Structural Analysis of Protein-Facilitated Cooperative Folding in the bI3 RNP

Caia Deborah Suzanne Duncan

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2009

Approved by,
Kevin Weeks
Linda Spremulli
Gary Pielak
Mike Jarstfer
Marcey Waters

Abstract

Caia Deborah Suzanne Duncan

Structural Analysis of Protein-Facilitated Cooperative Folding in the bI3 RNP

(Under the direction of Kevin M. Weeks)

Most large RNAs require protein facilitators to achieve their active structures. Despite the importance of RNA structure for understanding the functions of associated proteins, initial and intermediate structures of the RNA are often assumed or not fully characterized. This gap in structural knowledge has lead to an over-simplification of the roles of many proteins in ribonucleoprotein complexes. Protein facilitators are currently categorized into two broad classes, cofactors and chaperones. Cofactors bind tightly an RNA to stabilize tertiary structures, while chaperones interact transiently with an RNA to facilitate acquisition of the most stable secondary structure. In this research, I comprehensively interrogate the structure of an RNA throughout the stages of protein-facilitated folding. In principle, the yeast bI3 group I intron has to potential to fold into the active secondary structure conserved among group I introns. The bI3 RNA requires two proteins (the bI3 maturase and Mrs1) for splicing. These proteins seem to have characteristics of cofactors and were not expected to affect the secondary structure. In contrast, by developing a high-throughput SHAPE experiment, I find that approximately half the bI3 RNA is not in the catalytically active secondary structure prior to protein

binding. I develop and test a structural model for the misfolded RNA using SHAPE analysis of point mutations. I find that three conserved elements form stable, extensively mispaired, non-native structures. Solvent accessibility experiments show that these non-native structures are incapable of forming native group I intron tertiary interactions. I next demonstrate that binding by either the Mrs1 or maturase protein alone promotes the formation of distinct tertiary structures. However, neither individual protein binds to or affects the non-native secondary structures in the misfolded RNA. Strikingly, simultaneous binding of both proteins enables the RNA to achieve its active conformation. These results highlight a large-scale cooperative folding process between the bI3 RNA and the Mrs1 and maturase proteins in which the final RNA structure is drastically different than the sum of individual protein-bound states. The bI3 RNP thus represents a new category of ribonucleoprotein assembly mechanism in which binding by multiple proteins drives non-hierarchical RNA folding.

“Evolution is cleverer than you.”

~ Francis Crick

To Effe May Duncan

Acknowledgements

First and foremost I'd like to thank Kevin Weeks, who gave me independence until he somehow knew I needed help. He offered me the opportunity to work on a great project and took the time to teach me about all aspects that make a great scientist. For that I am deeply appreciative. I'd also like to thank all the faculty of the chemistry department, especially Linda Spremulli and Dorothy Erie, for their deep insight and open door policies.

To the Weeks lab, past and present, you challenged my science, my intelligence and my sanity, I am the scientist I am today because of you. To those I leave behind, stay strong, work hard, play nicely and keep your sense of humour… you'll need it!

I'd also like to thank all my friends on the 7th and 8th floors and at Duke. Christie, thank you for years of scientific and non-scientific discussions alike. Liz, Erika, Sheryl and Vanessa, I would have been lost without your warped humour and twisted distractions. I shake my fist at you.

To my family and friends here and far away, thank you for all the support, especially when it wasn't asked for. To Risa, without your constant nagging and random insults I would have completely lost touch with the real world. And to Jo, Pauline, Helena, Saz and Alison, we've vented though 9 years of education together, despite that pesky ocean and now, we're finally done!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1M7	1-methyl-7-nitroisatoic anhydride
Å	Angstrom
A	adenosine
ATP	adenosine triphosphate
C	cytosine
cDNA	complementary DNA
Cys	cysteine
CTP	cytosine triphosphate
°C	degrees Celsius
ddNTP	dideoxynucleotide triphosphate
ΔG	Gibbs free energy
dATP	deoxyadenosine triphosphate
ddATP	dideoxyadenosine triphosphate
ddTTP	dideoxythymine triphosphate
dITP	deoxyinosine triphosphate
dNTP	deoxynucleotide triphosphate
dTTP	deoxythymine triphosphate
DMD	descrete molecular dynamics
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiotreitol
EDTA	ethylene diaminetetraacetic acid

EtOH	ethanol
Fe	iron
$[Fe(II)(NH_4)_2]SO_4$	ferrous ammonium sulfate
G	guanosine
g	gram
ΩG	omega G; term for the group I intron 3′ splice site
GNRA	RNA loop motif of the sequence: (G)-(any nucleotide)-(G or A)-(A)
GTP	guanosine triphosphate
HEPES	4-(2-hydroxy-ethyl)-1-piperazine-ethanesulfonic acid
H_2O	water
H_2O_2	hydrogen peroxide
h	hour
IPTG	isopropyl β-D-1-thiogalactopyranoside
$K_{½}$	equilibrium binding constant where half the substrate is bound
k_{cat}	rate of catalysis
KCl	potassium choloride
K_d	equilibrium dissociation constant
k_{obs}	observed rate of catalysis
kDa	kilodalton
KoAc	potassium acetate
Mops	3-(N-Morpholino)propanesulfonic acid
M	molar
min	minute

min^{-1}	per minute; unit of catalysis
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
mRNA	messenger RNA
ml	milliliter
µl	microliter
mM	millimolar
µM	micromolar
NaCl	sodium chloride
NaOH	sodium hydroxide
nM	nanomolar
nt	nucleotide
NTP	nucleotide triphosphate
·OH	hydroxyl radical
^{32}P	phosphorous isotope
PAGE	polyacrylamide gel electrophorsis
PCR	polymerase chain reaction
PDB	protein dilution buffer
pG	guanosine monophosphate
pmol	picomol
PNK	polynucleotide kinase
RMSD	root mean squared deviation
RNA	ribonucleic acid

RNP	ribonucleoprotein
rRNA	ribosomal RNA
s	second
SHAPE	selective 2′-hydroxyl acylation analyzed by primer extension
T	thymidine
Taq	*Thermus aquaticus*
TBE	90 mM Tris-borate, 2 mM EDTA
TEMED	N,N,N′,N′-tetramethylethylenediamine
Tris	tris(hydroxymethyl) aminomethane
tRNA	transfer RNA
UTP	uridine triphosphate
v/v	volume/volume
W	watt
w/v	weight/volume

Chapter 1

Structural Characterization of RNA-Protein Interactions

1.1 RNA Structure

RNA acts not only as passive carrier of genetic information, but is also actively involved many cellular processes, including protein synthesis and gene regulation[1]. Essential to the function of any RNA is its structure and, in some cases, the ability to achieve multiple stable structures. Many RNAs form specific structures that selectively bind proteins and function as ribonucleoproteins (RNPs). RNPs such as the spliceosome and signal recognition particle play vital roles in mRNA processing and protein localization, respectively[1]. The most abundant RNAs in the cell, and arguably the most essential, are the ribosomal RNAs (rRNAs), which form the ribosome. The ribosome is a ribozyme, where proteins are synthesized by a catalytic center composed of RNA[1,2]. Additional catalytic ribozymes, such as group I and II introns, RNase P, and the hammerhead ribozyme, function to cleave RNA during mRNA processing[3-5]. In order to be catalytically active, ribozymes have highly structured three-dimensional architectures[6-9]. Other RNAs, such as riboswitches, can fold into multiple stable conformations. These multiple conformations allow riboswitches to operate as metabolite sensors that regulate gene expression. Since RNA structure is vital to its function, the RNA folding process must be accurate and robust for biological importance.

1.1.1 RNA Folding. RNA folding is driven by both kinetics and thermodynamics. The folding process is predominantly hierarchical, such that secondary structure forms first, followed by the tertiary structure (Figure 1.1)[10,11]. The stability of RNA secondary structure is established by base pairs, base stacking and favorable electrostatic interactions made as helices form. Tertiary structures are considered to be the association of two secondary structure elements[10,11]. In this traditional folding model,

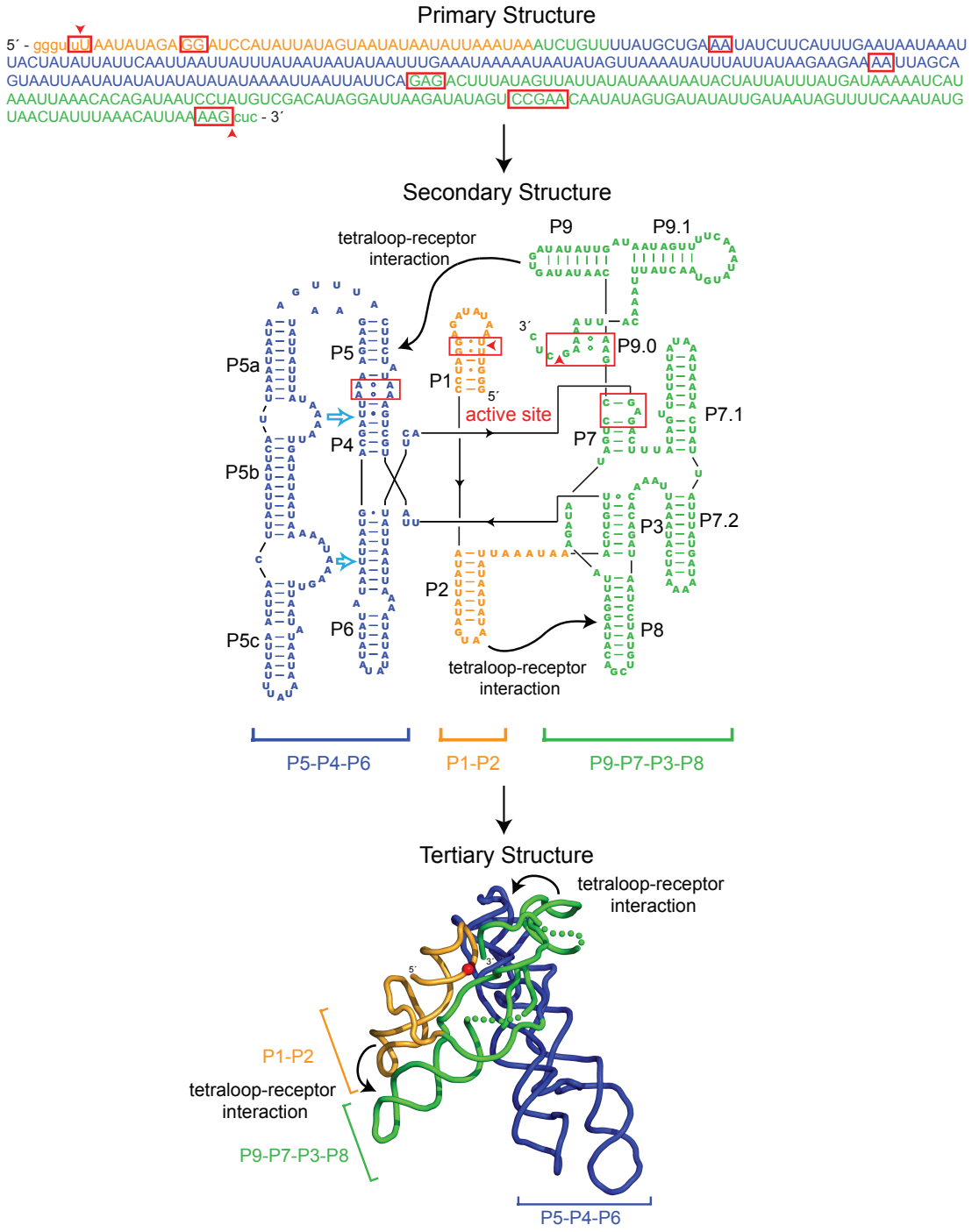

Figure 1.1. Hierarchical folding of the bI3 group I intron. Primary structure folds back on itself to form base pairing interactions in the secondary structure. Tertiary interactions, such as tetraloop-receptor motifs (black arrows) and A-minor interactions (blue arrows), hold the RNA in a compact and tight structure. Nucleotides are colored by domain. The active site is highlighted in red boxes. The splice sites are highlighted with red arrows or as a red spheres in the tertiary structure. P7.1 and P9.1 helices are replaced by a dashed backbone in the tertiary structure for clarity.

tertiary structural motifs are smaller than secondary structures and, therefore, contribute proportionately less to the energetic stability of the RNA.

Not all RNAs follow this model of hierarchical folding. In a few well documented examples, especially stable tertiary structures fold concurrently with secondary structure[12-14]. These stable tertiary structures influence the final conformation of the folding secondary structure[15,16]. In some cases, the formation of a specific tertiary structure acts to rearrange the preformed stable secondary structure[17-19]. The study of RNA folding is a difficult undertaking and many RNA folding pathways have not been fully characterized. Therefore, the prevalence of non-hierarchal folding in RNA is unknown.

The energetic landscape of RNA folding often leads to structural intermediates and 'kinetic traps'[10,11]. Misfolded or trapped RNAs are disadvantageous to the cell. To bolster fast and accurate folding, RNAs have evolved peripheral domains[20,21] and have recruited protein facilitators[22-24].

1.1.2 Protein-Assisted RNA Folding. Large RNAs often recruit proteins to facilitate fast and accurate folding in the cell. Protein facilitators are classified into two general categories, chaperones and cofactors[22-24] (Figure 1.2). RNA chaperones use diverse mechanisms to rearrange RNA secondary structure[22,25]. In general, chaperones dissociate stable RNA structures to enable the RNA to re-fold and achieve its most stable structure. The interaction between chaperone and RNA is often non-specific and transient.

In contrast, cofactor proteins remain must stably bound to an RNA to create a functional RNP complex[23,25]. Cofactors often stabilize one specific tertiary structure motif from a population of diverse structures of equal thermodynamic stability. Cofactors recognize specific structures or sequences in an RNA and, therefore, bind one particular

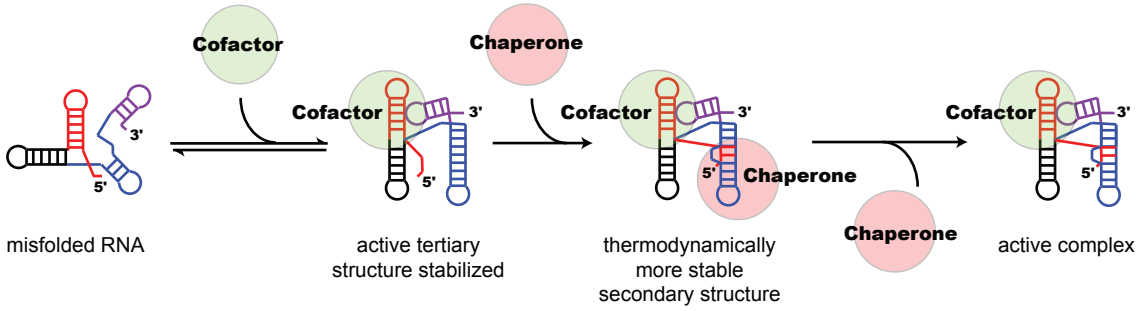

Figure 1.2. Protein-assisted RNA folding. Cofactor proteins (green circle) promote the formation of tertiary structure (the purple loop associates with the red helix) and must remain tightly bound to the RNA. Chaperones (red circle) actively rearrange RNA structure (the red RNA sequence pairs into the blue helix) to form a more stable RNA structure. The chaperone protein does not need to remain bound in the active complex.

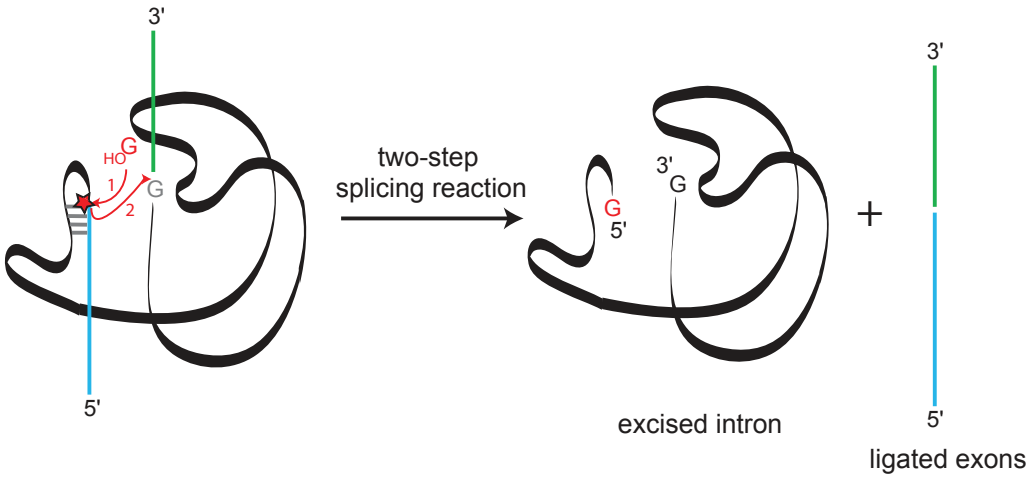

Figure 1.3. Splicing reaction of a group I inron. The structure of the intron RNA (black ribbon) positions an exogenous guanosine (red G) to attack the 5' splice site (red star). The resulting cleavage between the intron and 5' exon (blue line) induces a structural rearrangement so that the 5' exon is in position to attack the 3' splice site (black G). The resulting reaction ligates the exons (blue and green lines) and liberates the intron.

RNA or a small subset of similar RNAs. In the conventional model, RNA structure is not substantially altered upon cofactor binding. Thus, the RNA to returns to a mixture of conformations if the cofactor dissociates.

1.2 Group I Introns

Group I introns are self-splicing ribozymes that have inserted themselves into coding regions of genes[26]. Group I introns are commonly found in mRNAs, rRNAs and tRNAs of plant mitochondria and chloroplasts, but have also been found in the genomes of lower eukaryotes, bacteria and eubacteria[3].

Splicing involves two sequential transesterification reactions that are RNA catalyzed (Figure 1.3)[3,27,28]. In order to splice, group I introns form a guanosine binding pocket to bind an exogenous guanosine. Splicing is initiated when the 3′-hydroxyl of the exogenous guanosine cleaves the bond between the intron and the 5′-exon. After this first cleavage, the intron undergoes an obligate structural rearrangement. The 3′-splice site, termed ΩG, now occupies the active site[28,29]. The second transesterification reaction occurs as the 5′-exon attacks the 3′ splice site. This final step ligates the two exon sequences and liberates the intron.

1.2.1 Group I Intron Structure. Group I introns have a highly conserved structure essential for catalysis[1,20]. The structure of the intron core has been extensively characterized through phylogenetic sequence analysis[20], biochemical techniques[30,31], and high-resolution structures[7,32]. The core is composed of nine Watson-Crick base paired elements (termed Px) that are joined by single stranded regions (termed Jx/x) and capped by loops (termed Lx)[20] (Figure 1.1). These elements are arranged into three domains, the

P1-P2, the P5-P4-P6 and the P9-P7-P3-P8. The active site includes elements from all three domains. The 5′- and 3′-splice sites are located in the P1 and P9.0 helices, respectively. The 5′-splice site docks with two highly conserved A·A pairs in the P4 helix. The G binding pocket is composed of several base triplets in the P7 helix, which is part of a pseudoknotted structure with the P3 helix. Numerous tertiary interactions hold the three RNA domains together. Two highly conserved tertiary structure interactions are the GNRA tetraloop-receptor motifs formed between the L2 and L9 loops and their respective receptors in the P8 and P5 helices[20,33,34].

1.2.2 Proteins Associated with Group I Introns. A few minimal group I introns splice autonomously *in vitro*, however, most require protein assistance to achieve a catalytically active structure. Both cofactors and chaperones appear to facilitate group I intron splicing[25,35]. Cofactors can generally bind multiple introns through recognition of the catalytic core[36-38] or may be intron specific through recognition of idiosyncratic RNA elements[39-41]. Tight cofactor binding turns these self-splicing ribozymes into catalytically active ribonucleoproteins (RNPs). Chaperones, on the other hand, are non-specific and transient RNA binders which enable misfolded introns to achieve the conserved structure of the catalytic core[35,42,43].

1.3 The bI3 Ribonucleoprotein

In this research, I examine the protein assisted RNA folding of the *Saccharomyces cerevisiae* bI3 group I intron. This RNA is the third intron in the mitochondrial cytochrome *b* gene[44]. In order to splice *in vitro*, the bI3 RNA requires the tight binding of two proteins, the bI3 maturase and Mrs1[44-46]. The active complex is composed of six

subunits; the intron RNA, a monomer of the maturase and two Mrs1 dimers[47]. The complete bI3 RNP is ~420 KDa and splices with a k_{cat} of 0.3 min^{-1} [47]. Each protein binds the naked RNA without affecting subsequent binding of the other protein[48]. Therefore it has been assumed that each protein functions independently to facilitate splicing. Here, I provide evidence that the maturase and Mrs1 proteins bind distinct RNA elements, yet RNA splicing is a cooperative effort by the proteins. It is clear that separately each protein functions as a cofactor, but together, they cannot be classified as either cofactors or chaperones.

1.3.1 The bI3 Group I Intron RNA. The bI3 intron RNA is 372 nucleotides, approximately twice that of minimal self-folding introns. The bI3 RNA is comprised of the conserved elements of the group I intron structure, P1-P9, and four peripheral helices, P5abc, P7.1, P7.2 and P9.1 (Figure 1.1). In this research, I use a ΔL8 construct that replaces a 1,752 nucleotide protein encoding domain in the L8 loop of the intron with 6 nucleotides[44,47]. The RNA is AU rich (85% of nucleotides) which likely contributes to a higher propensity to form misfolded structures.

1.3.2 The bI3 Maturase. The bI3 maturase is encoded in L8 of the intron, in frame with the third exon[44]. It contains a LAGLIDADG-motif conserved in a large family of DNA homing endonucleases[49]. However, in its evolution into an RNA cofactor, the bI3 maturase has lost its ability to cleave DNA[47]. Additionally, the maturase has evolved from binding DNA major-grooves to binding RNA minor-grooves, specifically the idiosyncratic bI3 P5c helix[41]. The maturase binds tightly to the P5c helix as a monomer with a K_d of 1.0 nM[48]. Maturase binding stabilizes the entire P5-P4-P6 domain and, thus, results in the stabilization the intron active site, over 50 Å away.

1.3.3 Mrs1. The Mrs1 protein is encoded in the nucleus and targeted to yeast mitochondria[45,46]. It has evolved from DNA junction-resolving enzymes and, as such, forms a homodimer in solution[50,51]. Like the bI3 maturase, Mrs1 has lost its ability to cleave DNA[47]. Two Mrs1 dimers bind the bI3 RNA cooperatively with a $K_{½}$ of 11 nM and a Hill coefficient of 2.0 [47]. In the absence of magnesium, Mrs1 binds the bI3 RNA in two individual events, suggesting the existence of two distinct binding sites[48]. In this study, I identify the Mrs1 binding sites to be the two conserved tetraloop-receptor interactions: L2 with the P8 helix and L9 with the P5 helix.

1.4 RNA Structure Mapping

The first step in analyzing the protein assisted folding of the bI3 RNP was selecting structure probing techniques. In order to interrogate local nucleotide flexibility and the global RNA fold, I used SHAPE chemistry and hydroxyl radical footprinting, respectively (Figure 1.4). Both techniques are adaptable to a variety of reaction conditions, including the presence of proteins. Additionally, both techniques are insensitive to nucleotide identity and react with every position in an RNA. This allows every nucleotide in the ~540 nucleotide bI3 transcript to be characterized in a single high-throughput experiment. By using this combination of techniques, I comprehensively probed all levels of RNA structure in the bI3 RNP at nucleotide resolution.

1.4.1 SHAPE Chemistry. Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) chemistry is a robust and comprehensive method to analyze RNA secondary structure[13,52]. SHAPE chemistry targets the ribose 2′-hydroxyl, which is accessible on every nucleotide in an RNA. In a SHAPE reaction, the RNA is treated with

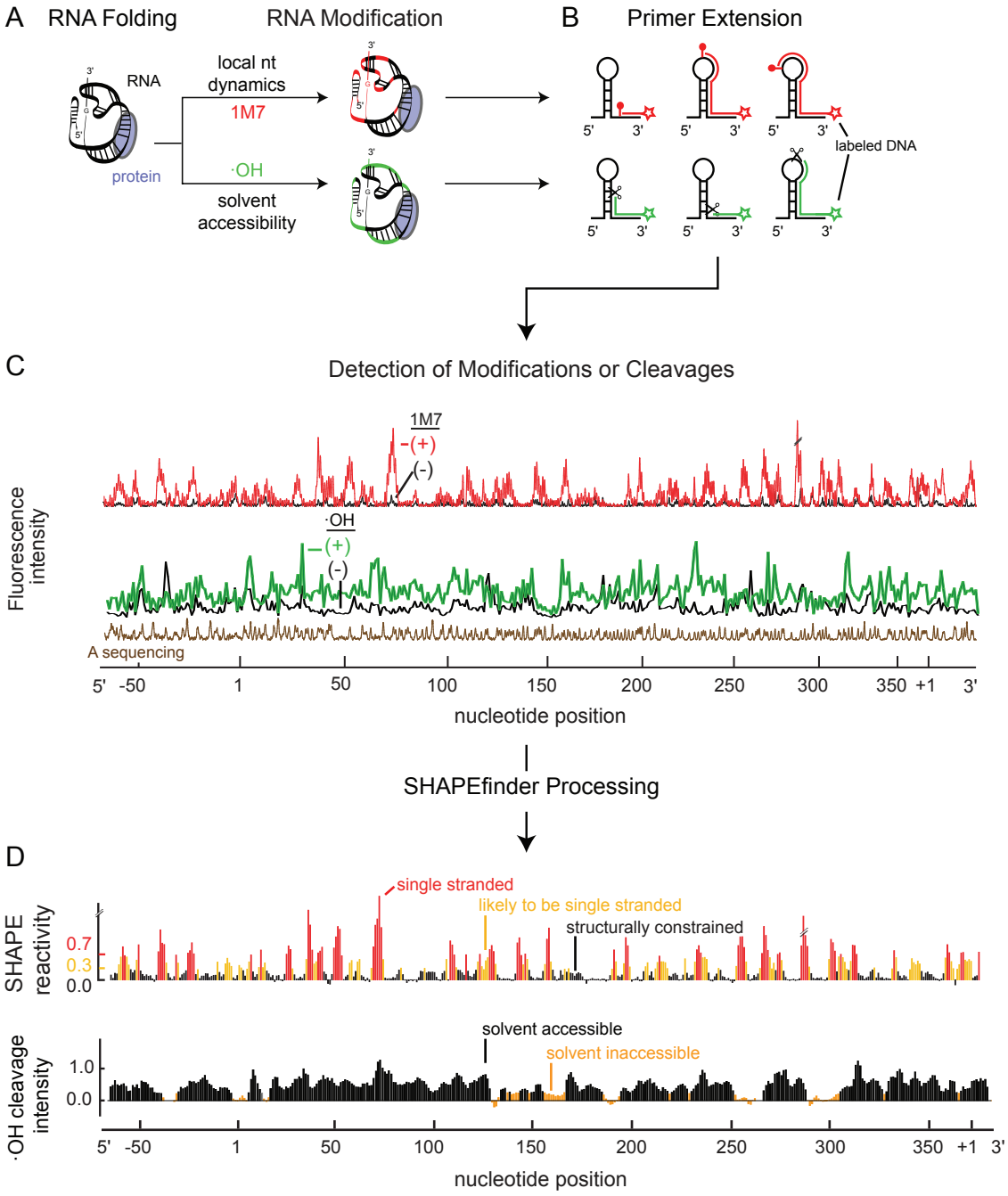

Figure 1.4. High-throughput RNA structure mapping. (A) The RNA is folded and modified with 1M7 or cleaved with hydroxyl radicals (·OH). (B) RNA modifications or cleavages result in stops to primer extension, which are detected using color-coded, fluorescently labeled primers. (C) Resulting cDNAs are separated by capillary electrophorsis. (D) After quantification of each peak, the data are normalized on a scale of 0 to ~2. Each nucleotide is assessed for structural flexibility and solvent accessibility.

an electrophilic reagent (1M7)[53] that selectively reacts with flexible nucleotides (Figure 1.4B). The nucleophilic strength of the ribose hydroxyl is gated by the local structure of the nucleotide. Nucleotides in base paired conformations react poorly with 1M7. Flexible nucleotides (those in single stranded conformations) are more likely to sample conformations that react with 1M7 to form 2′-*O*-adducts. These bulky adducts are detected as stops in a primer extension reaction. High-throughput experiments use fluorescently labeled primers and the resulting cDNAs are separated by capillary electrophoresis (Figure 1.4B,C). In addition to RNA modification, a no reagent control and sequencing experiments are performed concurrently. After processing, each nucleotide has a quantified SHAPE reactivity which correlates to structure. Here, single stranded nucleotides are reactive and base paired nucleotides are unreactive (Figure 1.4D). These SHAPE data can then be used as pseudo-energies to constrain the output of a thermodynamic structure prediction program to produce highly accurate RNA secondary structure models[54,55].

1.4.2 Hydroxyl Radical Footprinting. Hydroxyl radical footprinting can be used to estimate the solvent accessibility of the RNA backbone. Short-lived radicals are generated *in situ* in a reaction between Fe^{2+} and H_2O_2 [56,57]. The iron atom is chealated by a large EDTA moiety which sequesters the generation of hydroxyl radicals to the periphery of the RNA. Backbone cleavages occur in positions accessible to solvent, regardless of local nucleotide structure. High-throughput footprinting assays are analyzed as described above for SHAPE experiments. Here, cleavages are detected as fluorescent peaks and are inferred to be solvent accessible (Figure 1.4D).

1.5 Research Overview

My research goal was to characterize the protein-assisted folding of the bI3 RNA. I used SHAPE and hydroxyl radical footprinting to assess the secondary and tertiary structure of the RNA in distinct stages in the folding process.

1.5.1 The Structure of the Free RNA. When studying RNA-protein interactions, often the initial structure of an RNA is assumed or poorly characterized. On paper, the bI3 RNA is capable of folding into its phylogenetically predicted and active secondary structure. Moreover, both the bI3 maturase and Mrs1 proteins have characteristics of cofactors and, therefore, could be assumed to have a limited effect on secondary structure. In Chapter 2, however, I perform SHAPE experiments on the protein-free RNA and find that the data are not compatible with the phylogenetically determined active structure. I develop a new model for the free RNA by using SHAPE data as pseudo-energy constraints in a structure prediction program. In this model, roughly half the RNA is misfolded compared to the phylogenetic structure, including the P1 and P9.0 splice site helices, the P7-P3 pseudoknot, and the P7.1 and P9.1 peripheral helices. Additionally, I predict a novel paring interaction between the 5′ and 3′ exons.

I test this model of the free RNA by designing mutations in three elements, the exon stem, P1 and P9.1. SHAPE analysis of these mutated RNAs supports the formation of the alternate interactions predicted in the free RNA model. In addition, I subject the mutant RNAs to splicing assays and determine that the alternate structure is not catalytically active. I then perform a SHAPE experiment on the bI3 RNA bound by both the maturase and Mrs1 proteins. Significant changes in SHAPE reactivity indicate that the bI3 RNA undergoes dramatic structural rearrangements to achieve a catalytically active

conformation. Furthermore, protein dissociation produces a relaxation of the RNA to a structural state similar to the free RNA. I infer that the active structure of the bI3 RNA is not the thermodynamically most stable structure in the absence of proteins.

1.5.2 Locating the Mrs1 Binding Site. Knowledge of the protein binding site is a crucial factor in understanding the role a protein plays in RNA folding. In Chapter 3, I identify two tetraloop-receptor interactions as the binding site for the Mrs1 protein on the bI3 RNA. Initially, I compare hydroxyl radical data for the free RNA to that of the Mrs1-bound RNA. I find significant differences in cleavage intensities localized to the two distinct GNRA tetraloop-receptor interactions in the bI3 RNA. These tetraloop-receptor interactions are the interaction of the L2 loop with the P8 helix and the L9 loop with the P5 helix (indicated with arrows on the secondary and tertiary structures in Figure 1.2). I then test the ability of Mrs1 to bind tetraloop-receptor motifs in other RNAs. I show that Mrs1 binds tetraloop-receptor motifs with a similar protection pattern in both the *Azoarcus* and RNase P RNAs.

The location of the Mrs1 proteins provided the last piece of information needed to construct a model of the bI3 holoenzyme. I first construct models of the bI3 RNA and Mrs1 protein. Combined with the crystal structure of the bI3 maturase, I develop a working model of the bI3 RNP. Strikingly, every long-range tertiary interaction is stabilized by a protein. In addition, no protein comes within a 20 Å radius of the scissile phosphate or within 50 Å of any other protein.

1.5.3 Cooperative Folding of the bI3 RNP. In Chapter 4, I analyze the protein assisted folding of the bI3 RNA by structurally interrogating the four bound states of the RNA: free, maturase-bound, Mrs1-bound and the final six-component complex. SHAPE

and hydroxyl radical experiments indicate that the maturase and Mrs1 proteins primarily stabilize RNA tertiary structures. However, the effect of protein binding is complex. Both techniques detect structural fluctuations or 'torsional stress' propagating through the RNA upon binding by either protein. In the final-complex, the RNA achieves a dramatically different secondary structure compared to the free RNA. Structural rearrangements occur distal from the protein binding sites. I infer that secondary structure rearrangements in the bI3 RNA is a cooperative effect that results from binding by both the maturase and Mrs1 proteins.

1.5.4 Perspective. The cooperative rearrangement of secondary structure in the bI3 RNA leads to an ambiguous classification of the bI3 maturase and Mrs1 proteins. Individually they are traditional cofactors; each tightly binds the bI3 RNA to stabilize specific tertiary structures. Together, however, they induce a large-scale structural rearrangement, which would be considered chaperone-like if the structure formed was thermodynamically more stable. The active secondary structure is not the thermodynamically most stable structure, as evidenced by the relaxation of the RNA upon protein dissociation. Additionally, these structural rearrangements occur in RNA structures that do not directly contact the proteins, which is uncharacteristic of either chaperones or cofactors. These studies imply that general classifications of RNA-protein interactions are not sufficient to make assumptions about the RNA structure or protein function.

1.6 References

1. Gesteland, R.F., Cech, T.R. & Atkins, J.F. *The RNA World*, (Cold Spring Harbor Laboratory Press, New York, 2006).

2. Noller, H.F., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416-9 (1992).

3. Saldanha, R., Mohr, G., Belfort, M. & Lambowitz, A.M. Group I and group II introns. *FASEB J.* **7**, 15-24 (1993).

4. McKay, D.B. Structure and function of the hammerhead ribozyme: an unfinished story. *RNA* **2**, 395-403 (1996).

5. Smith, J.K., Hsieh, J. & Fierke, C.A. Importance of RNA-protein interactions in bacterial ribonuclease P structure and catalysis. *Biopolymers* **87**, 329-38 (2007).

6. Pley, H.W., Flaherty, K.M. & McKay, D.B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74 (1994).

7. Golden, B.L., Gooding, A.R., Podell, E.R. & Cech, T.R. A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* **282**, 259-64 (1998).

8. Krasilnikov, A.S., Yang, X., Pan, T. & Mondragon, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760-4 (2003).

9. Toor, N., Keating, K.S., Taylor, S.D. & Pyle, A.M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77-82 (2008).

10. Brion, P. & Westhof, E. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113-37 (1997).

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

12. Gluick, T.C. & Draper, D.E. Thermodynamics of folding a pseudoknotted mRNA fragment. *J. Mol. Biol.* **241**, 246-62 (1994).

13. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in $tRNA^{Asp}$ transcripts. *J. Am. Chem. Soc.* **127**, 4659-67 (2005).

14. Chauhan, S. et al. RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. *J. Mol. Biol.* **353**, 1199-209 (2005).

15. LeCuyer, K.A. & Crothers, D.M. The Leptomonas collosoma spliced leader RNA can switch between two alternate structural forms. *Biochemistry* **32**, 5301-11 (1993).

16. Chauhan, S. & Woodson, S.A. Tertiary interactions determine the accuracy of RNA folding. *J. Am. Chem. Soc.* **130**, 1296-303 (2008).

17. Wu, M. & Tinoco, I., Jr. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11555-60 (1998).

18. Silverman, S.K., Zheng, M., Wu, M., Tinoco, I., Jr. & Cech, T.R. Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA* **5**, 1665-74 (1999).

19. Zheng, M., Wu, M. & Tinoco, I., Jr. Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the Tetrahymena group I ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3695-700 (2001).

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

21. Johnson, T.H., Tijerina, P., Chadee, A.B., Herschlag, D. & Russell, R. Structural specificity conferred by a group I RNA peripheral element. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10176-81 (2005).

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

23. Weeks, K.M. Protein-facilitated RNA folding. *Curr. Opin. Struct. Biol.* **7**, 336-42 (1997).

24. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5**, 908-19 (2004).

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

26. Lang, B.F., Laforest, M.J. & Burger, G. Mitochondrial introns: a critical view. *Trends Genet* **23**, 119-25 (2007).

27. Cech, T.R. The chemistry of self-splicing RNA and RNA enzymes. *Science* **236**, 1532-9 (1987).

28. Cech, T.R. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* **59**, 543-68 (1990).

29. Adams, P.L. et al. Crystal structure of a group I intron splicing intermediate. *RNA* **10**, 1867-87 (2004).

30. Burke, J.M. Molecular genetics of group I introns: RNA structures and protein factors required for splicing--a review. *Gene* **73**, 273-94 (1988).

31. Cech, T.R. Conserved sequences and structures of group I introns: building an active site for RNA catalysis--a review. *Gene* **73**, 259-71 (1988).

32. Golden, B.L., Kim, H. & Chase, E. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.* **12**, 82-9 (2005).

33. Woodson, S.A. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**, 324-30 (2005).

34. Vicens, Q. & Cech, T.R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**, 41-51 (2006).

35. Mohr, G., Caprara, M.G., Guo, Q. & Lambowitz, A.M. A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the Tetrahymena ribozyme. *Nature* **370**, 147-50 (1994).

36. Bousquet, I., Dujardin, G., Poyton, R.O. & Slonimski, P.P. Two group I mitochondrial introns in the cob-box and coxI genes require the same MRS1/PET157 nuclear gene product for splicing. *Curr. Genet.* **18**, 117-24 (1990).

37. Weeks, K.M. & Cech, T.R. Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* **271**, 345-8 (1996).

38. Rho, S.B., Lincecum, T.L., Jr. & Martinis, S.A. An inserted region of leucyl-tRNA synthetase plays a critical role in group I intron splicing. *EMBO J.* **21**, 6874-81 (2002).

39. Valcarcel, J. & Gebauer, F. Post-transcriptional regulation: the dawn of PTB. *Curr. Biol.* **7**, R705-8 (1997).

40. Solem, A., Chatterjee, P. & Caprara, M.G. A novel mechanism for protein-assisted group I intron splicing. *RNA* **8**, 412-25 (2002).

41. Longo, A. et al. Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* **12**, 779-87 (2005).

42. 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).

43. 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).

44. Lazowska, J. et al. Protein encoded by the third intron of cytochrome b gene in Saccharomyces cerevisiae is an mRNA maturase. Analysis of mitochondrial mutants, RNA transcripts proteins and evolutionary relationships. *J. Mol. Biol.* **205**, 275-89 (1989).

45. Kreike, J., Schulze, M., Pillar, T., Korte, A. & Rodel, G. Cloning of a nuclear gene MRS1 involved in the excision of a single group I intron (bI3) from the mitochondrial COB transcript in S. cerevisiae. *Curr. Genet.* **11**, 185-91 (1986).

46. Kreike, J., Schulze, M., Ahne, F. & Lang, B.F. A yeast nuclear gene, MRS1, involved in mitochondrial RNA splicing: nucleotide sequence and mutational analysis of two overlapping open reading frames on opposite strands. *EMBO J.* **6**, 2123-9 (1987).

47. Bassi, G.S., de Oliveira, D.M., White, M.F. & Weeks, K.M. Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 128-33 (2002).

48. Bassi, G.S. & Weeks, K.M. Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst. *Biochemistry* **42**, 9980-8 (2003).

49. Dalgaard, J.Z. et al. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.* **25**, 4626-38 (1997).

50. Wardleworth, B.N., Kvaratskhelia, M. & White, M.F. Site-directed mutagenesis of the yeast resolving enzyme Cce1 reveals catalytic residues and relationship with the intron-splicing factor Mrs1. *J. Biol. Chem.* **275**, 23725-8 (2000).

51. Lilley, D.M. & White, M.F. Resolving the relationships of resolving enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9351-3 (2000).

52. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223-31 (2005).

53. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144-5 (2007).

54. Wilkinson, K.A. et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, 883-899 (2008).

55. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 97-102 (2009).

56. Latham, J.A. & Cech, T.R. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276-82 (1989).

57. Tullius, T.D. & Greenbaum, J.A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127-34 (2005).

Chapter 2

SHAPE Analysis of Long-Range Interactions Reveals Extensive and Thermodynamically Preferred Misfolding in a Fragile Group I Intron RNA

2.1 Introduction

Large RNA molecules fold into intricate three-dimensional structures that are essential for many fundamental biological processes, including translation, mRNA processing, and viral replication[1]. In almost all cases, large RNAs achieve their functional structures only through interactions with a large variety of protein facilitators. Two prominent and widely applicable models emphasize that protein facilitators function either to bind stably to an RNA and, thereby, promote formation of the active tertiary structure, or to interact relatively transiently and non-specifically to facilitate rearrangement of incorrect secondary structures[2-4]. These proteins are commonly called cofactors and chaperones, respectively. In some cases, RNA folding and function may require facilitation by both classes of proteins.

An assumption implicit in the mechanisms proposed for both cofactors and chaperones is that the functionally active secondary structure is also the thermodynamically most stable secondary structure. Here, I introduce a system that does not conform to either of these models but, instead, is one in which the most stable secondary structure for the free RNA is very different from the catalytically active structure. Thus, one role for protein facilitators must be to promote large-scale rearrangements in both RNA secondary and tertiary structure.

Group I introns are good models for complex RNA folding reactions. These catalytic RNAs are composed of two roughly coaxially stacked domains that interact through extensive tertiary interactions to form a highly structured catalytic core and dock with a third helical domain to form a sophisticated active site[5-8]. Binding by a guanosine cofactor initiates a two-step splicing reaction to yield ligated exons and the excised intron[6].

The self-splicing activity of group I introns also provides a robust readout for correct RNA folding.

Although a few group I introns splice autonomously, most introns in this class require the participation of protein facilitators to fold into a catalytically active structure that splices efficiently. The yeast bI3 group I intron holoenzyme is a large and complex ribonucleoprotein (RNP) composed of the bI3 intron RNA and two proteins, the bI3 maturase and Mrs1 (Figure 2.1A). The active complex consists of six components, one RNA, the intron encoded bI3 maturase, and two dimers of the nuclear encoded Mrs1 protein[9-11]. The RNA appears to require that both proteins be stably bound in order to be catalytically active. The bI3 RNP is, thus, an excellent model for understanding the mechanism by which proteins facilitate RNA folding.

An obligatory first step in understanding any RNP assembly reaction is an accurate understanding of the initial state of the free RNA, prior to protein binding. This free RNA state often lacks significant tertiary structure and has been difficult to characterize fully for any large RNA. A useful starting point for developing models of RNA folding and RNP assembly reactions is to posit that the structure of the free RNA reflects a loose version of the same secondary structure present in the final, native complex[3,4,12,13]. In some cases, the free RNA appears to undergo small, local secondary structure rearrangements to form the native state[14-16]. However, the widespread assumption that the secondary structure of a free RNA closely resembles that of the native state has not been broadly tested. Furthermore, even when it is clear that a large fraction of an RNA misfolds relative to the catalytically active structure, very little information is available regarding the specific alternative secondary structures that populate the misfolded state.

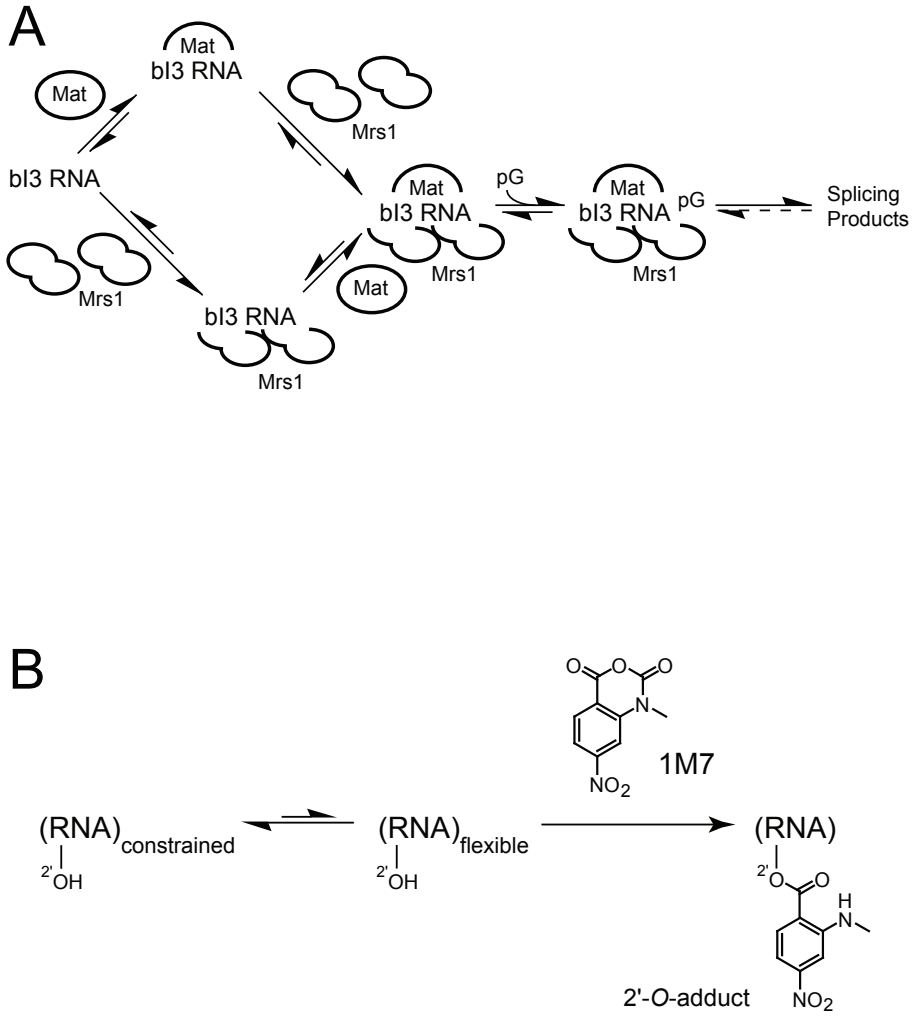

Figure 2.1. Schemes for (A) assembly of the bI3 ribonucleoprotein complex and (B) RNA SHAPE chemistry.

A powerful approach for analyzing RNA secondary structure is single nucleotide resolution RNA SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) chemistry[17,18]. SHAPE chemistry monitors local flexibility at every nucleotide in an RNA using an hydroxyl selective electrophile to probe the reactive state of the ribose 2'-OH group. The 2'-OH is accessible to solution in most RNA conformations, especially prior to formation of tertiary interactions. SHAPE does not monitor solvent accessibility but, instead, the propensity of a nucleotide to adopt relatively rare, but highly reactive, conformations that increase the nucleophilic reactivity of the 2'-hydroxyl group[17,19]. Thus, SHAPE chemistry maps local nucleotide flexibility in RNA. Sites of 2'-*O*-adduct formation, and thus conformationally flexible RNA nucleotides, are then detected as stops in a primer extension reaction. cDNA fragments generated during the primer extension step are resolved by capillary electrophoresis using fluorescently labeled primers in a high-throughput SHAPE experiment[19,20].

High-throughput SHAPE has two attributes that are important when analyzing the structure of a large RNA like the bI3 group I intron. First, I routinely obtain reads of over 450 nucleotides such that the entire bI3 RNA structure can be interrogated in a single experiment. Second, I obtain local nucleotide flexibility information at single nucleotide resolution throughout the read, which means that the effects of single point mutations can be evaluated at all possible structures in the RNA, also in a single experiment.

Using single nucleotide resolution SHAPE analysis, I find that the structure of the bI3 RNA prior to binding by the maturase and Mrs1 protein facilitators is dramatically different from its phylogenetically determined secondary structure. This alternatively folded and catalytically inactive structure is nonetheless the thermodynamically most

stable structure. SHAPE experiments show that the bI3 maturase and Mrs1 cofactors function to promote large-scale rearrangements in the secondary structure as well as to stabilize the catalytically active tertiary structure of the RNA.

2.2 Results

2.2.1 Free bI3 RNA Structure at Single Nucleotide Resolution. An absolute prerequisite for understanding the role of protein facilitators in any RNP complex is an accurate model for the structure of the RNA prior to protein binding. I analyzed the structure of the free bI3 RNA using high-throughput RNA SHAPE technology. In this two step process, I first treated the RNA with 1-methyl 7-nitroisatoic anhydride (1M7)[20] (Figure 2.1B). Subsequently, sites of 2'-*O*-adduct formation were detected by primer extension using fluorescently labeled primers, resolved by capillary electrophoresis. Sites of specific 2'-*O*-adducts were located by comparison with a reaction performed in parallel in which the reagent was omitted (Figure 2.2A, compare red and black traces). Sites of modification were identified by comparison with a dideoxy nucleotide sequencing reaction.

I obtain the absolute SHAPE reactivities for every nucleotide in the RNA by integrating the peaks in the (+) and (–) reagent traces, calculating the net reactivity after subtracting background, and placing all intensities on a normalized scale. On this scale, nucleotides with normalized intensities less than 0.3 are defined as unreactive; whereas, nucleotides with reactivities greater than 0.7 are deemed highly reactive (Figure 2.2B, black columns and red columns, respectively). Nucleotides with reactivities between 0.3 and 0.7 are judged to be moderately reactive (Figure 2.2B, yellow columns). In a typical

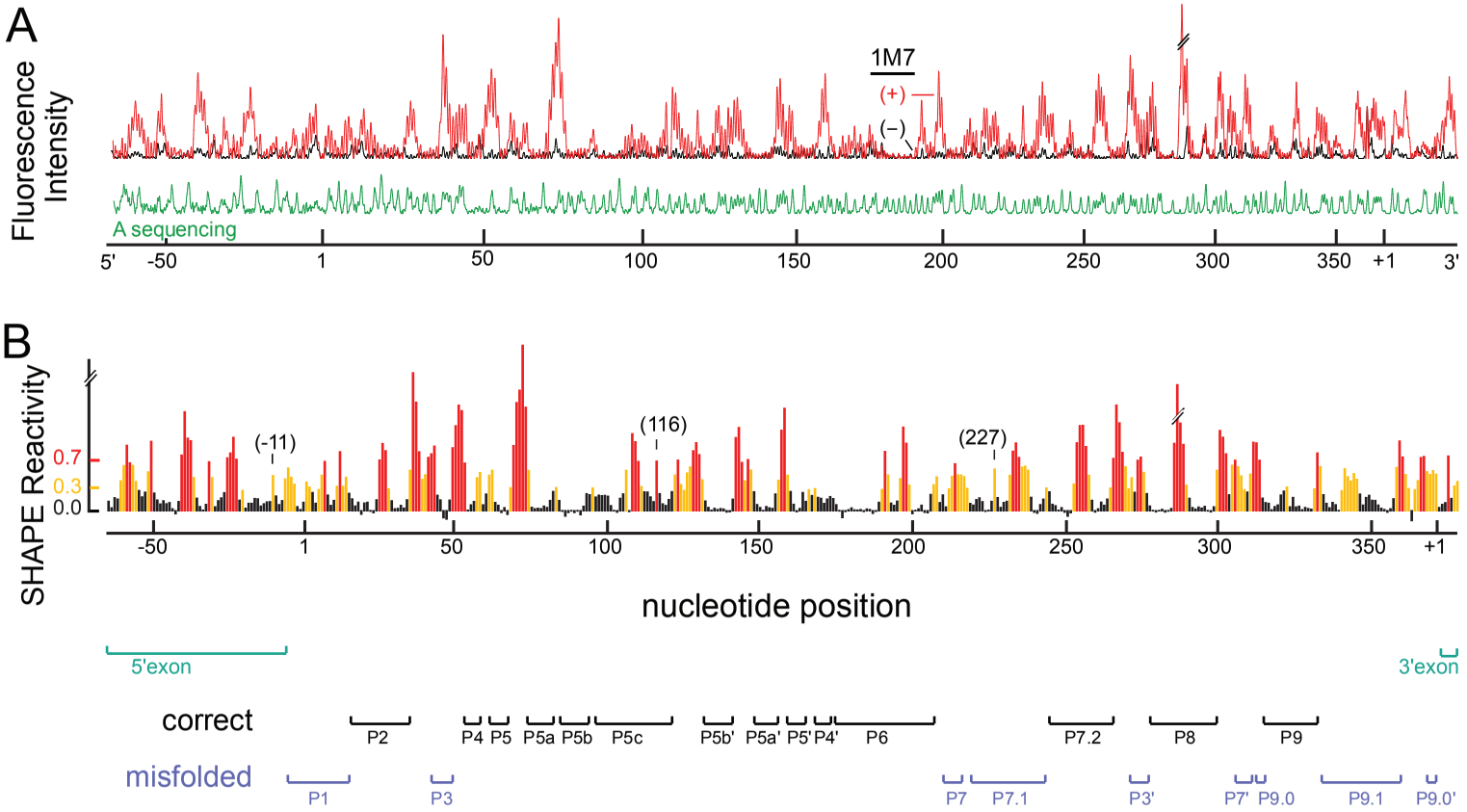

Figure 2.2 SHAPE analysis of the bI3 group I intron over a single 480 nucleotide read. (A) Processed fluorescence intensities versus position resolved by capillary electrophoresis. (+) and (–) 1M7 reagent traces are red and black, respectively; the sequencer marker is green. (B) Histogram of scaled SHAPE reactivities. Positions with high, moderate and low reactivities are shown in red, yellow and black, respectively. Bars below the histogram identify paired elements in the RNA. SHAPE reactivities that are incompatible with phylogenetically conserved secondary structure are emphasized in blue; exons are green.

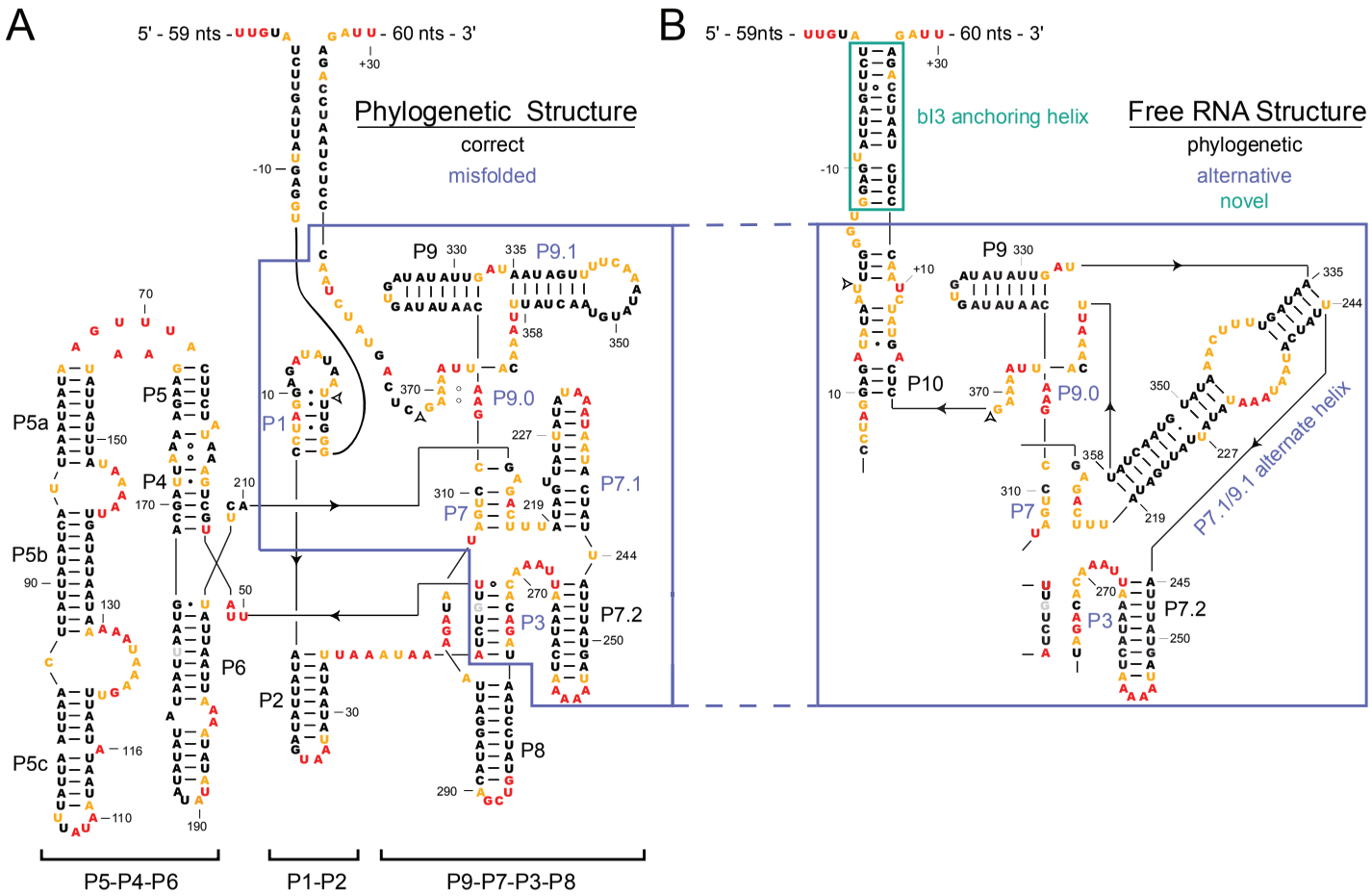

Figure 2.3. SHAPE reactivities superimposed on (A) phylogenetic and (B) free bI3 secondary structure models. Nucleotides are colored (red, yellow and black) according to their SHAPE reactivities using the same scale as in Figure 2B. Paired elements (Px) are labeled in black or blue depending on whether SHAPE reactivities are compatible with the secondary structure determined by comparative sequence analysis. Splice sites are denoted with open arrows.

single SHAPE experiment, I analyze the entire 372 nucleotides of the bI3 intron, 73 nucleotides of the 5' exon, and 30 nucleotides of the 3' exon for a total read length of 475 nucleotides (Figure 2.2).

2.2.2 Free bI3 RNA Does Not Fold into the Phylogenetically Conserved or Catalytically Active Structure. The SHAPE reactivity data indicate that many regions of the free bI3 RNA fold in a way that is consistent with the group I intron structure determined by comparative sequence analysis[5] (emphasized with black bars, bottom of Figure 2.2B). For example, in the P5-P4-P6 domain, all base paired regions are unreactive, whereas nucleotides which connect these regions are reactive (compare the reactive red and yellow nucleotides with the unreactive black nucleotides, Figure 2.3A). Other structures that appear to form the phylogenetically conserved and catalytically active secondary structure include P2, P8 and P9 (all such elements are identified with black labels, Figure 2.3A).

In contrast, I also identify extensive regions in which the SHAPE data do not recapitulate the phylogenetically determined structure. One clear example is the P1 helix, which contains the 5' splice site. SHAPE reactivities are inconsistent with formation of P1 in two ways. First, many nucleotides in the P1 stem are reactive and therefore conformationally flexible (Figure 2.2B). Second, superimposition of these reactivities on the phylogenetic secondary structure juxtaposes reactive bases with unreactive base pairing partners (see P1 helix, Figure 2.3A).

A second example of disagreement between SHAPE data and the catalytically active structure occurs in the conserved helical structures of the P3 and P7 helices. Many nucleotides in these helices are reactive (Figure 2.2B). When these SHAPE reactivities are

superimposed on the secondary structure, again, reactive nucleotides are formally paired with unreactive nucleotides (Figure 2.3A).

I also identify well defined regions in the 5' and 3' exons in which nucleotides are unreactive towards SHAPE chemistry. These unreactive nucleotides have the potential to form a structure which I call the bI3 anchoring helix (Figure 2.3B, green box). I will explicitly test this model and the broad ability of SHAPE to predict new structures below.

In sum, single nucleotide SHAPE analysis shows that about one-half the bI3 RNA folds into the conserved group I intron secondary structure as established by comparative sequence analysis and confirmed by crystallographic analysis[5,8]. These regions include the P5-P4-P6 domain and the P2, P8, P9, and P7.2 helices (identified with black labels, Figures 2.2B and 2.3A). In contrast, the SHAPE clearly shows the remaining half of the RNA folds into a structure that is incompatible with the phylogenetic structure (emphasized with bold blue labels, Figures 2.2B and 2.3A).

2.2.3 Structural Model for the Free bI3 RNA. Given that the SHAPE reactivities are inconsistent with the phylogenetically established secondary structure, I sought to create a new and accurate secondary structure model for the free bI3 RNA. I used a new experimentally-constrained secondary structure prediction algorithm to develop plausible, and testable, models for the bI3 RNA structure. In this approach, SHAPE reactivities are used as pseudo-free energy change constraints to augment lowest free energy predictions based on thermodynamic parameters alone. Reactive nucleotides are penalized if they are base paired, whereas unreactive nucleotides receive a modest energetic bonus for being paired.

My working model for the free bI3 RNA (Figure 2.3B) has three important features. First, regions of the RNA whose SHAPE reactivities were consistent with the phylogenetic structure, like the P5-P4-P6 domain, were directly predicted by the SHAPE pseudo-energy approach. Second, SHAPE reactivities strongly support the formation of an anchoring helix that links the intron to the flanking mRNA sequences (highlighted by the green box, Figure 2.3B). Third, SHAPE data are consistent with a model in which the remaining sequences form two distinct elements that are misfolded relative to the catalytically active structure. The P7.1 and P9.1 helices form an extended pairing interaction, which I call the P7.1/9.1 alternate helix, and sequences in P1 base pair with sequences in the 3' exon to form a helical structure that contains the predicted three base pair P10 helix (formed after splicing of the 5' splice site)[5,6] and also additional base pairing interactions (Figure 2.3B). SHAPE data are now almost exactly consistent with this alternate secondary structure model. For example, nucleotides proposed to pair in the P7.1/9.1 alternate helix are unreactive; whereas, nucleotides in the central internal loop and at the bulged U^{227} position are reactive. Similarly, unreactive nucleotides in the 3' exon are now predicted to base pair, while reactive nucleotides, like U^{+10}, remain in unpaired bulge structures (compare boxed regions of Figure 2.3A with 2.3B).

2.2.4 Mapping Long-Range RNA Interactions by SHAPE Analysis of Point Mutations. SHAPE reactivity information is clearly sufficient to falsify incorrect RNA secondary structure models, including many regions of the bI3 group I intron. Using this same SHAPE information, I have been able to propose an alternate secondary structure that is consistent with SHAPE reactivities (Figure 2.3B). A striking feature of SHAPE chemistry is its ability to unambiguously detect single nucleotide features in RNA at high

signal to noise. For example, each of the single nucleotide bulges at positions -11, 116 and 227 in the bI3 intron are readily detected as nucleotides with reactivities above that of their surrounding base paired nucleotides (see labeled peaks, Figure 2.2B). I, therefore, developed a new method, involving SHAPE analysis of point mutations, to directly assess the validity of the alternate model for the bI3 RNA.

2.2.5 The bI3 Anchoring Helix. I initially applied SHAPE analysis of point mutations to characterize the proposed bI3 anchoring helix (Figure 2.4A). I focused on the bulge at position -11 where the overall pattern of SHAPE reactivity is exactly consistent with the proposed secondary structure. U^{-11} comprises a reactive single-nucleotide bulge in the middle of a largely structured region (black trace, Figure 2.4C). Based on this model, I designed two single nucleotide exon mutants (termed E1 and E2) to test the long-range base pairing interactions that form in the proposed anchoring helix. An inserted A (E1) between nucleotides +17 and +18 should base pair with U^{-11} and eliminate the bulge; whereas, a double mutation (E2) creates a new bulge on the opposite strand of this same helix (Figure 2.4A, red and blue arrows respectively). Both RNA mutants, in the presence of protein cofactors and guanosine, spliced with kinetics that were indistinguishable from the native RNA (Figure 2.4B and Table 2.1).

Over the entire 547 nucleotide RNA, SHAPE reactivity patterns for both mutants were essentially identical to the native RNA (see comparison of entire mutant traces with the native sequence RNA, Figure 2.4D). For the E1 mutant, the only significant SHAPE reactivity difference greater than 2-fold was a strong decrease in reactivity precisely at nucleotide U^{-11} (compare red and black traces, Figure 2.4C). Analogously, overall reactivity patterns for the E2 double mutant are identical to the E1 mutant with the

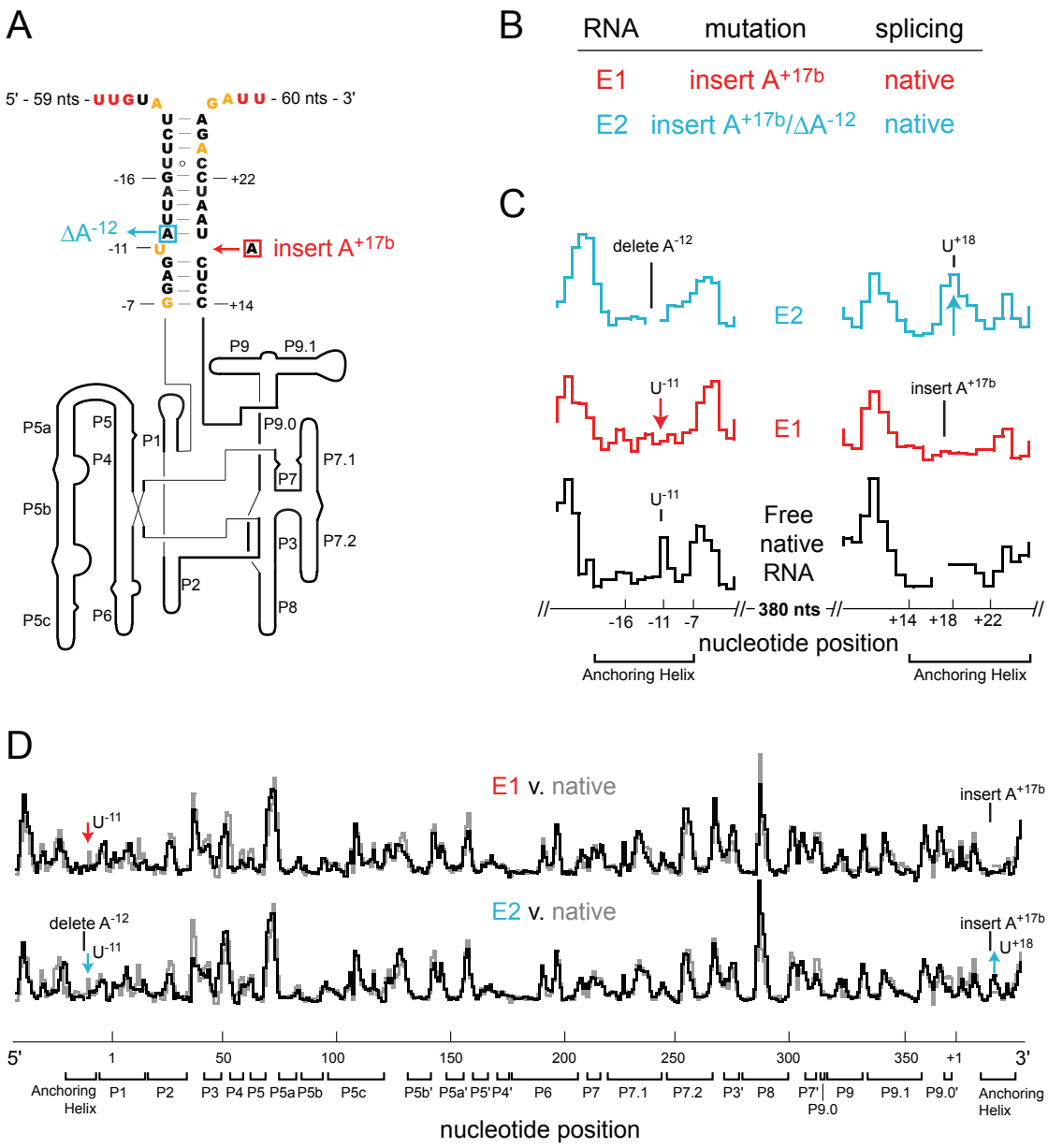

Figure 2.4. Long-range interactions in the bI3 anchoring helix, confirmed by SHAPE analysis of point mutations. (A) Secondary structure model of the bI3 anchoring helix (nucleotides are colored as in Figure 3). Point mutations are highlighted with colored boxes. (B) RNA mutations and their effect on splicing. (C) Expanded SHAPE reactivity histograms for native and mutant RNAs as a function of nucleotide position. Sites of strong reactivity increases or decreases in the mutants, as compared with the native sequence RNA, are emphasized with colored arrows. (D) Histograms of the complete SHAPE reactivity profiles showing that changes in reactivity for the E1 and E2 mutants (black traces) are precisely localized to the anchoring helix region. Complete profiles for all RNAs are provided in Figure 2.9.

exception that there is an additional increase in nucleotide reactivity centered at position +18 (compare red and blue traces, Figure 2.4C).

These experiments show that SHAPE analysis of single point mutations robustly identifies long-range base pairing interactions involving sequences that lie ~400 nucleotides apart (Figure 2.4D). Moreover, SHAPE analysis of point mutants strongly supports the model that the bI3 intron is linked to the flanking mRNA sequences via a stable anchoring helix.

2.2.6 Disruption of Pairing in the Anchoring Helix Interferes with Intron Folding. I next explored the functional role of the bI3 anchoring helix by mutating the base paired exon regions neighboring the intron sequences. The first two mutants (E3 and E4) disrupt the first four base pairs of the anchoring helix (in magenta and blue, Figure 2.5A). The third mutant (E5) contains both E3 and E4 mutations and yields exon sequences that could, in principle, reform the anchoring helix. Together these mutants test, first, whether a stable mRNA exon structure facilitates intron folding and catalysis, and second, if the sequence itself is important.

I tested the ability of the three mutants to form the catalytically active structure using splicing assays performed in the presence of both bI3 maturase and Mrs1 proteins. Approximately one half of the native bI3 RNA splices in a fast phase characterized by k_{obs} at 0.26 min^{-1} with the remaining RNA splicing in a slow phase of 0.006 min^{-1}, consistent with previous work[9]. In contrast, the three anchoring helix mutants each exhibited compromised splicing profiles (Figure 2.5B and Table 2.1). The poor splicing activity observed for the E3 and E4 mutants emphasizes that base pairing in the anchoring helix plays a functional role in bI3 RNA folding and splicing. Moreover, the E5 double mutant

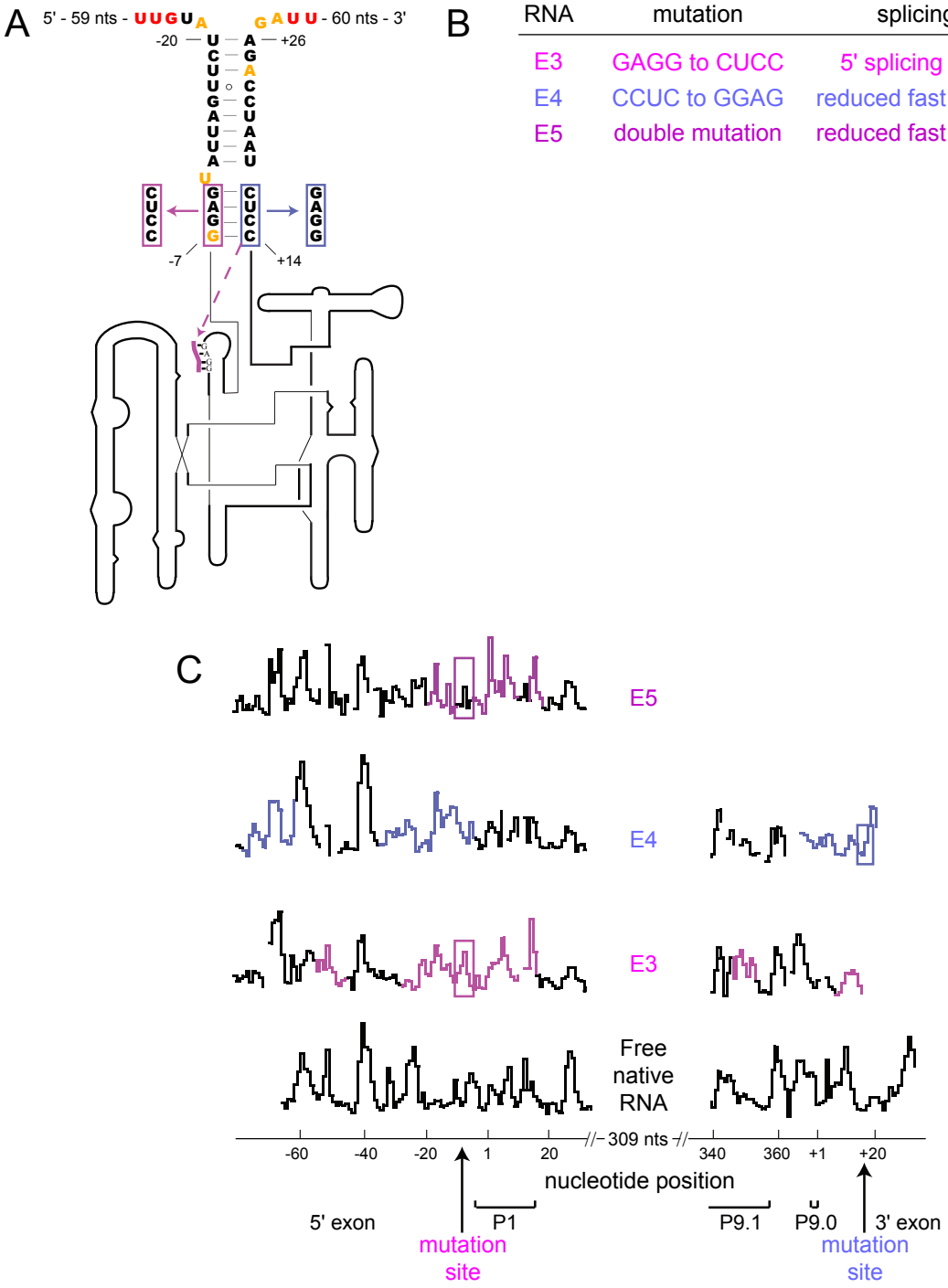

Figure 2.5. SHAPE analysis of base pairing mutations in the bI3 anchoring helix. (A, B) Mutations and their splicing activities. The dashed line in (A) indicates an alternate pairing proposed for the E3 mutation. (C) SHAPE reactivities versus position for the native and mutant RNAs. Regions in the RNA mutants that show the largest differences in SHAPE reactivities as compared to the native RNA are emphasized in color. Regions corresponding to mutated nucleotides are boxed.

did not restore native anchoring helix function, indicating that both the structure and sequence of the anchoring helix are important for achieving a catalytically active RNA structure.

As judged by SHAPE reactivities, the E3 mutation produced extensive changes as compared to the native RNA sequence. The mutated nucleotides show increased reactivity, as expected. However, their former base pairing partners remain unreactive, suggesting that these nucleotides form new non-native interactions in the RNA (compare magenta regions of the E3 trace with the native trace, Figure 2.5C). The complete SHAPE reactivity profiles for all mutants and the native sequence RNA are provided in the Supporting Information (Figure 2.9). Together, these changes are consistent with a structural rearrangement in the 5' exon such that the 3' exon pairs with the P1 helix and is accompanied by extensive RNA misfolding in the 5' exon and in the P1, P2, P7.1 and P9.1 helices (dashed magenta line, Figure 2.5A; and see Figure 2.9).

Differences in SHAPE reactivities between the E4 mutant and the native RNA are largely confined to the exons but also include the J3/4, P4 and J6/7 elements (Figure 2.9). The exon regions in the E4 mutant have substantially altered SHAPE reactivity profiles as compared to the native RNA (note the dramatic differences in SHAPE profiles highlighted in blue, Figure 2.5C). The overall pattern of SHAPE reactivity is consistent with significant structural rearrangements in both the 5' and 3' exons.

In principle, the E5 mutant had the potential to restore a native-like base pairing pattern to the anchoring helix. In contrast to this expectation, the SHAPE reactivity profile shows that exon structure in this mutant is significantly different from the native sequence

RNA (Figure 2.5C, compare purple regions of the E5 trace with the black native trace): the E5 sequence thus induces significant misfolding in the exons.

I conclude from these experiments that the bI3 anchoring helix is important both in structure and in sequence to allow the intron to fold to its native and catalytically active structure. Moreover, comparison of results obtained with point mutations (Figure 2.4) versus with the 4-8 nucleotide mutations in the E3-E5 RNAs (Figure 2.5) emphasizes that even four nucleotide mutations can induce unexpectedly large changes to an RNA structure (summarized in Figure 2.9). SHAPE analysis of single point mutations thus represents a novel and robust approach for analyzing long-range structure in large RNAs.

2.2.7 The P7.1/9.1 Alternate Helix. I next used SHAPE analysis of point mutations to probe the proposed P7.1/9.1 alternate helix (Figure 2.3B). My strategy was to analyze mutants that afford distinct changes in SHAPE reactivity depending on whether the phylogenetic or alternate structure was formed.

My first target was the reactive nucleotide U^{227}. U^{227} forms a single nucleotide bulge in the P7.1/9.1 alternate helix model, but is predicted to pair with the unreactive A^{136} in the phylogenetic model (compare U^{227} in upper and lower panels of Figure 2.6A). The insertion of a single A between nucleotides 350 and 351 (mutant A1) provides a pairing partner for, and should decrease the reactivity of, U^{227} in the alternate helix model.

The SHAPE reactivity profile of the A1 mutant was virtually identical to native RNA over 400 nucleotides of intron and exon sequences except for a clear decrease in reactivity at U^{227} (see Figure 2.9; compare blue and black traces, Figure 2.6C). The mutation itself (the inserted A) also exhibits low SHAPE reactivity (Figure 2.6C). Low

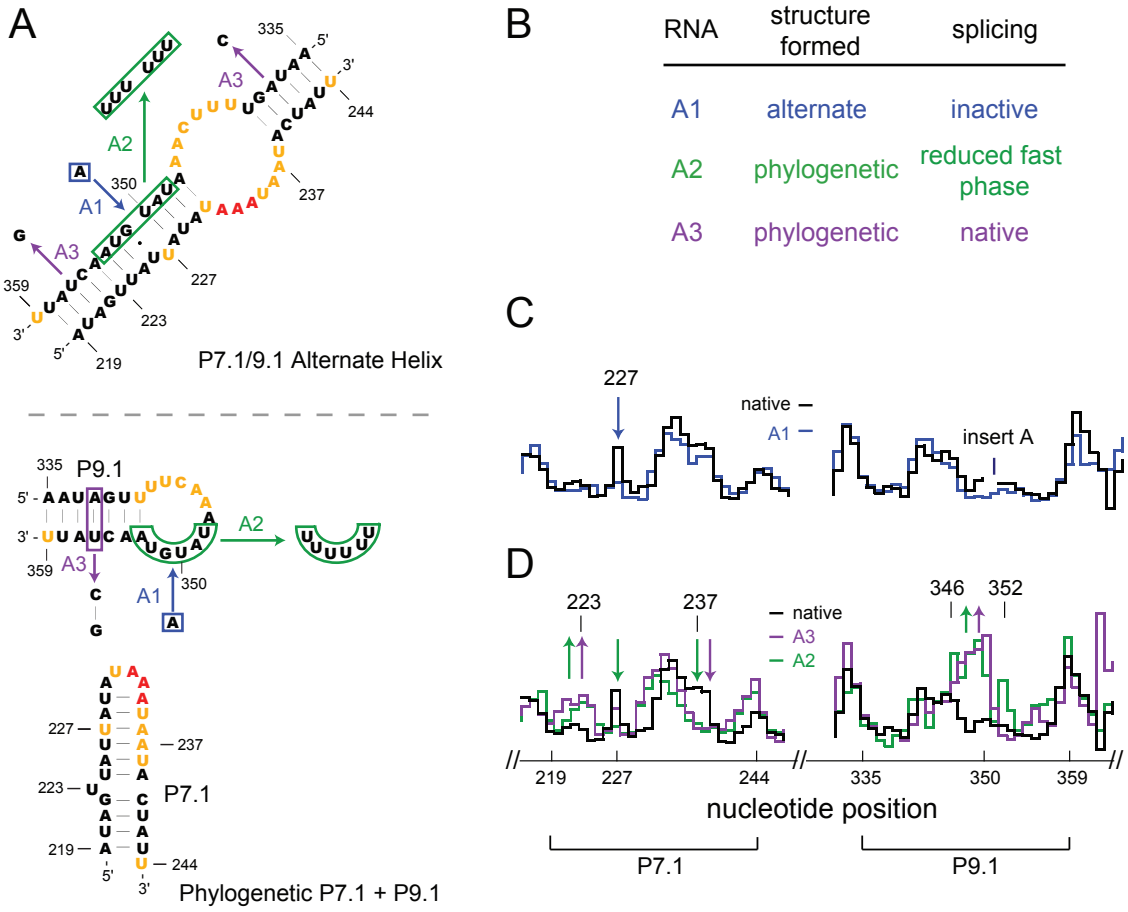

Figure 2.6. SHAPE analysis of point mutations in the P7.1/9.1 alternate helix. (A) Secondary structure models for the alternate (upper panel) and phylogenetically conserved helices (lower panels). Experimental SHAPE reactivities corresponding to free RNA are superimposed on both structures. Mutations are highlighted in blue, green and purple. (B) Summary of mutant structures and splicing activity. (C,D) SHAPE reactivities for the P7.1 and P9.1 regions of the native and mutant sequence RNAs.

reactivities at these two nucleotides strongly support formation of the P7.1/9.1 alternate helix.

An alternate strategy for testing misfolding in the free bI3 RNA is to introduce mutations to the RNA that would leave the phylogenetic structure intact or severely disrupt the formation of the alternate helix. The first mutant (A2) targets loop nucleotides in the phylogenetic P9.1 helix and will prevent pairing interactions that form in the alternate helix model (green boxes, Figure 2.6A). The second mutation (A3) replaces an A-U with a G-C base pair in the native P9.1 (in purple, Figure 2.6A). This mutation creates two mismatch pairs in the alternate helix model.

SHAPE analysis of the A2 and A3 mutants shows both to be nearly identical to the native RNA except in the P7.1 and the P9.1 helices and the P2 loop nucleotides, where both mutants exhibited similar changes (compare green and purple traces, Figure 2.6D and also the A2 and A3 traces, Figure 2.9). In the P7.1 helix, nucleotides near U^{223} show increased SHAPE reactivities while nucleotide U^{227} and those around A^{237} decrease in reactivity as compared to the native sequence. These differences indicate that the A2 and A3 mutants stabilize formation of P7.1. In the P9.1 helix, nucleotides from 346 to 352 show a large increase in reactivity (compare green and purple traces to the black trace at labeled nucleotides, Figure 2.6D). These data show that the A2 and A3 mutants strongly stabilize the phylogenetically accepted structure for P9.1. In addition, loop nucleotides at the end of P2 show low SHAPE reactivities in the A2 and A3 mutants (Figure 2.9), consistent with formation of GNRA tetraloop–receptor interactions, typical of many group I introns [7,9].

Splicing assays performed with the A2 and A3 mutants showed that both RNAs spliced at the same rate as the native RNA, although the A2 mutation exhibited a reduced fraction of RNA molecules that spliced in the fast phase (Figure 2.6B and Table 2.1). In contrast, introducing the single nucleotide A1 mutant renders the RNA incapable of splicing. Taken together, these SHAPE and splicing experiments strongly support a model in which P7.1 and P9.1 form a stable, non-native structure in the free RNA.

2.2.8 The P1 Helix. SHAPE data for the free RNA also show poor agreement with the structure of the phylogenetically proven P1 helix. The alternative model for this region shows better, but still imperfect, agreement with SHAPE information (Figure 2.7A,B). A likely explanation is that this region of the RNA samples at least two conformations prior to protein binding. I performed two classes of SHAPE experiments to characterize the predominant structure in this region. First, I stabilized the native P1 structure by binding both the Mrs1 and bI3 maturase proteins to the RNA and then subjected the entire complex to SHAPE analysis. Second, I mutated the reactive A^{7} nucleotide to U (mutant S1) so that this position would form a stable base pair with A^{+4} of the 3' exon (Figure 2.7B, mutation is highlighted with magenta box).

The protein-bound RNA has a SHAPE reactivity profile consistent with formation of the phylogenetically expected structure for the P1 stem-loop: loop nucleotides in P1 are reactive while the base paired stem is unreactive (gray dashed box, Figure 2.7A). This profile is roughly the opposite of that seen in the free RNA where loop nucleotides, such as position 4, are unreactive and nucleotides in the stem, such as -1 and 7, are reactive (compare gray and black traces, Figure 2.7D). In addition, reactivities near nucleotide +5 in the 3' exon become much more reactive in the protein-bound structure relative to the

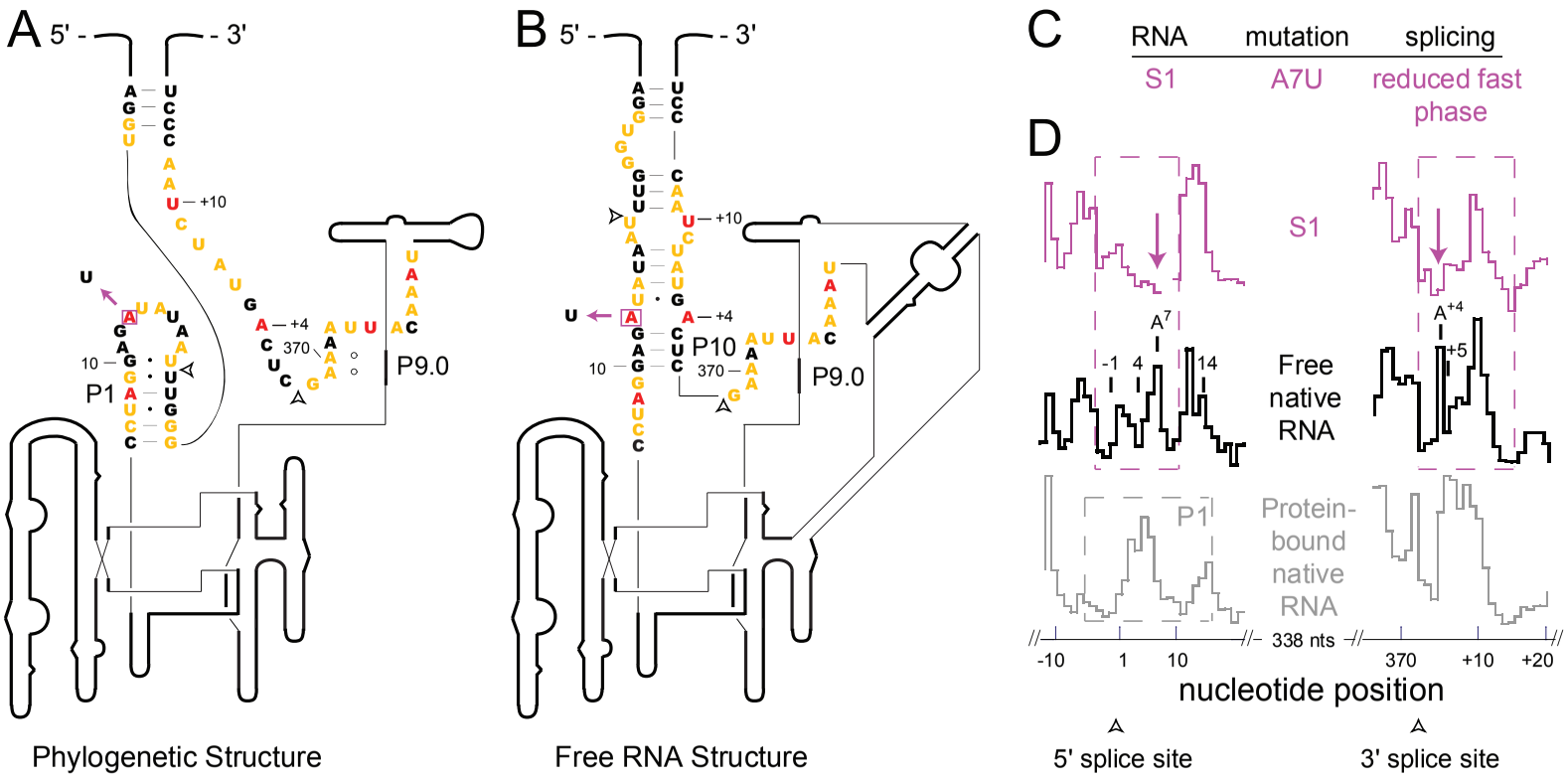

Figure 2.7. SHAPE analysis of the 5'and 3′ splice sites. (A,B) Secondary structure models for the phylogenetic and free bI3 RNA in the P1 helix and in interacting regions of the 3′ exon. (C) Splicing activity for the S1 mutation. (D) SHAPE reactivity histograms. Gray and magenta dashed boxes highlight nucleotides involved in the catalytically active P1 helix and the alternate structures, respectively.

free RNA, again consistent with formation of the flexible, phylogenetically accepted, linker in this region. These data indicate that the P1 helix is not stably formed prior to protein binding.

In contrast, SHAPE reactivities in the S1 mutant are similar to that of the free RNA. For example, nucleotides 4 and +5 are unreactive in both traces (compare black and magenta traces, Figure 2.7D). The largest direct reactivity differences between the S1 mutant relative to the free RNA are lower reactivities at the mutated nucleotide (U^{7}) and its predicted pairing partner (A^{+4}) (Figure 2.7D). The S1 mutant, however, exhibits additional direct changes visualized as increased reactivity around nucleotides 14 and -1 (compare black and magenta traces, Figure 2.7D). Comparatively, the S1 mutant is in better agreement with the structure shown in Figure 2.7B than the native sequence RNA. These data are consistent with a model in which the free RNA exists as a mixture of structures in this region that includes the structure shown in Figure 2.7B, plus additional components.

2.2.9 The Free Misfolded RNA Structure is Thermodynamically Preferred. SHAPE analysis of the free bI3 RNA indicates that this RNA does not form the expected catalytically active secondary structure, as judged by its similarity to the phylogenetically conserved structure. However, the RNA does achieve an active or near-active secondary structure when bound by the bI3 maturase and Mrs1 protein cofactors. Protein binding promotes large-scale changes in the bI3 RNA secondary structure, especially in the P1 and P7.1-P9.1 regions (Figures 2.7 and 2.8). I sought to test whether the catalytically active secondary structure is stable once formed upon protein binding.

I therefore formed the native complex in the presence of the bI3 maturase and Mrs1, removed the proteins by proteolysis, and then assessed both RNA structures by

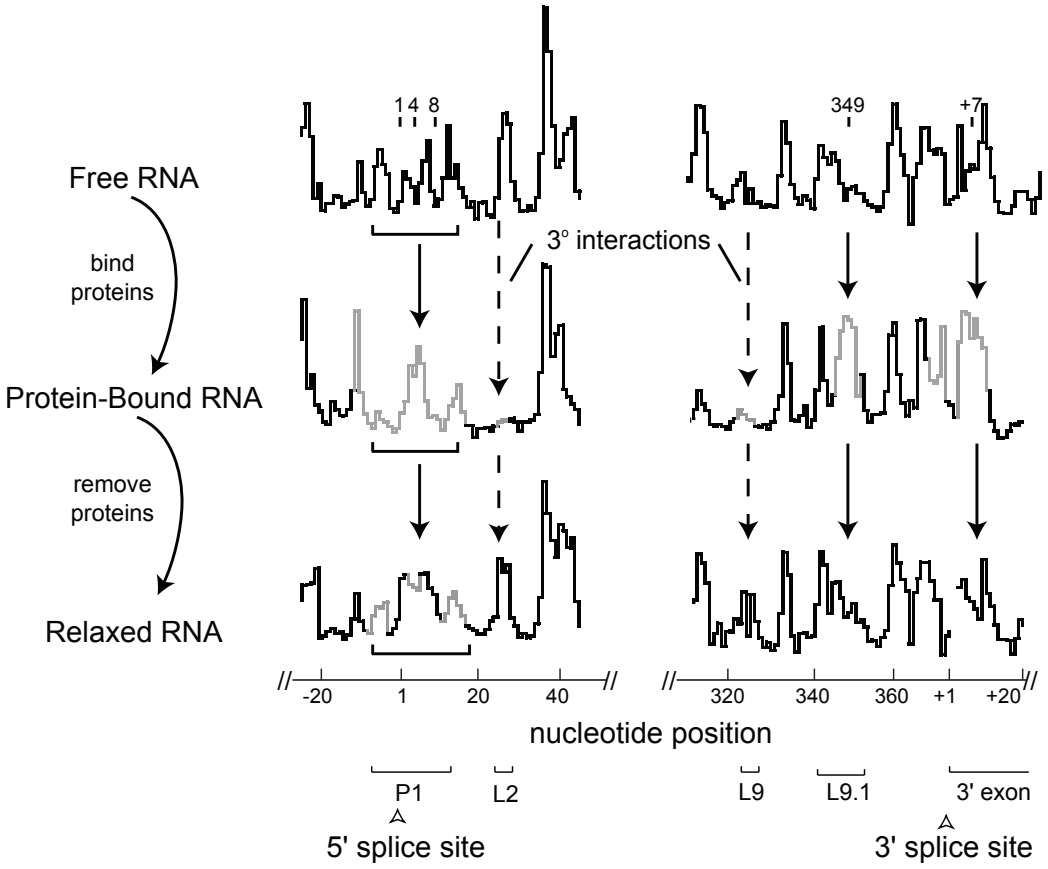

Figure 2.8. SHAPE analysis of bI3 RNA secondary structure rearrangements upon protein binding. Reactivity versus nucleotide position histograms are shown for free, bound and relaxed bI3 RNAs. Solid and dashed line arrows emphasize secondary and tertiary structure differences, respectively. Nucleotides that exhibit significant reactivity differences as compared with the free RNA are gray. Important structural landmarks in this region are shown with brackets.

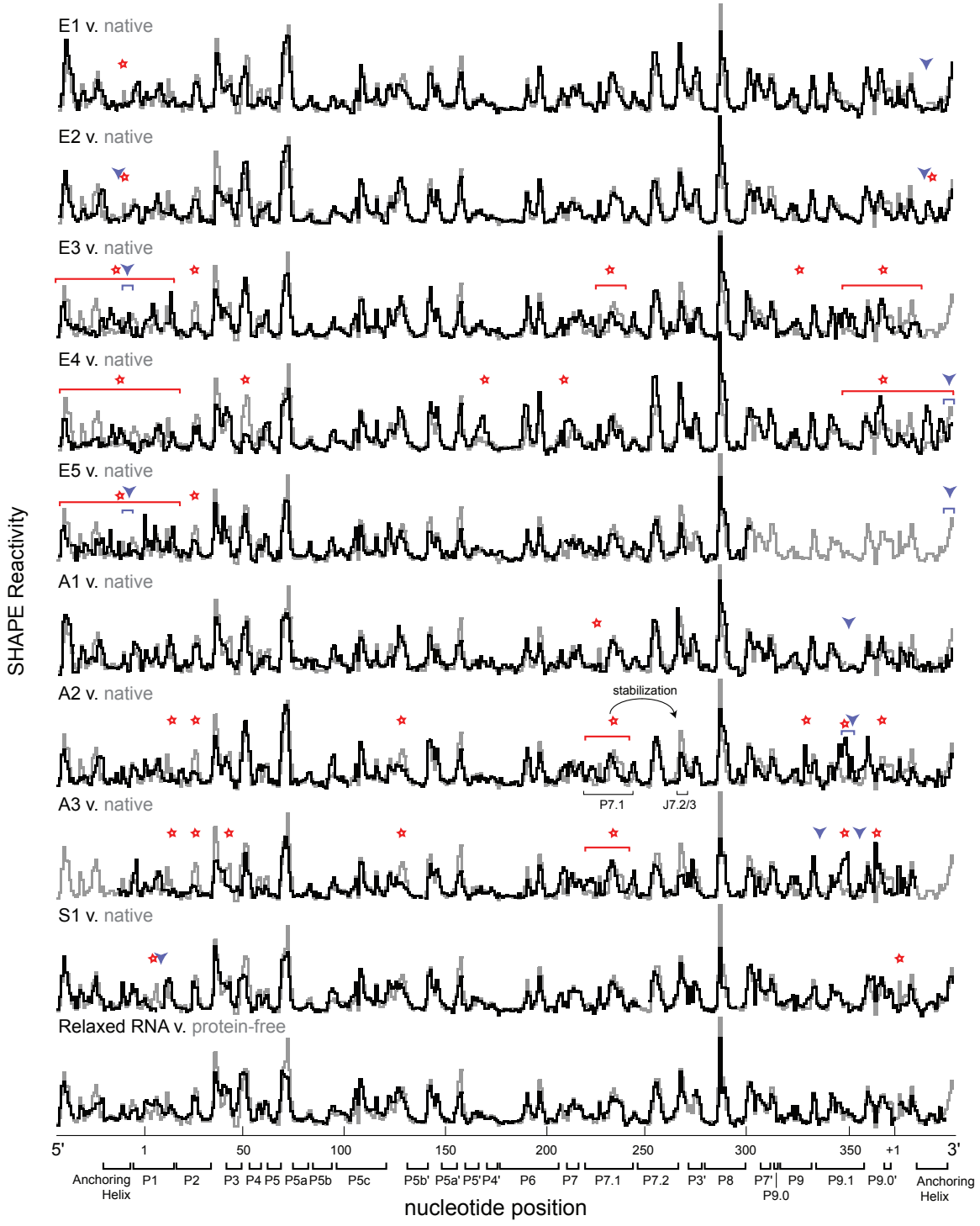

Figure 2.9. Histograms of SHAPE reactivity versus nucleotide position for all bI3 RNA constructs. Each mutant RNA (black trace) is compared with the protein-free native sequence RNA (gray trace). Sites of mutation are indicated with blue arrowheads. Significant reactivity differences (defined as a greater than 2-fold increase or decrease) between mutant and native sequence RNAs are highlighted with red stars. Bars below the histogram indicate structural landmarks in the RNA.

RNA	k_{obs} min^{-1} (fast phase)	Fraction in fast phase*
Native	0.26	0.44
E1	0.22	0.45
E2	0.22	0.40
E3	**< 0.01**‡	
E4	**0.19**	**0.10**
E5	**0.21**	**0.16**
A1	**nd**	
A2	0.39	**0.17**
A3	0.32	0.44
S1	**0.065**	0.30

For the RNA mutants, values considered to be significantly different from the native sequence are highlighted in bold. Splicing experiments were performed in the presence of both bI3 maturase and Mrs1 proteins.

*For all constructs, the remaining fraction of all RNAs spliced in a slow phase characterized by k^{obs} ~0.006 min^{-1}.

‡Splicing occurred primarily by cleavage at the 5' splice site.
nd, not detectable.

Table 2.1 Splicing of native and mutant bI3 RNAs

SHAPE. Consistent with a role in promoting the native structure, addition of the maturase and Mrs1 proteins stabilizes structures in P1 and P7.1-P9.1 that are consistent with the catalytically active structure of the bI3 RNA (compare gray regions of the protein-bound RNA trace to the free RNA trace, Figure 2.8). Dissociating the proteins from the RNA by proteolysis yields a SHAPE reactivity profile that is nearly superimposable with that of the free RNA (compare relaxed RNA to free RNA traces, Figure 2.8). Misfolding in P1 and formation of the P7.1/9.1 alternate helix are thus clearly not artifacts of the folding procedure nor do they reflect formation of a kinetically trapped state. Instead, the P1, P7.1 and P9.1 helices are simply not stable in the absence of binding by the bI3 maturase and Mrs1.

2.3 Discussion

2.3.1 Long-Range and Non-Phylogenetic RNA Structure Mapped by SHAPE Analysis of Point Mutations.

Understanding the initial state of an RNA is an absolute prerequisite for determining the role of protein cofactors, chaperones, and other effectors in facilitating RNA folding and ribonucleoprotein assembly reactions. For most large RNAs, this basic step has remained elusive. In general, large assumptions have had to be made about the structure of the free RNA in most experimental systems. *A priori*, the sequence of the bI3 RNA appears to be fully capable of folding to the well-established and catalytically active structure for group I introns RNAs. In strong contrast to this expectation, analysis by SHAPE chemistry shows extensive regions within the free bI3 RNA are incompatible with the conserved group I intron secondary structure (Figure 2.3). The structure of the free bI3 RNA was, therefore, unknown and phylogenetic information

alone was inadequate for developing a useful model for the secondary structure of the free RNA.

To develop testable candidate models for the structure of the free bI3 RNA, I first used single nucleotide resolution SHAPE data as a pseudo-free energy change term to constrain the output of a thermodynamic-based folding algorithm [19,21]. My approach for evaluating plausible structural models takes advantage of the quantitative, comprehensive, and high resolution information obtained from a SHAPE experiment. For example, I obtained strong evidence in support of the bI3 anchoring helix via analysis of a single point mutation (mutant E1, Figure 2.4). Strikingly, I detected the strong effect of this mutation due to formation of a single new base pair involving residues 400 nucleotides distant in sequence. At the same time, SHAPE reactivities were essentially unchanged for the remaining 450 nucleotides in the bI3 RNA (Figure 2.4D).

In contrast, other mutations caused unanticipated global changes in the secondary structure, which were readily detected as large-scale changes in SHAPE reactivities (Figure 2.9). Information from these mutants nonetheless provided strong structural evidence for an alternative model for the free bI3 RNA structure. For example, the two-nucleotide A3 mutation in P9.1 caused large-scale structural rearrangements in both the P9.1 and P7.1 helices (Figure 2.6) which strongly supported a pairing interaction between the P9.1 and P7.1 sequences in the free RNA.

Comparative sequence analysis is the most successful approach for determining the conserved structure required for function in an evolutionarily related group of RNAs. However, covariation analysis gives no information on whether any one representative RNA forms the functional structure characteristics of the family as a whole. In a single

experiment, SHAPE appears to be sufficient to identify whether a large RNA folds to a phylogenetically conserved structure. If an RNA does not, as in the case of the bI3 RNA, SHAPE analysis of point mutations represents a powerful and general technology for developing models for long-range interactions in any RNA.

2.3.2 Model for the Free bI3 RNA. My model of the protein-free bI3 RNA contains three classes of secondary structures as judged by compatibility of the SHAPE data with the phylogenetically proven and catalytically active secondary structure. These classes are phylogenetically correct, alternate (phylogenetically incorrect), and novel. Correct structures include the entire P5-P4-P6 domain and the P2, P7.2, P8 and P9 stem-loop helices (Figure 2.3A).

The largest alternate structural feature I identify is an extensive base pairing interaction between the peripheral P7.1 and P9.1 helices that I term the P7.1/9.1 alternate helix. In addition, sequences in the catalytic core that should form the P1 helix instead interact with the 3' exon to form an extended helical region. Also in the catalytic core, SHAPE data indicate that most of the nucleotides in the P3, P7 and P9.0 helical elements are flexible and do not form stable pairings (Figure 2.3B).

My model of the bI3 RNA also contains a novel interaction between the 5' and 3' exons that I term the bI3 anchoring helix. Disruption of this interaction by any of three mutations yields an RNA with severely compromised splicing activity. SHAPE analysis indicates that these mutations in the exon sequences cause the RNA to misfold in ways that cannot be rescued by protein binding. SHAPE analysis and splicing assays thus indicate that the bI3 anchoring helix plays a crucial role in sequestering the bI3 RNA so that it can fold independent of interference with exon sequences.

2.3.3 A Structurally Fragile RNA and a New Role for Protein Effectors. Many large RNAs require protein cofactors to bind and stabilize the active three-dimensional structure of the RNA. Some of the best studied systems include folding of group I intron RNAs facilitated by the CBP2 and CYT-18 proteins[3,22-25] and stabilization of 16S ribosomal RNA structure by small subunit proteins to form the active 30S ribosomal particle[26-28]. In these systems, the free RNA appears to fold to a largely correct secondary structure, even in the absence of protein binding. Binding by protein cofactors then functions largely to stabilize formation of the catalytically active tertiary structure.

In contrast, the conformation of the bI3 RNA is extremely fragile and requires binding by protein cofactors to effect large-scale rearrangements in its secondary structure prior to stabilization of tertiary folding by these same proteins. Use of single nucleotide resolution SHAPE chemistry has made it possible to identify a new role for protein facilitators in RNA folding reactions. The bI3 maturase and Mrs1 function to induce large-scale rearrangements in RNA secondary structure prior to also stabilizing the catalytically active tertiary structure. These studies also highlight the critical importance of understanding the structure of the free RNA state prior to assembly with protein effectors.

2.4 Experimental Procedures

2.4.1 bI3 RNA Constructs and Protein Expression. PCR templates for the bI3-ΔL8 native sequence RNA[9] and for the mutant constructs contained 7 nts of 5'-vector sequence, the entire 77 nt 5'-exon, the 372 nt intron, 30 nts of the 3'-exon, and 60 nts of 3'-vector sequence, including the primer binding site. The bI3-ΔL8 construct contains a large deletion of an open reading frame in L8 that improves folding behavior of this RNA but

does not change the structure of the catalytic core or protein binding [9]. RNAs were generated by run-off transcription using T7 RNA polymerase [1 mL, 37 °C, 6 h; containing 40 mM Hepes-NaOH (pH 7.5), 20 mM dithiotreitol (DTT), 0.5 mM spermidine, 10 mM $MgCl_2$, 0.005% (v/v) Triton X-100, 0.5 μg pyrophosphatase (Roche), 2 mM each nucleotide triphosphate, ~10 μg DNA template, 200 units SUPERNase-In (Ambion), 70 μg T7 RNA polymerase] and purified by gel electrophoresis. Mrs1 and ΔCys-bI3 maturase proteins were expressed and purified as described[9] with the exception that Mops was used as the buffer, [DTT] was 5 mM, and glycerol was reduced to 10% (v/v).

2.4.2 SHAPE Analysis of Native and Mutant RNAs. For each RNA construct, 5 pmol of RNA (0.1 μM final) was heated for 1 min at 95 °C, snap-cooled 1 min on ice, and folded at 37 °C for 10 min in reaction buffer [40 mM Mops-NaOH (pH 8.0), 80 mM potassium acetate (KOAc) (pH 8.0), 20 mM $MgCl_2$]. The reaction mixture was treated with a one-tenth volume of 5 mM 1M7 (dissolved in anhydrous DMSO) at 37 °C for 70 s (equal to 5 hydrolysis half lives)[20]. Concurrently, a no-reagent reaction was performed omitting 1M7. The RNA was subsequently precipitated with ethanol and resuspended in a mixture of water and a fluorescently labeled primer [4.5 pmol of a Cy5 labeled primer for the (+) 1M7 reactions or 6 pmol of a Cy5.5 labeled primer for (–) 1M7 reactions] to a total volume of 13 μL. Primers were annealed by heating at 65 °C for 6 min and 37 °C for 1 min. Extension buffer [5 mM DTT, 0.5 mM each dNTP, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, and 100 units SuperScript III Reverse Transcriptase (Invitrogen)] was added on ice and the reactions were incubated for 5 min at 37 °C, 20 min at 52 °C, and 5 min at 60 °C. Reactions were quenched by addition of a stop solution (4 μL) consisting of 50 mM EDTA, 1.5 M sodium acetate (pH 5.3), and 0.8 μg glycogen and placed on ice.

Dideoxy sequencing ladders were generated by primer extension using unmodified RNA and primers labeled with IR800 or WellRED D2 in the presence of 0.25 mM dTTP or ddATP. (+) and (–) reagent and sequencing reactions were combined and recovered by precipitation with ethanol; pellets were dried and resuspended in 40 μL deionized formamide; fluorescently labeled DNAs were resolved by capillary electrophoresis using a Beckman Coulter CEQ 2000XL DNA (capillary electrophoresis) analysis system. SHAPE performed in the presence of proteins contained 0.5 μM (5 molar excess over the RNA concentration) of both the Mrs1 dimer and bI3 maturase proteins bound to the RNA 20 min prior to modification. RNA was treated with 1M7 either before or immediately after digestion with proteinase K [60 μg (Invitrogen); 10 min, 37 °C] followed by extraction with phenol:chloroform:isoamyl alcohol (25:24:1) prior to primer extension. This proteolysis procedure completely eliminated detectable protein binding as judged by nitrocellulose filter binding using radiolabeled bI3 RNA.

2.4.3 Data Analysis. Raw traces from the CEQ 2000XL capillary electrophoresis instrument were analyzed using ShapeFinder[19]. In brief, ShapeFinder was used to adjust the fluorescent baseline (window of 40 pixels), separate dye intensities into discrete channels, perform a mobility shift to account for different dye mobilities, correct for signal decay, and scale the (+) and (–) reagent traces to make them equal to each other in non-reactive regions. Peaks in the (+) and (–) reagent traces were quantified by whole trace Gaussian integration and reactivities were scaled by discarding the top 2% of the most reactive peaks and dividing by the average intensity of the next 8% of peaks. This calculation places the data on a scale of 0 to ~2 where 1.0 is the average intensity of highly reactive peaks. RNAstructure[21] was used to generate structure models for the RNA using

SHAPE data as pseudo-free energy change terms (slope and intercept were 25 and -6, respectively)[19].

2.4.4 Splicing Assays and Protein Binding. RNA splicing reactions were performed using ~3 nM 5'-^{32}P-end-labeled RNA, 0.1 μM unlabeled RNA, 0.5 μM Mrs1 dimer, 0.5 μM ΔCys-bI3 maturase[9]. The precursor RNA was incubated at 95 °C for 1 min, snap-cooled on ice for 1 min, and refolded at 37 °C for 10 min in reaction buffer [40 mM Mops (pH 8.0), 80 mM potassium acetate (pH 8.0), 20 mM $MgCl_2$]. Proteins were incubated with the RNA for 30 min at 37 °C prior to addition of guanosine 5'-monophosphate (GMP) to a final concentration of 3 mM to initiate splicing. Control experiments indicated that no detectable splicing occurred without pG or both proteins present. Reactions were quenched on ice with the addition of EDTA to a final concentration of 100 mM, resolved on denaturing 12% polyacrylamide gels, and quantified by phosphorimaging (Molecular Dynamics). Fast and slow phases of splicing were observed, therefore, reaction rates were determined by a double exponential fit: fraction precursor RNA = $f_A e^{-k_A t} + (1-f_A)e^{-k_B t}$ where k_A and k_B are the observed rates for the fast and slow phases of splicing, respectively, and f_A is the fraction reactive in the fast phase. RNA-binding assays were performed by filter partitioning[9]. Both proteins bind the mutant RNAs with binding constants within 2-fold of that for the native sequence RNA, with the exception of the E2, A1, and S1 mutants, which bound within 6-fold of native sequence binding affinity.

2.5 References

1. Gesteland, R.F., Cech, T.R. & Atkins, J.F. *The RNA World*, (Cold Spring Harbor Laboratory Press, New York, 2006).

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

3. Weeks, K.M. Protein-facilitated RNA folding. *Curr. Opin. Struct. Biol.* **7**, 336-42 (1997).

4. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5**, 908-19 (2004).

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

6. Cech, T.R. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* **59**, 543-68 (1990).

7. Woodson, S.A. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**, 324-30 (2005).

8. Vicens, Q. & Cech, T.R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**, 41-51 (2006).

9. Bassi, G.S., de Oliveira, D.M., White, M.F. & Weeks, K.M. Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 128-33 (2002).

10. Bassi, G.S. & Weeks, K.M. Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst. *Biochemistry* **42**, 9980-8 (2003).

11. Longo, A. et al. Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* **12**, 779-87 (2005).

12. Tinoco, I. & Bustamante, C. How RNA folds. *J. Mol. Biol.* **293**, 271-81 (1999).

13. Woodson, S.A. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* **57**, 796-808 (2000).

14. Wu, M. & Tinoco, I., Jr. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11555-60 (1998).

15. Nikolcheva, T. & Woodson, S.A. Facilitation of group I splicing in vivo: misfolding of the Tetrahymena IVS and the role of ribosomal RNA exons. *J. Mol. Biol.* **292**, 557-67 (1999).

16. Silverman, S.K., Zheng, M., Wu, M., Tinoco, I., Jr. & Cech, T.R. Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA* **5**, 1665-74 (1999).

17. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223-31 (2005).

18. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610-6 (2006).

19. Wilkinson, K.A. et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, 883-899 (2008).

20. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144-5 (2007).

21. Mathews, D.H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7287-92 (2004).

22. Lambowitz, A.M. & Perlman, P.S. Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem. Sci.* **15**, 440-4 (1990).

23. Webb, A.E., Rose, M.A., Westhof, E. & Weeks, K.M. Protein-dependent transition states for ribonucleoprotein assembly. *J. Mol. Biol.* **309**, 1087-1100 (2001).

24. Garcia, I. & Weeks, K.M. Structural basis for the self-chaperoning function of an RNA collapsed state. *Biochemistry* **43**, 15179-86 (2004).

25. Paukstelis, P.J., Chen, J.H., Chase, E., Lambowitz, A.M. & Golden, B.L. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* **451**, 94-7 (2008).

26. Moazed, D., Stern, S. & Noller, H.F. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.* **187**, 399-416 (1986).

27. Powers, T., Daubresse, G. & Noller, H.F. Dynamics of in vitro assembly of 16S rRNA into 30S ribosomal subunits. *J. Mol. Biol.* **232**, 362-74 (1993).

28. Talkington, M.W., Siuzdak, G. & Williamson, J.R. An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628-32 (2005).

Chapter 3

The Mrs1 Splicing Factor Binds the RNA Tetraloop-Receptor Motif

3.1 Introduction

RNA and proteins have co-evolved to form the ribonucleoproteins (RNPs) that now carry out many of the fundamental steps of gene regulation, including mRNA processing and protein biogenesis[1]. Core functions of complexes like the spliceosome and ribosome are performed in active sites composed of RNA; however, these RNA elements also require extensive participation by protein facilitators[1]. Similarly, most group I introns likely require protein cofactors to catalyze their own excision from flanking exons and splice efficiently. Group I introns, therefore, represent ideal models for testing the role of protein recruitment into ribonucleoprotein complexes.

The group I intron active site is comprised of RNA. The catalytic core is formed at the interface of three RNA domains, termed the P1-P2, the P5-P4-P6, and the P9-P7-P3-P8 domains. These domains are held in a precise and catalytically active three-dimensional architecture by inter-domain tertiary interactions[2-5]. In a few minimal group I introns, these tertiary interactions involve direct and compact interactions between RNA domains. However, most group I introns are relatively complex and have evolved large peripheral RNA elements or have recruited a wide range of protein cofactors to stabilize their active conformations, or both[2,6-8].

Protein cofactors use diverse strategies to stabilize group I intron RNA tertiary structure. Proteins like the *Neurospora crassa* Cyt-18 bind multiple RNAs by recognizing conserved elements in the group I intron catalytic core[9]. Alternatively, proteins like CBP2[10,11] and intron maturase proteins[12,13] recognize specific introns through interactions with idiosyncratic peripheral elements. Many group I intron splicing factors have been co-opted or evolved from proteins that perform other nucleic acid binding functions. In the

cases of the maturase proteins and Pet54, an existing nucleic acid binding surface is reused to accommodate the new group I intron substrate[12,14]. Alternately, proteins like CYT18 have evolved separate binding surfaces to perform distinct functions as a group I intron cofactor and as a tRNA synthetase[9,15].

The yeast mitochondrial bI3 group I intron is a good example of an RNA that has become dependent on proteins to fold and function correctly. bI3 RNA splicing requires specific binding by two proteins, the bI3 maturase and two dimers of the Mrs1 protein[16,17]. The free RNA is extensively misfolded and binding by the maturase and Mrs1 proteins induces large conformational rearrangements in both secondary and tertiary structure[18]. The bI3 maturase protein binds to the P5-P4-P6 domain and promotes formation of long-range tertiary interactions to stabilize the P5 and P4 components of the catalytic core[12]. The Mrs1 protein facilitates splicing for both the bI3 and aI5β introns in yeast mitochondria[19], even though these two introns are not especially similar. Mrs1 is related to the RuvC family of DNA junction resolvases and, in evolutionary terms, appears to have acquired an RNA binding activity only recently and in a small subset of organisms[20]. Mrs1 may have retained its nucleic acid binding site but is no longer capable of cleaving DNA[16,20]. At present, the RNA binding site and molecular function of Mrs1 in group I intron splicing are unexplored.

In this work, I use high-throughput hydroxyl radical footprinting to identify the RNA binding sites for Mrs1. Mrs1 binds at each of two conserved GNRA tetraloop-receptor interactions in the bI3 RNA. The tetraloop-receptor motif was one of the first long-range tertiary interactions to be identified for RNA[21]. This interaction involves hydrogen bonding between GNRA loop nucleotides with functional groups in the minor

groove of the receptor helix[21-23]. Mrs1 also binds structurally homologous, but non-cognate, tetraloop-receptor elements in the *Azoarcus* group I intron and in the *Bacillus subtilis* ribonuclease P (RNase P) specificity domain RNA. Thus, Mrs1 has evolved from a DNA-binding resolvase to a protein that binds to and stabilizes GNRA tetraloop-receptor motifs found ubiquitously in large RNAs. This work thus emphasizes the strong evolutionary pressure to usurp RNA-only structures with RNA-protein interactions. Every long-range RNA tertiary interaction in the bI3 ribonucleoprotein is stabilized by either the maturase or Mrs1 proteins, both of which have been co-opted from earlier DNA-binding functions.

3.2 Results

3.2.1 Mrs1 Binds and Stabilizes Tetraloop-Receptor Interactions in the bI3 RNA.

I identified RNA interaction sites for the Mrs1 protein using hydroxyl radical footprinting. Hydroxyl radicals are generated *in situ* from H_2O_2 in the presence of a Fe(II)-EDTA catalyst[24]. The hydroxyl radicals then cleave the RNA backbone in a way that is roughly correlated with solvent accessibility[25]. I created a high-throughput hydroxyl radical experiment by analyzing cleavage positions using primer extension performed with fluorescently labeled primers, resolved by capillary electrophoresis. I obtained single nucleotide resolution cleavage information for the entire 540 nucleotide bI3 splicing precursor in a single experiment (Figure 3.1A). Cleavage intensities were normalized to a scale from 0 to ~1.5, where 1.0 is defined as the average intensity of highly reactive nucleotides. On this scale, nucleotides with reactivities that are one-half the mean or less are classified as solvent inaccessible (Figure 3.1A, orange columns).

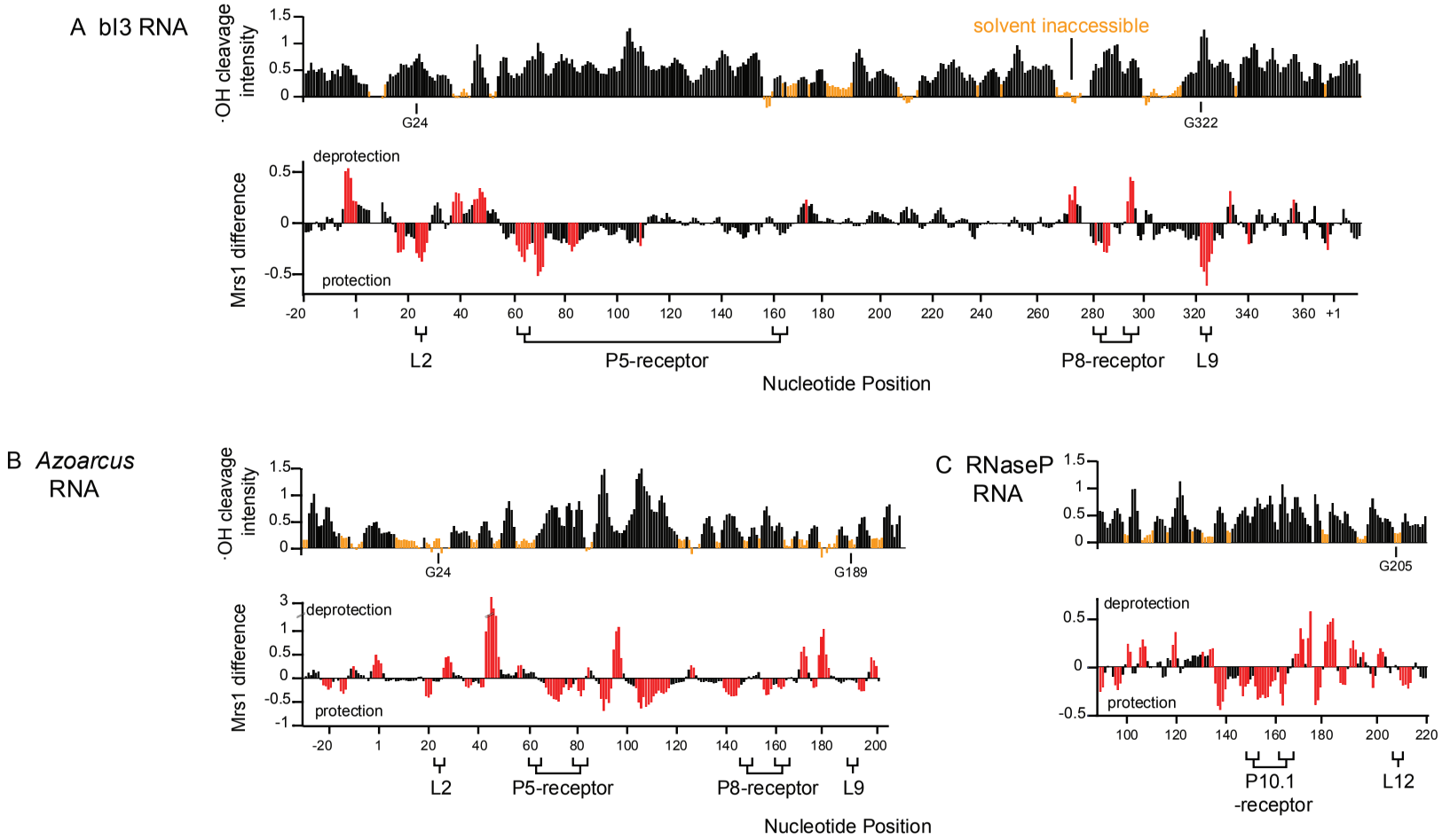

Figure 3.1. High-throughput hydroxyl radical footprinting analysis of the (A) bI3 group I intron, (B) Azoarcus group I intron and (C) RNase P specificity domain. (Top panels) Histograms of cleavage intensity versus nucleotide position. Solvent inaccessible nucleotides (intensities ≤ one-half the mean) are orange. (Bottom panels) Difference plots reporting cleavage intensities for the Mrs1-bound RNA minus the free RNA. Significant protections are emphasized in red. The first guanosine residue in each GNRA tetraloop is labeled explicitly. Bars below the axes indicate tetraloop and receptor sequences.

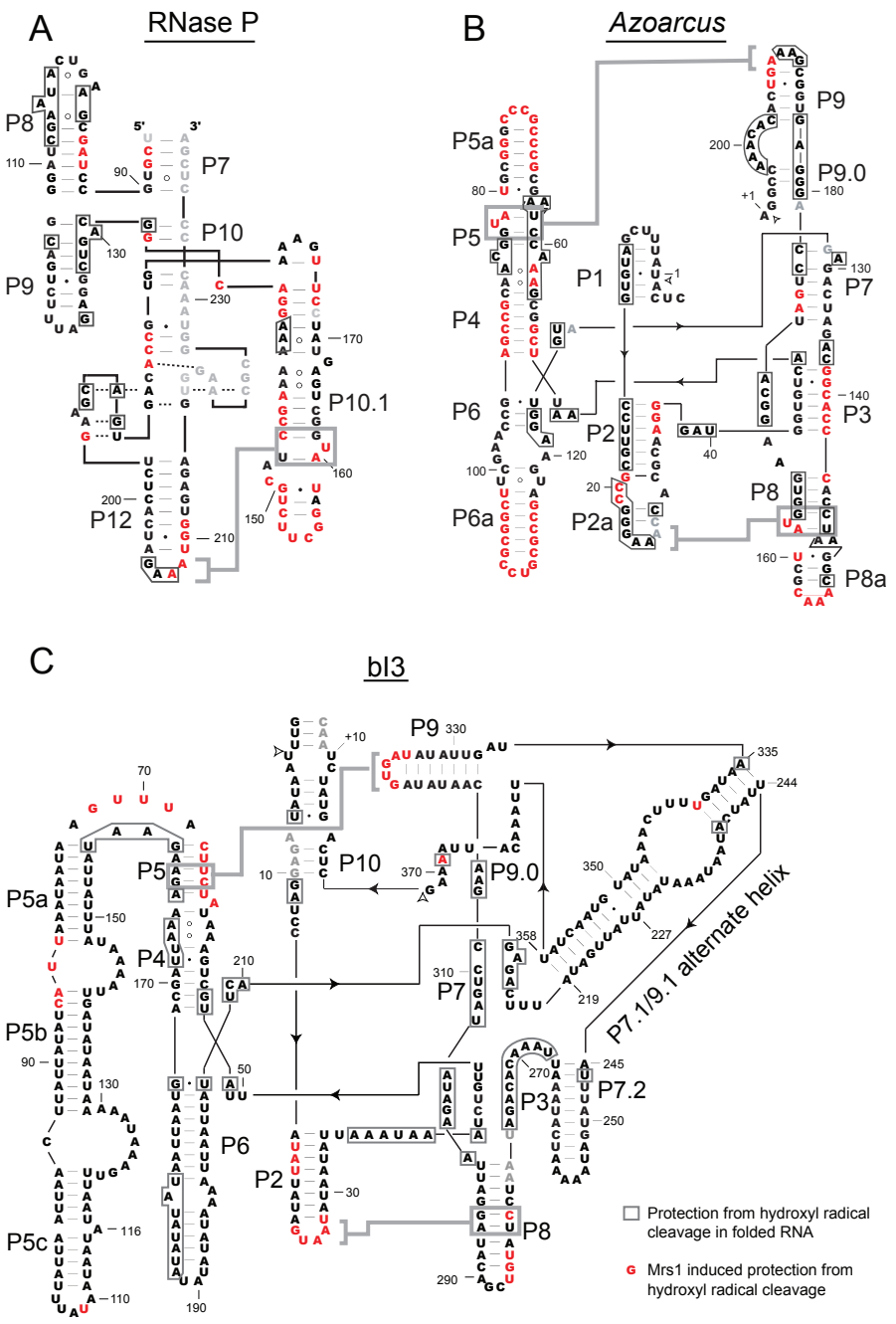

Figure 3.2. Secondary structure models of Mrs1 protections for the RNase P[32], *Azoarcus*[33], and bI3[18] RNAs. Nucleotides that are protected from hydroxyl radical cleavage upon Mrs1 binding are colored red. Nucleotides that are solvent inaccessible prior to Mrs1 binding are boxed. Tetraloop-receptor interactions are emphasized by gray lines. Group I intron splice sites are indicated with arrows.

In the free bI3 RNA, only ~ 20% of nucleotides are protected from cleavage prior to binding by the Mrs1 protein (top panel, Figure 3.1A). Most RNA elements expected to form tertiary contacts are reactive, including the entire P5-P4-P6 domain, the GNRA tetraloops at the ends of the P2 and P9 helices (G of each tetraloop is labeled in Figure 3.1A), and their respective receptors in the P8 and P5 helices (labeled, Figure 3.1A). These results indicate that, prior to protein binding, the bI3 RNA contains little tertiary structure and is not folded in a catalytically active structure.

Upon addition of Mrs1, extensive regions in the RNA became protected from cleavage. I quantified the effect of Mrs1 binding using a difference plot in which the hydroxyl radical cleavage intensities for the free RNA were subtracted from those for the Mrs1 bound RNA (lower panel, Figure 3.1A). Mrs1 protections and enhancements are reported as negative and positive differences, respectively. Almost all significant changes, taken to be an absolute reactivity difference of 0.2 (2-fold above background[12]) or greater, occur in RNA structures that participate in one of two tetraloop-receptor interactions in the bI3 RNA (highlighted in red, Figures 3.1 and 3.2). These tetraloop-receptor interactions link the L2 loop to the P8 helix and the L9 loop to the P5 helix. Combined, these two tertiary interactions structurally link the three group I intron RNA domains. Upon formation of tetraloop-receptor interactions, protections would be limited to loop and receptor nucleotides if these were only RNA-RNA contacts. The extent of protections observed on the tetraloop-receptor motifs, especially on elements on the exterior of the RNA, suggest that these protections include RNA-protein interactions. In addition, Mrs1 binding caused a few RNA elements to become more reactive, consistent with protein-induced conformational changes (Figure 3.1). These results provide strong evidence that

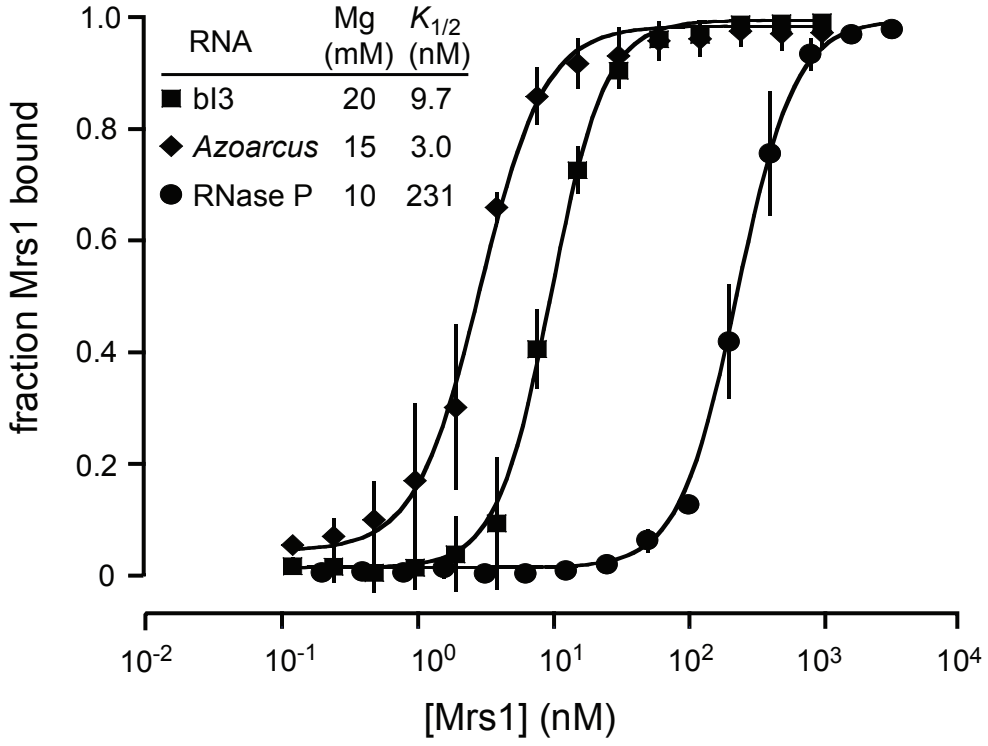

Figure 3.2. Mrs1 binding to the bI3, *Azoarcus,* and RNase P RNAs. Binding experiments are performed in duplicate under conditions that are optimal for RNA function[17, 32, 38]

Mrs1 binds to and stabilizes both GNRA tetraloop-receptor interactions in the otherwise misfolded bI3 RNA.

3.2.2 Mrs1 Generally Recognizes GNRA Tetraloop-Receptor Interactions. I tested whether Mrs1 might generally be able to bind tetraloop-receptor motifs by evaluating binding to two other RNAs that contain this interaction. The *Azoarcus* group I intron and RNase P specificity domain RNAs contain two and one GAAA tetraloop-receptor motif, respectively. These tetraloop-receptor elements fold independently and accurately in the presence of magnesium ion[23,26]. I first showed that all three RNAs bind Mrs1 under conditions that support accurate folding for each RNA. Consistent with previous work[16], two Mrs1 dimers cooperatively bind the bI3 RNA with a $K_{½}$ of 9.7 nM (Figure 3.2). The *Azoarcus* group I intron binds Mrs1 more tightly than does the cognate bI3 RNA: the $K_{½}$ is 1.7 nM. RNase P binds more weakly, the $K_{½}$ is 280 nM (Figure 3.3).

Mrs1 binding to all three RNAs is consistent with cooperative binding by two Mrs1 dimers. Cooperative binding to the *Azoarcus* intron likely reflects the presence of two tetraloop-receptor interactions in this RNA, as is also the case for the bI3 RNA. In contrast, the RNase P RNA contains only a single tetraloop-receptor motif. The apparent cooperativity may reflect a second, weak, Mrs1 binding site as reflected by the larger $K_{½}$ for RNase P as compared to the two group I intron RNAs.

High-throughput hydroxyl radical experiments for the free *Azoarcus* and RNase P RNAs show that both have proportionately more tertiary structure than does the bI3 RNA (orange columns in top panels of Figures 3.1B,C). Significant protection from hydroxyl radical cleavage in the elements that comprise the tetraloop motifs in the *Azoarcus* and RNase P RNAs indicates that these interactions are formed in the two RNAs (position of

the first G in each tetraloop is labeled in Figures 3.1B,C).

Binding by Mrs1 protects extensive regions from cleavage in both the *Azoarcus* and RNase P RNAs. I again visualized Mrs1 protection using difference plots (red columns, Figure 3.1B,C). Even though these RNAs are highly structured, Mrs1 binding yields modest protection from hydroxyl cleavage at all three tetraloop-receptor motifs in the *Azoarcus* and RNase P RNAs. The loop elements are P2 and P9 in the *Azoarcus* RNA and the P12 loop in RNase P. The receptor helices are P5 and P8 in *Azoarcus* RNA and the P10.1 in the RNase P RNA (Figures 3.1B,C and 3.2 A,B). Cleavage protections seen throughout the *Azoarcus* and RNase P RNAs indicate only a modest specificity of the Mrs1 for the tetraloop-receptor interaction in these RNAs.

3.2.3 Visualizing Mrs1 Binding to Tetraloop-Receptor Motifs. I superimposed the hydroxyl radical protection information on the three-dimensional structures for the five tetraloop-receptor interactions studied in this work. The bI3 RNA is largely unstructured in the absence of Mrs1 whereas these elements are pre-formed in the *Azoarcus* and RNase P RNAs. Thus, I evaluated the entire pattern of protection, corresponding to both RNA-RNA and RNA-protein interactions, for each RNA. Regions that are solvent inaccessible in the absence of Mrs1 versus those that become protected upon protein binding are emphasized in orange and red, respectively (Figure 3.4).

The five tetraloop-receptor interactions present in the three RNAs are structurally homologous but are in different local structural contexts. Despite these differences, the net effect of Mrs1 binding is similar in each case. Protected regions include (*i*) the tetraloop itself, (*ii*) the 5′ strand of the receptor helix, and (*iii*) flanking regions on the same face of the RNA as the tetraloop backbone (Figure 3.4). Compared to the *Azoarcus* and RNase P

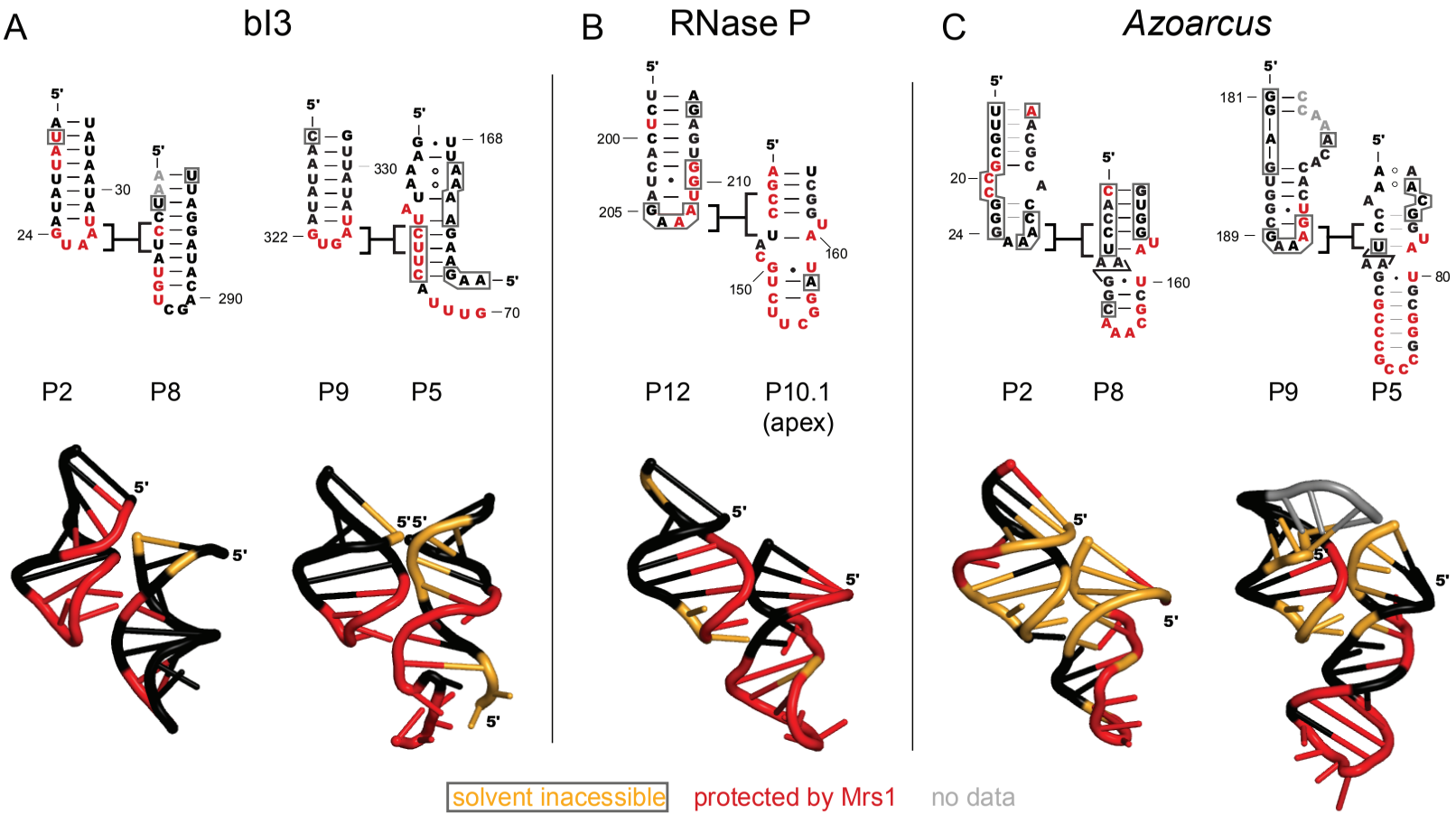

Figure 3.4. Secondary structure and three-dimensional representations for tetraloop-receptor interactions in the (A) bI3, (B) RNase P specificity domain, and (C) *Azoarcus* RNAs. Nucleotides specifically protected upon Mrs1 binding are red. Solvent inaccessible nucleotides are emphasized with boxes on the secondary structures and are orange in the three-dimensional models.

RNAs, a much larger fraction of the two tetraloop-receptors interactions in the bI3 RNA become protected once Mrs1 binds. I infer Mrs1 binding induces significant RNA-RNA interactions in the bI3 intron.

3.3 Discussion

The Mrs1 protein specifically binds to and promotes the formation of two distinct and conserved GNRA tetraloop-receptor interactions in the bI3 group I intron RNA. Mrs1 also binds at structurally homologous, but non-cognate, tetraloop-receptor interactions in the *Azoarcus* group I intron and RNase P specificity domain RNAs. Thus, Mrs1 appears to be a general tetraloop-receptor binding protein.

Mrs1 binding is also required for splicing of a second group I intron, the aI5β intron from the COX1 pre-mRNA in yeast mitochondria[19]. Similar to the bI3 RNP, two Mrs1 dimers bind the (1574 nt) aI5β RNA; however, in contrast to the bI3 RNA, binding is not cooperative[27]. The aI5β intron contains the L9-P5 tetraloop-receptor interaction but lacks the second P2-P8 interaction found in the bI3 RNA, which may explain the absence of cooperative binding. Given its large size, other tetraloop-receptor interactions or secondary binding sites may exist in the aI5β intron.

Identification of two tetraloop-receptor interactions as the Mrs1 binding site both explains why two dimers of Mrs1 are required for full folding and catalytic activity in the bI3 intron and also provides a basis for understanding the global architecture of the six-component bI3 RNP.

I developed a model of the bI3 RNP holoenzyme complex in order to better understand the strong evolutionary pressure for catalytic RNAs to recruit protein cofactors.

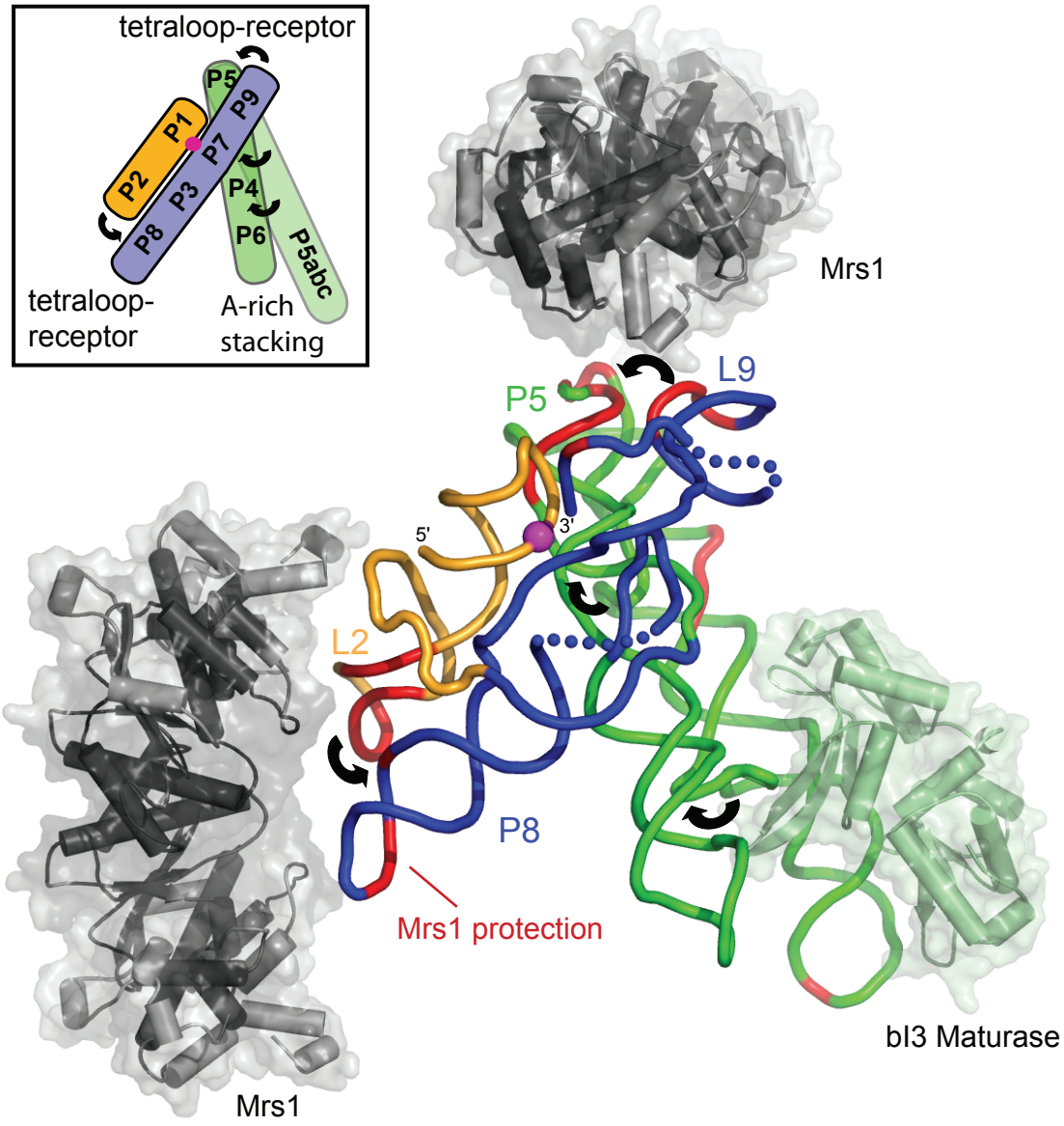

Figure 3.5. Structural model for the six-component bI3 ribonucleoprotein complex. The RNA backbone is shown as a tube and is colored by domain. P1-P2 is orange, P9-P7-P3-P8 is blue, P5-P4-P6 is green. Backbone positions protected by Mrs1 binding are red. Mrs1 dimers are colored dark and light gray in regions of high and low structural homology, respectively, relative to its Ydc2 homolog. The bI3 maturase[12] is in light green. The scissile phosphate at the active site is shown with a sphere

A model for the bI3 intron RNA in its catalytically active structure was created by grafting the bI3 structure onto group I introns with known structures[23,28,29] (Figure 3.5). The bI3 maturase protein was previously shown to bind to the peripheral P5c helix[12] (in green, Figure 3.5).

I first generated a model for Mrs1 by threading[30] the Mrs1 sequence against its nearest homologue, Ydc2[31]. Strong structural homology is apparent for the α-helices at the dimer interface and for the central β-sheet in each monomer (emphasized in dark gray, Figure 3.5). Residues whose positions are less well established by threading all lie on the periphery of the protein (in light gray, Figure 3.5). Modeling supports the view that, like its homologues, Mrs1 is an extended dimer with an axial ratio of ~ 2:1. The nucleic acid binding site is located on one side of the long face.

Nucleotides protected by Mrs1 occur predominantly at the tetraloop-receptor interactions involving L2-P8 and L9-P5 (in red, Figure 3.5). I assume Mrs1 uses approximately the same nucleic acid binding cleft for RNA as its homologues do for DNA. The tetraloop-receptor RNA elements fit best in the Mrs1 binding site when placed parallel to the long axis of the protein. The two tetraloop-receptor motifs are related by a ~90° rotation; thus, the long axes of the two Mrs1 dimers are roughly perpendicular (Figure 3.5).

The catalytic active site is located roughly at the center of the RNP complex (the scissile phosphate cleaved in the first step of splicing is shown as a magenta sphere, Figure 3.5). This active site is held in place by three long-range tertiary interactions located 30-50 Å from the scissile phosphate (illustrated schematically with curved arrows, Figure 3.4). All three bound proteins share critical features: (i) each protein stabilizes a crucial long-range RNA tertiary interaction, (ii) no protein binds within ~20 Å of the scissile phosphate

and (iii) no protein binds within ~50 Å of another protein.

The six-component bI3 complex represents a catalytic RNA caught in the act of becoming an obligate ribonucleoprotein. The catalytic active site is still composed of RNA and no protein appears to approach closer than ~20 Å to the scissile phosphate. However, in the bI3 RNP, every long-range RNA interaction is stabilized by a bound protein cofactor. Each protein has been co-opted from a prior DNA-binding function, suggesting that evolution to an RNA splicing factor is a recent event in evolutionary terms. Similar to much larger RNPs, like the ribosome, these essential proteins bind distally from the catalytic core and induce significant effects on RNA structures up to 50 Å away. This work thus provides a clear example of the strong evolutionary pressure to recruit protein cofactors to facilitate formation of RNA tertiary structure for RNA-centered reactions in biology.

3.4 Experimental Procedures

3.4.1 RNA Constructs and Protein Expression. The bI3 intron (with flanking exon sequences of 84 and 90 nucleotides) and the RNase P specificity domain (with flanking structure cassette sequences) RNAs were generated *in vitro* as described[18,32]. The *Azoarcus* RNA included the tRNA exon[33] and 5′ and 3′ structure cassette sequences[34] and was generated by *in vitro* transcription [1 mL, 25 °C, 6 h; containing 40 mM Tris (pH 7.5), 5 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 0.001% (v/v) Triton X-100, 0.166 μg of pyrophosphatase (Roche), 2 mM each nucleotide triphosphate, ~10 μg double stranded DNA PCR template, 60 units SUPERNase-In (Ambion), 70 μg T7 polymerase] and purified by gel electrophoresis. Mrs1 was expressed and purified as described[16] except that

Mops was used as the buffer, the concentration of DTT was 5 mM, and glycerol was 10% (v/v).

3.4.2 High-Throughput Hydroxyl Radical Cleavage Experiments and Analysis. Each RNA (3 pmol) was prepared under conditions optimal for folding of the specific RNA. The bI3 RNA was incubated in water at 95 °C for 1 min, 4 °C for 1 min and 37 °C for 10 min in reaction buffer [40 mM Hepes (pH 8.0), 80 mM potassium acetate (pH 8.0), 20 mM $MgCl_2$][16]; the *Azoarcus* RNA was incubated in reaction buffer [25 mM Hepes (pH 7.5), 15 mM $MgCl_2$] at 52 °C for 5 min and 37 °C for 2 min[33]; and the RNase P RNA was incubated in water at 95 °C for 2 min, 4 °C for 2 min, and 37 °C for 20 min in the presence of reaction buffer [100 mM Mops (pH 8.0), 100 mM NaCl, 10 mM $MgCl_2$][32]. Mrs1 dimers were incubated with each RNA at 5-fold molar excess (5 μM final concentration), for 30 min at 37 °C. Cleavage reactions (30 μl final volume; 5 min at 37 °C) were initiated by the sequential addition of 3 μl each of 5.0 mM $(NH_4)Fe(SO_4)_2$– 7.5 mM EDTA (pH 8.0), 0.1% H_2O_2, and 50 mM sodium ascorbate [bI3 reactions included half this concentration of Fe(II)-EDTA]. Background was assessed using a reaction omitting Fe(II). Reactions were quenched with glycerol to a final concentration of 30% (v/v). Mrs1 was removed by proteinase K digestion [60 μg, 37 °C, 30 min] followed by phenol:cholorform:isoamyl alcohol (25:24:1) extraction. Reactions were recovered by precipitation with ethanol. Sites of cleavage were identified by primer extension using fluorescently labeled primers [6-FAM (+) Fe(II)-EDTA, HEX (–) Fe(II)-EDTA, NED (ddT, sequence ladder)] in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT, 0.5 mM each dNTP, and 100 units of Superscript III reverse transcriptase (Invitrogen). Primer extension reactions corresponding to the plus and minus Fe(II)-

EDTA and sequencing lanes were combined, precipitated with ethanol, dissolved in formamide and separated on an Applied Biosystems 3130 capillary electrophoresis instrument. Fluorescent cleavage data were analyzed using ShapeFinder[18,35]. Integrated intensities were normalized by dividing the data set by the average of the 8% most reactive nucleotides after first excluding the top 2% of reactivities. Cleavage intensities were smoothed over a three nucleotide window for visualization.

3.4.3 Mrs1 Binding Assays. Mrs1-RNA binding assays were performed by filter partitioning using preincubated nitrocellulose (Whatman) and Hybond (Amersham) filters[17]. 5′-^{32}P-end-labeled RNAs (~ 1 nM) were folded in their respective buffers and incubated with Mrs1 for 30 min at 37 °C before filtering. Wells were washed with 3 vol of reaction buffer before and after binding. Filters were quantified by phosphorimaging (Molecular Dynamics) and fit to an equation that accounts for cooperative binding by two dimers of Mrs1[16], fraction RNA bound = $A([P]^n / [P]^n + K_{1/2}{}^n)$, where A is the total fraction of RNA bound (fit to 0.98), P is the protein concentration, n is the apparent Hill coefficient and $K_{½}$ is the concentration of Mrs1 where half the RNA is bound.

3.4.4 Mrs1 and bI3 RNA Models. The Mrs1 protein model was generated with I-TASSER[30,36] using Ydc2[31] as the template structure. Core Mrs1 protein structures had RMSD values that differed by less than 3.4 Å between the five output models and less than 5.5 Å when compared to Ydc2 [calculated using lsqman[37]]. The center-most model, with the smallest RMSD compared to the other models, was chosen as the representative model for Figure 3.4. A model for the bI3 RNA was assembled from crystal structures[28,29] using lsqman and Sybyl (Tripos). The structure was further refined by discrete molecular dynamics[38] adding base pairing, harmonic constraints around core elements, and long-

range constraints for non-canonical base pairs, peripheral elements, and long-range interactions. Structure images were composed with Pymol (Delano Scientific).

3.5 References

1. Gesteland, R.F., Cech, T.R. & Atkins, J.F. *The RNA World*, (Cold Spring Harbor Laboratory Press, New York, 2006).

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

3. Cech, T.R. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* **59**, 543-68 (1990).

4. Woodson, S.A. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**, 324-30 (2005).

5. Vicens, Q. & Cech, T.R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**, 41-51 (2006).

6. Saldanha, R., Mohr, G., Belfort, M. & Lambowitz, A.M. Group I and group II introns. *FASEB J.* **7**, 15-24 (1993).

7. Lang, B.F., Laforest, M.J. & Burger, G. Mitochondrial introns: a critical view. *Trends Genet.* **23**, 119-25 (2007).

8. Vicens, Q., Paukstelis, P.J., Westhof, E., Lambowitz, A.M. & Cech, T.R. Toward predicting self-splicing and protein-facilitated splicing of group I introns. *RNA* **14**, 2013-29 (2008).

9. Paukstelis, P.J., Chen, J.H., Chase, E., Lambowitz, A.M. & Golden, B.L. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* **451**, 94-7 (2008).

10. Weeks, K.M. & Cech, T.R. Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5' splice site domain. *Cell* **82**, 221-30 (1995).

11. Webb, A.E., Rose, M.A., Westhof, E. & Weeks, K.M. Protein-dependent transition states for ribonucleoprotein assembly. *J. Mol. Biol.* **309**, 1087-1100 (2001).

12. Longo, A. et al. Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* **12**, 779-87 (2005).

13. Caprara, M.G., Chatterjee, P., Solem, A., Brady-Passerini, K.L. & Kaspar, B.J. An allosteric-feedback mechanism for protein-assisted group I intron splicing. *RNA* **13**, 211-22 (2007).

14. Kaspar, B.J., Bifano, A.L. & Caprara, M.G. A shared RNA-binding site in the Pet54 protein is required for translational activation and group I intron splicing in yeast mitochondria. *Nucleic Acids Res.* **36**, 2958-68 (2008).

15. Paukstelis, P.J. et al. A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface. *Mol. Cell* **17**, 417-28 (2005).

16. Bassi, G.S., de Oliveira, D.M., White, M.F. & Weeks, K.M. Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 128-33 (2002).

17. Bassi, G.S. & Weeks, K.M. Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst. *Biochemistry* **42**, 9980-8 (2003).

18. Duncan, C.D. & Weeks, K.M. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* **47**, 8504-13 (2008).

19. Bousquet, I., Dujardin, G., Poyton, R.O. & Slonimski, P.P. Two group I mitochondrial introns in the cob-box and coxI genes require the same MRS1/PET157 nuclear gene product for splicing. *Curr. Genet.* **18**, 117-24 (1990).

20. Wardleworth, B.N., Kvaratskhelia, M. & White, M.F. Site-directed mutagenesis of the yeast resolving enzyme Cce1 reveals catalytic residues and relationship with the intron-splicing factor Mrs1. *J. Biol. Chem.* **275**, 23725-8 (2000).

21. Jaeger, L., Michel, F. & Westhof, E. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **236**, 1271-6 (1994).

22. Cate, J.H. et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-85 (1996).

23. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J. & Strobel, S.A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45-50 (2004).

24. Tullius, T.D. & Greenbaum, J.A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127-34 (2005).

25. Latham, J.A. & Cech, T.R. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276-82 (1989).

26. Krasilnikov, A.S., Yang, X., Pan, T. & Mondragon, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760-4 (2003).

27. 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**, 667-82 (2008).

28. Guo, F., Gooding, A.R. & Cech, T.R. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell* **16**, 351-62 (2004).

29. Golden, B.L., Kim, H. & Chase, E. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.* **12**, 82-9 (2005).

30. Zhang, Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins* **69 (Suppl 8)**, 108-17 (2007).

31. Ceschini, S. et al. Crystal structure of the fission yeast mitochondrial Holliday junction resolvase Ydc2. *EMBO J.* **20**, 6601-11 (2001).

32. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144-5 (2007).

33. Tanner, M. & Cech, T. Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium *Azoarcus*. *RNA* **2**, 74-83 (1996).

34. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in $tRNA^{Asp}$ transcripts. *J. Am. Chem. Soc.* **127**, 4659-67 (2005).

35. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. & Giddings, M.C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979-90 (2008).

36. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).

37. Kleywegt, G.J. & Jones, T.A. Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, 525-45 (1997).

38. Ding, F. et al. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* **14**, 1164-73 (2008).

Chapter 4

Non-Hierarchical Assembly of a Six-Component Ribonucleoprotein

4.1 Introduction

Large RNAs rarely achieve a functional structure without the aid of a wide variety of protein facilitators[1]. Protein facilitators are generally classified into two groups, cofactors and chaperones[2-4]. Cofactors bind tightly to an RNA and generally appear to promote or enhance the formation of otherwise unfavorable tertiary interactions. Conversely, chaperones facilitate the rearrangement of RNA secondary structure through transient interactions and allow an RNA to more readily achieve a thermodynamically favorable structure.

Folding for some RNAs is strongly hierarchal, such that the secondary structure forms first, followed by tertiary structure[5,6]. However, there are substantial and accumulating examples in which RNA tertiary structure either folds in tandem with the secondary structure or changes the secondary structures sampled during folding[7-15]. A protein cofactor that stabilizes an RNA tertiary structure element might then also have a dramatic impact on secondary structure. Such a function would represent a "non-hierarchical" contribution of protein binding and would exemplify a new type of protein-facilitated RNA folding, distinct from either cofactor or chaperone functions.

Group I introns are ideal model systems to study the impact of protein facilitators on RNA structure. Distinct versions of group I intron ribozymes interact with both protein cofactors and chaperones to achieve a conserved and catalytically active structure[2-4]. The conserved catalytic core is composed of two roughly co-axially stacked domains (termed the P5-P4-P6 and P9-P7-P8-P3 domains), which dock with a third domain (the P1-P2 domain) to form a complex active site[16-19] (Figure 4.1A). The two-step splicing reaction is initiated by binding of an exogenous guanosine to the folded catalytic core.

The bI3 ribonucleoprotein (RNP) is a six-component complex consisting of a group I intron RNA, a monomer of the bI3 maturase protein and two dimers of the Mrs1 protein[20,21] (Figure 4.1A). Detectable splicing by the bI3 RNA requires stable binding by both proteins. The maturase protein binds as a monomer to a peripheral element in the P5-P4-P6 domain, over 50 Å away from the active site[22]. The Mrs1 protein is a dimer and two dimers bind cooperatively to the bI3 RNA. Mrs1 binding promotes the formation of two GNRA tetraloop-receptor interactions that mediate long-range RNA inter-domain tertiary[23] (Figure 4.1A).

In this study, I probe the secondary and tertiary structure of the bI3 RNA in four conformational states: the free RNA, maturase-bound, Mrs1-bound, and final complete holo-complex states. I monitor local nucleotide dynamics and solvent accessibility at nearly every nucleotide in the RNA by high-throughput SHAPE and hydroxyl radical approaches, respectively. Unexpectedly, the free RNA is substantially misfolded relative to the conserved, catalytically active, secondary structure for group I introns. Binding by either the maturase or Mrs1 proteins individually stabilizes long-range RNA tertiary structures, but does not repair the misfolded secondary structure. However, in forming the final six-component complex, the RNA undergoes large-scale secondary structure rearrangements to form the active group I intron structure. Assembly of the bI3 RNP thus demonstrates a cooperative folding process in which RNA secondary and tertiary structure is dramatically different than when bound by either individual protein. Additionally, I visualize and describe a new, and unanticipated, role for protein cofactors in which the stabilization of long-range tertiary structures induces extensive rearrangements in RNA secondary structure, including regions distinct from, and located far away from, the protein

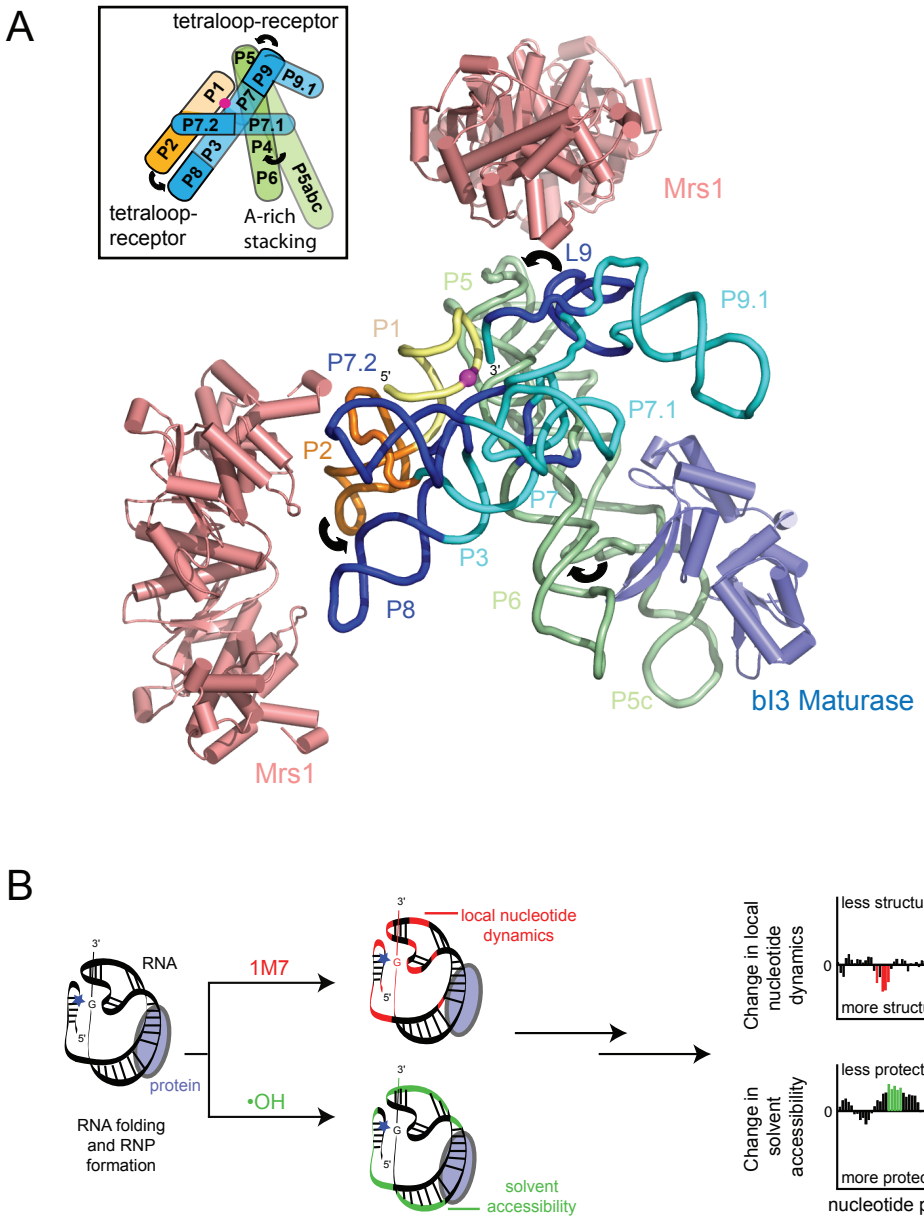

Figure 4.1. Comprehensive, high-throughput and high resolution structural analysis of the bI3 RNP. (A) Overview of the six-component bI3 architecture. RNA is colored by domain. Six misfolding elements (P1, P3, P7, P7.1, P9.1, and P9.0) are highlighted in light orange or cyan. The Mrs1 and maturase proteins are pink and blue, respectively. The scissile phosphate is shown as a magenta sphere. (B) Analysis of RNA structure as a function of protein binding using two different techniques. Local nucleotide flexibility and dynamics are probed using SHAPE chemistry (top, red). The hydroxyl-selective electrophile, 1M7, reacts preferentially at conformationally flexible nucleotides. Solvent accessibility at the RNA backone is approximated by hydroxyl radical (·OH) mediated cleavage (bottom, green). Changes in local nucleotide dynamics and solvent accessibility are detected at single nucleotide resolution in difference plots.

binding sites. In essence, the bI3 complex represents a clear example of non-hierarchical ribonucleoprotein assembly.

4.2 Results

4.2.1 Strategy. I use a combination of two high-throughput and high resolution techniques to comprehensively evaluate RNA structure and protein-assisted folding in the bI3 RNP (Figure 4.1B). I interrogate local nucleotide dynamics using SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) chemistry[24,25]. SHAPE uses a hydroxyl-selective electrophile that reacts at the ribose 2′-hydroxyl position and measures local nucleotide flexibility[24-26]. In a high-throughput SHAPE experiment, the bulky 2′-*O*-ester adducts are detected as stops to primer extension using color-coded fluorescent primers. After separation by capillary electrophoresis and performing a whole-trace Gaussian integration, I obtain absolute SHAPE reactivities for almost every nucleotide in the bI3 RNA. Data are normalized to a scale of 0 to ~2, where 1.0 is defined as the average intensity of highly reactive nucleotides. Nucleotides with moderate to high reactivities correlate strongly with single stranded nucleotides (orange and red nucleotides; Figure 4.2A). Weakly reactive nucleotides are conformationally constrained and are therefore likely either to be base paired or to participate in highly structured tertiary interactions (black nucleotides; Figure 4.2A).

In the second complementary approach, I measure the solvent accessibility of each nucleotide using high-throughput hydroxyl radical footprinting experiments. Hydroxyl radicals are generated in an *in situ* from H_2O_2 in a reaction catalyzed by Fe(II)-EDTA[27]. Backbone cleavage is roughly correlated with ribose solvent accessibility[28-30]. Hydroxyl

radical-induced cleavages at accessible nucleotides are, again, identified as stops to primer extension. Weakly cleaved nucleotides, taken to be cleavage intensities of less than one-half the mean, are solvent inaccessible and buried within the RNA or RNP structure (orange nucleotides, Figure 4.3A). High-throughput SHAPE and hydroxyl radical probing provide a quantitative, nucleotide-resolution, view of almost every nucleotide at distinct stages in the protein-facilitated folding of the bI3 intron.

4.2.2 The Free RNA is Misfolded Prior to Protein Binding. Previous analysis of the free bI3 RNA by SHAPE indicated that about half the RNA does not fold to the catalytically active and conserved group I intron secondary structure[31]. Six specific RNA elements are misfolded as compared to the secondary structure expected for the catalytically active structure (Figure 4.4A and emphasized in Figure 4.1A). Nucleotides in the conserved P3, P7 and P9.0 helices all have high SHAPE reactivities (nucleotides near positions 272, 213 and 313, respectively; Figures 4.2A, 4.4A). High SHAPE reactivities indicate that these nucleotides do not form stable, paired elements in the free RNA. In addition, nucleotides in the P1, P7.1 and P9.1 elements have SHAPE reactivity patterns consistent with alternate base pairing interactions. Instead of folding into individual helices, P7.1 and P9.1 sequences base pair with each other to form the P7.1/9.1 alternate helix (Figure 4.4A). Similarly, sequences in the highly conserved P1 helix base pair with nucleotides in the 3′ exon[31]. At a minimum, each of these six misfolded elements must undergo extensive secondary structure rearrangements to form a catalytically active group I intron complex.

I also assessed the degree of solvent accessible tertiary structure formed in the free RNA using hydroxyl radical footprinting. Most nucleotides in the free RNA are accessible

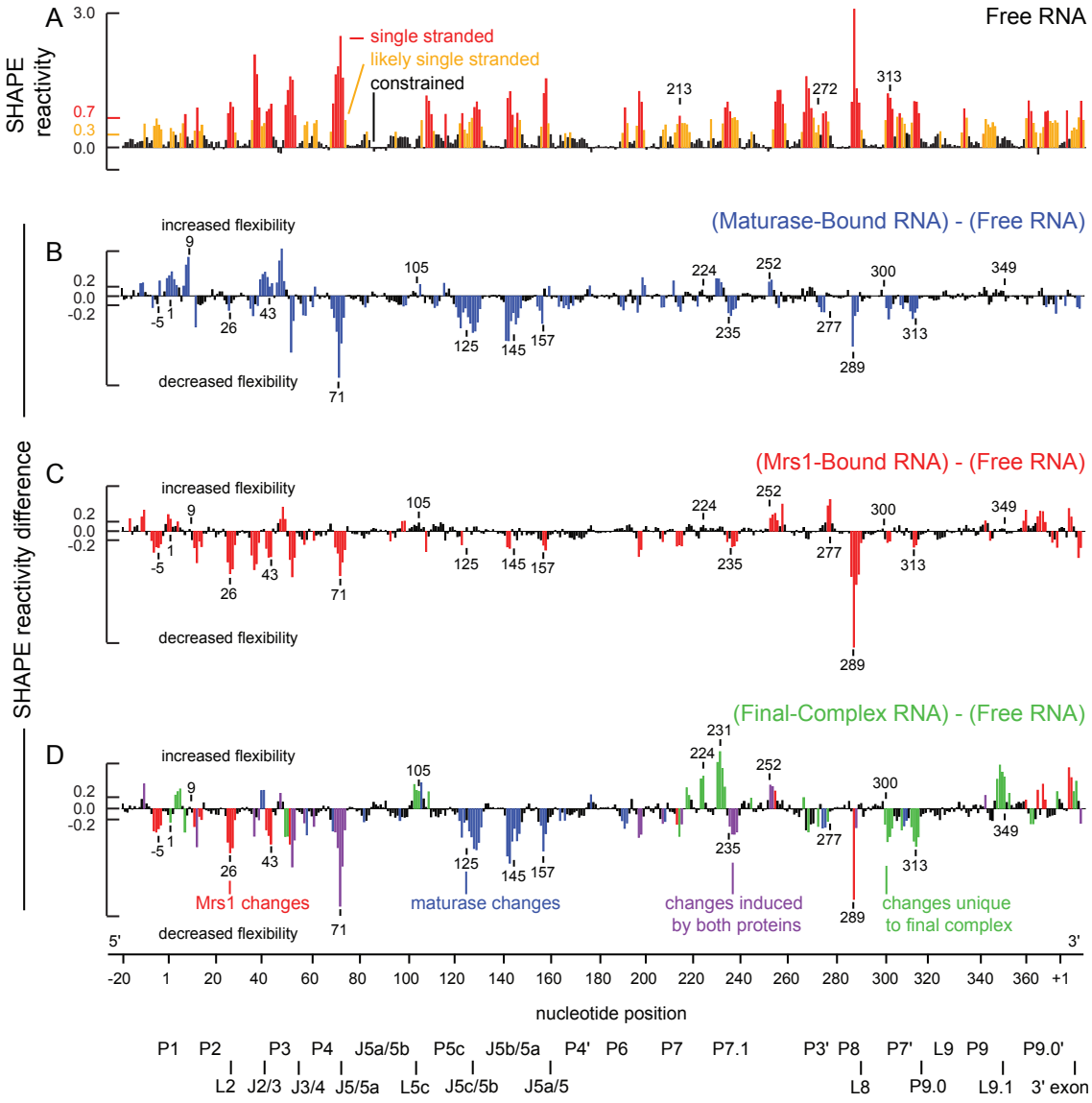

Figure 4.2. Protein-assisted folding of the bI3 RNA monitored by SHAPE. (A) Reactivity as a function of nucleotide position for the free RNA. Bars are colored by reactivity. (B-D) Difference plot histograms for maturase-bound, Mrs1-bound and the final complex. Nucleotides with significant changes in reactivities are colored by the protein that induces a given effect; blue for maturase, red for Mrs1, purple if affected by both proteins and green for unique changes seen only in the final complex. Structural landmarks are highlighted below the axis.

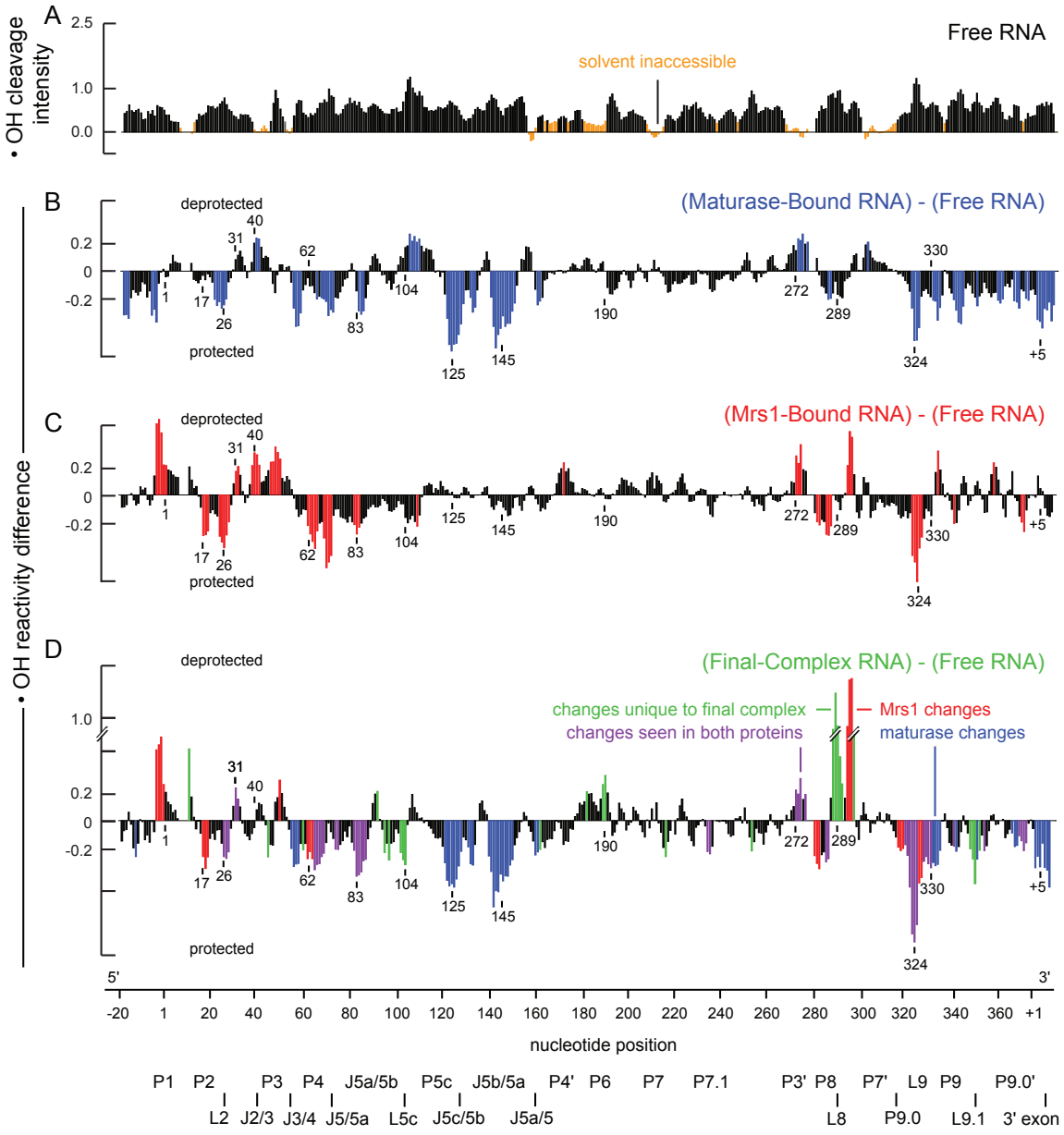

Figure 4.3. Protein-mediated changes in backbone solvent accessibility in the bI3 RNA. (A) Cleavage intensity versus nucleotide position for the free RNA. Nucleotides with reactivities less than one-half the mean are emphasized in orange. (B-D) Difference plot histograms for maturase-bound, Mrs1-bound and the final complex RNAs. Nucleotides with significant protein-induced changes in reactivity are colored as in Figure 2. Structural landmarks are highlighted below the axis.

to cleavage (back nucleotides, Figure 4.3A), indicating that little tertiary structure is formed in the free RNA prior to protein binding. A few regions are inaccessible to hydroxyl radical cleavage: those regions include the P1, J2/3, P3, P4, J8/7, P7 and P9.0 elements. Interestingly, several of these solvent inaccessible regions, including sequences corresponding to P1, P3 and P7 are elements that are clearly misfolded in the protein-free group I intron secondary structure. Thus, even where there is evidence for the formation of solvent-excluding tertiary interactions, it is not clear that those recapitulate native interactions. Taken together, these SHAPE and hydroxyl radical data suggest that the free RNA forms a stable alternate secondary structure which is incapable of making native tertiary interactions or folding into a compact and catalytically active structure.

4.2.3 The bI3 RNA Achieves a Dramatically Different Structure in the Final Complex. In the final complex, where both the bI3 maturase and Mrs1 proteins are stably bound, I observe complex and extensive SHAPE reactivity changes consistent with large-scale secondary structure rearrangements. I quantified these changes by subtracting the nucleotide-resolution reactivity pattern of the final complex from those for the free RNA in a difference plot. Absolute differences in reactivity greater than 0.2 are considered significant. In the P7.1 element, for example, nucleotides 224 and 231 increase in SHAPE reactivity, whereas, nucleotide 235 decreases in reactivity (labeled, Figure 4.2D). These reactivity changes are consistent with the interpretation that, in the final complex, the P7.1/9.1 alternate helix has been disrupted and the phylogenetically conserved P7.1 helix has formed (compare structural contexts of P7.1 nucleotides; Figure 4.4A, 4.4D). All of the six previously identified misfolded elements (P1, P3, P7, P7.1, P9.1 and P9.0) now have SHAPE reactivities consistent with formation of the accepted group I intron

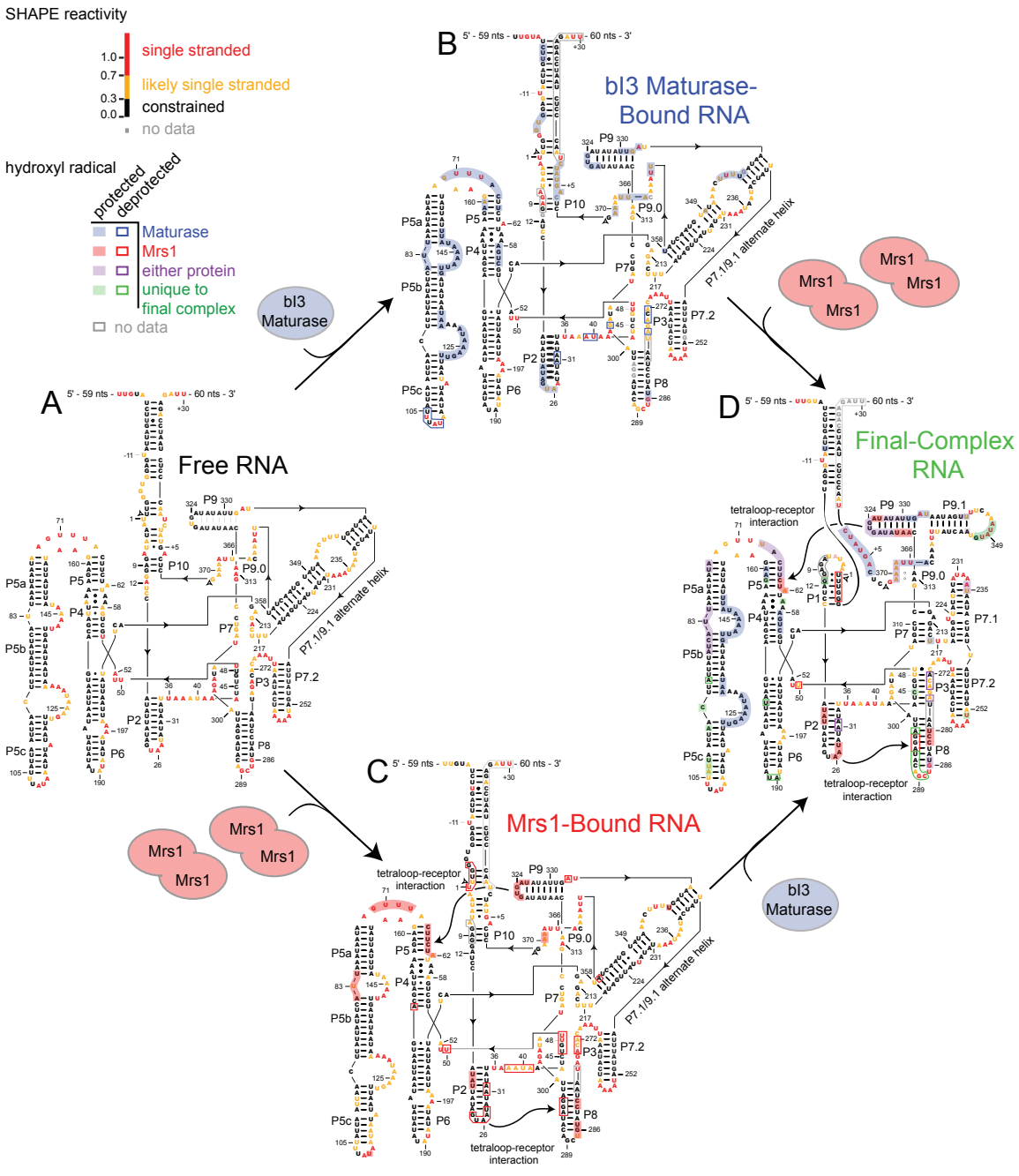

Figure 4.4. RNA folding in the six-component bI3 RNP. Secondary structural models for the (A) free RNA, (B) maturase-bound RNA, (C) Mrs1-bound RNA, and (D) final complex RNA. Nucleotides are colored to reflect their SHAPE reactivities in each RNA state. Filled and hollow boxes highlight hydroxyl radical cleavage protections and enhancements, respectively, and are colored according to their protein effector; blue, red, purple and green for maturase, Mrs1, either protein, or unique to the final complex, respectively. Px labels indicate paired elements in the RNA.

secondary structure (Figure 4.4D). In addition, several linker regions and loops have low reactivities in the final complex, suggesting that the RNA is forming the compact tertiary interactions necessary for group I intron splicing.

This conclusion is corroborated by analyzing nucleotide solvent accessibility in the final complex. Differences in hydroxyl radical cleavage were also visualized in a difference plot. Upon formation of the final complex, one-third of the nucleotides in the bI3 RNA become protected from cleavage (Figure 4.3D). In addition, several sets of nucleotides, such as nucleotides 190 and 289, show enhanced cleavage indicating that they become more exposed to solvent. I interpret these data to indicate that a significant structural rearrangement occurs in the RNA at these nucleotides. In all, structural changes are seen in nearly all major RNA structures (boxes, Figure 4.4D), with exception of the P6, P7 and P7.1 helices. These data support the interpretation that the RNA undergoes extensive conformational rearrangements in folding to the final state and that the final complex contains an RNA with a tightly folded and compact tertiary structure.

4.2.4 The bI3 Maturase Stabilizes Tertiary Structures in the P5-P4-P6 Domain. I then used SHAPE to probe the RNA bound by the maturase protein. My goal was to determine which of the observed structural rearrangements that distinguish the free RNA and the final complex can be attributed to maturase binding. The bI3 maturase binds in the P5c helix of the P5-P4-P6 domain and promotes the tertiary folding of this domain, including formation of two sets of A-minor interactions[22]. As expected, SHAPE reactivity decreases in the flexible joiner regions of the P5-P4-P6 domain, including in the J5/5a, J5c/5b, J5b/5a and J5a/5 elements (highlighted in blue at positions 71, 125, 145 and 157; Figure 4.2B). These decreases in reactivity are maintained in the final complex

(blue/purple bars, Figure 4.2D). Other changes, outside of the P5-P4-P6 domain, are also maintained in the final complex, including in the J3/4, P7.1, P3, P7.2, and J8/7 elements (blue/purple bars at positions 40, 235, 275, 252 and 307; Figure 4.2D). Maturase binding also induces changes in other regions of the RNA, many of which are not maintained in the final complex (see nucleotides 1, 9 and 43 in Figure 4.2B,D).

I further examined tertiary structure changes induced by maturase binding using hydroxyl radical footprinting. Maturase binding has dramatic effects in the P5-P4-P6 domain and at the 5′ and 3′ ends of the intron (blue bars, Figure 4.3B). As expected, several elements in the P5-P4-P6 domain become protected, including the two sets of A-minor interactions (nucleotides at positions 125 and 145; Figures 4.3B and 4.4B). Additional cleavage differences are seen in the 5′ and 3′ exons, and in the P2, P3, P8, P9 and P9.1 helices (blue nucleotides at positions 1, 26, 272, 289, 324 and +5; Figure 4.3B,D). Most of these tertiary changes induced by maturase binding are maintained in the final complex (blue/purple bars, Figure 4.3D). However, similar to the observation using SHAPE, a subset of changes are not maintained in the final complex. These regions include the 5′ exon, regions at the 3′ end of the intron and in the P5-P4-P6 domain near nucleotide 104 (compare Figures 3B and 3D). These data suggest, first, that maturase binding promotes tertiary structure folding in the P5-P4-P6 domain and, second, that some regions in the maturase-stabilized structure undergo additional conformational rearrangements to form the final complex.

4.2.5 Mrs1 Stabilizes Two Tetraloop-Receptor Interactions. A difference plot analysis of the SHAPE data for the Mrs1-bound and the free RNAs shows decreases in reactivity throughout the RNA (red bars, Figure 4.2C). These experiments emphasize that

Mrs1 induces extensive changes in the stability of individual secondary structure elements. However, as with the maturase, Mrs1 binding has little effect on the P1, P7.1 or P9.1 elements, indicating that these structures are sill misfolded. Most Mrs1 effects on the bI3 structure are maintained in the final complex (red/purple bars, Figure 4.2D).

Finally, hydroxyl radical footprinting of the Mrs1-bound RNA emphasizes that most changes are localized precisely at the two GNRA tetraloop-receptor interactions in the bI3 RNA (red bars, Figure 4.3C). Both the L2 and L9 tetraloops are strongly protected from cleavage upon Mrs1 binding (at positions 26 and 324, Figure 4.3C). In addition, receptor helices in P8 and P5 display both protections and enhancements from cleavage (see nucleotides 286, 295 and 62; Figure 4.3C). Elements near the tetraloop-receptor motifs are also affected by Mrs1 binding, including the P1, P3, and P4 helices, and the J2/3 and J5/5a joiner regions. Structural changes visualized by hydroxyl radical cleavage are corroborated by SHAPE because single stranded sequences proximal to the receptor helices and tetraloops, such as the J2/3, J3/4, J5/5a and L8, also show reduced reactivities (red nucleotides around 43, 51, 71, and 289, respectively; Figure 4.2C). That major changes in the hydroxyl radical cleavage pattern are localized to the two tetraloop-receptor motifs indicate that Mrs1 promotes the formation of this distinct tertiary structure in the RNA but does not directly modulate other tertiary structures in the RNA.

4.2.6 Cooperative Interactions in the bI3 RNP. Individually, both the maturase and Mrs1 proteins stabilize distinct tertiary structures (blue and red bars; Figures 4.2D, 4.3D). In general, tertiary structure changes induced by binding of either protein alone occur in peripheral structural elements located distant from the catalytic core. However, effects of protein binding are seen throughout the RNA, indicating that structural changes

propagate through connected RNA structures. There also exists a dynamic interplay between the maturase and Mrs1 proteins. For example, SHAPE reactivity around nucleotide 43 increases upon maturase binding, but decreases significantly upon Mrs1 binding (labeled, Figure 4.2). In the final complex, reactivity resembles that of Mrs1 binding alone, suggesting that Mrs1 has the stronger influence on RNA structure in this region. Thus, stabilization of tertiary structures upon protein binding influences the stability of local and distal secondary structures.

Both SHAPE and hydroxyl radical experiments also reveal reactivity differences that occur only in the final complex and not in either the maturase- or Mrs1-bound RNAs (green bars; Figures 4.2D, 4.3D). Strikingly, all six of the RNA elements that are misfolded in the free RNA only show significant SHAPE reactivity changes in the final complex (green nucleotides in P1, P3, P7, P7.1, P9.1 and P9.0, Figure 4.2D). These differences in SHAPE reactivity correlate directly with the formation of the active secondary structure for the bI3 RNA (see P1, P3, P7, P7.1, P9.1 and P9.0, Figure 4.4D). Hydroxyl radical data corroborate this observation and demonstrate clear RNA structure rearrangements unique to the final complex RNA (green bars, Figure 4.3D).

4.3 Discussion

The folding of the bI3 RNP is a strikingly cooperative process in which two protein cofactors, each of which stabilizes distinct tertiary interactions, work together to induce large-scale RNA secondary structure rearrangements. SHAPE and hydroxyl radical experiments show that both the maturase and Mrs1 proteins perform traditional RNA cofactor functions by stabilizing specific tertiary interactions. The maturase stabilizes the

U-shaped structure in the P5-P4-P6 domain and Mrs1 promotes the formation of two GNRA tetraloop-receptor interactions that link the three RNA domains conserved among group I introns. Neither protein alone repairs the extensive misfolding seen in the free RNA. Instead, the RNA undergoes a large-scale change in secondary structure only when bound by both proteins in the final complex. Overall RNP assembly is thus highly cooperative and starkly non-hierarchical: protein-mediated tertiary structure stabilization strongly influences the RNA secondary structure.

The proteins that facilitate RNA folding are commonly divided into the categories of cofactors and chaperones[2-4]. While the bI3 RNP proteins each behave as cofactors, whose stable binding mediate important tertiary interactions, their combined role to rearrange the secondary structure does not fit the classic definitions of either cofactor or chaperone. Both the maturase and Mrs1 stably bind the RNA (with binding constants of 1 nM and 20 nM respectively[21]) and function individually to stabilize the RNA tertiary structure, much like conventional cofactors. Despite this, the large-scale rearrangement in secondary structure induced by the simultaneous binding of both proteins is more characteristic of an RNA chaperone. In addition, structural changes in the bI3 RNA occur in elements that are not directly bound by either protein, a characteristic that does not fit the definition of either cofactor or chaperone.

In the case of the bI3 RNP, I propose that binding by both proteins induces significant tension in the connecting RNA segments to force secondary structure rearrangement in regions distal from the protein binding sites. In the individually bound states, both the maturase and Mrs1 proteins induce long-range structural changes throughout the RNA. A subset of these effects are not maintained in the final complex.

For instance, local structures in the misfolded P1 and P3 elements in the catalytic core are clearly modulated by the individual maturase or Mrs1 protein binding. These structural changes do not reflect native-like interactions because they are not represented in the final complex RNA (for example, positions near 1, 9, 43 and 277 in Figure 4.2). However, the structural effects of the individual protein bound states converge when both proteins bind the RNA. I propose that this accumulation of structural pressure propagates through the bI3 RNA, eventually causing large-scale structural rearrangements in the final complex RNA.

These structural studies of the cooperative folding of the bI3 RNP highlight the importance of obtaining comprehensive and high-resolution information when studying complex RNA-protein interactions. Several individual RNA cofactors have been shown to influence the formation of tertiary structures that lie 30-50 Å from their primary interaction sites[32-34]. It was only possible to generate strong hypotheses regarding the impact of maturase and Mrs1 binding because local nucleotide dynamics and solvent accessibility were measured at >96% of all analyzed nucleotides in the bI3 RNA. As the high-throughput and high-resolution approaches described here are applied more broadly to large protein-RNA complexes, proteins previously characterized to make simple contributions to RNA folding are likely also to be found to exert highly complex, cooperative, and non-additive effects on RNA folding. These long-range protein-induced affects also add another level of control, and potential switching functions, in RNP complexes. In large and dynamic RNPs, these cooperative effects are likely to be important to stabilize alternate structures in the functional core of the complex, distant from protein binding sites including in large RNPs like the ribosome and splicesome.

4.4 Experimental Procedures

4.4.1 bI3 RNA, Maturase and Mrs1 Expression. The bI3-ΔL8 RNA was generated and purified as described[31]. The RNA sequence contained 7 nts of 5'-vector sequence, the entire 77 nt 5'-exon, the 372 nt intron, 30 nts of the 3'-exon, and 60 nts of 3'-vector sequence. The ΔCys-bI3 maturase and Mrs1 proteins were expressed and purified as described[20] with the exception that Mops was used as the buffer, dithiothreitol (DTT) was 5 mM, and glycerol was reduced to 10% (v/v).

4.4.2 High-Throughput SHAPE Experiments. The bI3 RNA (3 pmol) was heated in water at 1 min 95 °C, snap-cooled and placed on ice for 1 min, and re-folded at 37 °C for 10 min in reaction buffer [40 mM Mops-NaOH (pH 8.0), 80 mM potassium acetate (KOAc) (pH 8.0), 20 mM $MgCl_2$] in a final volume of 23 µl. Proteins were allowed to bind at 37 °C for 30 min. Experiments with the free RNA were supplemented with 4 µl of protein buffer [40 mM Mops-NaOH (pH 7.5), 400 mM NaCl, 5 mM DTT, 10 % (v/v) glycerol]; if present, 2 µl of 6 µM maturase and Mrs1 dimer proteins were added. Reaction mixtures were added to 3 µl of 5 mM 1M7 (dissolved in anhydrous DMSO) at 37 °C and allowed to react for 70 s (equal to 5 hydrolysis half lives)[25]. No-reagent control reactions were performed by omitting 1M7. Complexes were subsequently incubated with proteinase K [60 µg, 37 °C, 30 min] and the digested proteins were removed by phenol:cholorform:isoamyl alcohol (25:24:1) extraction. Modified RNA was subsequently precipitated with ethanol and resuspended in water. Sites of RNA modification were identified by reverse transcriptase-mediated primer extension[31] using fluorescently labeled primers [6-FAM, HEX, NED and VIC labeled primers were used for the (+) and (-) 1M7 reactions and dideoxy sequencing in the presence of 0.25 mM dTTP or ddATP,

respectively]. (+) and (–) reagent and dideoxy sequencing primer extension reactions were combined, recovered by precipitation with ethanol, and resuspended in 10 µL deionized formamide; fluorescently labeled DNAs were resolved by capillary electrophoresis using a Applied Biosystems 3130 capilliary electrophoresis instrument.

4.4.3 High-Throughput Hydroxyl Radical Cleavage Experiments. RNA-protein complexes were formed as described for SHAPE experiments in a volume of 21 µl. Reactions were initiated by sequential addition of 3 µl each of 2.5 mM $(NH_4)Fe(SO_4)_2$/3.75 mM EDTA (pH 8.0), 0.1% H_2O_2, and 50 mM sodium ascorbate. Reactions were quenched after 5 min at 37 °C with 10.5 µl 75% glycerol (v/v). The no-reaction control was performed in the absence of Fe(II)-EDTA. Reactions were treated with proteinase K and visualized by primer extension as described above for the SHAPE experiments.

4.4.4 Data Analysis. SHAPE and hydroxyl radical data from the Applied Biosystems 3130 were analyzed using ShapeFinder[35] as described[31]. After integration of all peaks in the (+) and (–) reagent traces, reactivities were scaled by discarding the top 2% most reactive peaks and dividing by the average intensity of the next 8% of reactive peaks. Hydroxyl radical cleavage data were additionally smoothed over a window of three nucleotides. For secondary structure predictions, SHAPE data were converted to a pseudo-free energy change term and used to constrain structure models (slope and intercept were 2.5 and -0.6, respectively)[36].

4.5 References

1. Gesteland, R.F., Cech, T.R. & Atkins, J.F. *The RNA World*, (Cold Spring Harbor Laboratory Press, New York, 2006).

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

3. Weeks, K.M. Protein-facilitated RNA folding. *Curr. Opin. Struct. Biol.* **7**, 336-42 (1997).

4. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5**, 908-19 (2004).

5. Brion, P. & Westhof, E. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113-37 (1997).

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

7. LeCuyer, K.A. & Crothers, D.M. The Leptomonas collosoma spliced leader RNA can switch between two alternate structural forms. *Biochemistry* **32**, 5301-11 (1993).

8. Gluick, T.C. & Draper, D.E. Thermodynamics of folding a pseudoknotted mRNA fragment. *J. Mol. Biol.* **241**, 246-62 (1994).

9. Wu, M. & Tinoco, I., Jr. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11555-60 (1998).

10. Buchmueller, K.L., Webb, A.E., Richardson, D.A. & Weeks, K.M. A collapsed non-native RNA folding state. *Nat. Struct. Biol.* **7**, 362-6 (2000).

11. Zheng, M., Wu, M. & Tinoco, I., Jr. Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the Tetrahymena group I ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3695-700 (2001).

12. Chauhan, S. et al. RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. *J. Mol. Biol.* **353**, 1199-209 (2005).

13. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in $tRNA^{Asp}$ transcripts. *J. Am. Chem. Soc.* **127**, 4659-67 (2005).

14. Chauhan, S. & Woodson, S.A. Tertiary interactions determine the accuracy of RNA folding. *J. Am. Chem. Soc.* **130**, 1296-303 (2008).

15. Wang, B., Wilkinson, K.A. & Weeks, K.M. Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry. *Biochemistry* **47**, 3454-61 (2008).

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

17. Cech, T.R. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* **59**, 543-68 (1990).

18. Woodson, S.A. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**, 324-30 (2005).

19. Vicens, Q. & Cech, T.R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**, 41-51 (2006).

20. Bassi, G.S., de Oliveira, D.M., White, M.F. & Weeks, K.M. Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 128-33 (2002).

21. Bassi, G.S. & Weeks, K.M. Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst. *Biochemistry* **42**, 9980-8 (2003).

22. Longo, A. et al. Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* **12**, 779-87 (2005).

23. Duncan, C.D. & Weeks, K.M. The Mrs1 splicing factor binds the RNA tetraloop-receptor motif. *Biochemistry Submitted*(2009).

24. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223-31 (2005).

25. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144-5 (2007).

26. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G. & Weeks, K.M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S^2) in RNA. *J. Am. Chem. Soc.* **130**, 12244-5 (2008).

27. Tullius, T.D. & Dombroski, B.A. Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469-73 (1986).

28. Latham, J.A. & Cech, T.R. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276-82 (1989).

29. Tullius, T.D. & Greenbaum, J.A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127-34 (2005).

30. Balasubramanian, B., Pogozelski, W.K. & Tullius, T.D. DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9738-43 (1998).

31. Duncan, C.D. & Weeks, K.M. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* **47**, 8504-13 (2008).

32. Kuglstatter, A., Oubridge, C. & Nagai, K. Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nat. Struct. Biol.* **9**, 740-4 (2002).

33. Maeder, C. & Draper, D.E. A small protein unique to bacteria organizes rRNA tertiary structure over an extensive region of the 50 S ribosomal subunit. *J. Mol. Biol.* **354**, 436-46 (2005).

34. Caprara, M.G., Chatterjee, P., Solem, A., Brady-Passerini, K.L. & Kaspar, B.J. An allosteric-feedback mechanism for protein-assisted group I intron splicing. *RNA* **13**, 211-22 (2007).

35. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. & Giddings, M.C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979-90 (2008).

36. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 97-102 (2009).