μM CYT-19 and 2 mM ATP-Mg^{2+} (blue). Values for the docking rate and equilibrium constants are given in **Table 2.2**.

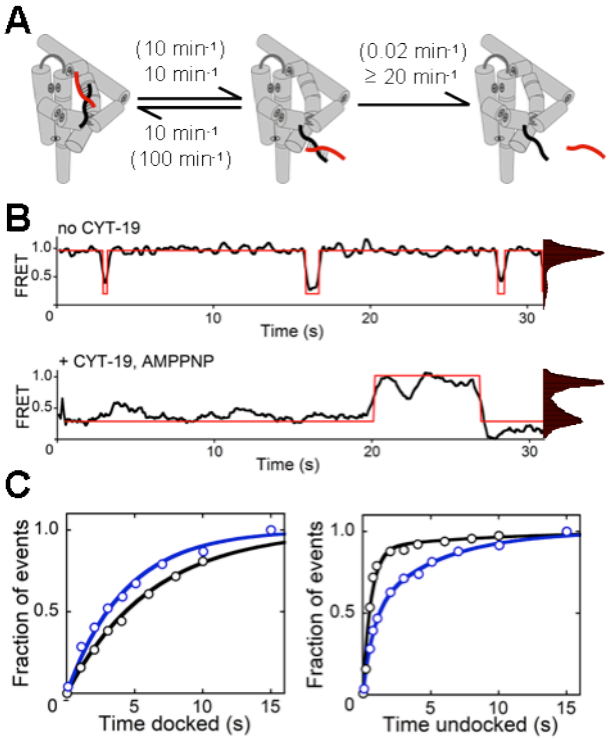

Figure 2.9: CYT-19-mediated unwinding of a shorter P1 helix (6 bp) is rate limited by spontaneous loss of tertiary contacts

(A) Cartoon representation showing docking, undocking and unwinding rate constants for the 6-bp P1 helix in the presence of CYT-19. Rate constants in the absence of CYT-19 are shown in parentheses and are similar to previous values[189]. **(B)** Representative FRET traces and histograms (transitions shown in red) in the absence of CYT-19 (top) and with 1 μM CYT-19 and AMP-PNP (bottom). **(C)** Lifetime distributions of the docked (left) and undocked (right) states in the absence of CYT-19 (black, 102 molecules) and with 1 μM CYT-19 and AMP-PNP (blue, 163 molecules).

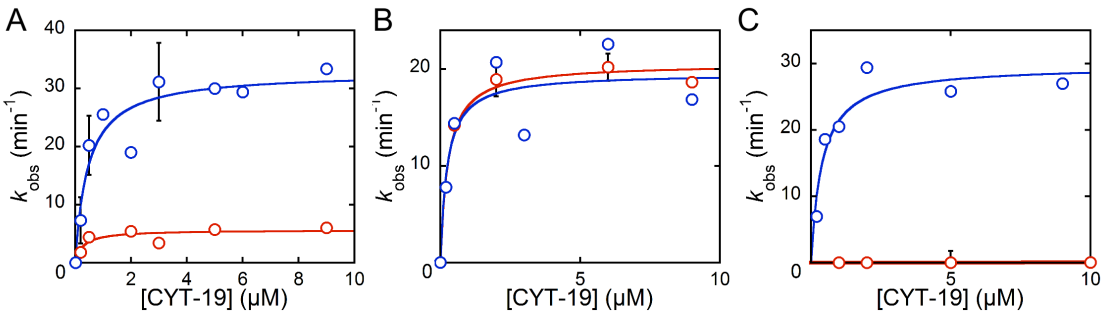

Figure 2.10: CYT-19-mediated unwinding of the 6-bp P1 helix is rate limited by spontaneous undocking of P1

To verify that the observed correlation between the maximum P1 unwinding rate and the undocking rate is due to P1 docking stability, ensemble experiments were performed with the native *Tetrahymena* ribozyme and its long-lived misfolded conformer, which does not stably dock the P1 helix[180]. (See **Table 2.1** for sequences and properties of substrate oligonucleotides.) **(A)** The CYT-19 concentration dependence for unwinding the 6-bp P1 helix formed with substrate –1d,rSA$_5$ by the native ribozyme shows a maximum unwinding rate constant (k_{max}) of 6 min^{-1} (red), which is comparable to the intrinsic undocking rate constant measured in single molecule experiments (**Figure 2.9**). When docking is inhibited by misfolding the ribozyme (blue), k_{max} is increased to ~30 min^{-1}. **(B)** With a substrate for which P1 docking is blocked by replacement of a 2′-hydroxyl group with a 2′-O-methyl group (–3m,rSA$_5$), the undocked state predominates and CYT-19-mediated unwinding is accelerated, with no difference between the native ribozyme (red) and the misfolded ribozyme (blue). We infer that the lower value for the k_{max} of this substrate compared to the standard substrate (–1d,rSA$_5$, **Figure 2.10A**) reflects an effect of the methoxy substitution on CYT-19-mediated unwinding. **(C)** CYT-19-mediated unwinding of the P1 duplex containing the 6-nt product (rP), which docks much more strongly than the helix formed with the standard substrate. As above, results from the native and misfolded ribozyme species are shown in red and blue, respectively. Error bars represent the standard deviation of at least two independent measurements.

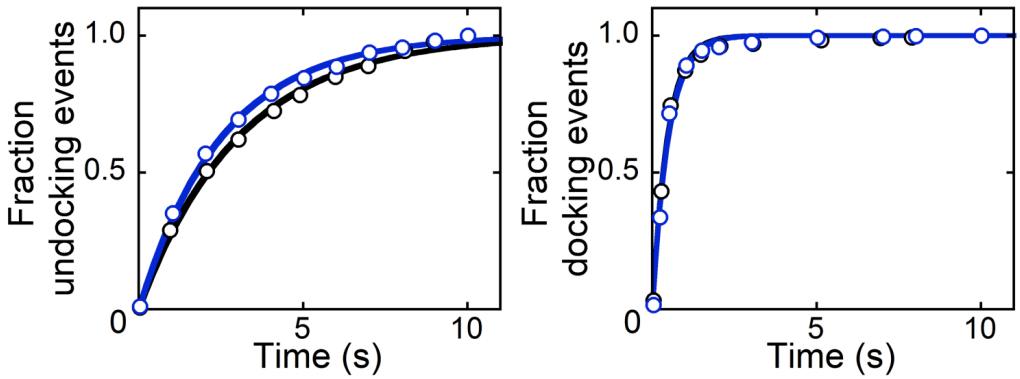

Figure 2.11: A C-tail truncation mutant of CYT-19 does not stably capture the P1 helix

Lifetime distributions of the docked (left) and undocked (right) states in the absence of CYT-19 (black, 103 molecules) and with 1 μM CYT-19 and ATP (blue, 138 molecules).

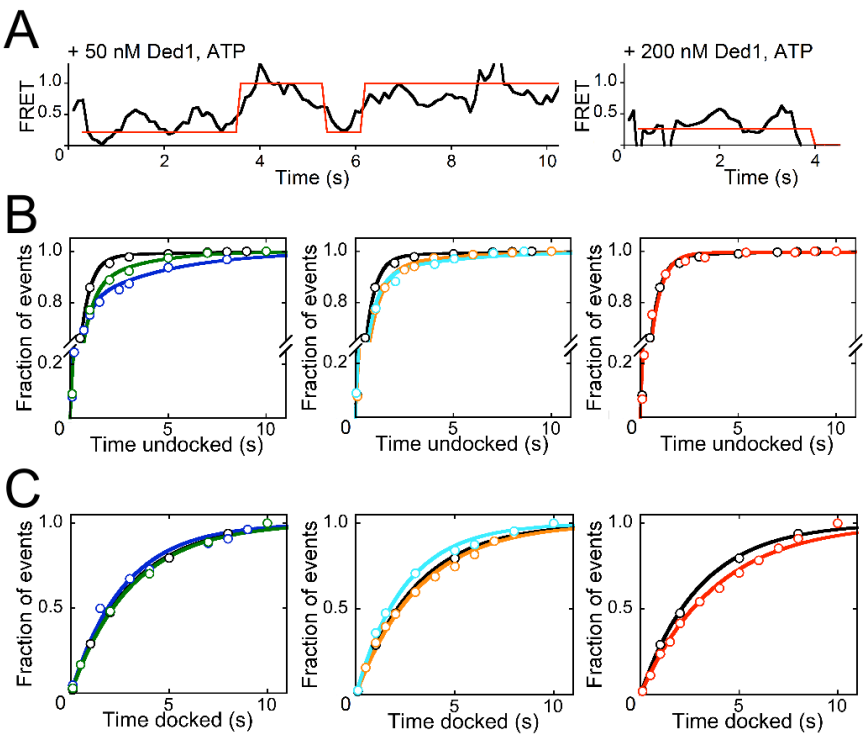

Figure 2.12: Ded1 destabilizes docking of the P1 helix

(A) Representative FRET traces showing extended undocked lifetimes before redocking (left) and unwinding (right) in the presence of Ded1 and ATP (transitions shown in red). **(B)** Lifetime plots of the undocked states in the absence of Ded1 (black, all panels), with 50 nM (blue) or 0.2 μM (green) Ded1 and 2 mM ATP (left panel), with 0.1 μM (cyan) or 0.9 μM (orange) Ded1 and 2 mM AMP-PNP (center panel), and with 0.9 μM Ded1 and no nucleotide (red, right panel). **(C)** Lifetime plots of the docked state of P1 under the same conditions and represented by the same color scheme as (B). The calculated k_{dock} and k_{undock} values for each condition are listed in **Table 2.4**.

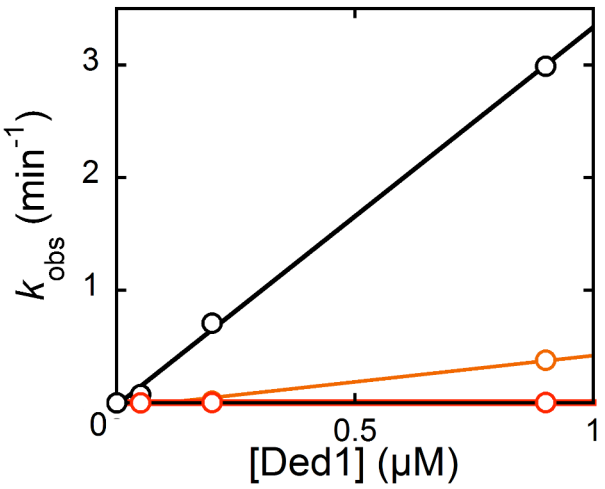

Figure 2.13: Unwinding of the standard 11-bp P1 helix by Ded1

Ded1 unwinds the P1 helix in the presence of 2 mM ATP with a second order rate constant of 3.4×10^6 $M^{-1}min^{-1}$ (black). Secondary structure disruption by Ded1 is reduced in the presence of 2 mM AMP-PNP (orange, 4.6×10^5 $M^{-1}min^{-1}$), and without nucleotide (red, 2.8×10^3 $M^{-1}min^{-1}$).

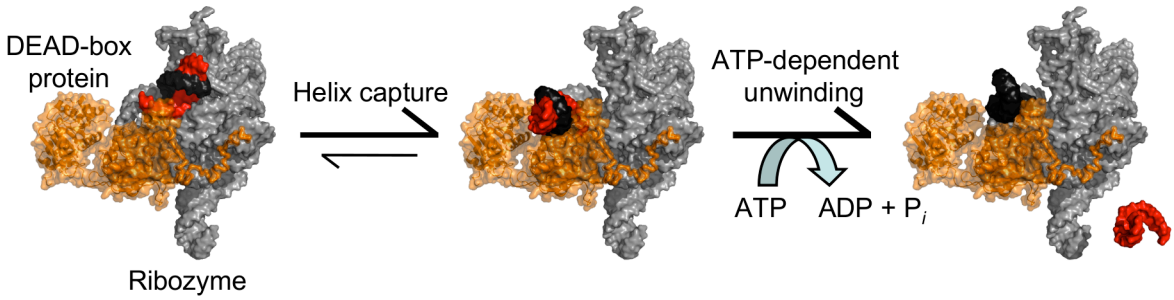

Figure 2.14: Model for RNA tertiary structure disruption by helix capture

DEAD-box proteins (orange) associate with structured RNAs non-specifically (left), which can result in the helicase core being positioned to interact with transiently exposed helices (center). This interaction prevents reformation of tertiary contacts by the bound helix, destabilizing the RNA tertiary structure and allowing DEAD-box proteins to use ATP to perform helix unwinding (right).

oligonucleotide	properties	sequence
-1d,rSA_5	minimizes 5' cleavage	CCCUCdUAAAAA (with and without 3' Cy3)
-3m,rSA_5	disrupts P1 docking	CCCmUCUAAAAA (with and without 3' Cy3)
rP	strengthens P1 docking	CCCUCU
-1d,rSA_3C_2	minimizes 5' cleavage	CCCUCdUAAACC (with and without 3' Cy3)
-1m,rSA_3C_2	destabilizes P1 docking	CCCUCmUAAACC (with and without 3' Cy3)
-3m,rSA_3C_2	disrupts P1 docking	CCCmUCUAAACC (with and without 3' Cy3)
tether-Cy5	immobilizes ribozyme to slide	biotin-d(TGTGTAAGTTTTAGGTTGATTTTG)-Cy5

Table 2.1: Sequences and properties of oligonucleotides used in ensemble and single molecule experiments

In order for the P1 helix to be visualized with smFRET, the indicated oligonucleotides were labeled on their 3' end with Cy3 dye and the DNA tether was labeled with its FRET pair, Cy5. For the first two oligonucleotides, Cy3 replaces the 3' nucleotide (resulting in –1d, rSA_4-Cy3 and –3m,rSA_4-Cy3).

Table 2.2

substrate	[CYT-19] (μM)	nucleotide	# molecules	K_{dock}	k_{undock} (min^{-1}) (amp)[a]	$\tau_{undocked\ state}$ (s) (amp)[a]	k_{dock} (min^{-1})	$k_{obs,unwind}$ (min^{-1})
-1d,rSA_3C_2	0		2448	8.4	20	0.52 (0.92) 4.3 (0.08)	120	0.0004[b]
	0.5	ATP	963	4.7	22	0.55 (0.62) 3.0 (0.38)	7.7	0.16±0.02
	0.5	AMP-PNP	545	4.5	17	0.56 (0.86) 2.9 (0.14)	8.6	0.007
	1	ATP	207	2.4	17	0.53 (0.58) 4.0 (0.42)	4.0	0.54±0.19
	1	AMP-PNP	862	2.8	17	0.50 (0.37) 4.2 (0.63)	3.6	0.07±0.03
	1	ADP	244	2.3	25	0.50 (0.86) 2.9 (0.14)	15	0.005
	1	no nt	189	1.7	22	0.53 (0.82) 3.2 (0.18)	10	0.014
	2	ATP	217	0.99	18	0.52 (0.54) 3.2 (0.46)	4.0	0.78±0.09
	2	ADP	206	4.1	16	0.43 (0.47) 2.7 (0.53)	9.3	0.03
	2	no nt	307	2.1	16	0.47 (0.63) 3.3 (0.37)	9.9	0.06
-1m,rSA_3C_2	0	ATP	293	0.54	150	0.62	97	0.0013[b]
	1	ATP	120	0.14	180	0.73 (0.58) 2.0 (0.42)	19	0.44
-3m,rSA_3C_2	0	ATP	111	0.22	190	0.79	76	
	2	ATP	170	0.03	204 (0.98) 4.8 (0.02)	0.80 (0.90) 3.8 (0.10)	6.5	0.63

Table 2.2: P1 docking kinetics and equilibria for the 11-bp P1 helix

Values were determined in single molecule fluorescence experiments except where indicated. The slow phase for P1 docking in the absence of CYT-19 is attributed to heterogeneous P1 docking behavior (**Figure 2.3**). The docking rate constant in the presence of CYT-19 (k_{dock}) was calculated as described in **Section 2.2.7.2** of the **Materials and Methods**. Except where indicated, the observed rate constant for unwinding ($k_{obs,unwind}$) was determined by single molecule fluorescence by monitoring the disappearance of substrate from the ribozyme over time, using multiple fields of view. Thus, $k_{obs,unwind}$ reflects the overall rate constant for the two-step process of undocking and helix unwinding. See **Table 2.1** for sequences and effects of each substrate. [a]Relative amplitudes for each phase of the docking kinetics were determined from the fit of the undocked lifetimes normalized by total number of transition events and are listed in parenthesis. [b]Rate constants for P1 unwinding in the absence of CYT-19 were measured in ensemble experiments.

		Redocking	Unwinding	Truncation by shutter	Total
[CYT-19] (μM)	**# molecules**	**k_{dock} (min^{-1}) (fraction)**	**k_{unwind} (min^{-1}) (fraction)**	**$k_{truncation}$ (min^{-1}) (fraction)**	**k_{obs} (min^{-1}) (fraction)**
0.5	945	7.7 (0.38)	6.8 (0.37)	4.9 (0.25)	20 (1)
1	206	4.0 (0.27)	3.3 (0.25)	7.2 (0.48)	15 (1)
2	217	4.0 (0.21)	9.7 (0.55)	4.5 (0.24)	19 (1)

Table 2.3: Rate constants of the various "fates" of undocked P1 helix

In the presence of CYT-19, the undocked P1 helix may redock or unwind. Additionally, the fluorescence signal may be artificially truncated by the shuttering of the excitation laser. For each CYT-19 concentration, the fractions of undocking events that ended with redocking, unwinding or were truncated by the shutter were determined and the corresponding rate constants (k_{dock}, k_{unwind} and $k_{truncation}$, respectively) were calculated by multiplying the observed rate constant (k_{obs}) by the probabilities of each outcome (see **Section 2.2.7.2** for details). To determine the unwinding rate constant (k_{unwind}), the calculated rate constant reflecting disappearance of Cy3 was further corrected by subtracting the rate constant for Cy3 photobleaching, as measured independently ($k_{photobleach} = 0.55\ min^{-1}$, **Figure 2.4**). Values reported in the text as the fraction of events that ended in unwinding or redocking express these outcomes relative to each other, *i.e.* normalized to 100%.

[Ded1] (μM)	nucleotide	# molecules	k_{dock} (min^{-1}) (amp)[a]	k_{undock} (min^{-1})
0.05	ATP	292	11 (0.27) 160 (0.73)	16
0.1	AMP-PNP	143	20 (0.14) 109 (0.86)	18
0.1	no nt	1337	105	17
0.2	ATP	397	20 (0.20) 120 (0.80)	19
0.9	AMP-PNP	323	20 (0.23) 130 (0.77)	24
0.9	no nt	125	116	16

Table 2.4: Docking kinetics for the 11-bp P1 helix as measured by single molecule fluorescence in the presence of Ded1 and the indicated nucleotides

The docking rate constant in the presence of Ded1 (k_{dock}) was calculated as for CYT-19 (see **Section 2.2.7.2**). [a]Amplitudes for each phase of the docking kinetics, listed in parentheses, were determined from the fit of the undocked lifetimes and normalized by the total number of transition events.

Appendix: Detection of expanded CGG repeats in Fragile X syndrome

A.1 INTRODUCTION

Fragile X syndrome is the most common cause of inherited mental retardation and is directly linked to expansion of CGG repeats within the *FMR1* gene[207]. An individual with 6-54 CGG repeats manifests none of the phenotypes characteristic of Fragile X syndrome such as speech delay, cognitive dysfunction and physical abnormalities. Carriers with repeats of 55-200 ("pre-mutation") are at risk for Fragile X tremor ataxia syndrome (FXTAS) and may develop Fragile X syndrome phenotypes later on in life[208,209]. The presence of more than 200 CGG repeats disrupts protein production of the Fragile X mental retardation protein (FMRP), which is essential for neural development during embryogenesis[207]. Progression from a "grey zone" of 45-54 CGG repeats to full mutation can occur within two generations[210,211].

Current diagnostic assays relay primarily on time-intensive lab techniques such as Southern blotting, and PCR and capillary electrophoresis methods have greatly improved the accuracy and range of testing[212–217]. Recently, research has focused on development of a rapid, high throughput assay to identify grey zone and pre-mutation patients using various single molecule methods[218–228]. The goal of our studies was to first optimize annealing of dye-labeled probes to repeat regions and then to accurately quantify the number of repeats. Single molecule fluorescence techniques would allow for fast analysis of DNA samples of heterogeneous populations of repeat lengths and even screen for

individual heterozygous patients that have one allele in the pre-mutation or full mutation range.

A.2 Materials and Methods

A.2.1 Materials

Dye-labeled and biotinylated DNA (including single-stranded model DNA targets, probes and primers) were purchased from IDT (Coralville, IA). The sequences for these oligonucleotides are listed in **Table A.1**.

A.2.2 Preparation of CGG repeat templates

DNA templates containing 23, 56, 96, 118 and 340 CGG repeats were prepared as PCR products from cell-line DNA with a master mix containing GC-rich Amp buffer, *FMR1* primers and GC-rich polymerase mix (Asuragen)[217]. The sequences for the primers used are listed in **Table A.1**.

The forward primer used to prepare single-stranded CGG template was biotinylated at the 5' end for immobilization on magnetic beads and the reverse primer was 5' phosphorylated for lambda exonuclease digestion (**Figure A.1A**). For double-stranded repeat templates used for single molecule electrokinetic stretching measurements, the forward primer was both biotinylated and labeled with Cy3 dye and the reverse primer was labeled with Cy5 dye (**Figure A.1B**). In addition, DNA rulers of defined lengths were designed as detailed in ref. 229 to calibrate the electrokinetic stretching method, which was performed by Dr. Brian Cannon.

The PCR protocol for amplification of these DNA templates is as follows: incubation at 98 °C for 4 min followed by 30 cycles of 97 °C at 35 sec, 64 °C for 35 sec, and 68 °C for 4 min with an additional 20 sec added for each of the last 20 cycles. The samples were then incubated at 68 °C for 10 min and stored at 4 °C. The PCR products were purified (Qiagen) and eluted with water. To prepare the ssDNA targets, the PCR products were then subject to digestion by lambda exonuclease at 37 °C for 15 min and the enzyme was inactivated by incubation at 75 °C for 10 min (see **Figure A.1A**). Sizes of templates were confirmed by gel electrophoresis on 1.5% agarose gels.

A.2.3 Hybridization of dye-labeled probes to model repeat templates

The DNA templates were annealed with Cy3-labeled probes and Cy5-reference probes (at ~100X excess) by incubation at 98 °C for 5 min, slow cooling to 50 °C over 30 min, incubation at 78 °C for 5 min, then slow cooling to 45 °C for 30 min. The hybridization buffer was made up of 20 mM Tris, 100 mM or 500 mM NaCl at pH 7.5. Some experiments were performed with the addition of 10 mM Mg^{2+}.

A.2.4 Cy3-dCTP incorporation into CGG target DNA

Cy3-labeled dCTP was added to the PCR protocol outlined in **Section A.2.2** (minimum final concentration of dye-labeled dCTP in PCR reaction was 19 μM). After purification (Qiagen), the sizes of PCR products were confirmed by gel electrophoresis on 1.5% agarose gels.

A.2.5 Measurement of fluorescence intensity and analysis of CGG repeat length using ensemble methods

DNA templates that were dye-labeled either by Cy3-dCTP incorporation or with annealed probes were incubated with excess streptavidin-coated magnetic beads (MyOne Strepavidin C1 Dynabeads, 400 pmol/μL binding capacity) at room temperature for 30 min (**Figure A.1A**). The beads were pulled down with a magnet and washed with the appropriate hybridization buffer (see above) at least 5 times. The fluorescence intensity of the DNA targets directly labeled with Cy3 was then measured with a fluorescence plate reader (POLARstar). The number of Cy3 dyes incorporated was determined from the fluorescence intensity and normalized by the calculated amount of DNA in solution. The length of the repeat region can then be calculated from the number of Cy3 per DNA target (with an additional correction for the decreased probability of Cy3 incorporation with increased repeat length).

After pull down by the magnetic beads, the DNA targets that were hybridized to Cy3-labeled probes were additionally incubated with 0.1 M NaOH at 95 °C for 5 min to release the dye-labeled probes into solution for measurement by fluorescence plate reader. Dye-labeled probes were also incubated with the magnetic beads in the absence of biotinylated template and washed in parallel to measure nonspecific binding onto the streptavidin-coated beads. A Cy5-labeled reference probe was also annealed to determine the amount of DNA in solution. The number of Cy3-labeled probes bound per target DNA was then calculated from the bulk Cy3 fluorescence intensity of each DNA sample

normalized by the Cy5 fluorescence signal. Calibration curves for fluorescence intensity were made for each plate by serial dilution of the appropriate Cy3 and Cy5 probes used.

A.2.7 Single molecule fluorescence measurements of DNA templates

The microscope setup for single molecule data acquisition is described in **Section 2.2.4**. Biotinylated DNA templates with annealed probes or incorporated Cy3-dCTP were immobilized onto PEG-treated microscope slides (see **Section 2.2.5**). The sample chamber was excited constantly at 532 nm and the Cy3 fluorescence intensity was monitored over time in order to visualize Cy3 photobleaching of single molecules. The number of photobleaching steps corresponds to the number of probes bound to each target DNA.

A majority of the photobleaching experiments as well as all results from the electrokinetic stretching method of length measurement was performed by Dr. Brian Cannon[229].

A.3 RESULTS

Accurate measurement of the repeat length is important for distinguishing between different stages in the progress of Fragile X syndrome. We employed multiple approaches to quantify the number of CGG repeats present on the DNA sample. Cy3 dye was used to visualize the CGG repeat region by either direct incorporation of Cy3-dCTP with PCR or hybridization of the DNA target with Cy3-labeled probes complementary to

the repeat region. In either case, the fluorescence intensity of Cy3 was expected to correspond to the length of the repeat region.

A.3.1 Incorporation of Cy3-dCTP

First, we labeled the CGG repeat region with Cy3 dye by amplifying the DNA target with Cy3-dCTP and then measured the fluorescence intensity of these samples with both a fluorescence plate reader and single molecule microscopy. As expected, longer repeat lengths resulted in higher fluorescence. However, the efficiency of Cy3-dCTP incorporation decreased with increased number of CGG repeats (**Figure A.2A**). The length measurements using ensemble and single molecule methods were in good agreement with each other as well as with simulations that took into account the reduced probability of dye incorporation (**Figure A.2B** and **C**). Nevertheless, the large expected variability in fluorescence intensity indicated that this approach to CGG repeat length determination would not yield accurate or precise results.

A.3.2 Hybridization of Cy3-labeled probes to CGG repeat region

Our second approach relied on targeting Cy3-labeled oligonucleotides for hybridization with specific repeat regions. The CGG repeat length could then be calculated from the fluorescence intensity of the annealed probes, which were composed of 5 repeats of CCG (**Table A.1**). Optimal annealing of Cy3-labeled oligonucleotides with their DNA targets is therefore a crucial component for developing an accurate diagnostic assay for Fragile X syndrome. Model DNA targets with 5, 10, 15, 20 and 25

CGG repeat regions were used to optimize hybridization conditions as well as establish the accuracy and precision of both ensemble and single molecule fluorescence measurements.

Bulk fluorescence measurements of the model templates annealed with Cy3-probes in 20 mM Tris (pH 7.5) and 100 mM NaCl demonstrated that this approach yielded relatively accurate measurements of the length of the repeat region (**Figure A.3**). However, DNA targets with greater repeat lengths (>10 binding sites), prepared from cell-line DNA, showed incomplete hybridization of the repeat region and varying the monovalent and divalent ion conditions in the hybridization buffer did not result in significantly improved probe annealing (**Figure A.4**). Furthermore, single molecule fluorescence measurements of model DNA targets hybridized with Cy3-labeled probes revealed a high degree of heterogeneity for probe annealing for longer repeat regions (**Figure A.5A,** see also ref. 229). For these experiments, the number of Cy3-labeled probes annealed was determined by observing photobleaching of single molecules of the DNA target, with the number of photobleaching steps corresponding to the number of probes hybridized to the repeat region. Probe annealing heterogeneity may reflect the possible formation of intramolecular structure within the CGG repeat region, which would decrease the binding sites available for probe hybridization[230–233]. For these experiments, increasing the divalent or monovalent ion concentration did allow a greater fraction of the probes to anneal to some molecules with longer repeat regions. However, the heterogeneity observed was not significantly affected (**Figure A.5B**).

The fluorescence of cyanine dyes such as Cy3 is known to be sensitive to their nucleobase environment[234–236]. The Cy3 dye molecule attached to the end of the probe could interact with the terminal nucleotide, which may reduce its fluorescence emission and additionally, may interfere with hybridization to the DNA target. Therefore, we then switched to an internally dye-labeled oligonucleotide (**Table A.1**) and found that these probes had increased hybridization fidelity (**Figure A.6**).

A.3.3 Dual-mode single molecule fluorescence assay

An assay developed by Dr. Brian Cannon, in collaboration with Asuragen, Inc., using Cy3-labeled probe hybridization paired with length measurements acquired with an electrokinetic stretching method[229]. The single-stranded target DNA for these experiments included a specific sequence at one end of the CGG repeat region that was hybridized with a Cy5-labeled reference probe (**Table A.1**). The number of Cy3-labeled probes hybridized to CGG repeat regions was quantified by fluorescence intensity with calculations analogous to those described above (and described fully in ref. 229). In addition, by stretching the DNA target across the sample chamber with an electric field, the physical distance was measured between the end of the DNA target tethered to the microscope slide and the Cy5 dye on the opposite end of the repeat region (methods and calculations detailed in ref. 229). By correlating these parallel and independent measurements of each individual DNA target, the results were biased to exclude fluorescence measurements where only partial hybridization occurred and led to confident length evaluations of CGG repeats[229]. For a homogeneous sample of $(CGG)_{118}$

DNA targets, ~40% of observed molecules gave good agreement between the dye-counting and distance measurements[229]. These independent measurements of the repeat region allowed screening of samples from heterogeneous populations of DNA targets (see **Figure A.7**) and may be adapted for other disorders where characterization of specific repeat sequences are important for diagnosis.

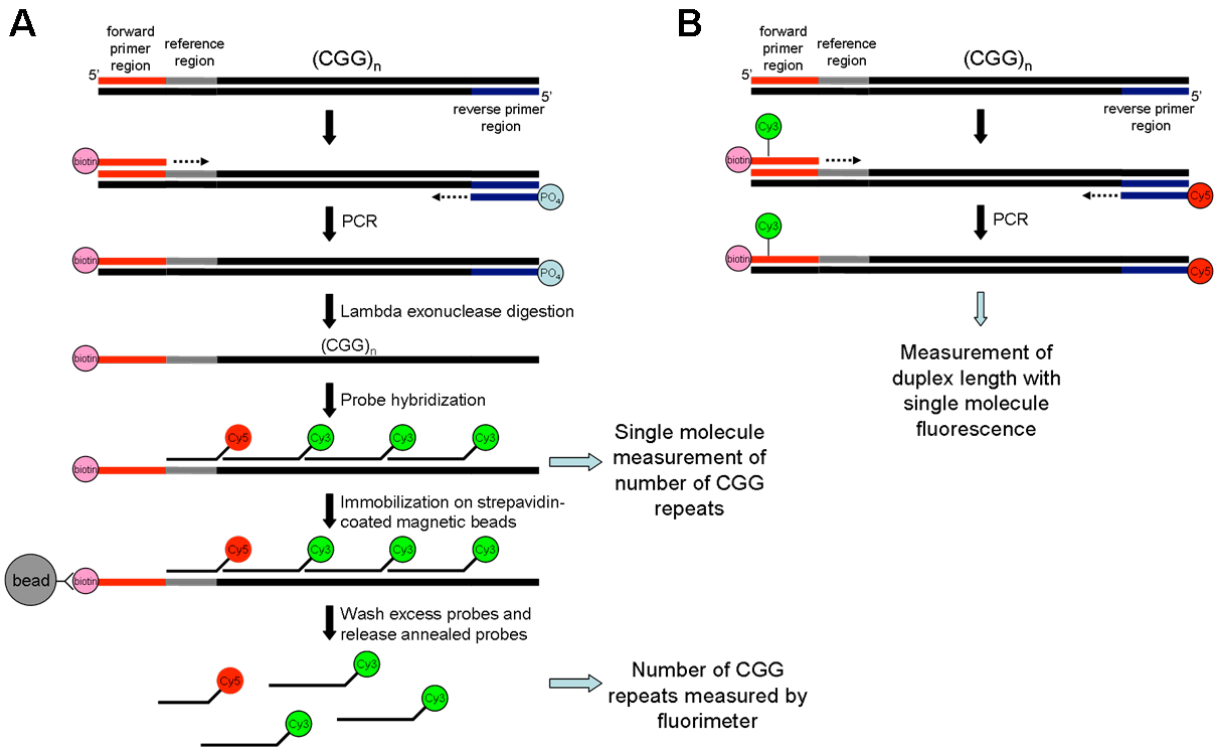

Figure A.1: Schematic outlining the preparation of the various DNA templates for both single molecule and bulk fluorescence experiments

(A) Digestion of the phosphorylated DNA strand by lambda exonuclease resulted in single-stranded target repeats used in probe-annealing experiments. **(B)** Double-stranded DNA templates were prepared with Cy3 and Cy5 dyes on opposite ends to directly measure the length of the repeat region using electrokinetic stretching methods[229].

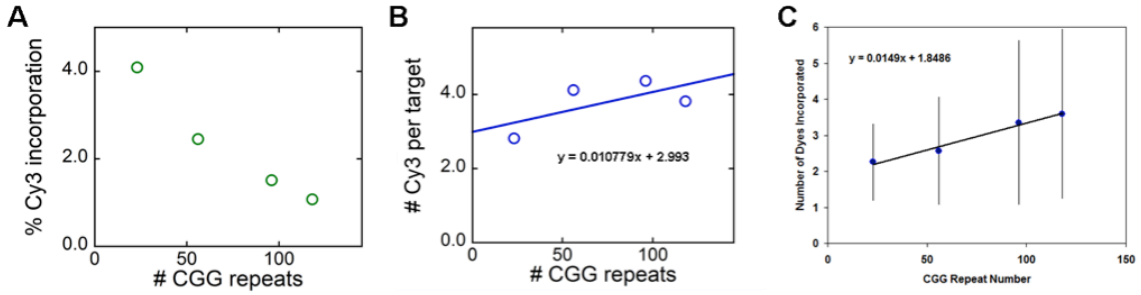

Figure A.2: Incorporation of Cy3-dCTP into DNA targets with varying CGG repeat lengths

(A) The efficiency of Cy3 incorporation into the DNA targets decreased as the number of CGG repeats increased. **(B)** Ensemble measurement of the fluorescence of incorporated Cy3-dCTP increases with CGG repeat length. **(C)** Single molecule measurement of fluorescence intensity is in good agreement with ensemble results and both are well within the expected spread of variability (shown as vertical bars).

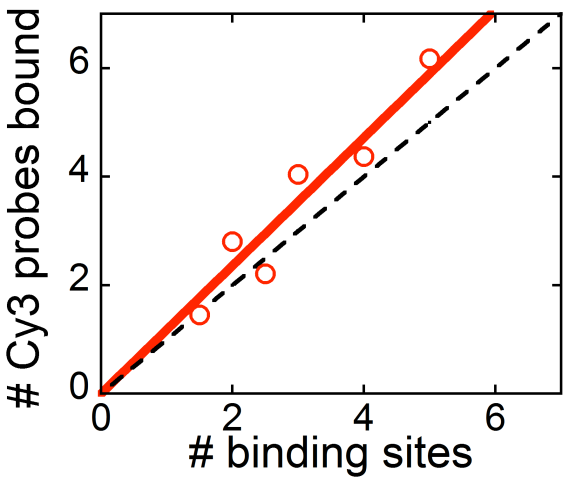

Figure A.3: Optimized hybridization of Cy3-labeled $(CCG)_5$ probes to model DNA templates as measured by bulk fluorescence

Model DNA templates with regions of 5, 10, 15, 20 and 25 CGG repeats were probed with 3' Cy3-labeled oligonucleotides. The hybridization buffer was made up of 20 mM Tris, 100 mM NaCl at pH 7.5. The number of Cy3 probes annealed was calculated by normalizing the signal for each of the model templates by that for the model template with one binding site (i.e. 5 CGG repeats). The black dashed line represents ideal hybridization.

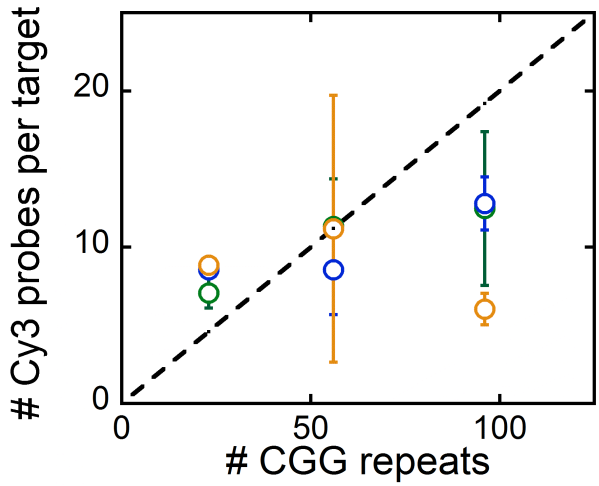

Figure A.4: Hybridization of Cy3-labeled $(CCG)_5$ probes to ssDNA templates prepared from cell-line DNA

DNA targets with 23, 56 and 96 CGG repeats were probed with 3' Cy3-labeled oligonucleotides. The salt conditions of the hybridization buffer (20 mM Tris at pH 7.5) were varied: 100 mM NaCl (green), 500 mM NaCl (blue) and 100 mM NaCl with 10 mM $MgCl_2$ (orange). The black dashed line represents ideal hybridization.

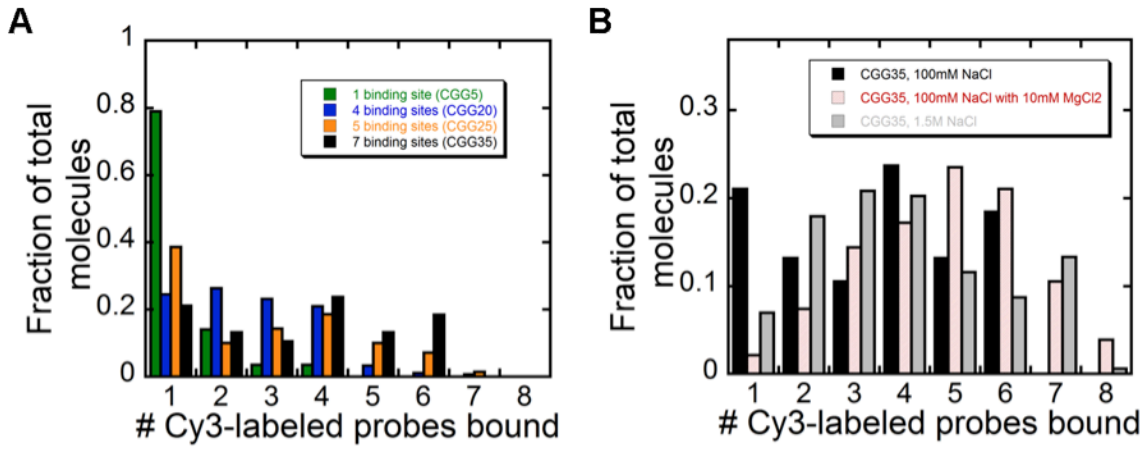

Figure A.5: The number of Cy3-labeled probes bound to target DNA determined using single molecule fluorescence.

(A) The probes were annealed to model DNA templates in 20 mM Tris (pH 7.5) and 100 mM NaCl. Disappearance of Cy3 fluorescence was monitored by single molecule microscopy with single-step photobleaching as a readout for the number of Cy3 dyes associated with each molecule. **(B)** Addition of 10 mM $MgCl_2$ or increasing the monovalent salt concentration improved probe hybridization to longer repeat regions but had little effect on annealing heterogeneity.

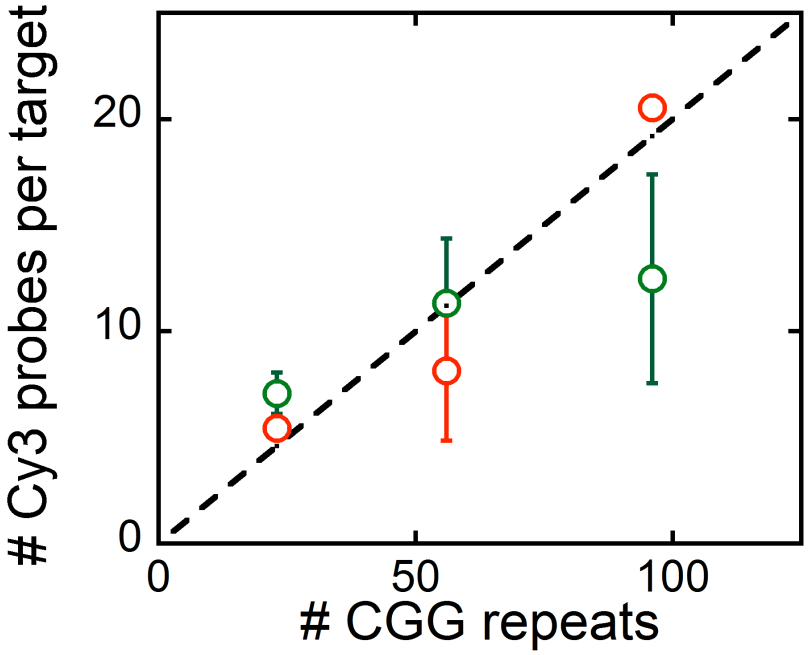

Figure A.6: Hybridization of $(CCG)_2$-Cy3-$(CCG)_3$ probes to ssDNA templates prepared from cell-line DNA.

DNA targets with 23, 56 and 96 CGG repeats were probed with oligonucleotides internally labeled with Cy3 (red) and compared to the 3' end labeled probes (green). The probes were annealed in 20 mM Tris buffer (pH 7.5) with 100 mM NaCl. The black dashed line represents ideal hybridization.

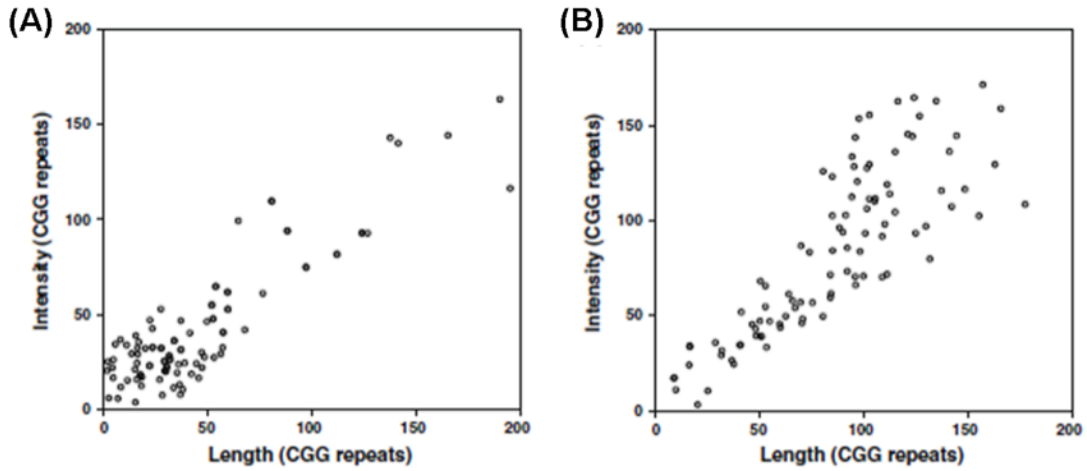

Figure A.7: Scatter plots of the correlated dual-mode measurements for the heterogeneous mixtures of DNA targets

(A) A 3:1 mixture of $(CGG)_{23}$ to $(CGG)_{118}$ DNA targets. 103 of 186 molecules analyzed (55.4%) had correlated measurements of the CGG repeat size. **(B)** A 1:3 mixture of $(CGG)_{23}$ to $(CGG)_{118}$ DNA targets. 98 of 223 molecules analyzed (43.9%) had correlated measurements of the CGG repeat size. Data that gave uncorrelated measurements were rejected from further analysis and are not shown in the plots. These experiments were performed by Dr. Brian Cannon and originally published in Cannon, et al (2012).

Model DNA targets:
5’ Biotin-AMeY-$(CGG)_n$, where n=5, 10, 15, 20, 25
Probes:
3’ Cy3-$(CCG)_5$
3’ $(CCG)_2$-Cy3-$(CCG)_3$
3’ Cy5-TGA GAA ACC ACT TTA TTT GG (“Cy5-AMeY”)
3’ Cy5-CAT CTT CTC TTC AGC CCT GCT A (“Cy5-ref”)
Primers:
For biotinylated ss $(CGG)_n$, where n=23, 56, 96, 118, 340, including a sequence complementary to the Cy5-ref probe:
5’ Biotin-GTC AGG CGC TCA GCT CCG TTT CGG TTT CAC TTC CGG T (forward primer)
5’ PO_4-AGC CCC GCA CTT CCA CCA CCA GCT CCT CCA (reverse primer)
For biotinylated ds dual-labeled $(CGG)_n$, where n=23, 56, 96, 118, 340:
5’ Biotin-GT-Cy3-C AGG CGC TCA GCT CCG TTT CGG TTT CAC TTC CGG T (forward primer)
5’ Cy5-AGC CCC GCA CTT CCA CCA CCA GCT CCT CCA (reverse primer)

Table A.1: Sequences of the oligonucleotides used for development of Fragile X diagnostic assay

Bibliography

1. Noller, H. F. RNA structure: reading the ribosome. *Science* **309,** 1508–1514 (2005).
2. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3,** (2011).
3. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16,** 1725–1747 (2010).
4. Wilson, R. C. & Doudna, J. A. Molecular mechanisms of RNA interference. *Annu Rev Biophys* **42,** 217–239 (2013).
5. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–655 (2009).
6. Puglisi, J. D., Blanchard, S. C. & Green, R. Approaching translation at atomic resolution. *Nat. Struct. Biol.* **7,** 855–861 (2000).
7. Korostelev, A. & Noller, H. F. The ribosome in focus: new structures bring new insights. *Trends Biochem. Sci.* **32,** 434–441 (2007).
8. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The signal recognition particle. *Annu. Rev. Biochem.* **70,** 755–775 (2001).
9. Nyathi, Y., Wilkinson, B. M. & Pool, M. R. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochim. Biophys. Acta* **1833,** 2392–2402 (2013).
10. Collins, K. The biogenesis and regulation of telomerase holoenzymes. *Nat. Rev. Mol. Cell Biol.* **7,** 484–494 (2006).
11. Egan, E. D. & Collins, K. Biogenesis of telomerase ribonucleoproteins. *RNA* **18,** 1747–1759 (2012).
12. Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **319,** 1787–1789 (2008).
13. Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mamm. Genome* **19,** 454–492 (2008).
14. Dinger, M. E., Amaral, P. P., Mercer, T. R. & Mattick, J. S. Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Brief Funct Genomic Proteomic* **8,** 407–423 (2009).
15. Clark, M. B. *et al.* The Reality of Pervasive Transcription. *PLoS Biol* **9,** e1000625 (2011).

16. Guttman, M. *et al*. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).

17. Weinberg, Z., Perreault, J., Meyer, M. M. & Breaker, R. R. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462,** 656–659 (2009).

18. Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **4,** (2012).

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

20. Potratz, J. P. & Russell, R. RNA catalysis as a probe for chaperone activity of DEAD-box helicases. *Meth. Enzymol.* **511,** 111–130 (2012).

21. Brion, P. & Westhof, E. Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* **26,** 113–137 (1997).

22. Russell, R. RNA misfolding and the action of chaperones. *Front Biosci* **13,** 1–20 (2008).

23. Serra, M. J. & Turner, D. H. Predicting thermodynamic properties of RNA. *Meth. Enzymol.* **259,** 242–261 (1995).

24. Tinoco, I., Jr & Bustamante, C. How RNA folds. *J. Mol. Biol.* **293,** 271–281 (1999).

25. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35,** 161–175 (2006).

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

27. Chadalavada, D. M., Cerrone-Szakal, A. L. & Bevilacqua, P. C. Wild-type is the optimal sequence of the HDV ribozyme under cotranscriptional conditions. *RNA* **13,** 2189–2201 (2007).

28. Baldwin, R. L. & Rose, G. D. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24,** 77–83 (1999).

29. Baldwin, R. L. & Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr. Opin. Struct. Biol.* **23,** 4–10 (2013).

30. Gartland, W. J. & Sueoka, N. Two interconvertible forms of tryptophanyl sRNA in E. coli. *Proc. Natl. Acad. Sci. U.S.A.* **55,** 948–956 (1966).

31. Adams, A., Lindahl, T. & Fresco, J. R. Conformational differences between the biologically active and inactive forms of a transfer ribonucleic acid. *Proc. Natl. Acad. Sci. U.S.A.* **57,** 1684–1691 (1967).

32. Uhlenbeck, O. C., Chirikjian, J. G. & Fresco, J. R. Oligonucleotide binding to the native and denatured conformers of yeast transfer RNA-3 Lea. *J. Mol. Biol.* **89,** 495–504 (1974).

33. Madore, E., Florentz, C., Giegé, R. & Lapointe, J. Magnesium-dependent alternative foldings of active and inactive Escherichia coli tRNA(Glu) revealed by chemical probing. *Nucleic Acids Res.* **27,** 3583–3588 (1999).

34. Weidner, H., Yuan, R. & Crothers, D. M. Does 5S RNA function by a switch between two secondary structures? *Nature* **266,** 193–194 (1977).

35. Woodson, S. A. & Cech, T. R. Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the Tetrahymena intervening sequence RNA. *Biochemistry* **30,** 2042–2050 (1991).

36. Woodson, S. A. & Emerick, V. L. An alternative helix in the 26S rRNA promotes excision and integration of the Tetrahymena intervening sequence. *Mol. Cell. Biol.* **13,** 1137–1145 (1993).

37. Pichler, A. & Schroeder, R. Folding problems of the 5' splice site containing the P1 stem of the group I thymidylate synthase intron: substrate binding inhibition in vitro and mis-splicing in vivo. *J. Biol. Chem.* **277,** 17987–17993 (2002).

38. Duncan, C. D. S. & Weeks, K. M. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* **47,** 8504–8513 (2008).

39. Sinan, S., Yuan, X. & Russell, R. The Azoarcus group I intron ribozyme misfolds and is accelerated for refolding by ATP-dependent RNA chaperone proteins. *J Biol Chem* **286,** 37304–12 (2011).

40. Chadalavada, D. M., Senchak, S. E. & Bevilacqua, P. C. The folding pathway of the genomic hepatitis delta virus ribozyme is dominated by slow folding of the pseudoknots. *J. Mol. Biol.* **317,** 559–575 (2002).

41. Cannone, J. J. *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3,** 2 (2002).

42. Leontis, N. B. & Westhof, E. Analysis of RNA motifs. *Curr. Opin. Struct. Biol.* **13,** 300–308 (2003).

43. Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16,** 279–287 (2006).

44. Chojnowski, G., Walen, T. & Bujnicki, J. M. RNA Bricks--a database of RNA 3D motifs and their interactions. *Nucleic Acids Res.* **42,** D123–131 (2014).

45. Hodak, J. H., Downey, C. D., Fiore, J. L., Pardi, A. & Nesbitt, D. J. Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 10505–10510 (2005).

46. Fiore, J. L. & Nesbitt, D. J. An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* **46,** 223–264 (2013).

47. Zarrinkar, P. P. & Williamson, J. R. Kinetic intermediates in RNA folding. *Science* **265,** 918–924 (1994).

48. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. & Woodson, S. A. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* **279,** 1940–1943 (1998).

49. 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).

50. Russell, R. & Herschlag, D. New pathways in folding of the Tetrahymena group I RNA enzyme. *J Mol Biol* **291,** 1155–67 (1999).

51. 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).

52. Russell, R. *et al*. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci U S A* **99,** 155–60 (2002).

53. Laederach, A., Shcherbakova, I., Jonikas, M. A., Altman, R. B. & Brenowitz, M. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 7045–7050 (2007).

54. Shcherbakova, I., Mitra, S., Laederach, A. & Brenowitz, M. Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol* **12,** 655–66 (2008).

55. Wan, Y., Suh, H., Russell, R. & Herschlag, D. Multiple unfolding events during native folding of the Tetrahymena group I ribozyme. *J Mol Biol* **400,** 1067–77 (2010).

56. Mitchell, D., 3rd & Russell, R. Folding Pathways of the Tetrahymena Ribozyme. *J. Mol. Biol.* (2014). doi:10.1016/j.jmb.2014.04.011

57. Pan, J. & Woodson, S. A. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* **280,** 597–609 (1998).

58. Russell, R. *et al*. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* **363,** 531–44 (2006).

59. Mitchell, D., 3rd, Jarmoskaite, I., Seval, N., Seifert, S. & Russell, R. The long-range P3 helix of the Tetrahymena ribozyme is disrupted during folding between the native and misfolded conformations. *J. Mol. Biol.* **425,** 2670–2686 (2013).

60. 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).

61. Rangan, P., Masquida, B., Westhof, E. & Woodson, S. A. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 1574–1579 (2003).

62. Chauhan, S. *et al.* RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. *J. Mol. Biol.* **353,** 1199–1209 (2005).

63. Chauhan, S., Behrouzi, R., Rangan, P. & Woodson, S. A. Structural rearrangements linked to global folding pathways of the Azoarcus group I ribozyme. *J. Mol. Biol.* **386,** 1167–1178 (2009).

64. Roh, J. H. *et al.* Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle X-ray scattering. *J. Am. Chem. Soc.* **132,** 10148–10154 (2010).

65. Mitra, S., Laederach, A., Golden, B. L., Altman, R. B. & Brenowitz, M. RNA molecules with conserved catalytic cores but variable peripheries fold along unique energetically optimized pathways. *RNA* **17,** 1589–1603 (2011).

66. Jiang, Y.-F., Xiao, M., Yin, P. & Zhang, Y. Monovalent cations use multiple mechanisms to resolve ribozyme misfolding. *RNA* **12,** 561–566 (2006).

67. Duncan, C. D. S. & Weeks, K. M. Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding. *Biochemistry* **49,** 5418–5425 (2010).

68. Roth, A. & Breaker, R. R. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78,** 305–334 (2009).

69. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43,** 880–891 (2011).

70. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92,** 315–26 (1998).

71. Chen, W. & Moore, M. J. The spliceosome: disorder and dynamics defined. *Curr. Opin. Struct. Biol.* **24C,** 141–149 (2014).

72. Herschlag, D. RNA chaperones and the RNA folding problem. *J Biol Chem* **270,** 20871–4 (1995).

73. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5,** 908–919 (2004).

74. Levin, J. G., Guo, J., Rouzina, I. & Musier-Forsyth, K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **80,** 217–286 (2005).

75. Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4,** 118–30 (2007).

76. Clodi, E., Semrad, K. & Schroeder, R. Assaying RNA chaperone activity in vivo using a novel RNA folding trap. *EMBO J.* **18,** 3776–3782 (1999).

77. Rajkowitsch, L., Semrad, K., Mayer, O. & Schroeder, R. Assays for the RNA chaperone activity of proteins. *Biochem. Soc. Trans.* **33,** 450–456 (2005).

78. 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,** 2300–2312 (2002).

79. 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,** 1175–1180 (2002).

80. Schroeder, R., Grossberger, R., Pichler, A. & Waldsich, C. RNA folding in vivo. *Curr. Opin. Struct. Biol.* **12,** 296–300 (2002).

81. 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).

82. 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).

83. Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20,** 313–324 (2010).

84. Linder, P. Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res* **34,** 4168–80 (2006).

85. Jankowsky, E. & Fairman, M. E. RNA helicases--one fold for many functions. *Curr Opin Struct Biol* **17,** 316–24 (2007).

86. Fairman, M. E. *et al.* Protein displacement by DExH/D 'RNA helicases' without duplex unwinding. *Science* **304,** 730–4 (2004).

87. 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).

88. Rössler, 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,** 2088–2096 (2001).

89. Jarmoskaite, I. & Russell, R. RNA Helicase Proteins as Chaperones and Remodelers. *Annu. Rev. Biochem.* (2014). doi:10.1146/annurev-biochem-060713-035546

90. Cordin, O. & Beggs, J. D. RNA helicases in splicing. *RNA Biol* **10,** 83–95 (2013).

91. Chang, T.-H., Tung, L., Yeh, F.-L., Chen, J.-H. & Chang, S.-L. Functions of the DExD/H-box proteins in nuclear pre-mRNA splicing. *Biochim. Biophys. Acta* **1829,** 764–774 (2013).

92. 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).

93. Chen, J. Y. *et al*. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol. Cell* **7,** 227–232 (2001).

94. Yang, F. *et al*. Splicing proofreading at 5' splice sites by ATPase Prp28p. *Nucleic Acids Res.* **41,** 4660–4670 (2013).

95. Mefford, M. A. & Staley, J. P. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* **15,** 1386–1397 (2009).

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

97. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8,** 847–855 (1998).

98. Noble, S. M. & Guthrie, C. Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* **143,** 67–80 (1996).

99. Edwalds-Gilbert, G. *et al*. Dominant negative mutants of the yeast splicing factor Prp2 map to a putative cleft region in the helicase domain of DExD/H-box proteins. *RNA* **6,** 1106–1119 (2000).

100. Lardelli, R. M., Thompson, J. X., Yates, J. R., 3rd & Stevens, S. W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16,** 516–528 (2010).

101. Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. The DEAD-box protein family of RNA helicases. *Gene* **367,** 17–37 (2006).

102. Del Campo, M. & Lambowitz, A. M. Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol Cell* **35,** 598–609 (2009).

103. 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).

104. 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,** 713–725 (2006).

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

106. Andersen, C. B. F. *et al.* Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313,** 1968–1972 (2006).

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

108. Tanner, N. K., Cordin, O., Banroques, J., Doère, 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).

109. Cordin, O., Tanner, N. K., Doère, 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).

110. Banroques, J., Doère, M., Dreyfus, M., Linder, P. & Tanner, N. K. Motif III in superfamily 2 'helicases' helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J Mol Biol* **396,** 949–66 (2010).

111. Mallam, A. L., Del Campo, M., Gilman, B., Sidote, D. J. & Lambowitz, A. M. Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* **490,** 121–5 (2012).

112. 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).

113. Polach, K. J. & Uhlenbeck, O. C. Cooperative binding of ATP and RNA substrates to the DEAD/H protein DbpA. *Biochemistry* **41,** 3693–702 (2002).

114. Theissen, B., Karow, A. R., Köhler, J., Gubaev, A. & Klostermeier, D. Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc Natl Acad Sci U S A* **105,** 548–53 (2008).

115. Karow, A. R. & Klostermeier, D. A conformational change in the helicase core is necessary but not sufficient for RNA unwinding by the DEAD box helicase YxiN. *Nucleic Acids Res.* **37,** 4464–4471 (2009).

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

117. Henn, A., Cao, W., Hackney, D. D. & De La Cruz, E. M. The ATPase cycle mechanism of the DEAD-box rRNA helicase, DbpA. *J. Mol. Biol.* **377,** 193–205 (2008).

118. 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).

119. Mohr, G. *et al*. Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. *J Mol Biol* **375,** 1344–64 (2008).

120. Diges, C. M. & Uhlenbeck, O. C. Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J.* **20,** 5503–5512 (2001).

121. Klostermeier, D. & Rudolph, M. G. A novel dimerization motif in the C-terminal domain of the Thermus thermophilus DEAD box helicase Hera confers substantial flexibility. *Nucleic Acids Res.* **37,** 421–430 (2009).

122. 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,** 959–967 (2006).

123. Mallam, A. L. *et al*. Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *Proc Natl Acad Sci U S A* **108,** 12254–9 (2011).

124. 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,** 3619–3626 (1993).

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

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

127. 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).

128. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76,** 23–50 (2007).

129. Lohman, T. M., Tomko, E. J. & Wu, C. G. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.* **9,** 391–401 (2008).

130. Chamot, D., Colvin, K. R., Kujat-Choy, S. L. & Owttrim, G. W. RNA structural rearrangement via unwinding and annealing by the cyanobacterial RNA helicase, CrhR. *J. Biol. Chem.* **280,** 2036–2044 (2005).

131. Rogers, G. W., 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).

132. Rogers, G. W., Lima, W. F. & Merrick, W. C. Further characterization of the helicase activity of eIF4A. Substrate specificity. *J Biol Chem* **276,** 12598–608 (2001).

133. Chen, Y. *et al*. DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc Natl Acad Sci U S A* **105,** 20203–8 (2008).

134. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W. & Weber, P. C. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. *Structure* **7,** 1353–1363 (1999).

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

136. He, Y., Andersen, G. R. & Nielsen, K. H. Structural basis for the function of DEAH helicases. *EMBO Rep.* **11,** 180–186 (2010).

137. Jackson, R. N. *et al*. The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J.* **29,** 2205–2216 (2010).

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

139. Jarmoskaite, I. & Russell, R. DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdiscip Rev RNA* **2,** 135–52 (2011).

140. Rogers, G. W., 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).

141. Bizebard, T., Ferlenghi, I., Iost, I. & Dreyfus, M. Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* **43,** 7857–7866 (2004).

142. Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell* **28,** 253–63 (2007).

143. Garcia, I., Albring, M. J. & Uhlenbeck, O. C. Duplex destabilization by four ribosomal DEAD-box proteins. *Biochemistry* **51,** 10109–10118 (2012).

144. Huang, Y. & Liu, Z.-R. The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. *J. Biol. Chem.* **277,** 12810–12815 (2002).

145. 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).

146. Bowers, H. A. *et al*. Discriminatory RNP remodeling by the DEAD-box protein DED1. *RNA* **12,** 903–12 (2006).

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

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

149. Lund, M. K. & Guthrie, C. The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol Cell* **20,** 645–51 (2005).

150. Tran, E. J., Zhou, Y., Corbett, A. H. & Wente, S. R. The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. *Mol. Cell* **28,** 850–859 (2007).

151. Stewart, M. Ratcheting mRNA out of the nucleus. *Mol. Cell* **25,** 327–330 (2007).

152. Linder, P. mRNA export: RNP remodeling by DEAD-box proteins. *Curr Biol* **18,** R297–9 (2008).

153. Strahm, Y. *et al*. The RNA export factor Gle1p is located on the cytoplasmic fibrils of the NPC and physically interacts with the FG-nucleoporin Rip1p, the DEAD-box protein Rat8p/Dbp5p and a new protein Ymr 255p. *EMBO J.* **18,** 5761–5777 (1999).

154. Weirich, C. S. *et al*. Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. *Nat. Cell Biol.* **8,** 668–676 (2006).

155. 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).

156. Bifano, A. L. & Caprara, M. G. A DExH/D-box protein coordinates the two steps of splicing in a group I intron. *J. Mol. Biol.* **383,** 667–682 (2008).

157. 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).

158. Séraphin, B., Simon, M., Boulet, A. & Faye, G. Mitochondrial splicing requires a protein from a novel helicase family. *Nature* **337,** 84–87 (1989).

159. 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,** 1575–1588 (1994).

160. Bhaskaran, H. & Russell, R. Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* **449,** 1014–8 (2007).

161. Del Campo, M. *et al*. Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol Cell* **28,** 159–66 (2007).

162. 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).

163. Landry, S. J., Jordan, R., McMacken, R. & Gierasch, L. M. Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature* **355,** 455–457 (1992).

164. Zhu, X. *et al*. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272,** 1606–1614 (1996).

165. Schmid, D., Baici, A., Gehring, H. & Christen, P. Kinetics of molecular chaperone action. *Science* **263,** 971–973 (1994).

166. Alfano, C. & McMacken, R. Heat shock protein-mediated disassembly of nucleoprotein structures is required for the initiation of bacteriophage lambda DNA replication. *J. Biol. Chem*. **264,** 10709–10718 (1989).

167. Diamant, S., Ben-Zvi, A. P., Bukau, B. & Goloubinoff, P. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. *J. Biol. Chem*. **275,** 21107–21113 (2000).

168. Bukau, B., Weissman, J. & Horwich, A. Molecular chaperones and protein quality control. *Cell* **125,** 443–451 (2006).

169. Saibil, H. R. Chaperone machines in action. *Curr. Opin. Struct. Biol*. **18,** 35–42 (2008).

170. Barends, T. R. M., Werbeck, N. D. & Reinstein, J. Disaggregases in 4 dimensions. *Curr. Opin. Struct. Biol*. **20,** 46–53 (2010).

171. Narlikar, G. J. & Herschlag, D. Isolation of a local tertiary folding transition in the context of a globally folded RNA. *Nat Struct Biol* **3,** 701–10 (1996).

172. 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).

173. Pyle, A. M. & Cech, T. R. Ribozyme recognition of RNA by tertiary interactions with specific ribose 2'-OH groups. *Nature* **350,** 628–631 (1991).

174. Bevilacqua, P. C. & Turner, D. H. Comparison of binding of mixed ribose-deoxyribose analogues of CUCU to a ribozyme and to GGAGAA by equilibrium dialysis: evidence for ribozyme specific interactions with 2' OH groups. *Biochemistry* **30,** 10632–10640 (1991).

175. Herschlag, D., Eckstein, F. & Cech, T. R. Contributions of 2'-hydroxyl groups of the RNA substrate to binding and catalysis by the Tetrahymena ribozyme. An energetic picture of an active site composed of RNA. *Biochemistry* **32,** 8299–311 (1993).

176. Strobel, S. A. & Cech, T. R. Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the Tetrahymena ribozyme. *Biochemistry* **32,** 13593–604 (1993).

177. Pyle, A. M. *et al*. Replacement of the conserved G.U with a G-C pair at the cleavage site of the Tetrahymena ribozyme decreases binding, reactivity, and fidelity. *Biochemistry* **33,** 13856–13863 (1994).

178. Knitt, D. S., Narlikar, G. J. & Herschlag, D. Dissection of the role of the conserved G.U pair in group I RNA self-splicing. *Biochemistry* **33,** 13864–79 (1994).

179. Strobel, S. A. & Cech, T. R. Minor groove recognition of the conserved G.U pair at the Tetrahymena ribozyme reaction site. *Science* **267,** 675–679 (1995).

180. Zhuang, X. *et al*. A single-molecule study of RNA catalysis and folding. *Science* **288,** 2048–51 (2000).

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

182. Chen, J., Tsai, A., O'Leary, S. E., Petrov, A. & Puglisi, J. D. Unraveling the dynamics of ribosome translocation. *Curr. Opin. Struct. Biol.* **22,** 804–814 (2012).

183. Henn, A., Bradley, M. J. & De La Cruz, E. M. ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annu Rev Biophys* **41,** 247–67 (2012).

184. Linder, P. & Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* **12,** 505–16 (2011).

185. Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends Biochem. Sci*. **36,** 19–29 (2011).

186. Russell, R., Jarmoskaite, I. & Lambowitz, A. M. Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol* **10,** 44–55 (2013).

187. Karunatilaka, K. S., Solem, A., Pyle, A. M. & Rueda, D. Single-molecule analysis of Mss116-mediated group II intron folding. *Nature* **467,** 935–9 (2010).

188. Rogers, G. W., Komar, A. A. & Merrick, W. C. eIF4A: the godfather of the DEAD box helicases. *Prog Nucleic Acid Res Mol Biol* **72,** 307–31 (2002).

189. Bartley, L. E., Zhuang, X., Das, R., Chu, S. & Herschlag, D. Exploration of the transition state for tertiary structure formation between an RNA helix and a large structured RNA. *J Mol Biol* **328,** 1011–26 (2003).

190. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463,** 681–4 (2010).

191. Levitus, M. & Ranjit, S. Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Q Rev Biophys* **44,** 123–51 (2011).

192. Iqbal, A. *et al*. Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc Natl Acad Sci U S A* **105,** 11176–81 (2008).

193. Iqbal, A., Wang, L., Thompson, K. C., Lilley, D. M. & Norman, D. G. The structure of cyanine 5 terminally attached to double-stranded DNA: implications for FRET studies. *Biochemistry* **47,** 7857–62 (2008).

194. Narlikar, G. J., Gopalakrishnan, V., McConnell, T. S., Usman, N. & Herschlag, D. Use of binding energy by an RNA enzyme for catalysis by positioning and substrate destabilization. *Proc Natl Acad Sci U S A* **92,** 3668–72 (1995).

195. Betterton, M. D. & Jülicher, F. Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys Rev E Stat Nonlin Soft Matter Phys* **71,** 011904 (2005).

196. Manosas, M., Xi, X. G., Bensimon, D. & Croquette, V. Active and passive mechanisms of helicases. *Nucleic Acids Res* **38,** 5518–26 (2010).

197. Sattin, B. D., Zhao, W., Travers, K., Chu, S. & Herschlag, D. Direct measurement of tertiary contact cooperativity in RNA folding. *J Am Chem Soc* **130,** 6085–7 (2008).

198. 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).

199. Shi, X. *et al*. Roles of long-range tertiary interactions in limiting dynamics of the tetrahymena group I ribozyme. *J. Am. Chem. Soc*. **136,** 6643–6648 (2014).

200. Jarmoskaite, I., Bhaskaran, H., Seifert, S. & Russell, R. The DEAD-box protein CYT-19 is activated by exposed helices in a group I intron RNA (submitted).

201. Ding, Y. *et al*. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505,** 696–700 (2014).

202. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505,** 701–705 (2014).

203. Shajani, Z., Sykes, M. T. & Williamson, J. R. Assembly of bacterial ribosomes. *Annu Rev Biochem* **80,** 501–26 (2011).

204. Cordin, O., Hahn, D. & Beggs, J. D. Structure, function and regulation of spliceosomal RNA helicases. *Curr Opin Cell Biol* **24,** 431–8 (2012).

205. Koodathingal, P. & Staley, J. P. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA Biol* **10,** 1073–9 (2013).

206. Solomatin, S. V., Greenfeld, M. & Herschlag, D. Implications of molecular heterogeneity for the cooperativity of biological macromolecules. *Nat Struct Mol Biol* **18,** 732–4 (2011).

207. Penagarikano, O., Mulle, J. G. & Warren, S. T. The pathophysiology of fragile x syndrome. *Annu Rev Genomics Hum Genet* **8,** 109–129 (2007).

208. Hagerman, R. J. *et al*. High functioning fragile X males: demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. *Am. J. Med. Genet*. **51,** 298–308 (1994).

209. Wang, J. Y. *et al*. Fragile X-associated tremor/ataxia syndrome: influence of the FMR1 gene on motor fiber tracts in males with normal and premutation alleles. *JAMA Neurol* **70,** 1022–1029 (2013).

210. Terracciano, A. *et al*. Expansion to full mutation of a FMR1 intermediate allele over two generations. *Eur. J. Hum. Genet*. **12,** 333–336 (2004).

211. Fernandez-Carvajal, I. *et al*. Expansion of an FMR1 grey-zone allele to a full mutation in two generations. *J Mol Diagn* **11,** 306–310 (2009).

212. Chen, L. *et al*. An information-rich CGG repeat primed PCR that detects the full range of fragile X expanded alleles and minimizes the need for southern blot analysis. *J Mol Diagn* **12,** 589–600 (2010).

213. Lyon, E. *et al*. A simple, high-throughput assay for Fragile X expanded alleles using triple repeat primed PCR and capillary electrophoresis. *J Mol Diagn* **12,** 505–511 (2010).

214. Hantash, F. M. *et al*. Qualitative assessment of FMR1 (CGG)n triplet repeat status in normal, intermediate, premutation, full mutation, and mosaic carriers in both sexes: implications for fragile X syndrome carrier and newborn screening. *Genet. Med*. **12,** 162–173 (2010).

215. Khaniani, M. S., Kalitsis, P., Burgess, T. & Slater, H. R. An improved Diagnostic PCR Assay for identification of Cryptic Heterozygosity for CGG Triplet Repeat Alleles in the Fragile X Gene (FMR1). *Mol Cytogenet* **1,** 5 (2008).

216. Todorov, T., Todorova, A., Georgieva, B. & Mitev, V. A unified rapid PCR method for detection of normal and expanded trinucleotide alleles of CAG repeats in huntington chorea and CGG repeats in fragile X syndrome. *Mol. Biotechnol*. **45,** 150–154 (2010).

217. Filipovic-Sadic, S. *et al*. A novel FMR1 PCR method for the routine detection of low abundance expanded alleles and full mutations in fragile X syndrome. *Clin. Chem*. **56,** 399–408 (2010).

218. Pushkarev, D., Neff, N. F. & Quake, S. R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol*. **27,** 847–850 (2009).

219. Eid, J. *et al*. Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).

220. Schlapak, R., Kinns, H., Wechselberger, C., Hesse, J. & Howorka, S. Sizing trinucleotide repeat sequences by single-molecule analysis of fluorescence brightness. *Chemphyschem* **8,** 1618–1621 (2007).

221. Jo, K. *et al*. A single-molecule barcoding system using nanoslits for DNA analysis. *Proc. Natl. Acad. Sci. U.S.A*. **104,** 2673–2678 (2007).

222. Xiao, M. *et al*. Determination of haplotypes from single DNA molecules: a method for single-molecule barcoding. *Hum. Mutat*. **28,** 913–921 (2007).

223. Burton, R. E. *et al*. A microfluidic chip-compatible bioassay based on single-molecule detection with high sensitivity and multiplexing. *Lab Chip* **10,** 843–851 (2010).

224. Chan, E. Y. *et al*. DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome Res*. **14,** 1137–1146 (2004).

225. Protozanova, E. *et al*. Fast high-resolution mapping of long fragments of genomic DNA based on single-molecule detection. *Anal. Biochem*. **402,** 83–90 (2010).

226. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 7702–7707 (2009).

227. Levine, P. M., Gong, P., Levicky, R. & Shepard, K. L. Real-time, multiplexed electrochemical DNA detection using an active complementary metal-oxide-semiconductor biosensor array with integrated sensor electronics. *Biosens Bioelectron* **24,** 1995–2001 (2009).

228. Rothberg, J. M. *et al*. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475,** 348–352 (2011).

229. Cannon, B., Pan, C., Chen, L., Hadd, A. G. & Russell, R. A Dual-Mode Single-Molecule Fluorescence Assay for the Detection of Expanded CGG Repeats in Fragile X Syndrome. *Mol Biotechnol* (2012). doi:10.1007/s12033-012-9505-z

230. Nadel, Y., Weisman-Shomer, P. & Fry, M. The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.* **270,** 28970–28977 (1995).

231. Paiva, A. M. & Sheardy, R. D. Influence of sequence context and length on the structure and stability of triplet repeat DNA oligomers. *Biochemistry* **43,** 14218–14227 (2004).

232. Jarem, D. A., Huckaby, L. V. & Delaney, S. AGG interruptions in (CGG)(n) DNA repeat tracts modulate the structure and thermodynamics of non-B conformations in vitro. *Biochemistry* **49,** 6826–6837 (2010).

233. Völker, J., Klump, H. H. & Breslauer, K. J. DNA energy landscapes via calorimetric detection of microstate ensembles of metastable macrostates and triplet repeat diseases. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 18326–18330 (2008).

234. Harvey, B. J. & Levitus, M. Nucleobase-Specific Enhancement of Cy3 Fluorescence. *J Fluoresc* **19,** 443–448 (2009).

235. Agbavwe, C. & Somoza, M. M. Sequence-Dependent Fluorescence of Cyanine Dyes on Microarrays. *PLoS ONE* **6,** e22177 (2011).

236. Harvey, B. J., Perez, C. & Levitus, M. DNA sequence-dependent enhancement of Cy3 fluorescence. *Photochem. Photobiol. Sci.* **8,** 1105–1110 (2009).