



universität
wien

DISSERTATION

Titel der Dissertation

„DEAD-box protein facilitated folding of the group II intron
ai5 γ in vivo“

Verfasser

Mag. Andreas Liebeg

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Studienkennzahl lt. Studienblatt:

A 094 490

Dissertationsgebiet lt. Studienblatt:

Molekulare Biologie

Betreuerin / Betreuer:

Mag. Dr. Christina Waldsich, Privatdoz.

Table of Contents

1 Introduction.....	8
1.1 RNA structure and folding.....	8
1.1.1 Secondary structure elements of RNA.....	8
1.1.2 Common tertiary structure motifs in RNA.....	9
1.1.3 Metal ions and RNA folding.....	11
1.2 Group II introns.....	12
1.2.1 Biological relevance of group II introns.....	12
1.2.2 The best studied group II intron: ai5 γ	12
1.2.3 Group II introns are capable of self-splicing.....	14
1.2.4 Structural organization of group II introns.....	15
1.2.5 Group II intron folding.....	19
1.3 DEAD-box proteins.....	24
1.3.1 Structural organization of DEAD-box proteins.....	25
1.3.2 Function of DEAD-box helicases.....	26
1.3.3 The DEAD-box protein Mss116p.....	31
2 Material and Methods.....	41
2.1 Mutation of MSS116.....	41
2.2 RNA preparation.....	46
2.3 cDNA synthesis.....	47
2.4 Data analysis.....	50
2.5 Dot spots.....	50
3 Results.....	52
3.1 Introducing mutations in chromosomal copy of the mss116 gene.....	52
3.2 Conserved motifs of Mss116p and their role in respiratory growth of yeast...54	
3.3 Dissecting the motifs of Mss116p essential for splicing of ai5 γ in vivo.....56	
3.4 The role of ATPase activity as well as RNA binding and unwinding of Mss116p in folding of ai5 γ	58
3.4.1 General approach.....	58
3.4.2 The intracellular structure of ai5 γ	60

3.4.3 Mss116p ATPase activity and ai5 γ folding.....	62
3.4.4 The basic tail of Mss116p and ai5 γ folding.....	82
3.4.5 Helicase activity of Mss116p and ai5 γ folding.....	98
3.4.6 Exons are accessible to DMS in all different strains.....	114
4 Discussion.....	116
5 References.....	126
6 Appendix.....	147
6.1 Publications.....	147
6.2 Curriculum vitae.....	148

Figure Index

Figure 1.1: Influence of ions on RNA folding.....	13
Figure 1.2: Graphical representation of ai5 γ	15
Figure 1.3: Secondary structure and implied tertiary interactions of D5 and U6 RNA..	21
Figure 1.4: In vitro folding pathway of the D135 ribozyme.....	23
Figure 1.5: Mss116 with bound RNA oligonucleotide and ATP analog.....	26
Figure 1.6: DII of Mss116p and C-terminal extension bound to dsRNA.....	30
Figure 1.7: Unwinding mechanism of DEAD-box proteins.....	32
Figure 1.8: Splicing efficiency of the ai5 γ intron in vivo in the absence and presence of Mss116p.....	33
Figure 1.9: Important amino acids of Mss116p.....	36
Figure 1.10: The two paradigms of Mss116p facilitated ai5 γ folding.....	38
Figure 1.11: Mss116p-induced conformational changes in ai5 γ	40
Figure 2.1: Overview of the 2-step method for introducing of point mutations in the yeast protein Mss116p.....	45
Figure 3.1: Verification of correct location of integration cassette harboring the desired mutation of mss116 gene and the marker gene NAT1 in chromosome IV locus.....	55
Figure 3.2: Aerobic versus anaerobic growth of yeast strains with different genetic background regarding Mss116p.....	56
Figure 3.3: Splicing of ai5 γ in yeast strains expressing wild type Mss116p or its various variants.....	59
Figure 3.4: DMS chemical probing. (A) DMS methylates N1 of adenosines and N3 of cytosines.....	60
Figure 3.5: In vivo DMS modification pattern of ai5 γ	63
Figure 3.6: DMS accessibility of the 5' part of the κ - ζ element in the K158R strain.....	66
Figure 3.7: DMS accessibility around the 3' part of the κ - ζ element in the K158R strain.	68
Figure 3.8: DMS accessibility around EBS2 and coordination loop in the K158R strain.	71

Figure 3.9: DMS accessibility of the very 5' end of ai5 γ in the K158R strain.....	72
Figure 3.10: DMS accessibility around stem d3 in the K158R strain.....	74
Figure 3.11: DMS accessibility around stem c1 in the K158R strain.....	75
Figure 3.12: DMS accessibility of J2/3 and D3 in the K158R strain.....	77
Figure 3.13: DMS accessibility of D4 in the K158R strain.....	78
Figure 3.14: DMS accessibility of D5 in the K158R strain.....	80
Figure 3.15: Differential summary map.....	83
Figure 3.16: DMS accessibility of the 5' part of the κ - ζ element in the DXTD strain..	85
Figure 3.17: DMS accessibility around the 3' part of the κ - ζ element in the Δ CTD strain.....	87
Figure 3.18: DMS accessibility around EBS2 and the coordination loop in the Δ CTD strain.....	88
Figure 3.19: DMS accessibility of the very 5' end of ai5 γ in the Δ CTD strain.....	89
Figure 3.20: DMS accessibility around the d3 stem in the Δ CTD strain.....	91
Figure 3.21: DMS accessibility around the c1 stem in the Δ CTD strain.....	92
Figure 3.22: DMS accessibility of D5 in the Δ CTD strain.....	94
Figure 3.23: DMS accessibility of J2/3 and D3 in the Δ CTD strain.....	95
Figure 3.24: DMS accessibility of D4 in the Δ CTD strain.....	96
Figure 3.25: Differential summary map.....	99
Figure 3.26: DMS accessibility of the 5' part of the κ - ζ element in the Q412A strain.	102
Figure 3.27: DMS accessibility around the 3' part of the κ - ζ element in the Q412A strain.....	103
Figure 3.28: DMS accessibility around EBS2 in the Q412A strain.....	105
Figure 3.29: DMS accessibility of the very 5' end of ai5 γ in the Q412A strain.....	106
Figure 3.30: DMS accessibility around the d3 stem in the Q412A strain.....	107
Figure 3.31: DMS accessibility around the c1 stem in the Q412A strain.....	109
Figure 3.32: DMS accessibility of J2/3 and D3 in the Q412A strain.....	110
Figure 3.33: DMS accessibility of D4 in the Q412A strain.....	111
Figure 3.34: DMS accessibility of D5 in the Q412A strain.....	112
Figure 3.35: Differential summary map.....	115

Figure 3.36: DMS accessibility of the exons.....	117
Figure 4.1: Comparison of DMS modification pattern of ai5 γ in different genetic backgrounds.....	121

Zusammenfassung

Das DEAD-box protein Mss116p ist unabdingbar für das aerobe Wachstum von Hefe, indem es alle Gruppe I und II Introns im mitochondriellen Genom beim Splicen unterstützt. Das Gruppe II intron ai5 γ verbleibt ungefalted, falls Mss116p *in vivo* nicht präsent ist. Es ist jedoch rätselhaft wie Mss116 ai5 γ dabei hilft, seine aktive Struktur zu erlangen. Eine Möglichkeit besteht darin, dass das DEAD-box Protein seine Helikaseaktivität dazu nutzt, falsch gefaltete Regionen wieder aufzulösen, um eine korrekte Faltung zu erlauben. Im Gegensatz dazu wurde vorgeschlagen, dass Mss116p eine Intermediärstruktur stabilisiert, die essenziell für die weitere Faltung ist. Mit einer PCR-basierte Methode führten wir unterschiedliche Mutationen in das Genom von Hefe ein, welche wichtige Aktivitäten von Mss116p ausschalten. Diese Aktivitäten bestanden in der spezifischen Bindung des RNA-Substrats, in der Bindung und Hydrolyse von ATP sowie im Aufwinden von doppelsträngiger RNA. Anschließend kartierten wir die Struktur von ai5 γ in den unterschiedlichen Hefestämmen mit chemischer Modifizierung *in vivo*. Die Struktur von ai5 γ ohne ATP-Bindung und –Hydrolyse, wo ein wiederholtes Binden von Mss116p nicht mehr möglich ist, ähnelt stark der Situation ohne Kofaktor. Bei Entfernen des C-terminalen Schwanzes ist das Gruppe II Intron ebenso stark entfaltet. Bei Inaktivierung der Helikaseaktivität waren die geringsten Effekte beobachtbar, wobei diese sich hauptsächlich auf Konstituenten des aktiven Zentrums beschränkten. Mss116p scheint ein multifunktionelles Enzym zu sein, welches seinen C-terminalen Schwanz zur spezifischen Bindung des RNA-Substrats nutzt und mittels mehreren Runden von Protein-RNA-Bindungsvorgängen eine intermediäre Struktur stabilisiert. Außerdem stellt es mittels seiner Helikaseaktivität sicher, dass sich native Strukturelemente nicht vorzeitig bilden.

Abstract

The DEAD-box protein Mss116p is essential for aerobic growth of yeast assisting in splicing of all group I and II introns found in the mitochondrial genome. The group II

intron ai5 γ remains largely unfolded if Mss116p is not present *in vivo*, but it is enigmatic how Mss116p helps ai5 γ adopting its native structure. It was proposed that the DEAD box protein uses its helicase activity to unwind misfolded species thus allowing native folding. In contrast, a smooth folding landscape was postulated where Mss116p does not act as a chaperone but by stabilizing an on-pathway intermediate at the core of the ribozyme. By applying a PCR based method we generated yeast strains harboring different chromosomal mutations which disrupt specific activities of Mss116p, namely ATP binding and hydrolysis, RNA binding and RNA unwinding. We then mapped the structure of ai5 γ in these different genetic backgrounds using an *in vivo* chemical probing technique. The structure of ai5 γ with disrupted ATP binding and hydrolysis activity, where protein turnover is abolished, highly resembles the state in a *Dmss116* strain. Deleting the C-terminal tail, which is important for specific RNA binding, also results in a largely unfolded group II intron. Disrupting the helicase activity of Mss116p showed the least impact on ai5 γ 's structure, being mostly clustered around active site constituents. Thus Mss116p appears to be a multifunctional enzyme using specific binding by its C-terminal tail as well as protein turnover exerted by ATP binding and hydrolysis to stabilize an on-pathway intermediate while ensuring correct active site assembly by unwinding prematurely formed native structure elements.

1 Introduction

1.1 RNA structure and folding

When Francis Crick first introduced his “central dogma of molecular biology” consisting in an information flow from DNA via RNA to proteins, it was widely acknowledged that proteins exist in distinct three dimensional configurations¹. In contrast RNA was thought to be a short lived information carrier without the necessity for a specific structure. The crystallization of tRNA in the late 60s of the last century revealed that RNA molecules also adopt complex conformations in space² and more than a decade later it was shown that RNA does not only have the ability to fold into a defined tertiary structure but that RNA molecules are also able to perform catalytic activities just as proteins do^{3,4}. These two findings triggered a paradigm shift to a cellular world in which RNA has an inevitable part in all processes with its ability to adopt sophisticated structures laying the foundation for a vast variety of functions. Folding of RNA into these structures differs tremendously from protein folding.

1.1.1 Secondary structure elements of RNA

Proteins have 20 building blocks, showing great difference in charge, polarity and size. In RNA there are only 4 basic building blocks, which are in comparison quite similar in the mentioned characteristics and folding results in a large electrostatic potential due to its backbone phosphates, a fact that makes positively charged ions crucial for RNA folding. The RNA sequence defines the primary structure of RNA molecules. Complementary regions in this primary sequence form Watson-Crick base pairs resulting in helices that are connected by single stranded regions⁵. These RNA secondary structure elements are inherently stable which can lead to intermediate non-native structures interfering with native folding⁶. Helices in RNA have special properties compared to their counterparts found in DNA molecules. In contrast to the B-

form helices found in DNA, stems present in RNA adopt an A-form geometry⁷. This is a direct result of the RNA 2'-OH group which favors the C3'-endo conformation of the ribose as the C2'-endo conformation would result in a steric clash with the 3'OH-group. The resulting helix has a wide and shallow minor groove while its major groove is narrow and deep.

The 3'-and 5'-ends of a double helix are often connected by hairpin loops that must close with a Watson-Crick pairing and consist of 2 to 14 nucleotides⁸. The GNRA tetraloop is the most common and well studied type of closing loop and it plays an important role in tertiary structure stabilization⁹. Internal loops are introduced in a RNA helix by nucleotides that do not form Watson-Crick base pairs in at least one of the involved strands. Non-canonical base pairing is frequent in these nucleotides. This feature in symmetric loops for example allows introducing a distortion in a normally very symmetric RNA helix by unwinding, unstacking or kinking¹⁰. An important property of asymmetrical internal loops is the introduction of sharp turns such as K-turns that are very important for the tertiary structure of a RNA molecule^{11,12}.

Multi-way junctions describe a region in which three or more helices meet linking these with none or more nucleotides. They play an important role in constraining the possible orientations of the involved helices¹³. In three-way junctions, for example, it is proposed that coaxial stacking happens opposite of the longest stretch of junction nucleotides¹⁴.

1.1.2 Common tertiary structure motifs in RNA

In secondary structure maps of RNA molecules junctions, bulges and loops are depicted as single-stranded regions but in fact they are often paired in non-canonical ways¹⁵. Large RNAs, as the model RNA of this study, consist of many helices that have to be formed in order to allow tertiary folding. Compaction of these molecules is enabled by counterions *in vitro* but tertiary interactions have to be established to stabilize this compact state¹⁶.

Formation of the tertiary structure is driven mostly by coaxial stacking of helices. Two different helices align their axes by stacking of their bases forming a contiguous helix, a

highly stabilizing tertiary structure¹⁷. Coaxial stacking is mainly influenced by sequence but ionic conditions and junction topology also play a major role in its formation^{13,18}.

Base triple interactions orient three bases in a co-planar way by hydrogen bonding of a central base to the two other involved bases. This central base forms a canonical Watson-Crick base pair with one of the other nucleotides and uses either major or minor groove constituents to interact with the third nucleotide. Base triples are found in a variety of tertiary structure motifs, such as kink turns, and stabilize tertiary interactions, like tetraloop-receptor interactions¹⁹. Another very abundant and thus important tertiary interaction involving a minor groove base triple is found in the A-minor motif. Although all nucleotides are able to be the minor groove constituent in this interaction it is mostly an adenosine²⁰. Its N1-C3-N3 edge is interacting with one or both 2'OH-groups as well as with the bases of the host duplex, which is often a G-C base pair.

The 2'-OH group of the ribose moiety of the nucleotides can be the acceptor as well as the donor of hydrogen bonds and therefore interact with many residues positioning helical elements in the overall structure²¹. These 2'-hydroxyl contacts usually form networks of interactions. For example, the so called ribose zippers, in which two backbone strands are brought closely together by 2'-OH interactions, are the most commonly found arrays of this kind²².

Tetraloop-receptor motifs were found in almost every large RNA crystal structure so far. It always involves a hairpin tetraloop with the most prominent sequence GNRA²³ and a receptor motif with different levels of structural complexity, tandem G-C pairs being a more simple example²⁴. In group II introns another conserved 11 nucleotide long receptor motif was found lying in the heart of DI and interacting with the closing tetraloop of DV²⁵.

Long-range tertiary interactions involving Watson-Crick base pairing include kissing loops and pseudoknots. The latter connects nucleotides of a loop with complementary bases outside of the closing loop²⁶. Kissing loops are formed by Watson-Crick base pairing of two hairpins with complementary sequence in their closing loop²⁷.

The structure of RNA molecules often depends on the availability of metal ion both for establishing as well as maintaining secondary and tertiary structure motifs.

1.1.3 Metal ions and RNA folding

Metal ions play a crucial role in both RNA folding and catalysis. The high negative charge of the RNA backbone poses a tremendous challenge in RNA folding. Diffuse ions that retain their hydration sphere are attracted by this electrostatic field and by charge shielding allow secondary structure formation²⁸. This ion atmosphere is created by monovalent ions, in particular K^+ . Chelated ions, which are found less frequently, establish direct contact with the RNA by coordinating with polar RNA atoms via both inner-sphere or outer-sphere coordination. Monovalent ions like K^+ are generally interpreted as diffuse ions and Mg^{2+} as to interact site specifically (Figure 1.1) but there are many different roles for different mono- and divalent ions in RNA structures²⁹. For example K^+ was found to be also coordinated at tertiary structure level³⁰ as well as divalent ions to be diffusely bound and exchanging rapidly²⁹. Additionally to their role in RNA folding inner-sphere coordinated ions play an important role in ribozyme catalysis by activating the attacking nucleophile as well as stabilizing the leaving group during the transition state^{31–35}.

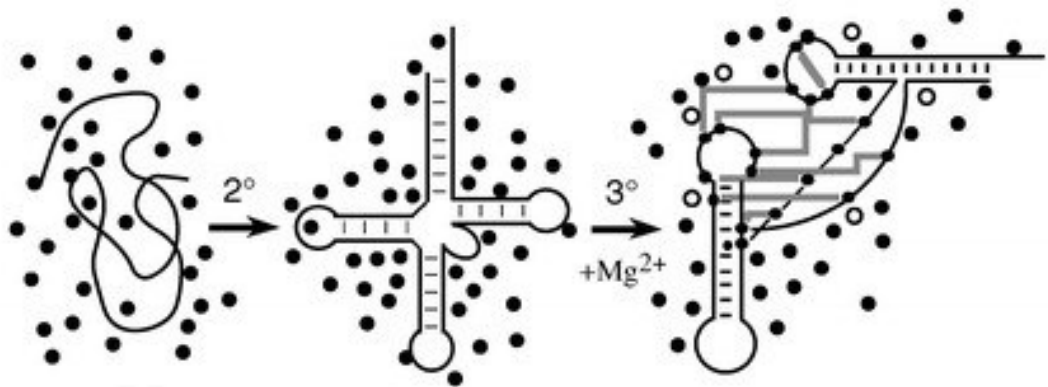


Figure 1.1: Influence of ions on RNA folding. Monovalent ions trigger secondary structure formation, while adding of divalent ions generally induces tertiary folding. This figure has been adapted from²⁸.

1.2 Group II introns

Large noncoding RNAs with highly complex structures are important in all of terrestrial life. Fundamental cellular processes like protein translation by the ribosome or pre-mRNA splicing by the spliceosome rely on their ribonucleic part for catalysis^{36,37}. Group II introns are some of the largest ribozymes in nature and found in bacteria³⁸, archaea³⁹, as well as the organelles of eukaryotic fungi and plants⁴⁰ and even most recently in certain animals⁴¹. Catalysis, structure and folding of this class of self-splicing ribozymes were studied intensely although mostly *in vitro* and under non-physiological conditions.

1.2.1 Biological relevance of group II introns

The diverse family of group II introns not only played an important role in genome organization, but these are also evolutionarily connected to the eukaryotic spliceosome, with which it shares its splicing mechanism³⁵. As this class of ribozymes is found exclusively in organelles of eukaryotes they descend evolutionarily from endosymbionts⁴². Fragmentation of the incorporated group II intron might have led to the spliceosome⁴³, components of the telomerase, LINE elements and most nuclear introns, thereby having a great influence on the human genome⁴⁴. Their ability to perform reverse splicing and therefore to act as mobile elements can accelerate genomic reorganizations⁴⁵ and it was shown that environmental factors can influence group II intron mobility⁴⁶, hinting to an epigenetic influence of group II introns.

Group II introns are divided into three main families – group IIA, IIB and IIC^{38,47}. Group IIC introns are only found in bacteria and are considered as the ancestor of the larger and more complex IIA and IIB families³⁸.

1.2.2 The best studied group II intron: *ai5γ*

Ai5γ, the model RNA molecule of this study, is part of the IIB family. It is found in the *COX1* gene which is embedded in the mitochondrial genome of *Saccharomyces cerevisiae*. Its folding pathway, structure and catalytic activity have been studied

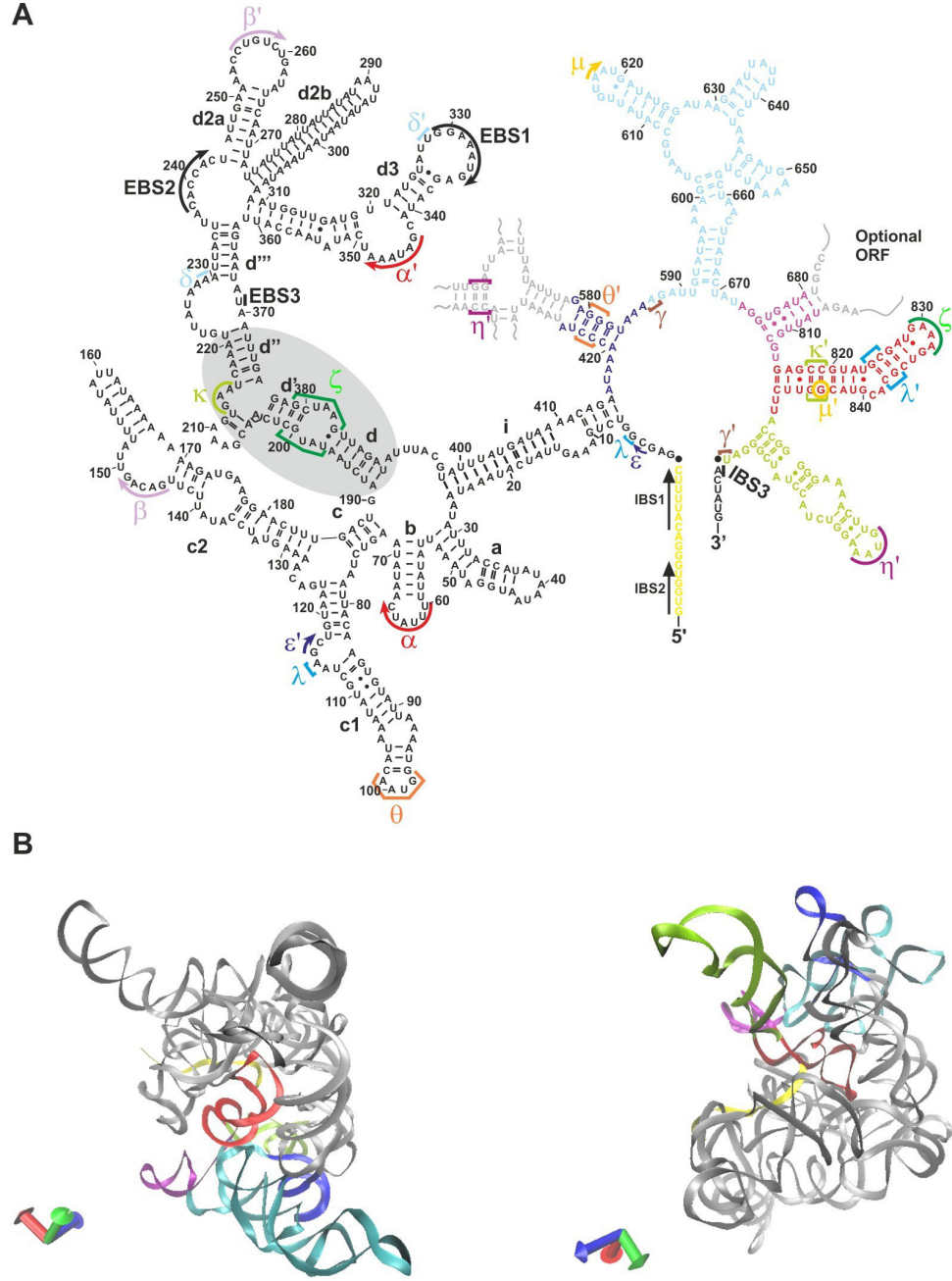


Figure 1.2: Graphical representation of *ai5γ*. (A) Secondary structure map. Tertiary interactions are shown in colored greek letters. The 5' exon sequence is shown in yellow, D1 in dark grey, D2 in blue, D3 in light blue, D4 in pink, D5 in red and D6 in green. Exon binding sites (EBS) and intron binding sites (IBS) are indicated with black arrows. Gray ellipse indicates the folding control element. (B) Tertiary fold of *ai5γ* modelled on the crystal structure of *Oi*. Group IIC intron. Color code follows A. Coordinates from⁶⁹.

intensely *in vitro* rendering it a perfect candidate for studying *in vivo* folding of large RNAs. Its efficient splicing is crucial for aerob yeast growth although cells are still viable under anaerob conditions due to fermentation a fact that adds to the overall practicability of this ribozyme for functional studies *in vivo*.

1.2.3 Group II introns are capable of self-splicing

Group II introns are able to promote forward as well as reverse splicing⁴⁸ by catalyzing phosphodiester cleavage and ligation reactions in both RNA and DNA molecules⁴⁹. In this chemical mechanism a nucleophilic oxygen attacks the phosphate in the phosphodiester bond of the polynucleotide⁵⁰. During this process two metal ions coordinate both the 3'-OH leaving group and the attacking nucleophile in an intermediate state⁵¹. In the prevalent pathway for self splicing in group II introns the nucleophilic oxygen is the 2'-hydroxyl group of a highly conserved bulged adenosine in domain VI. The 3'-OH of the released 5' exon subsequently attacks the phosphate at the 3' splice site, resulting in ligated exons and an excised lariat intron⁵². Interestingly, the very same lariat structure as a result of intron splicing occurs in spliceosomal introns⁵³. Alternatively, the nucleophile can be the oxygen of a water molecule leading to a linear excised intron⁵⁴. This pathway has been observed in most group II introns that have been studied and is present both *in vitro* and *in vivo*^{54–56}. Both the lariat intron as well as the linear intron were shown to be able to perform reverse splicing^{57,58}, targeting not only RNA but also DNA molecules⁵⁹. In order to act as a mobile element a maturase, which is intron encoded, needs to be present⁶⁰. This protein typically consists of a RNA binding domain, a reverse transcriptase as well as an endonuclease domain⁴⁵. This maturase is not present in all group II introns, including the model RNA of this study ai5γ. Nevertheless it was shown that the excised ai5γ intron retains its structure *in vivo*, a prerequisite of reverse splicing, underlining the importance of understanding group II intron folding⁶¹.

1.2.4 Structural organization of group II introns

Early phylogenetic studies showed that despite little primary sequence conservation group II introns show high similarity at the secondary structure level⁴⁷. A secondary structure map was established for group II introns in the late 1980ies showing main features of this large multidomain RNA. Their size ranges between 400 and 1000 nucleotides and their secondary structure is organized in 6 domains protruding from a central wheel (Figure 1.2). In contrast to group I introns in which different domains interact by base pairing or coaxial stacking⁶² group II introns have a network of long-range tertiary interactions stabilizing the native fold and establishing the catalytic core of the intron^{31,63–68}. A major accomplishment was the crystallization of the *Oceanobacillus iheyensis* group IIC intron³¹. Despite its small size the x-ray structure of this ribozyme provided deep insights into the architecture of group II introns in general. More recently the data of this study was used to model the tertiary structure of the ai5γ group II intron⁶⁹.

Domain I

Domain I (D1) is by far the largest of all 6 domains. As it is found at the 5'-end of group II introns it is transcribed first, suggesting that it also folds independently from other domains⁷⁰. It serves as a scaffold for docking of the subsequently transcribed domains, establishing many inter- and intradomain tertiary interactions. Folding of D1 is the rate-limiting step in native structure formation and it is essential for ribozyme catalysis together with D5⁷¹. D1 harbours two exon binding sites EBS1 and EBS2 which are complementary to the intron binding sites IBS1 and IBS2 in the 5'-exon⁷². In addition, EBS3 in D1 binds to the intron binding site IBS3 found in the 3'-exon⁷³. Stabilized by the formation of the β – β' interaction³⁸ all these exon-intron interaction sites find their place within the so called coordination loop, an asymmetrical loop that coordinates all essential parts for lariat formation: the 5' and 3' splice sites as well as the bulged adenosine, whose 2'-hydroxyl group initiates the branching pathway of forward splicing⁷⁴.

There are two folding events in D1 that are fundamental for its structural organization: the local folding of helices at the D1 core and coaxial stacking network at the conserved

five-way junction. The latter includes the formation of a T-loop motif that involves the junction between stems d and i as well as stem a⁷⁵. The T-loop is an often recurring five-nucleotide motif in noncoding RNA molecules that facilitates long-range tertiary interactions⁷⁶. The second essential folding step relates to the formation of the κ - ζ folding control element, which is a rate-limiting event in forming the native ai5 γ structure^{77,78}. The formation of the κ - ζ element might introduce a bend bringing the two lobes of D1 closer together, thereby allowing the establishment of the α - α' and the β - β' interactions^{55,77}. These two kissing loop interactions are important for the compaction of D1 and might also facilitate binding of the 5'-exonic substrate⁷².

The catalytic center of ai5 γ is established by docking of D5 onto D1, both of which are essential and sufficient for the catalytic activity of group II introns⁷¹. This docking is stabilized by the ζ - ζ' and κ - κ' tertiary contacts^{25,67}, whereas λ - λ' and ε - ε' are themselves required for catalysis, as these position D5 close to the 5'-exonic substrate^{64,79}.³¹ The tetraloop-receptor interaction θ - θ' seems to add further stabilization to the native fold of group II introns²³. A motif most recently found in the crystal structure of the *Oceanobacillus iheyensis* group IIC intron, the so-called z-anchor, involves the λ - λ' and ε - ε' tertiary interactions, the internal loop of stem i, D5 and the 5' end of the intron⁷⁵. By forming an intricate network of interactions the z-anchor stabilizes the catalytic core of the intron by arranging most of the interactions around the two-nucleotide bulge in D5.

Harboring many inter- and intradomain interaction sites as well as active site constituents D1 plays a crucial role in both the structural organization of the intron as well as its catalytic activity.

Domain II

It was shown that deletion of domain II (D2) has no effect on *cis*-splicing activity⁷² though the basal stem with its important θ - θ' interaction was not deleted in this study. This basal stem was found to be involved in helical stacking with stem i of D1 and interacts with the internal loop of D3⁷⁵. The resulting structure forms a shell around active site elements and therefore plays an important role in native structure

formation. The basal stem of D2 also orchestrates folding of J2/3, which is one of the most conserved elements of group II introns and part of the active site⁶⁵.

The second tertiary contact η - η' lies in the heart of D2²³. It is assumed to play an important role in establishing a silent conformation of D6⁵⁵.

Domain III

Domain III (D3) was shown to be a catalytic effector enhancing the rate of catalysis dramatically^{80–82}. Its appearance varies among the different group II intron families but the basal stem always progresses into an internal loop that shows high levels of phylogenetic conservation⁸³. In addition, this adenosine rich loop is highly internalized in the D135 ribozyme, indicating that it is part of the active site center in the intron $\alpha 5\gamma$ ⁶³. This is in line with the earlier observation that this internal loop plays a major role in enhancing catalytic activity⁸⁴. The available crystal structure of a group IIC intron shows that the internal loop adopts a near-helical structure with non-canonical base pairs widening the major groove making it a receptor for binding the basal stems of D1 and D2 by coaxial stacking⁵⁵. In group IIB introns D3 is extended forming the tertiary contact μ - μ' between the D3a terminal loop and D5, thereby fine tuning the κ - κ' interaction⁶⁶. Additionally, the single-stranded region between D3a and D3b was proposed to be part of the θ - θ' interaction site⁶⁵. In group IIB introns the conserved internal loop is close to the ε - ε' interaction within D1. Hence in IIB introns this loop might be oriented slightly differently than observed in the group IIC crystal structure facilitating interactions with J2/3, the bulge of D5 and ε - ε' ⁸⁵.

Domain IV

Domain IV (D4) is pointing away from the catalytic core in the group IIC intron of *Oceanobacillus iheyensis*³¹, making it perfect for harboring open reading frames that encode a maturase⁸⁶. This protein is necessary for intron mobility and in some cases for splicing *in vivo*.

Domain V

Domain V (D5) is the most internalized of all domains⁶³ and is positioned at the heart of group II introns³¹. It is anchored in D1 by the very important tertiary contacts ζ - ζ' and κ - κ' ^{125,67}. The overall structure of D5 tremendously resembles that of the U6 snRNA of the spliceosome, suggesting little change in its evolution from bacteria to human⁴³ (Figure 1.3). D5 consists of a highly conserved hairpin loop structure harboring a dinucleotide bulge⁸⁷. Upon docking of D5 onto the D1 scaffold this bulge is twisted, creating a high negative potential due to condensed backbone phosphates³¹. Two Mg^{2+} ions coordinate at this site, suggesting that these two metal ions participate in the catalytic reaction of group II introns^{50,55}. The catalytic center is stabilized by a triple helix formed by the linker region J2/3 and one residue of the di-nucleotide bulge with the most highly conserved nucleotides in D5, the so-called catalytic AGC triad, whose backbone also contributes to the negative potential at the metal binding site. D5 has long been known to have a catalytic face on one side and a binding face on the other⁸⁸. As such, the invariant G of the catalytic triad contributes to catalysis with its major groove atoms, while its minor groove constituents ensure binding to the D1 core⁶⁸. Although D5 is the catalytic center of group II introns the intronic environment plays a crucial role for assembling the active site. In addition to the mentioned J2/3, ζ - ζ' and κ - κ' interactions, λ - λ' is formed with D1⁶⁴ and μ - μ' with D3⁶⁶. D5 alone shows no activity, whereas together with D1 catalysis takes place⁷¹. Efficient catalysis though requires the presence of D3⁸⁹⁻⁹¹ as in the ribozyme D135, in which all crucial elements for the establishment of the catalytic center are present.

Domain VI

Domain VI (D6) harbors the highly conserved bulged adenosine, whose 2'-OH is the nucleophile in the branching pathway of group II intron splicing⁹²⁻⁹⁴. There are two proposed positions of D6 in the overall conformation of group II introns. Site-directed cross linking studies showed that D6 aligns along the D1c stem positioning it close to the coordination loop allowing the bulged adenosine to come close to the 5'-splice site^{74,95}. Additionally, UV-crosslinking studies showed that D6 and D5 lie side by side in the native conformation⁹⁶, so that the 5'-exon would be sandwiched between those two

domains, allowing lariat formation. Nevertheless this is not the only conformation D6 can adopt. In its second configuration it is twisted 90° settling outside of the intron core and it is stabilized by the tetraloop-receptor interaction η - η' with D2²³. In this conformation the branching pathway is no longer an option. This might happen after the first step of splicing, when the presence of the bulged adenosine is no longer needed⁹⁷. However, crosslinking studies showed that D5 and D6 are positioned close to each other before, during and after splicing^{74,95}. Thus, it has been proposed that the η - η' interaction might protect D6 from degradation or allows regulation of different splicing pathways⁵⁵.

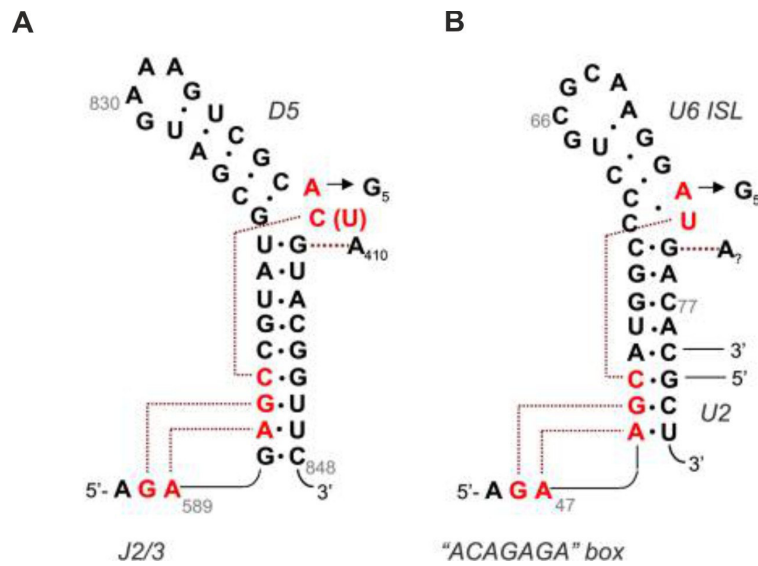


Figure 1.3: Secondary structure and implied tertiary interactions of D5 and U6 RNA. (A) D5 of ai5g with tertiary interactions involving the two nucleotide buldge, J 2/3 and the catalytic triad. (B) Structural representation of U6 helix 1b model for the human spliceosomal active site with assigned tertiary interactions from ai5 γ . Figure was adapted from⁵⁵.

1.2.5 Group II intron folding

Folding of an RNA molecule describes the process of structure formation from its extended, unstructured state to the native, catalytically active conformation. RNA

folding is hierarchical meaning that secondary structure elements form first, an event that highly depends on addition of monovalent ions^{5,28,98}. Divalent ions then trigger tertiary structure formation and a second phase of compaction of the RNA molecule. During folding a variety of folding intermediates may appear defining the energy landscape of the folding process. In some cases these intermediate states are off-pathway and stable, trapping the RNA molecule in a non-native conformation as observed for the *Tetrahymena* group I intron^{99–102} or the RNase P RNA¹⁰³. Resolving those kinetically trapped structure then becomes the rate-limiting step in native structure formation. Additionally to this phenomenon, intermediate states can be on-pathway defining a smooth energy landscape as observed in the RNase P C-domain¹⁰⁴ and the group II ribozyme ai5γ^{70,77,105,106}. In these cases an on-pathway intermediate structure has to be stabilized to allow native structure formation.

Although folding landscapes of ribozymes are quite diverse, for many RNA molecules compaction is an early event in folding and subsequent native structure formation is the actual rate-limiting step. Interestingly the group II intron ribozyme D135 is quite different in that its collapse is slow and subsequent progression to its native conformation happens rapidly (Figure 1.4).

Global and local folding events of ai5γ

Hydroxyl radical footprinting studies on D135 showed that this remarkably large ribozyme adopts a very compact structure, in which almost half the molecule is protected⁶³. The regions of protection appear all at the same Mg²⁺ concentration, indicating cooperative folding¹⁰⁷. The level of needed counterions was very high and no regions of protections appeared before-hand. In analytical ultracentrifugation experiments the same Mg²⁺ concentration that led to hydroxyl radical protection gave rise to compaction of D135¹⁰⁵. The free energy of folding is very similar to that of unfolding and repeated unfolding and folding events do not impair the catalytic activity of the ribozyme. Furthermore, D135 becomes catalytically active at the same time compaction happens⁷⁰. All these findings initially indicated that D135 folds to a specific tertiary conformation in a cooperative 2-state way. However, the rate constant for folding of D135 is quite slow compared to other large RNA molecules which undergo

two-state folding¹⁰⁸. This could be explained by the appearance of kinetically trapped conformations that have to be resolved before native structure formation as seen for the *Tetrahymena* group I intron¹⁰⁰. But neither the rate constants for the appearance of catalytic activity nor that for compaction are increased by the addition of urea, which should resolve misfolded regions^{63,105}. Also no areas of internalization were observed at levels of Mg^{2+} lower than needed for general compaction and no stable misfolded species were observed on native gel analysis, excluding the existence of kinetic traps in the folding pathway of ai5 γ ^{70,105}. Strikingly also *in vivo* there are no indications of misfolding when monitored by DMS chemical probing¹⁰⁹.

In contrast to the rugged energy landscape of large RNA molecules like RNase P and the *Tetrahymena* group I intron folding of the D135 ribozyme follows a smooth energy landscape where an on-pathway intermediate is stabilized by Mg^{2+} *in vitro*.

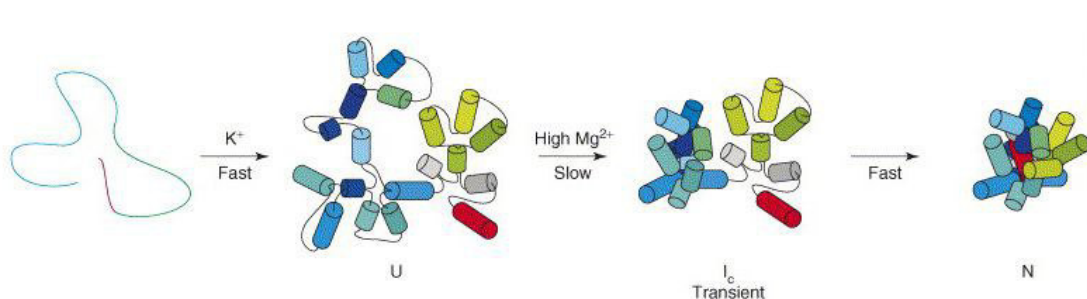


Figure 1.4: *In vitro* folding pathway of the D135 ribozyme. After addition of K^+ secondary structure elements are formed (U). Addition of Mg^{2+} triggers the collapse of D1 (I_c) which leads to fast formation of the native structure (N). Figure was adapted from¹⁰⁸.

Formation of the κ - ζ folding control element is crucial for the collapse of ai5 γ

The rate constant of folding of D1, D13, D15 and D135 formation are very similar indicating that D1 contains the on-pathway intermediate whose formation is rate-limiting for the collapse of the intron⁷⁰. By combining Nucleotide Analog Interference Mapping (NAIM), in which nucleotide analogs with single atom modifications are incorporated randomly into the RNA molecule, with native gel electrophoresis functional groups important for intron folding can be identified^{77,78}. All NAIM effects

were found in D1, highlighting the importance of its structure formation for the collapse of the intron. The strongest NAIM signals were clustered around the d, d' and d'' stems which harbor the two motifs, κ and ζ , involved in tertiary interactions with D5. Interestingly, the addition of K^+ ions does not trigger secondary structure formation of this so-called κ - ζ element in contrast to other helical regions of D1. Folding of this substructure is dependent on Mg^{2+} *in vitro* and it was shown to form a divalent ion binding pocket⁷⁸. Different concentrations of Mg^{2+} influence the ratio of collapsed to unfolded species of D135, but the rate constant of compaction is unaffected, hinting at a conformational capture mechanism responsible for the slow folding of this large ribozyme. Since the κ - ζ element also forms a docking site for D5, the dependence of folding of the ai5 γ group II introns on the formation of a substructure that is crucial for active site formation might be a genius way of nature to ensure the proper function of a large ribozyme at an early stage of folding.

Folding of ai5 γ at near-physiological conditions

Efficient folding of the group II intron ai5 γ requires unusual high concentrations of mono- and divalent ions and elevated temperature, conditions that are not present in the cellular environment. Under near-physiological conditions the compaction of the D135 ribozyme is still observed by native gel electrophoresis but happens on a much larger timescale¹⁰⁶. The resulting collapsed species adopts a near-native conformation in which D1 is formed, but docking of D3 and D5 has yet to happen. Only at high concentration of Mg^{2+} native structure formation is triggered. Subsequent dilution of Mg^{2+} to a low concentration shifts natively folded ribozymes back to the near-native state, meaning that high Mg^{2+} concentration is needed for stabilization of the native state *in vitro*. Since the intracellular concentration of Mg^{2+} is well below that, *in vivo* native state formation and stabilization are likely to depend on protein-cofactors.

Protein cofactors assisting RNA folding *in vivo*

Aside from the mentioned differences in ion concentration and temperature folding conditions *in vivo* are far more complex than those *in vitro*. While folding *in vitro* starts with a fully transcribed RNA molecule the folding process happens co-transcriptionally *in vivo* and the directionality and velocity of transcription influences folding^{110,111}. Also the flanking sequence context differs *in vivo* and was shown to have an influence on the folding rate constants of RNA molecules¹¹² as well as on native structure formation as found in reduced level of misfolded species of the *Tetrahymena* group I intron in presence of the native flanking sequence¹¹³.

But most importantly, protein-cofactors assist in RNA folding in the cell. Since metal ions are crucial for RNA folding, proteins influencing ion homeostasis naturally play a role in RNA folding. For example, Mrs2p a Mg²⁺ transporter was shown to be important for folding of group II introns¹¹⁴.

As mentioned before large RNA molecules often show a rugged folding landscape with kinetic traps lying in the way of native structure formation. RNA chaperones are a class of proteins that resolve such misfolded intermediates although it is still unclear how these proteins actually act on the RNA molecule^{115–117}. RNA chaperons bind RNA in a non-specific way and are ATP independent. Interestingly, weakening of RNA binding by an introduced mutation in the RNA chaperone StpA even enhances its chaperone activity¹¹⁸, making it unlikely that RNA chaperones play a role in native structure stabilization.

In contrast to RNA chaperones, specific RNA-binding cofactors stabilize the native state or an on pathway intermediate^{60,119–125}. For example the page T4 *td* intron core is stabilized in its native state by the Cyt18 splicing factor¹²⁶, contributing to the compactness of this group I intron. Cbp2 assists in native state formation of the bI5 intron by capturing the collapsed folding intermediate¹²⁷.

Another member of this class of specific RNA-binding cofactors is the DEAD-box protein Mss116p, which is crucial for the splicing of all mitochondrial introns in *Saccharomyces cerevisiae*^{128,129}.

1.3 DEAD-box proteins

Almost all processes in RNA metabolism throughout the different cellular life forms and even in many viruses are connected to the activity of RNA helicases^{130,131}. Defects in their function have been linked to different diseases as cancer, developmental defects or neurodegenerative diseases^{132,133}. Phylogenetic studies and comparative functional and structural analysis resulted in the classification of 6 superfamilies (SF) of helicases¹³⁴. All RNA helicases bind and hydrolyze ATP. Most but not all of them do also perform ATP dependent unwinding on appropriate substrates *in vitro*¹³⁵, but these findings do not necessarily imply that this unwinding activity is part of the mechanistics of a helicase in its intracellular function.

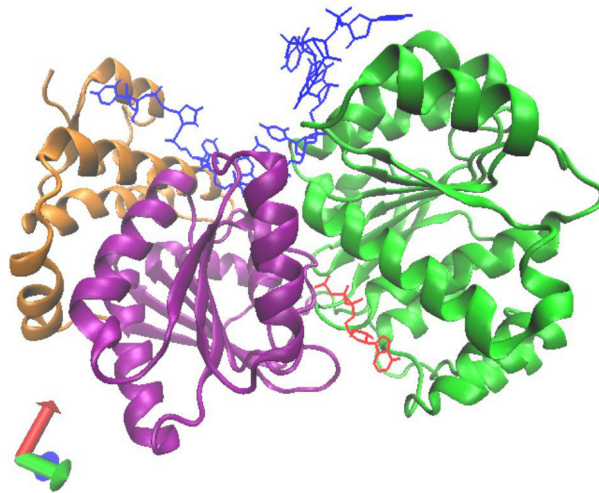


Figure 1.5: Mss116 with bound RNA oligonucleotide and ATP analog. Domain 1 (green) and Domain 2 (purple) build the helicase core establishing the ATP-binding cleft with bound ATP shown in red. The poly U RNA (blue) is bent by the two crimping recA-folds with a second bent introduced by interactions with the C-terminal extension in Mss116. Rendered using the pdb file 3I5X¹⁴⁶.

1.3.1 Structural organization of DEAD-box proteins

In eukaryotic cells RNA helicases are exclusively members of the SF1 or SF2 which do not form oligomeric ring systems such as SF3 to SF6 that are found in bacteria and viruses. They share a structurally conserved helicase core formed by two very similar domains each of which can be superimposed on the bacterial recombination protein RecA¹³⁶. Within these two RecA-like domains at least 12 conserved sequence motifs can be found in both SF1 and SF2 members. Sequence conservation though is limited to members of one superfamily¹³¹.

The largest family of RNA helicases is part of the superfamily 2 with 37 members found alone in humans up until now¹³¹. This so-called DEAD-box RNA helicase family is named after a highly conserved Asp-Glu-Ala-Asp motif found in motif II. Members of this family are important in as different aspects of RNA metabolism as RNA decay, mRNA transport, ribosome biogenesis, RNA splicing and even RNP disassembly^{137,138}.

The two recA folds that build the catalytic core remain flexible concerning their relative positions in the absence of bound ligands^{139–141}. Upon binding ATP and its target RNA the domains come into close proximity, forming two clefts on opposite sites of the helical core domains^{142–146}(Figure 1.5). On one side ATP is bound by the polar side chains of numerous conserved amino acids found in the motifs Q, I, II, III, Va and VI which form a complex network of interactions with the phosphate groups (I, II, III, VI) as well as the ribose (Va) and adenine (Q, VI)^{138,143–148}. On the other face of the helicase core domains side chain and backbone constituents of motifs Ia, Ib, Ic, IV, IVa and V bind the target RNA. Strikingly only the backbone of the RNA molecule contacts the DEAD-box helicase constituents, showing a binding interaction with low sequence specificity. The bound RNA itself is severely bent in contrast to RNA molecules bound by most other SF1 and SF2 helicases^{143,144,146,147}. Communication between ATP-binding and RNA-binding sites remains enigmatic since up to now the majority of information about the mechanistics of ATP and RNA binding is derived from crystal structures of various DEAD-box helicases trapped in a state, in which single stranded RNA is bound together with a non-hydrolyzable ATP analog.

DEAD-box proteins generally have no other domains inserted between the two lobes of the helicase core but all of them have C- and N-terminal domains of varying lengths¹³¹.

These domains play a role in protein-RNA or protein-protein interactions. Most studies with DEAD-box proteins were carried out on truncated variants of the respective protein, thus providing little information on these auxiliary domains. Nevertheless some terminal domains have been studied separately showing for example an RNA recognition motif (RRM) in the C-terminal domain of HerA or YxiN^{149,150}. Some DEAD-box proteins as Ded1p, Cyt19, SrmB or Mss116p show an abundance of basic amino acids in their C-terminal domain that is thought to interact non-specifically with structured RNA^{151,152}. In the case of Mss116p a truncated variant was used harboring part of the C-terminal domain¹⁴⁶. This so-called C-terminal extension was shown to extend the RNA binding site, introducing a second bend in the single stranded RNA.

1.3.2 Function of DEAD-box helicases

DEAD-box proteins are structurally highly conserved and display a variety of activities from RNA duplex unwinding and protein displacement to RNA strand annealing^{130,153,154}. In contrast to helicases of the other superfamilies DEAD-box proteins are non-processive. In the cellular environment these are usually building complexes with dozens to even hundreds of other proteins. Interestingly, many DEAD-box proteins were shown to be associated with the spliceosome^{155,156}. Most studies concerning biochemical features of DEAD-box proteins were done *in vitro* and, if at all, than only with a small fraction of their intracellular complex cofactors¹³⁸.

ATPase activity

All members of the DEAD-box protein family that have been biochemically characterized *in vitro* show ATP binding and hydrolysis activity¹⁵⁷⁻¹⁶¹. The two RecA-like domains are connected with a flexible linker and can move relative to each other if ATP and RNA are not bound¹⁵⁴. This is not case with processive helicases in which additional domains keep the helicase core domains in place¹⁵⁴. The comparatively long distance between DI and DII in the unbound state might reflect different binding properties of the two domains for ATP and RNA molecules¹⁶². Different crystal structures with bound ATP analogs and RNA substrate show highly similar conformations in which both domains are kept close together binding ATP in a cleft on

one side and the RNA molecule on the other side of the wedge^{143–148}. More precisely, ATP shows a similar K_d to DI alone and in complex with DII, whereas DII alone does not display significant affinity to the triphosphate¹⁶². These findings suggest that DI establishes the ATP-binding site in DEAD-box proteins^{162–164}. It was shown that ATP is bound with and without the presence of RNA, but hydrolysis is barely detectable without the RNA substrate^{165–168}. Crystallization in the presence of different ATP analogs, an ATP ground state analog that resembles the bound unhydrolyzed triphosphate and an ATP transition state analog that mimics the highest energy state of the not yet hydrolyzed triphosphate, revealed only minor changes in the ATP binding and none in the RNA binding site^{146,147}. That the RNA substrate binds tightly to both DEAD-box helicase/ATP-analog complexes correlates nicely with these findings¹⁶⁹. Considering the very low affinity to RNA with bound ADP both structures most likely represent the protein after RNA strand separation^{146,170}. Several DEAD-box proteins were shown to be still able to perform duplex unwinding with a bound non-hydrolyzable ATP ground state analog, showing that binding but not hydrolysis of ATP is necessary for unwinding activity¹⁶⁹. Nevertheless, hydrolysis of ATP is a common feature of DEAD-box proteins, but its proposed function lies in the highly reduced affinity to its RNA substrate after hydrolysis, allowing repeated cycles of RNA binding and/or unwinding^{165,169,171,172}.

RNA binding

A double-stranded RNA molecule shows the same K_d for DII and the DI/DII complex¹⁶². However, DI has no significant affinity towards the RNA substrate, indicating that DII establishes the RNA-binding site of DEAD-box proteins. The crystal structure of Mss116p-DII with a bound dsRNA substrate shows that DII forms a positively charged pocket establishing hydrogen bonds mainly with phosphate groups of the RNA. Most interestingly, these contacts only involve one strand of the bound dsRNA (Figure 1.6). RNA binding is supported by interactions of the C-terminal extension with 2'-hydroxyl groups of the RNA backbone of the other strand¹⁶². Together with the ATP-binding properties of DI this might reflect domain specificity for substrate capture in DEAD-box proteins.

Single stranded RNA is bound along a crevice that involves residues of both helicase core domains, establishing interactions with the RNA backbone exclusively, including several hydrogen bonds with the 2'-hydroxyl groups of the ribose part of the nucleotides^{142–146,173}. Interestingly, ssRNA also establishes contacts with DI being bent 3' of the central nucleotides. It is thought that this mechanical stress results in duplex destabilization. In Mss116p the truncated C-terminal, the so-called C-terminal extension, introduces another bent at the 5' end of the RNA, further supporting this hypothesis¹⁴⁶. Ded1p and Mss116p are both able to load on an internal region of a RNA duplex and perform strand separation¹⁷⁴ which would also be consistent with a model of strand separation by inducing mechanical stress via conformational changes in the DEAD-box protein that is bound to both its substrates.

RNA specificity is established on the one hand by involvement of 2'-OH groups in RNA binding of mainly the C-terminal extension. On the other hand the shape of the RNA-binding pocket in DII does not allow binding of B-form helices^{138,162}.

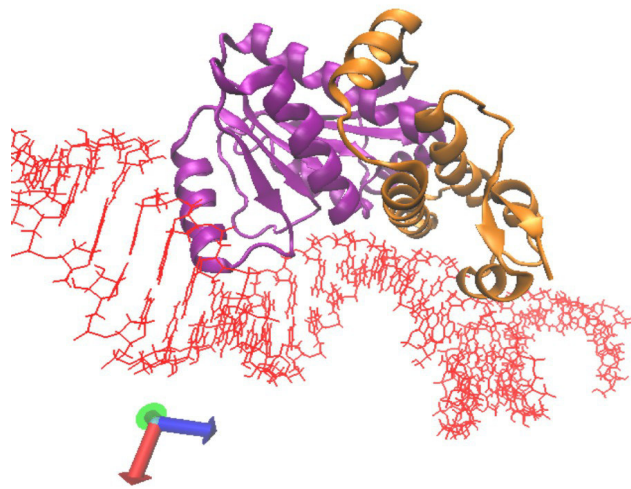


Figure 1.6: DII of Mss116p and C-terminal extension bound to dsRNA. DII is shown in purple, CTE in orange and 14bp dsRNA in red. Rendered using the pdb file 4DB2¹⁶².

Unwinding activity

Although not all RNA helicases actually show unwinding activity¹⁷⁵, it is their most prominent feature. The core domains of processive RNA helicases are held tightly together by other domains, allowing ATPase activity even without bound RNA¹⁵⁴. Translocation and unwinding in these proteins depends on ATP binding and hydrolysis, whereby the phosphate release provides the energy for duplex unwinding^{130,154,176}. In contrast to those translocating SF1 or SF2 RNA helicases DEAD-box proteins that unwind RNA duplexes *in vitro* do not show processivity^{174,177}. Consistently, unwinding by DEAD-box proteins is restricted to helices less than 25 nucleotides long^{159,178}. Additionally, the efficiency of unwinding decreases with increasing stability of the helix^{179,180}. These findings that are in contrast to all RNA helicases other than DEAD-box proteins led to defining the unwinding activity as local strand separation¹⁸¹. ATP hydrolysis and RNA strand separation are highly coupled in DEAD-box proteins, but it is still not fully understood how the ATPase cycle and destabilization of an RNA duplex are connected. Like for other DEAD-box proteins, DpbA is tightly bound to its RNA substrate when ATP is bound¹⁶⁵. Strand separation occurs during ATP hydrolysis when the ADP.Pi state evolves. Subsequent release of the now single stranded RNA is induced by Pi release. On the other hand, Ded1p catalyzes strand separation *in vitro* with the bound non-hydrolyzable ATP analog ADP-BeFx with an efficiency modestly lower to that in the presence of ATP¹⁶⁹, showing that unwinding activity *per se* is not dependent on hydrolysis of ATP. It was shown that the number of ATP molecules hydrolyzed by the DEAD-box protein *Cyt19* depends on Mg²⁺ concentration and duplex length¹⁸². At 2mM Mg²⁺ a 6 base pair long RNA duplex needed approximately half an ATP molecule for strand separation. For longer duplexes and/or higher [Mg²⁺] the events of ATP hydrolysis exceeded those of strand separation which would be consistent with ATPase activity enabling the protein to undergo conformational changes for repeated ATP-dependent dsRNA binding and strand separation activity. The correlation of Mg²⁺ concentration and RNA duplex length with ATP molecules consumed was also tested for Mss116p and Ded1p, showing the same interdependence¹⁸².

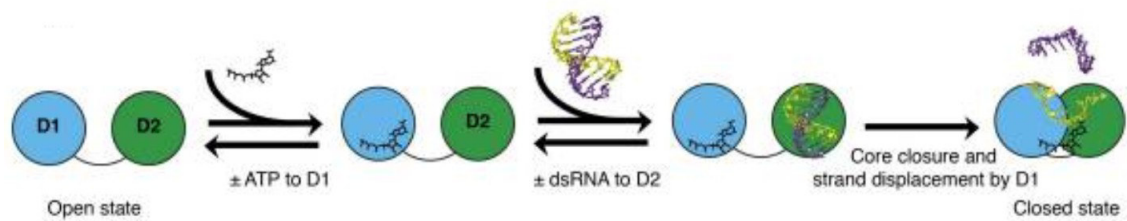


Figure 1.7: Unwinding mechanism of DEAD-box proteins. Independent binding of ATP and dsRNA by D1 and D2, respectively, in the open state. Upon substrate binding the helicase core shifts to the closed state, thereby separating the RNA strand by mechanical force. Adapted from¹⁶².

The recently published crystal structure of Mss116p-DII bound to dsRNA gave insight into the unwinding mechanistics of DEAD-box proteins¹⁶²(Figure 1.7). In the proposed model DI and DII bind ATP and dsRNA independently. Binding of the substrates then triggers the closed state of the helicase core and contacts between the RNA substrate and DI are established^{143–146}. This conformation is sterically incompatible with binding a dsRNA. Furthermore, the interactions of the RNA substrate with DI and the C-terminal extension introduce two bends upon helicase core closure which contribute to the driving force of RNA unwinding.

Strand annealing activity

Interestingly, several DEAD-box proteins also show annealing besides unwinding activity *in vitro*^{183,184}. Notably, the two DEAD-box proteins Ded1p and Mss116p are among the most potent strand annealers^{183,185}. ATP is not required for their annealing activity. It was proposed that binding of DII to one strand and of the C-terminal extension to the second might allow strand annealing, as in the absence of ATP the ATP-dependent interactions of DI do not introduce the bend in ssRNA¹⁶². It is striking that those two proteins also share a distinct basic tail, distinguishing them from other members of the DEAD-box protein family¹⁵².

1.3.3 The DEAD-box protein Mss116p

Mss116p was first identified in a screen for nuclear genes that are involved in splicing of yeast mitochondrial introns which are integrated in genes encoding for respiratory chain proteins¹²⁹. Subsequently this protein was related to the RNA helicase family due to sequence homology¹⁸⁶. Mss116p is involved in splicing of all mitochondrial group I and group II introns as well as in mitochondrial translation and RNA processing¹²⁸. While group I introns still retained self-splicing activity without Mss116p *in vivo* of 40-80%, hinting to involvement of other cofactors, the splicing level of all group II introns was decreased to less than 10%. *in vivo* we showed that splicing of the ai5 γ group II intron strain is not distinguishable from background noise in a $\Delta mss116$ whereas approximately 70-80% of the population is spliced in a wt strain (Figure 1.8)⁶¹. Mss116p was shown to induce splicing of ai5 γ *in vitro* under near-physiological conditions namely at 30°C and low Mg²⁺ concentration^{183,187}. The catalytic activity of ai5 γ does not depend on the assistance of any cofactors *per se* but the underlying formation of its native structure is undoubtedly dependent on one or more activities of Mss116p.

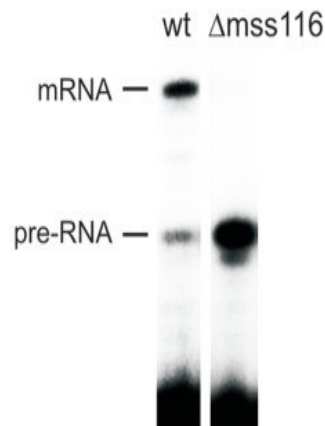


Figure 1.8: Splicing efficiency of the ai5 γ intron *in vivo* in the absence and presence of Mss116p. The upper band is referring to spliced and the lower band to unspliced population. Adapted from¹⁰⁹.

Biochemical properties of Mss116p

Mss116p binds group I and II introns non-specifically and with a similar K_d ranging from 1 to 11 nM¹⁸³. Strikingly the binding affinities are only moderately ATP- or sequence-dependent, which might reflect the importance of the basic C-terminal domain in RNA binding. Mss116p displays RNA-dependent ATPase activity *in vitro*, although very different values for k_{cat} (min⁻¹) have been reported. In these studies different protein purifications and experimental setups were used, resulting in values ranging from around 29.78±3.62 min⁻¹¹⁸⁸ and 56.7±13 min⁻¹¹⁸⁷ to 112±6 min⁻¹¹⁸³ and up to 168±10 min⁻¹¹⁸⁹. This ATPase activity is indispensable for its function as splicing cofactor^{183,187,188}. Mss116p is able to unwind short duplex RNA with blunt ends as well as with 3' or 5' overhang, therefore showing no directionality^{183,187-189}. Interestingly, aside from unwinding activity this DEAD-box protein also displays strand annealing with and without ATP¹⁸³. In the presence of ATP Mss116p and Ded1p seem to promote a steady state between these two activities¹⁸⁵.

To dissect the importance of the different properties of Mss116p in splicing and thus folding of ai5γ and other group I and II introns, *in vitro* and *in vivo* studies with mutations in different motifs disrupting one or another activity were carried out^{152,183,187-190}.

K158

The K158R mutant in DI was shown to neither bind nor to hydrolyze ATP, but retains its ability to bind RNA¹⁸⁷ (Figure 1.9A). As expected this ATPase-deficient variant of Mss116p is unable to promote splicing of the ai5γ group II intron *in vitro*^{187,190}. Furthermore If lysine 158 is changed to an alanine instead of an arginine only ATP hydrolysis but not ATP binding is lost^{187,188}. It has to be mentioned that ATPase activity was still slightly retained¹⁸⁷, however ATP binding was not tested¹⁸⁸. This K158A mutation was not able to rescue the phenotype of a *mss116*-knock out strain, which cannot grow on non-fermentable carbon sources¹⁸⁸. Accordingly, splicing of the ai5γ intron was very low *in vitro*^{187,188,190}.

SAT

Motif III was proposed to play an important role in coupling the ATPase activity of Mss116p and conformational changes that are necessary for dsRNA strand separation¹⁹¹. The conserved threonine 307 found in this motif interacts with a conserved histidine in motif VI^{142–147} and thus might help to stabilize the closed conformation of the two RecA-domains after RNA and ATP binding. Additionally to this interdomain interaction T307 and the conserved S305 form intradomain contacts with an aspartic acid residue in motif II, which subsequently is thought to be important for positioning the catalytic water near the ATP molecule^{142–147}. Initially, the S305A/T307A mutant was reported to have wt-like ATP and RNA binding activities, while being unable to unwind a 12 bp long RNA duplex¹⁸⁷. In this study the ATPase activity was reduced to 20% compared to wild type Mss116p. In another study it showed more than 30% of wild type ATPase activity¹⁸⁹. In the latter the ATP-dependent RNA-unwinding activity was still observed for a 6 bp long RNA duplex and abolished when using a length of 13 bp. Also splicing, efficiency was reported differently in both studies. Showing a 2-fold reduction in one being explained by the reduced ATPase efficiency¹⁸⁷ this was blamed on the long incubation time by the other and contrasted by an 8- to 28-fold reduction in splicing¹⁸⁹. Yeast cells harboring the very same double mutation were not able to grow on a non-fermentable carbon source¹⁹², showing the importance of this motif for the establishment of the respiratory chain *in vivo*. A third group was dissecting this mutation introducing both separately and involving *in vivo* experiments¹⁸⁸. They showed that contrary to the T307A the S305A mutation can still rescue the growth phenotype on glycerol plates. Yeast cells with the S305A Mss116p variant showed only slightly reduced splicing of ai5 γ , whereas in cells harboring the T307A mutation no splicing was detectable *in vivo*. ATP binding was similar to wild type Mss116p in both mutations. ATPase activity though was reduced to 65% in S305A and to 12% in T307A. The S305A mutant was able to unwind RNA with similar efficiency as wt Mss116p. The amplitude (4- to 10-fold) and rate of unwinding (2-fold) was in contrast reduced in the T307A mutant.

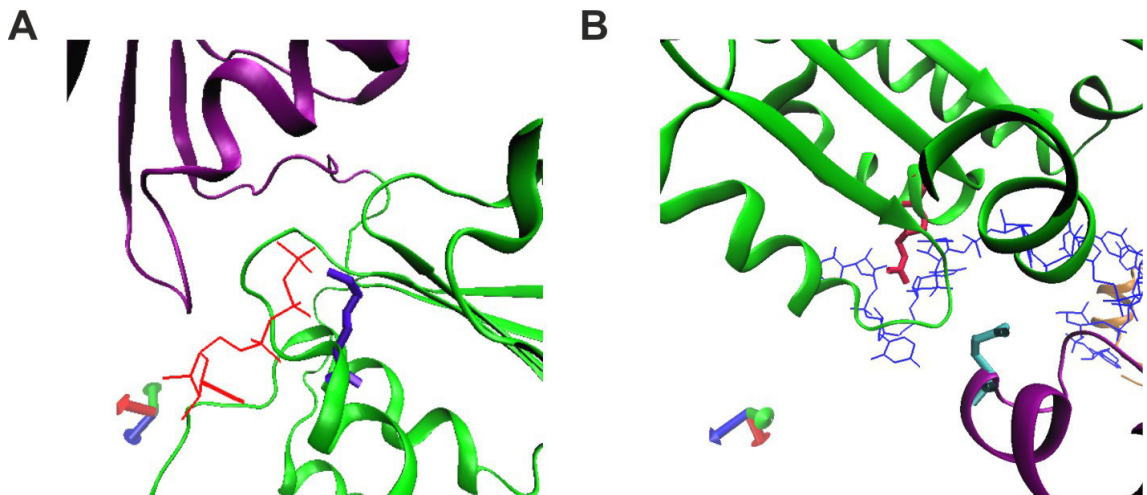


Figure 1.9: Important amino acids of Mss116p. (A) K158 shown in blue contacts the ATP analog in red. (B) R245, shown in red, is involved in bending the RNA interacting with it directly whereas Q412, shown in cyan, is not contacting the RNA but bridges domain 1 and domain 2. The CTE shown in yellow introduces a bend into the polyU RNA molecule shown in blue. Domain 1 is shown in green, domain 2 in purple. Rendered using pdb file 3I5X¹⁴⁶.

Q412

The residue Q412 in the QxxR motif forms an interdomain interaction with a conserved arginine in motif Ia, which is proposed to be necessary for establishing a continuous RNA binding surface and stabilizing the bent conformation of the substrate¹⁸⁸(Figure 34B). The reduction of unwinding activity of a Q412A mutated Mss116p is very similar to the one caused by the T307A mutation, whereas its ATPase activity and stimulation of ai5γ splicing resembles that of the wt protein *in vitro*. Also in contrast to the T307A mutant a yeast strain harboring the Mss116p mutation Q412A is still able to grow on a non-fermentable carbon source.

R245

The residue R245 in motif Ib was predicted to be important for establishing the bend in the RNA substrate by establishing contacts with two phosphates and a 2'OH^{143,146,147} (Figure 1.9B). Introducing the R245E mutation abolished respiratory growth as well as ai5γ splicing¹⁸⁸. Also, ATP binding was weakened and ATPase activity 3.6-fold reduced

in this mutant. Unwinding activity resembled that of the wt protein in the absence of ATP¹⁸⁸.

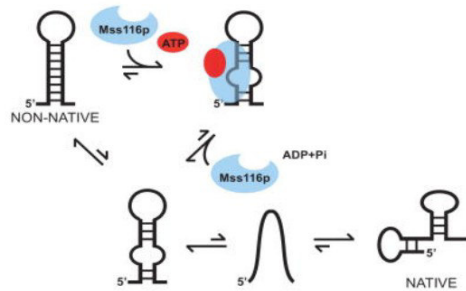
CTD

Splicing of ai5 γ can be efficiently rescued in a Δ mss116 strain by two other DEAD-box proteins namely the *N. crassa* mitochondrial helicase Cyt19 and the *S. cerevisiae* non-mitochondrial DEAD-box protein Ded1p^{183,187}. Structural comparison of Mss116p, Cyt19 and Ded1p show that only the first two proteins share the α -helical C-terminal extension, whereas all three DEAD-box proteins have a characteristic basic tail at the very C-terminus¹⁵². This tail was proposed to interact with structured RNA tethering the DEAD-box protein to its target molecule¹⁹³. Deletion of amino acids 569-664, which include the basic tail of Mss116p, do not interfere with respiratory growth of *S. cerevisiae*¹⁵². Splicing efficiency of ai5 γ was reduced about 2 fold and the ATPase activity resembled that of wt protein *in vitro*¹⁵². In contrast, both unwinding and annealing activities were significantly reduced in the Δ 569-664 strain. The K_d of the truncated Mss116p to ai5 γ was reduced about 10 fold compared to the wild type, whereas it remained comparable for a 38nt long single or double stranded RNA molecule which indicates that the C-terminal domain is responsible for binding of the specific RNA substrate. Hence available studies on the CTD suggest that reduced splicing, RNA-binding and RNA unwinding efficiencies do not result in reduced aerobic growth of a yeast strain with a truncated version of Mss116p.

Mss116p as a cofactor in folding of a group II intron

How Mss116p assists ai5 γ in folding is subject of an ongoing debate about two different paradigms: Mss116p has been proposed to act as a RNA chaperone resolving misfolded elements and in turn to allow proper intron folding^{188–190}. Alternatively, this DEAD-box protein was found to stabilize an on-pathway intermediate that is essential for native structure formation^{187,188,194} (Figure 1.10).

A



B

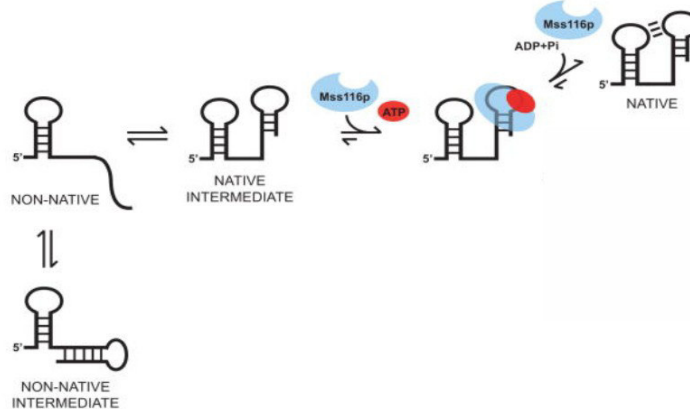


Figure 1.10: The two paradigms of Mss116p facilitated ai5 γ folding. (A) Mss116p destabilizes a nonnative structure upon binding ATP. Subsequent ATP hydrolysis releases the protein and allows refolding to the native structure of the RNA molecule. (B) Mss116p captures a short lived intermediate structure of its RNA substrate upon ATP binding and triggers tertiary folding by stabilization. ATP hydrolysis leads to substrate release enabling subsequent folding events. Figure was adapted from¹⁸⁸.

Although it is generally assumed that Mss116p stimulates folding of ai5 γ and thus splicing, in most studies it is not the structure formation but the splicing of the intron which is monitored. To reveal the mechanism of DEAD-box proteins involved in native structure formation of large RNA molecules, it is very important to look at the structural changes that are triggered by the protein cofactor. *In vitro* Mss116p stimulates the collapse of D1, which was shown to be the rate limiting step in ai5 γ folding and can also be triggered by a [Mg²⁺] of 100mM^{70,195}. More precisely, a complex substructure at the heart of D1, the so-called κ - ζ element has to fold properly to trigger subsequent folding steps⁷⁸. Compaction of a D1356 construct with assistance of Mss116p is very similar to D1 compaction in rate and amplitude, indicating that mainly D1 folding depends on the cofactor. The dependence on either high [Mg²⁺] or a cofactor suggests that electrostatic properties play a crucial role in the folding pathway^{196,197}. Interestingly, compaction was efficient even without the addition of ATP¹⁹⁵, which in turn was found to be essential for duplex unwinding and protein recycling¹⁶⁹. These findings suggest that Mss116p assisted compaction of D1 is indeed independent from ATP and that it is protein turnover that depends on ATP. Furthermore, the efficiency of splicing and compaction decrease at low [Mg²⁺] despite the presence of Mss116p^{183,195}. However, this is supposed to be exactly the opposite for kinetically trapped RNAs^{198,199}. It was shown that the DEAD-box protein Cyt19, which resolves misfolded states of heterologous introns, can also efficiently unfold the native state of an RNA substrate at low [Mg²⁺]²⁰⁰. Interestingly, Mss116p does not unfold the compact state of D1 RNA at low [Mg²⁺]^{183,195}. Hence, in compaction of D1 Mss116p acts independent from ATP in stabilizing a small but crucial substructure at the heart of this domain. However, the compact state of this ribozyme is not its native structure. Native structure formation is in contrast dependent on ATP.

In a single molecule study it was shown that Mss116p lowers the energy barriers between different folding states of the D135 ribozyme and that only native state formation was ATP dependent²⁰¹. The compaction of D135 was shown to depend on the formation of the κ - ζ substructure in D1 which also serves as docking site for the active site center during later events of the folding pathway^{77,78}. Importantly, nucleotide analog interference mapping showed that Mss116p stabilizes this κ - ζ element, thereby

To determine the effect of Mss116p on the structure of ai5 γ in yeast cells we established DMS chemical probing *in vivo*^{109,203}. In the wt yeast strain ai5 γ adopts a structure that highly resembles the *in vitro* fold of the D123456 ribozyme. In contrast, ai5 γ could not adopt its tertiary fold without its cofactor although most but not all secondary structure elements were formed *in vivo*¹⁰⁹ (Figure 1.11). Notably, the strongest differences were seen in the κ - ζ element and in regions that are involved in assembling the active site center like, the κ - κ' and ζ - ζ' tertiary interaction or the internal loop in D3. Also, the stem which connects D1 to the central wheel in group II introns, is not formed in the *mss116* knock out strain. As ai5 γ is largely unfolded in the absence of the cofactor these findings suggest that Mss116p assists ai5 γ folding at an early step being crucial for formation of the D1 scaffold *in vivo*.

Since the unwinding activity is a central function of DEAD-box proteins actually defining them as RNA helicases it is not surprising that strand separation was proposed to be of utmost importance for Mss116p assisted ai5 γ folding^{183,189,190,204}. Ded1p and Cyt19 as mentioned above are able to stimulate splicing of ai5 γ *in vitro* and *in vivo*^{70,183,187,204}, both of which were shown to destabilize misfolded structures of the *Tetrahymena* group I intron^{200,204}. Interestingly, the most striking common feature of these DEAD-box proteins is not found in their helicase activity, but in a shared specific C-terminal domain^{152,193}. This domain was shown to play an important role in ai5 γ binding¹⁵². In contrast to the results stated above¹⁹⁵ it was shown that proteolysis of Mss116p in the absence of ATP does not lead to D135 native structure formation though it did so at a higher protein to RNA ratio¹⁹⁰. This indicates that ATPase activity is not solely needed for protein turnover and might hint to ATP dependent unwinding activity being important for later stages of ai5 γ folding. Additionally, the Mss116 S305A/T307A mutant shows a reduced ability to promote splicing of ai5 γ ^{187,189}, but this effect can be explained by the lower RNA binding affinity and ATPase activity of this mutant and does not necessarily relate to its unwinding deficiency^{187,194}. By adding Mss116p to the pre-folded D135 with and without ATP it was shown that this ribozyme loses its splicing activity over time¹⁹⁰. This result was interpreted to reflect the ability of Mss116p to disrupt the native structure of D135. However, Mss116p is known to bind

RNA efficiently in the absence and presence of ATP, thus the observed effect might be due to the protein competing with the ribozyme for substrate binding^{183,187,205}.

Interestingly, exonic sequences change the effect of Mss116p on ai5 γ folding *in vitro*. Mss116p promotes splicing of ai5 γ with long and short exons with a stronger stimulating effect with long exons^{194,190}. Additionally, the unwinding deficient Mss116p S305A/T307A mutant has a greatly reduced ability to induce splicing of ai5 γ with long exons *in vitro*. Maybe the unwinding activity of Mss116p is needed to resolve exonic substructures that are detrimental for ai5 γ native structure formation as observed for the *Tetrahymena* group I intron²⁰⁶. It has to be mentioned though that these might not be true *in vivo*, as folding takes place co-transcriptionally and exons are coated with proteins that would prevent them from interfering with intron folding¹⁹⁴.

Although the ai5 γ group II intron has been intensely studied since many years *in vitro*, it is still unclear how native structure formation is mediated *in vivo*. The DEAD-box protein Mss116p is a multifunctional enzyme, which is not only involved in ai5 γ folding but also plays an important role in mitochondrial translation and RNA processing¹²⁸. It is able to bind and hydrolyze ATP, to bind RNA with and without ATP and to unwind RNA in an ATP-dependent non-processive manner. Introducing mutations that interfere with ATP binding, ATP hydrolysis, RNA binding and RNA unwinding, it was the aim of this study to elucidate the roles of those activities of Mss116p in ai5 γ folding by monitoring structural changes in the intron population by *in vivo* DMS structural probing.

2 Material and Methods

2.1 Mutation of MSS116

Overview of point mutations introduced in MSS116

Table 1: Mutations introduced in MSS116.

domain	mutation	function
Motif I	K158R	ATP binding and hydrolysis
Motif I	K158A	ATP hydrolysis
Motif Ib	R245E	ATP binding and hydrolysis, RNA unwinding
QxxR	Q412A	RNA unwinding
ΔCTD	stop codons at AA569/570	RNA binding

MSS116 sequence

5' AGAGTTGCACGTTAGGCTGATAAATGTTGACCTCTATATTGATAAAAGGTCGCACACCTGTTCTTGCAAGCAGAAA
TTTATTAGCAGCTCTCTCGAATTGCAACCATATTACTTGGGCTGTTTCAAGAAGATTATACAATGATGGAAACAGAGA
TCAAAGAAATTTTGGTAGGAATCAGAGAAACAACAACAGTAACAGGTATAGAACTCCCGTTTCAATTCTCGCCCAAG
GACAAGGTCTAGAGAAGACGATGATGAAGTGCACCTTTGATAAGACCACCTTTTCCAAACTAATTCACGTCCCAAAAGA
AGATAATAGTAAAGAAGTAACGCTGGACTCTTTACTGGAGGAAGGAGTCCTTGATAAAGAAATACATAAAGCCATTAC
CAGGATGGAATTTCCCGTTTAACTCCTGTACAGCAAAAGACTATCAAACCAATATTATCCAGTGAAGACCATGATGT
TATCGCAAGAGCTAAAACCGGTACAGGTGACATTTGCCCTTCTTGATTCCAATTTTCCAGCATTTGATAAATACGAA
GTTTCGATTCTCAATATATGGTAAAAGCTGTTATTGTCTGCACCAACAAGAGATTGTCCTTGCAAATCGAAGCAGAGGT
GAAAAAATCCACGATATGAATTATGGCTTGAAAAAGTACGCTTGTGTCTCTTTAGTTGGTGGTACTGACTTCAGGGC
AGCTATGAATAAAATGAATAAGTTGAGA CCAAACATTGTAATTGCCACTCCAGGGA GATTGATTGATGTTTTGGAAAA
ATACTCCAACAAATTTTTAGGTTCTG TAGATTACAAAGTGCTAGATGAAGCTGACAGATTGCTAGAAATTGGATTTCAG
GGATGATTTAGAACTATTTCTGGTATTTTAAATGAAAAGAACTCCAAATCAGCCGATAATATCAAGACTTTGTTGTT
CTCTGCTACTCTGGACGACAAAGTACAAAAGTTGGCAAACAATATCATGAATAAAAAAGAATGTCTCTTTTTAGACAC
AGTAGATAAGAATGAACCTGAAGCACATGAAAGAATTGATCAGTCTGTCTCATTTTCAGAAAAGTTTGCCAAATAGTAT
ATTTGCTGCTGTGGAGCATATAAAGAAACAAATAAAGGAAAGAGATTCTAATTACAAAGCCATTATTTTTGCACCAAC
TGTTAAATTCACATCCTTTTTGTGTCAGTATTTTAAAGAATGAATTTAAGAAAGATTTA CCAATTTTGGAATTTTCATGG
TAAATTTACT CAAAATAAAAGAACAAAGTTTAGTTAAAGATT CAAAAAGATGAATCCGGTATTTTGGTTTGTACGGA
TGTGGGAGCCCGTGGTATGGATTTCCCAAACGTCCACGAAGTTTTGCAAATTGGTGTTCATCTGAGTTGGCTAATTA
TATTCATAGAATTGGTAGAACAGCCAGAAGTGGGAAAGAAGGATCTTCTGTGCTATTTCATTGTAAAGATGAATTGCC
ATTTGTTAGAGAGTTGGAAGATGCAAAAATATAGTCATTGCCAAACAAGAAAAATATGAGCCAGTGAAGAAATAAA
ATCTGAGGTCCTAGAGGCAGTGACCGAAGAACCGGAAGACATATCAGATATTGTAATAAGTTTAATCTCATCCTACAG
ATCCTGTATCAAGGAATACAGGTTTTCTGAAAGGCGCATACTTCTGAAATTGCTTCCACTTATGGTGTACTGTTAAA
TGATCCTCAATTAAGATTCCAGTATCCAGAAGGTTTTTAGATAAATTAGGTTTGAGCCGTAGTCCAATTGGTAAGGC
AATGTTTGAAATTAGAGACTATAGTAGTAGAGATGGAATAACAAGTCCTACGATTATGATGATTTCAGAAATTTTC
TTTCAGAGGCAACAAAACATAATAATAGGAGCCAAAATAGAGATTACGATGACGAACCGTTTACGCGCAGCAACAA
CAATAGACGCTCATTCTCCCGTAGTAATGATAAAAATAATTACTCCAGTAGAAACAGCAACATATATTAGAAAAGATA
AAAAAGGAGGACCAAGAGATGTGGGTTGATTTTTAAATAACAATTCCTCTATTAAGACGCTGTATTTTAGACCCCAT
CTTGATTCAATACTTTACGCTGTAAATACCTAGTATATAAATATTAACAAAAATTTTTTTCTTCTCCTATTCTCTT
AAAATGATAACTATACGAAGATCCTACTGGTATATTTATTCGTCAATGCTTATTTACGACCGGTACCTCCATTCTTAC
TTTCGCTCTTCTCTCTCAAAGGAAAAGTAATTCTAACATTGTAATTC3'

Grey region shows MSS116 ORF. Colored regions are referring to forward primers (colors as in table), italic letters to reverse primers to introduce mutations (Table 3). Colored letters show nucleotides that were mutated. Underlined regions show homologues regions for NAT1 genomic insertion.

Table 2: *Saccharomyces cerevisiae* strains.

name	characteristics	Parent strain
TN16	GII-0 ai5 γ , Mat α ade1 lys1 ura3-51	a161
HRH189	MSS116 Δ ::kan ^R	a161
yCW08	MSS116-3'UTR::CloNat ^R	TN16
yCW01	MSS116::K158R, CloNat ^R	TN16
yCW05	MSS116::Q412A, CloNat ^R	TN16
yCW10	MSS116::stop codon at AA569/570, CloNat ^R	TN16
yCW11	MSS116::R245E, CloNat ^R	TN16
yCW12	MSS116::K158A, CloNat ^R	TN16

yCW strains were cloned by myself. Strains TN16 and HRH189 were kindly provided by Philip Pearlman¹²⁸

Table 3: Primers used in MSS116 mutagenesis.

name	Sequence 5'-3'
yCW08F_1	<u>TAACAATTCTCTATTAAGACGCCTGTATTTTAGACCCCATCTTGATTCAATACTTT</u> <u>CACAGCTTGCCTTGTCCCCGCCG</u>
yCW08R_1	<u>AAAGTAAGAATGGAGGTACCGGTCGTAAATAAGCATTGACGAATAAATATACCAGTA</u> <u>GGATCGACACTGGATGGCGGCGT</u>
yCW01F	<u>TATCGCAAGAGCTAAAACCGGTACAGGTAGGACATTTCCTTCTTGATTCCAATTTT</u> <u>CCA</u>
yCW01R	CTTTGAGAGGAGGAAGAGCGAAAGTAAGAATGGAGGTACCGGTCGTAAATAAGCATT GAC
yCW05F	<u>CCAAATTTTGAATTTTCATGGTAAAATTACTGCAAATAAATAAAGAACAAGTTTAG</u> <u>TTAAAAGATTCT</u>
yCW10F	<u>CCTGAAATTGCTTCCACTTATGGTGTACTGTTAAATGATCCTCAATTAAUAGTAACCA</u> <u>GTATCCAGAAGGTTTTTAGATAAATTAGGTTTGAGCCGTAGTCCAATTGGTAAGG</u>
yCW11F	<u>CCAAACATTGTAATTGCCACTCCAGGGGAATTGATTGATGTTTGGAAAAATACTCC</u> <u>AACAAATTTTTTAGGTTTCG</u>
yCW12F	<u>TATCGCAAGAGCTAAAACCGGTACAGGTGGACATTTCCTTCTTGATTCCAATTTT</u> <u>CCA</u>

Mutations are shown in red letters.

Introduction of point mutations in MSS116

Point mutations were introduced in MSS116 using the 2-step method of Toulmay and Schneider²⁰⁷. First I introduced the NAT1 gene downstream of the MSS116 ORF, specifically in the 3'UTR of MSS116. To do that I did a PCR on PAG25 (Euroscarf plasmid) using the primers yCW08F_1 and yCW08R_1 amplifying NAT1 with flanking sequences homologous to regions in the 3'UTR of MSS116. This PCR fragment was then transformed into the TN16 strain (Figure 2.1), resulting in yCW08 in which NAT1 is inserted in the 3' UTR of MSS116. The genomic DNA of this strain was used as a template to obtain a PCR fragment that includes the desired mutation as well as the NAT1 gene for selection. Transformation into TN16 resulted in clones able to grow on selective media containing CloNAT (100 mg/ml). After isolation of genomic DNA from the transformants MSS116 was amplified by PCR and sent for sequencing.

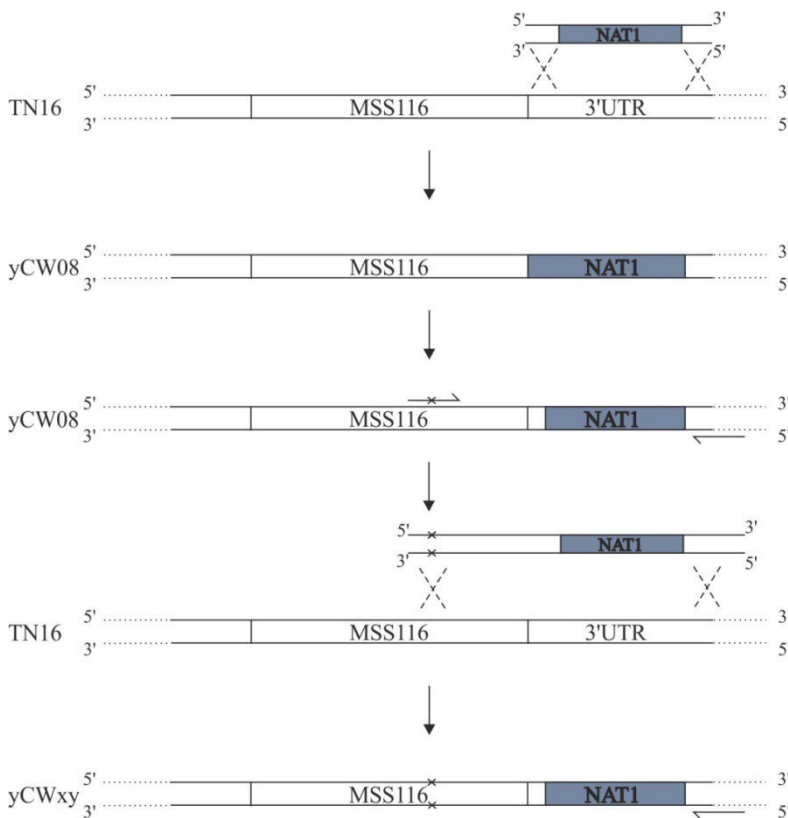


Figure 2.1: Overview of the 2-step method for introducing of point mutations in the yeast protein Mss116p.

PCR of homologous recombination cassette

For this PCR I used a mixture of GoTaq DNA Polymerase (Promega, 5U/μl) and Pfu DNA Polymerase (Promega, 2-3 U/μl) to ensure efficiency and accuracy in DNA polymerization. I typically needed a volume of 200-400 μl of PCR reaction (0.5 mM dNTPs, 0.5x GoTaq colourless reaction buffer, 0.5x Pfu reaction buffer, 1 mM yCWxyF primer, 1 mM yCW01R primer, 1.5 mM MgSO₄, 1/20 volume DMSO, 1/100000 (w/v) genomic DNA, 2 U GoTaq DNA Polymerase, 2 U Pfu DNA Polymerase) to get the required amount of DNA for transformation. The PCR product was checked on a native 0.8 % agarose gel and purified with the Wizard® SV-Gel and PCR Cleanup System (Promega) following the instructions from the manufacturer.

Transformation in *Saccharomyces cerevisiae*

TN16 yeast cells were first grown o/n in 5 ml YPD medium (1 % yeast extract (Difco), 2 % peptone, enzymatic digest (Difco), 2 % glucose, 55 mg/L L-tyrosine, 55 mg/L adenine hemisulfate salt). 50 ml of YPD were inoculated with a volume of this preculture to yield an OD₆₀₀ of ~ 0.5. This main culture was harvested at an OD₆₀₀ between 0.4 and 0.6 by a 1 min centrifugation step at RT and 7000 rpm. Cells were washed in 10 ml ddH₂O (sterile) and then resuspended in 0.5 ml LiAc/TE (0.1 M LiAc, 10 mM Tris/HCl pH 8.0, 1 mM EDTA pH 8.0). Meanwhile ssDNA from salmon sperm (10 mg/ml) was boiled for 5 min. DNA for transformation was amplified by PCR using a forward primer harboring the desired mutation(s) (see) a reverse primer binding downstream of the inserted NAT1 gene and genomic DNA of yCW08 as a template. The transformation mix consisted of 600 μl 40 % PEG4000/LiAc/TE, 100 μl competent cells, 20 μl salmon sperm DNA and 10-20 μg of recombination cassette for transformation. This mix was then incubated at 30 °C for 1 h. 70 μl of DMSO were added. The cells were then heat shocked for 15 min at 42 °C and recovered for 1.5 h at 30 °C. After pelleting the cells at 7000 rpm and RT for 1 min they were plated on YPD/CloNat plates and grown for 3 days at 30 °C.

Genomic DNA preparation

Genomic DNA was extracted using the Wizard™ Genomic DNA Purification Kit from Promega following the manufacturer's instructions.

Sequencing

For sequencing transformants were picked and genomic DNA was prepared as described before. This DNA then served as a template for PCR (0.5 mM dNTPs, 1x Pfu reaction buffer, 1 μM CW01F primer, 1 μM CW01R primer, 1.5 mM MgSO₄, 1/20 volume DMSO, 1/100000 (w/v) genomic DNA, 2 U Pfu DNA Polymerase (Promega)). Samples were purified using the Wizard® SV-Gel and PCR Cleanup System (Promega) and sent to LGC genomics or Microsynth for sequencing using multiple primers to obtain the whole sequence (Table 4). Correct integration of the recombination cassette was checked by using primers binding outside the recombination cassette (CW61F/R).

Table 4: Primers for sequencing.

name	Sequence 5'-3'
CW01F	ACGCGTCGACATGTTGACCTCTATATTGATAAAAG
CW01R	CGGAATTCCTAATATATGTTGCTGTTTCTACTG
CW61F	ACCTCACTTACTCACACGATG
CW61R	ATTGACGAATAAATATACCAGTAG
yCW01-seq1	CCAGTGAAGACCATGATGTTATCGC
yCW01-seq3	CCTAGAGGCAGTGACCGAAGAACCG
CW01M	GACACAGTAGATAAGAATGAACCTGAAGCACATG
GCW01F	CTGGCGCCGGAGAGTCGTGTATCG
GCW01R	GGTATTTACACGTGAAAGTATTGAATCAAGATGGGGTC

Glycerol stocks

Cells harboring the desired mutation were grown in YPD-Clonat o/n and stored in YPD-Clonat with 15 % glycerol (f.c.) at -80°C.

2.2 RNA preparation

Cell growth

A colony was picked and inoculated in 5 ml YPD (TN16, HRH189) or 5 ml YPD/CloNat (yCW01/05/10/11/12) and incubated o/n. 2 ml (TN16, yCW05/10) or 4 ml (HRH189, yCW01/11/12) of this preculture were then used to inoculate 50 ml fresh YPR (1 % yeast extract (Difco), 2 % peptone, enzymatic digest (Difco), 2 % raffinose, 55 mg/L L-tyrosine, 55 mg/L adenine hemisulfate salt). This main culture was grown until an OD₆₀₀ of approximately 1. Cells were harvested by a 1 min centrifugation step at RT and 7000 rpm. Cells were then resuspended in 1.8 ml YPD and split in two 1.5 ml reaction tubes each now containing 1 ml of cell suspension.

DMS modification

DMS was added to one of the two samples from with a final concentration of 50mM. After brief vortexing the samples were put at 30 °C shaking with 300 rpm for 2 min. Immediately afterwards β-mercaptoethanol was added to a final concentration of 0.7 M in order to stop the reaction and 50 µl isoamyl alcohol to keep DMS from pelleting. This was carried out also with cellsthat were not treated with DMS as β-mercaptoethanol also affects cell disruption. The resulting sample was vortexed thoroughly for 15 s and the cells pelleted by a 1 min centrifugation at RT and 7000 rpm. Cells were washed in 1 ml YPD and β-mercaptoethanol was again added to DMS treated cells. After another pelleting step cells were either immediately disrupted or frozen at -80 °C.

Yeast total RNA extraction

Total yeast RNA was prepared using the Ambion RiboPure™ RNA purification kit for yeast. Individual steps were carried out following the protocol of the Ambion Kit. Elution volume was 100 µl. Therefor the elution was precipitated with 2.5x volume of 0.3 M NaOAc/EtOH at -20 °C o/n with subsequent centrifugation at 13200 rpm for 30 min (4°C). The total RNA pellet was dried for 10 min at RT and resuspended in 10 µl

ddH₂O. RNA concentration was determined using the Nanodrop spectrophotometer and the RNA solution diluted to a concentration of 16 µg/µl.

2.3 cDNA synthesis

Primers for Reverse Transcription

Table 5: Primers for reverse transcription.

name	Sequence 5'-3'
Sc-111	GCCCCAATGAAAATTAATCAG
Sc-6	GAAAATAGCACCCATTGATAATAC
Sc6	AAATATTATTTATGATAACTTTCAGA
Sc82	TGTTACCATTTTAATACACTTG
Sc160	GTTCTTCATCTTTTTTTTATAA
Sc234	GTTTTCAATTAGTGGTGTAAG
Sc312	TTCCAATACATAACATCAACC
Sc375	GTAAATATCTAACTTAGCTCTC
Sc440	TTGGTATTATTATTATTTTATTATTA
Sc520	TATTTCTTAATCCAATAATTATTTATA
Sc588	GCATTAGCTTTTTATACAATC
Sc656	CACCTATAGTATAAGTTAGCAG
Sc754	ATTTATCAGTTATATATAAACCTCC
Sc805	CGGCTCACGCAATATC
Sc866	CCCGATAGGTAGACCTTTAC

Primer purification

The 100 µM desalted primer solution was mixed 1:1 with 1x loading buffer (7 M Urea, 25 % sucrose, 0.25 % bromphenol blue, 0.25 % xylene cyanol) and applied on a 20 % denaturing polyacrylamide gel (40 ml 20 % polyacrylamide (7 M Urea, 20 % acrylamide/bisacrylamide 19:1, 0.089 M TRIS base, 0.089 M boric acid, 2 mM EDTA pH8.0), 20 µl TEMED, 200 µl 10 % APS). This gel was run at 25 W (limiting) until the xylene dye passed half the gel. The primer was detected by UV shadowing and the respective gel piece was cut out and rotated in 1ml elution buffer (250 mM NaOAc pH 5.4, 10 mM TRIS pH 7.5, 2 mM EDTA pH 8.0, 0.1 % SDS) o/n. The resulting solution

was then precipitated with 2.5x volume of 0.3 M NaOAc/EtOH at -20°C for 30 min with subsequent centrifugation at 13200 rpm for 30 min. The pellet was dried at RT for 10 min and resuspended in half the volume of ddH₂O initially applied on the gel. A₂₆₀ was measured at the Nanodrop Spectrophotometer and the molarity was determined using the following formula

$$c \text{ (mol/L)} = A_{260} / (\epsilon_{\text{total}} * \text{pathlength (cm)}) * \text{dilution}$$

in which $\epsilon_{\text{total}} = 0.9 * (15346.6 * n_{\text{adenine}} + 11751.8 * n_{\text{guanine}} + 9929.7 * n_{\text{thymidine}} + 7623.2 * n_{\text{cytosine}})$

Primer Kinase Reaction

4.5 µl ddH₂O were mixed with 1 µl of 10 µM oligonucleotide, 1 µl of 10x T4 polynucleotide kinase reaction buffer (NEB), 0.5 µl T4 polynucleotide kinase (NEB, 10 U/µl) and 3 µl of γ-P³²-ATP (6000 Ci/mmol). The resulting solution was incubated at 37 °C for 30-45 min and the reaction stopped with 0.5 µl 0.5 M EDTA pH 8.0 and incubation at 95 °C for 1 min. Afterwards the oligonucleotide was precipitated with 2.5x volume of 0.3 M NaOAc/EtOH at -20 °C for 30 min with subsequent centrifugation at 13200 rpm for 30 min (4°C). The pellet was dried at RT for 10 min and resuspended in 20 µl ddH₂O.

Reverse transcription reaction using AMV reverse transcriptase

2.5 µl of a 16 µg/µl total RNA solution were mixed with 1 µl 4.5x hybridization buffer (225 mM K-Hepes pH 7.0, 450 mM KCl) and 1 µl of the P³² labeled primer stock solution. This mix was incubated for 1 min at 95°C. The reactions were put on ice immediately afterwards for 2 min. 1.15 µl ddH₂O, 0.6 µl 6.7x extension buffer (1.3 M TRIS pH 8.0, 100 mM MgCl₂, 100 mM DTT), 0.3 µl dNTPs (2.5 µM each) and 0.15 µl AMV reverse transcriptase (Promega, 10 U/µl) were added to the reaction. To reveal adenosines and cytosines on the analytic PAGE following reverse transcription, 1 µl of either a 1 mM ddTTP or 1 mM ddGTP solution was added in addition to the extension mix. Reverse transcription was carried out for 1 h at 42°C. Then 1.5 µl 1 M NaOH was added and the remaining RNA hydrolyzed for 1 h at 42°C. The pH was neutralized with

1.5 µl 1 M HCl and 1 µl 0.5 M EDTA pH 8.0 was added. The cDNA was precipitated with 2.5x the volume of 0.3 M NaOAc/EtOH at -20°C for 30 min with subsequent centrifugation at 13200 rpm for 30 min. The pellet was dried at RT for 10 min and resuspended in 8 µl 1x loading buffer (0.089 M TRIS base, 0.089 M boric acid, 2 mM EDTA pH8.0, 7 M urea, 25 % sucrose, 0.25% bromphenol blue, 0.25 % xylene cyanol).

Reverse transcription reaction using Transcriptor™ reverse transcriptase

For primer Sc866 Transcriptor™ reverse transcriptase had to be used due to the high content of G-C pairs in the RNA region to be mapped with this primer. Hybridization is carried out as described in . For extension of the annealed primer 7.5 µl ddH₂O, 4 µl 5x reaction buffer (Roche Applied Science), 2 µl 10mM dNTPs, 1 µl 0.1 M DTT, 0.5 µl RNasin (Promega, 40 U/µl) and 0.5 µl Transcriptor reverse transcriptase (Roche Applied Science, 20 U/µl) were added. For sequencing lanes 5 µl ddH₂O were replaced with 5 µl of a 1mM ddTTP or ddGTP solution. The extension was carried out for 1 h at 50°C. For subsequent RNA degradation 3µl of a 1 M NaOH solution were used and samples were incubated at 50°C for 1 h followed by neutralization with 3µl of 1 M HCl. Precipitation and resuspension were done as described before.

Poisoned primer assay

To measure the ratio of unspliced to spliced ai5γ a poisoned primer assay was performed. The transcription itself was identical to with the only exception that instead of a 2.5 mM dNTP solution a poisoned mix (0.88 mM dATP, dGTP, dTTP and 3.68 mM ddCTP) was used for extension of primer Sc-6 leading to cDNAs of different length referring either to preRNA or mRNA.

Denaturing PAGE of cDNAs

For resolving the cDNA population resulting from reverse transcription a denaturing 8 % polyacrylamide gel was prepared (80 ml 8 % polyacrylamide (7 M urea, 8 % acrylamide/bisacrylamide 19:1, 0.089 M TRIS base, 0.089 M boric acid, 2 mM EDTA

pH8.0), 45 μ l TEMED, 450 μ l 10% APS). Best resolution was achieved by using a gel size of 33 cm x 51.5 cm and limiting PAGE at 75 W. 4 μ l of the resuspended cDNAs were loaded on the gel and run until the bromphenol blue dye reached the bottom end of the gel. After vacuum drying for 1 h at 80 °C the gel was exposed to a phosphorimager screen (GE healthcare) for 1 to 3 days depending on the intensity of the signal and finally scanned on a Storm 820 phosphorimager (GE healthcare).

2.4 Data analysis

For analysis of scanned gels the software ImageQuant was used. With this software the radioactive counts of individual bands can be determined. This data was then extensively analyzed using Microsoft Excel. First the differences in the overall intensity of different lanes had to be considered. Normalizations were done by balancing the counts of nucleotides that are accessible in ai5 γ in the different genetic backgrounds. To minimize any bias 2-3 nucleotides were used to calculate an average for normalization. Then a ratio of modification intensity for each nucleotide mapped in the wt and mutant strain was calculated to identify protections or enhancements in DMS modification. The changes in modification pattern between wild type and mutant strains were summarized in a differential map. To determine the sites of modification in one strain, the fraction of a specific nucleotide count compared to the overall counts of the nucleotides in one lane was calculated and plotted onto the secondary structure map.

2.5 Dot spots

To compare the effects of mutations in MSS116 on the respiratory growth, the respective yeast strains were grown on fermentable (YPD) and non-fermentable (YPG) carbon source using the wild type and knockout strain as controls. Single colonies of the different strains were picked and precultures were set up as described before. 30 ml of YPD was inoculated with this preculture to an OD₆₀₀ of 0.1. Cells were grown at 30 °C shaking with 180 rpm until the suspension reached an OD₆₀₀ between 0.4 and 0.6; importantly the OD of all different strains has to be the same at time of harvest. 2 μ l

drops of the cell suspensions were applied on both a YPD plate as well as a YPG plate (2% glycerol instead of glucose). Furthermore, 3 serial 1:10 dilutions were applied. Cells were grown at 30 °C for 3 days.

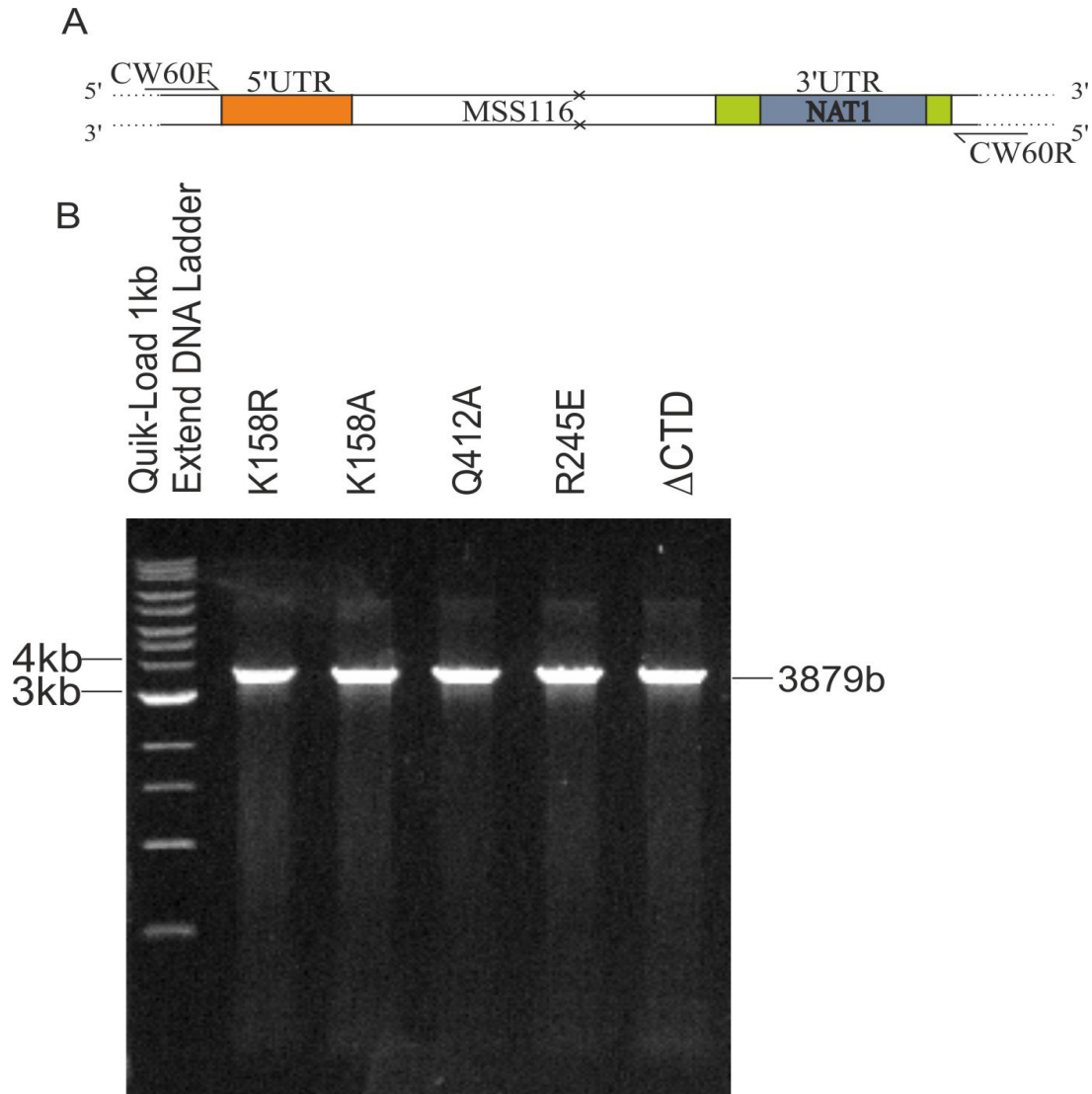
3 Results

3.1 Introducing mutations in chromosomal copy of the *mss116* gene

Mss116p plays a major role in splicing of all mitochondrial introns in yeast¹²⁸. Its activities, ATP binding and hydrolysis as well as RNA unwinding and binding, were subject to numerous studies^{152,183,187–190,192,194,204}. Nevertheless, the mode of action of this DEAD-box protein in RNA folding is still unclear. To reveal which activities are important for folding of the group II intron ai5 γ we introduced different mutations into the *mss116* gene to disrupt activities of interest. In order to do that we used a PCR based method which introduces the desired mutation in the chromosomal gene by homologous recombination²⁰⁷. The integrating PCR product harbors the desired mutation as well as a marker gene at the 3'-UTR of *mss116*, in our case the cloNat resistance gene NAT1. By directly mutating the chromosomal copy of *mss116* we ensure expression of the different mutated versions of Mss116p at endogenous levels which results in most biological relevant data, minimizing effects due to abundantly present proteins. This is a very important aspect of this study, since the effect of different levels of Mss116p in *in vitro* studies cannot be neglected and high levels of the protein can for example lead to ai5 γ splicing even in absence of an ATP substrate *in vitro*¹⁸³.

To assess the different activities of Mss116p we created mutant strains harboring the K158R, K158A, Q412A and R245E mutations, all of which have already been subjects of *in vitro* and *in vivo* studies^{187–189,192,195}. Additionally, we created a strain with two consecutive nonsense mutations at amino acid positions 569 and 570 leading to a truncated version of *mss116*, named Δ CTD strain in which the distinct basic tail has been deleted¹⁵². The incorporation of the correct mutation as well as the lack of additional mutations were confirmed by sequencing. To show that homologous recombination took place at the desired location we performed a PCR reaction using primers binding upstream of the 5'UTR of the *mss116* gene and downstream of the 3'UTR, which includes NAT1, respectively (Figure 3.1A). The resulting size of the PCR product of the different strains corresponds to a genotype in which the marker gene *nat1*

is integrated at the correct position in the 3'UTR of *mss116* (Figure 3.1B). Expression and mitochondrial localization of the strains that were subject to *in vivo* chemical probing (K158R, Q412A and Δ CTD) was verified by fluorescence microscopy of the GFP-tagged version of the respective mutated Mss116p (Heisig, unpublished data). All three versions of the DEAD-box protein localized in yeast mitochondria.



*Figure 3.1: Verification of correct location of integration cassette harboring the desired mutation of *mss116* gene and the marker gene *NAT1* in chromosome IV locus. (A) Scheme of the setup. (B) Agarose gel showing the correct size of the respective PCR products.*

3.2 Conserved motifs of Mss116p and their role in respiratory growth of yeast

The ai5 γ intron is a perfect candidate for *in vivo* studies of RNA folding: First, it has been studied extensively *in vitro*^{31,63,70,77,105–108}. Second, adopting its native structure leads to fulfillment of its catalytic function, self-splicing, which can be directly monitored with a

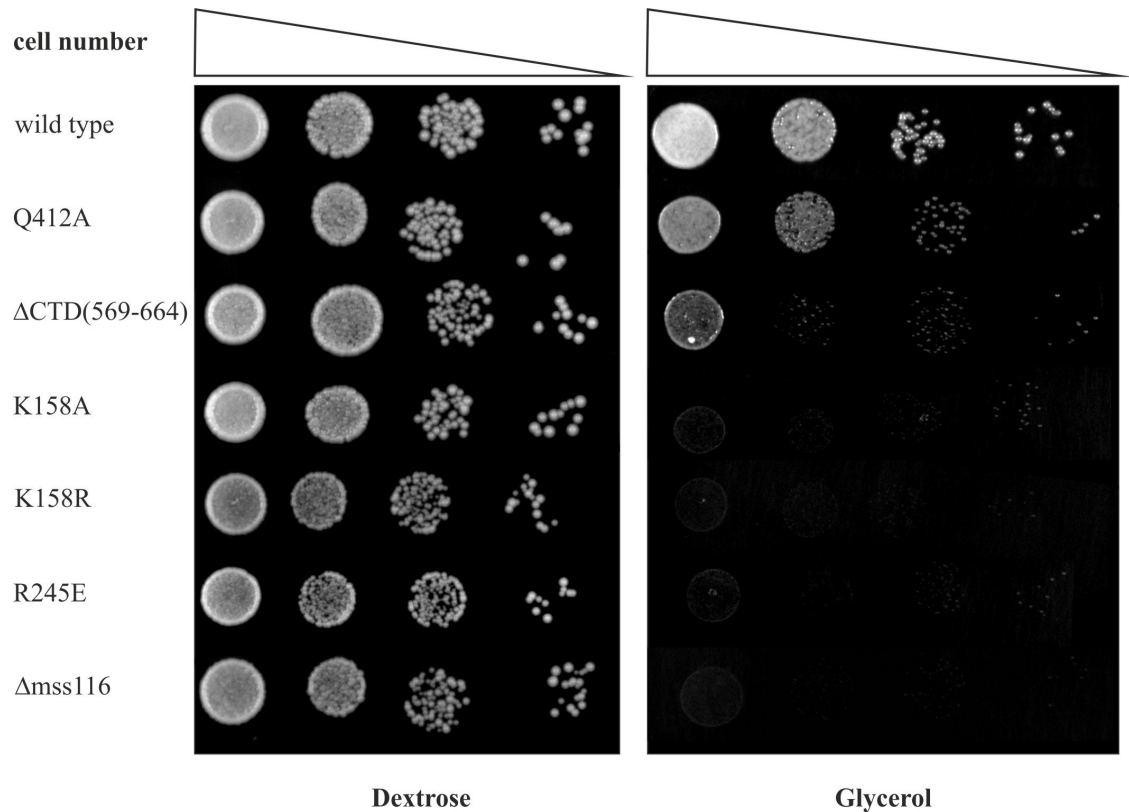


Figure 3.2: Aerobic versus anaerobic growth of yeast strains with different genetic background regarding Mss116p. Wild type yeast strain displayed on top and mss116-knockout strain at the bottom. On the left panel the yeast strains can grow by fermentation and respiration on YPD medium, on the right panel by respiration only on YP-glycerol medium. Cell number decreases in serial 1:10 dilution steps from left to right. Cells were grown in liquid medium, containing glucose as carbohydrate source o/n and then plated on plates containing either glucose or glycerol and grown at 30°C.

poisoned primer assay *in vivo*. Last but not least only if ai5 γ is able to fold properly and therefore excise itself out of the pre-RNA, the mature mRNA of COX1 can be translated resulting in a functional respiratory chain. By comparing growth of yeast on glucose as a fermentable carbon source and on glycerol which is non-fermentable, the functionality of the respiratory chain can be monitored. With this system the effect of the different mutations on the ability of yeast to grow aerobically can be monitored directly. As expected the wild type strain grows with comparable efficiency aerobically and anaerobically, whereas the knockout strain can only grow on a medium containing glucose (Figure 3.2). K158 which is found in motif I (Walker A) and is part of the ATP binding pocket was shown to be important for ATP binding and hydrolysis^{187,188}. The K158R mutation interferes with both leading to a loss of Mss116p's ability to assist in ai5 γ splicing¹⁸⁷. As expected the K158R shows a knockout-like phenotype in aerobic growth (Figure 3.2). Mutating K158R to an alanine also disrupts its ATPase activity^{187,188} and cells harboring this mutation are equally impaired in their growth on glycerol medium (Figure 3.2). The same is true for the R245E mutation that establishes several contacts to the RNA substrate and was proposed to play an important role in bending it^{143,144,146}. This mutation also disrupts ATP binding and hydrolysis¹⁸⁸. All three mutants were also shown to be unable to unwind a duplex RNA substrate^{187,188}.

The Δ CTD mutation results in a Mss116p version that is truncated at position 569 and therefore lacks its distinct basic tail. This version of the DEAD-box protein shows almost wild type-like ATPase activity (2-fold reduced), but significantly reduced RNA unwinding and annealing activity¹⁵². RNA binding was 10-fold decreased compared to wild type Mss116p. In our respiratory growth assay this yeast strain shows some aerobical growth, although the phenotype resembles more the knockout than the wild type strain (Figure 3.2). This is in contrast to previous studies, in which only a truncation at position 551 resulted in the inability to grow on YP-glycerol medium and the Δ 569-664 strain displayed wild type-like growth¹⁵². The only apparent difference in approach is the protein being expressed from a plasmid rather than, as in our case, from its endogenous locus on the chromosome. This might result in elevated levels of proteins, supressing the effects of a truncated basic tail.

Q412 found in the QxxR motif is indirectly involved in introducing the bent in the bound RNA-substrate which is thought to be incompatible with a helical conformation of dsRNA, making this bent crucial for RNA unwinding^{146,188}. In contrast to the previously mentioned mutations Q412A displays wild-type ATP binding and hydrolysis¹⁸⁸. However, its helicase activity is strongly reduced. Interestingly, yeast containing the Q412A mutation still grows aerobically (Figure 3.2).

In summary aerobic growth of yeast cells is impaired the strongest in strains harboring *mss116* mutations that have a strong effect on ATPase and/or ATP binding activity as in K158R, K158A and R245E. On the other hand the C-terminal truncation that impairs RNA unwinding and RNA binding shows respiratory growth although strongly reduced. If only RNA unwinding is disrupted as in the Q412A mutation, respiratory growth is only slightly reduced compared to the wild-type genotype. These findings suggest that ATP binding and hydrolysis are crucial for aerobic growth. Also, RNA binding of Mss116p is important for the ability of yeast cells to grow on a non-fermentable carbon source. In contrast, disrupted unwinding activity of Mss116p results in an almost wildtype-like phenotype on YP-glycerol medium.

3.3 Dissecting the motifs of Mss116p essential for splicing of ai5 γ *in vivo*.

To monitor ai5 γ splicing *in vivo* we established a poisoned primer assay. Here the dNTP mix for reverse transcription of the *cox1* group II intron ai5 γ contains ddCTP instead of dCTP together with the deoxy version of the other nucleotides. This terminates reverse transcription at the first encountered G. Using the Sc-6 primer that binds in the 5'-exon relative to the ai5 γ intron, primer extension proceeds either 8 nucleotides if the intron is still integrated in the pre-RNA or 14 nucleotides if splicing has occurred. By quantifying the respective bands the ratio of unspliced and spliced ai5 γ can be determined.

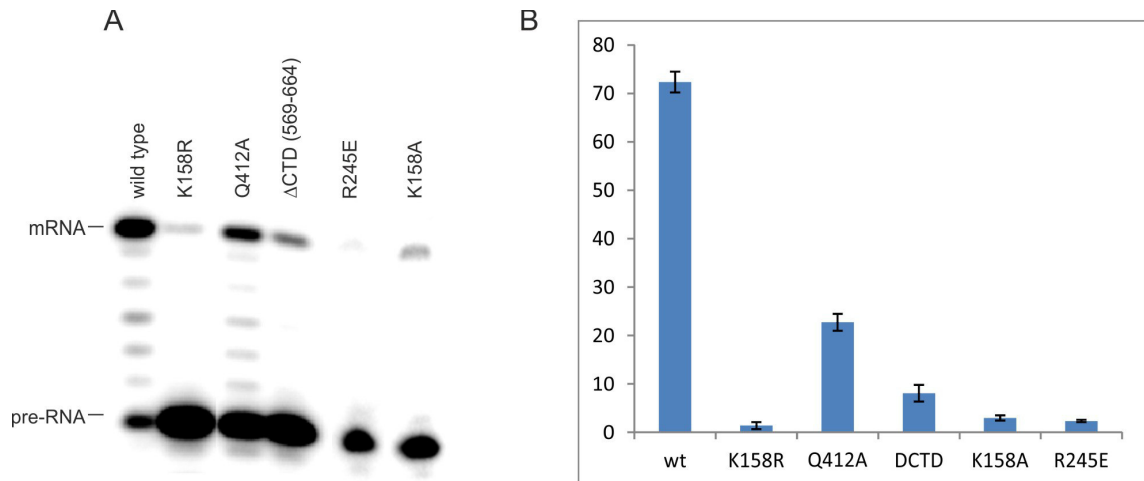


Figure 3.3: Splicing of *ai5γ* in yeast strains expressing wild type *Mss116p* or its various variants. (A) Representative gel a of poisoned primer assay. The lower band is referring to the fraction of pre-RNA the upper one to mRNA. (B) Average splicing efficiency of *ai5γ* in different strains was determined from 3 independent experiments with indicated standard deviation.

In wild type yeast approximately 75% of the *ai5γ* introns are spliced (Figure 3.3) and in the absence of *Mss116p* the level of spliced introns drops to <1% (Table 6) both results being consistent with *in vivo* splicing found in earlier experiments¹⁰⁹. In the K158R and K158A as well as the R245E mutant the splicing levels are comparable to the *mss116* knockout strain (Figure 3.3). This correlates with the observation that these mutants were also showing no respiratory growth on glycerol containing media. Truncation of the basic tail in the ΔCTD strain also tremendously reduced the self-splicing efficiency of *ai5γ*, resulting in approximately 10% of mature mRNA (Figure 3.3). This is again in contrast to the findings of a previous study using plasmid encoded *Mss116p* variants, resulting in a splicing activity that was indistinguishable from wild type levels¹⁵². The least, although still significant, difference to wild type *ai5γ* splicing was seen for the Q412A mutant, in which mRNA levels reached only 1/3 of those found in the wild type strain (Figure 3.3).

Table 6: Splicing levels of *ai5γ* in the presence of different *Mss116p* variants *in vivo*.

Mss116p variant	Motif	Impaired activities	Relative splicing (1/ k_{rel})
Wild type	-		1
Knockout	-		<0.01
K158R	Motif I	ATP binding and hydrolysis, RNA unwinding	<0.01
K158A	Motif I	ATP hydrolysis, RNA unwinding	0.02
R245E	Motif Ib	ATP binding and hydrolysis, RNA unwinding	<0.01
Δ CTD	Basic tail	RNA binding, RNA unwinding	0.1
Q412A	QxxR	RNA unwinding	0.33

Relative splicing levels are determined by $1/k_{rel}$. k_{rel} is the fraction of spliced *ai5γ* in a wild type genetic background divided by that of the respective mutant strain. Values are obtained from at least 3 individual experiments.

3.4 The role of ATPase activity as well as RNA binding and unwinding of *Mss116p* in folding of *ai5γ*

3.4.1 General approach

Up until now there are not many possibilities to obtain insight into the structure of RNA molecules *in vivo* let alone the high resolution data obtained by NMR and X-ray scattering *in vitro*. Despite the tremendous value of those two techniques for revealing the most intricate connection of structure and function in a variety of ribozymes *in vitro* both methods often require non-physiological conditions or changes in the molecule of interest itself. It is therefore important to compare those high resolution structures obtained *in vitro* as well as *in silico* approaches with data obtained *in vivo*.

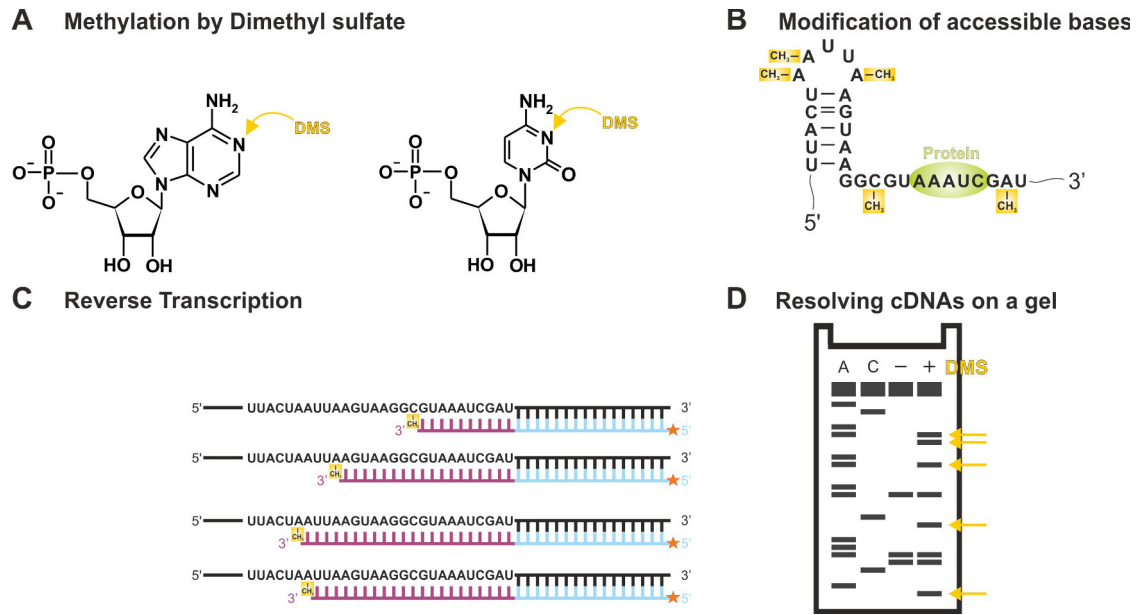


Figure 3.4: DMS chemical probing. (A) DMS methylates N1 of adenosines and N3 of cytosines. (B) Modifications only take place if the respective nitrogen is not involved in hydrogen bonding. (C) Methyl groups lead to a premature stop of reverse transcription resulting in a pool of cDNA of different lengths. (D) This cDNA pool can then be resolved on a denaturing polyacrylamide gel. Sequencing lanes show the exact location of the respective DMS modification. A control lane, in which RNA that was isolated from untreated cells was reverse transcribed, reveals natural stops of the reverse transcriptase.

In order to obtain information on the structure of ai5 γ *in vivo* we established chemical probing of RNA molecules in yeast cells using dimethyl sulfate²⁰³. DMS is a small molecule that readily penetrates the cell and all of its compartments. It adds a methyl group to N1 of adenosines and N3 of cytosines, if they are not involved in hydrogen bonding. Revealing these modifications by reverse transcription allows identification of adenosines and cytosines that are involved in secondary or tertiary interactions and correlating the obtained data with the secondary structure information as well as recently established three-dimensional structures of a group IIC intron and a modeled structure of D1356 based on this crystallographic data^{69,75}.

By quantifying the stops resulting from methylation we cannot only identify accessible nucleotides but also gain information about the degree of accessibility. This proves useful for dissecting the effects of the different Mss116p variants on the structure of ai5 γ *in vivo* by comparing its modification pattern in different genetic backgrounds. The resulting differential map provides us another puzzle piece to solve the question of how DEAD-box proteins assist large RNAs in adopting their native structure.

3.4.2 *The intracellular structure of ai5 γ*

We have already shown that the DMS accessibility of nucleotides in ai5 γ in wild type yeast is highly consistent with its phylogenetically established secondary structure as well as the crystal structure of the *O. iheyensis* group IIC intron^{31,47,109}. The DMS pattern is also almost identical to that of the D123456 ribozyme *in vitro*¹⁰⁹.

To be able to interpret the DMS modification pattern of ai5 γ in different genetic backgrounds it is of utmost importance to have a positive control showing the DMS pattern in a wild type situation. Therefore structural probing of mutant strains was always carried out in parallel with the wild type strain.

The data on DMS accessibilities of nucleotides in the ai5 γ group II intron confirms previous results¹⁰⁹. Ai5 γ adopts a catalytically active conformation, in which the active site center is properly formed and all tertiary contacts are in place (Figure 3.5). Adenosines and cytosines in helical regions are protected from DMS modification, revealing that all stems are established. Hence the secondary structure as well as the tertiary structure of ai5 γ in a wild type genetic background are in line with a fully catalytically active ribozyme. Since ai5 γ splices efficiently in wt yeast, the chemical probing data reflects mostly already spliced intron molecules.

3.4.3 *Mss116p* ATPase activity and *ai5γ* folding

Disruption of the ATPase activity of *Mss116p* consequently leads to defective group II intron splicing^{187,188}. Interestingly, although *in vitro* at near-physiological conditions ATP is needed for *Mss116p*-assisted native structure formation of the *ai5γ* ribozyme, compaction of D1 was shown to happen even in the absence of ATP¹⁹⁵. The ATPase activity was assumed to allow repeated cycles of RNA binding, since the ADP·*Mss116p* complex has very low affinity to its RNA substrate^{146,169,170}. In a single molecule study it was shown that in fact the D135 ribozyme reaches a near-native state in the presence of *Mss116p* even without ATP²⁰¹. Nevertheless, disrupting the ATPase activity with different mutations in motifs I and Ib all lead to an unspliced *ai5γ* intron and subsequently to a disfunctional respiratory chain¹⁸⁸.

The κ - ζ folding control element is not formed without ATPase activity of *Mss116p*

The collapse of the *ai5γ* group II intron depends on the formation of a small but very complex substructure, the so-called κ - ζ element⁷⁷. This region of D1 is also very important for establishing the active site center, rendering its structure formation a very crucial step in both *ai5γ* folding and catalysis³¹. We found that indeed this region of the intron shows some of the strongest enhancements in DMS accessibility when the ATPase activity of *Mss116p* is abolished due to the K158R mutation¹⁸⁷. The stems d, d' and d'' that are part of the κ - ζ element and whose formation is strongly Mg^{2+} dependent *in vitro*, do not form in the K158R strain. A191, C193, A195 and A196 (Figure 3.6) as well as A387 and A389 (Figure 3.7) are all more accessible to DMS, showing that stem d is not formed. Enhanced accessibility of A205 and C206 (Figure 3.6) reveal that also these Watson-Crick base pairs are not established and the respective helix is not formed. Methylation of A378 (Figure 3.7) indicate disruption of stem d' and enhancements in C217, A218, A219 and A220 (Figure 3.8) an unformed stem d'', respectively.

A63, C65, A344 and A346-A348 (Figure 3.9, Figure 3.10) are more accessible to DMS than in wild type yeast cells showing that the intradomain α - α' interaction is not formed. Enhancements in the nucleotides A146, C147 and A148 (Figure 3.6) as well as in C259 (Figure 3.8) indicate that also the β - β' tertiary contact remains open. Both of these interactions, which are kissing loop interactions established by Watson-Crick base

pairing, are playing an important role in D1 compaction^{38,55,72,77}. Folding of the κ - ζ element not only triggers the collapse of the intron, but formation of the active site also depends on this region lying at the very heart of the intron. Two very important tertiary interactions that also give this element its name, the κ - κ' and ζ - ζ' interactions, play a fundamental role in establishing the active site^{25,67}. Enhanced DMS-accessibilities of A214 and A215 as well as A204 (Figure 3.6) show that the κ - κ' interaction is not present. Not surprisingly strong DMS modifications in A198 (Figure 3.6) and A382 (Figure 3.7) as well as A830-A832 (Figure 3.14) found in the D5 tetraloop are present revealing that the ζ - ζ' tertiary contact is not established. These findings show that D5 is not docked into D1, an event which is triggered by D1 folding⁷⁰ and absolutely crucial for catalysis.

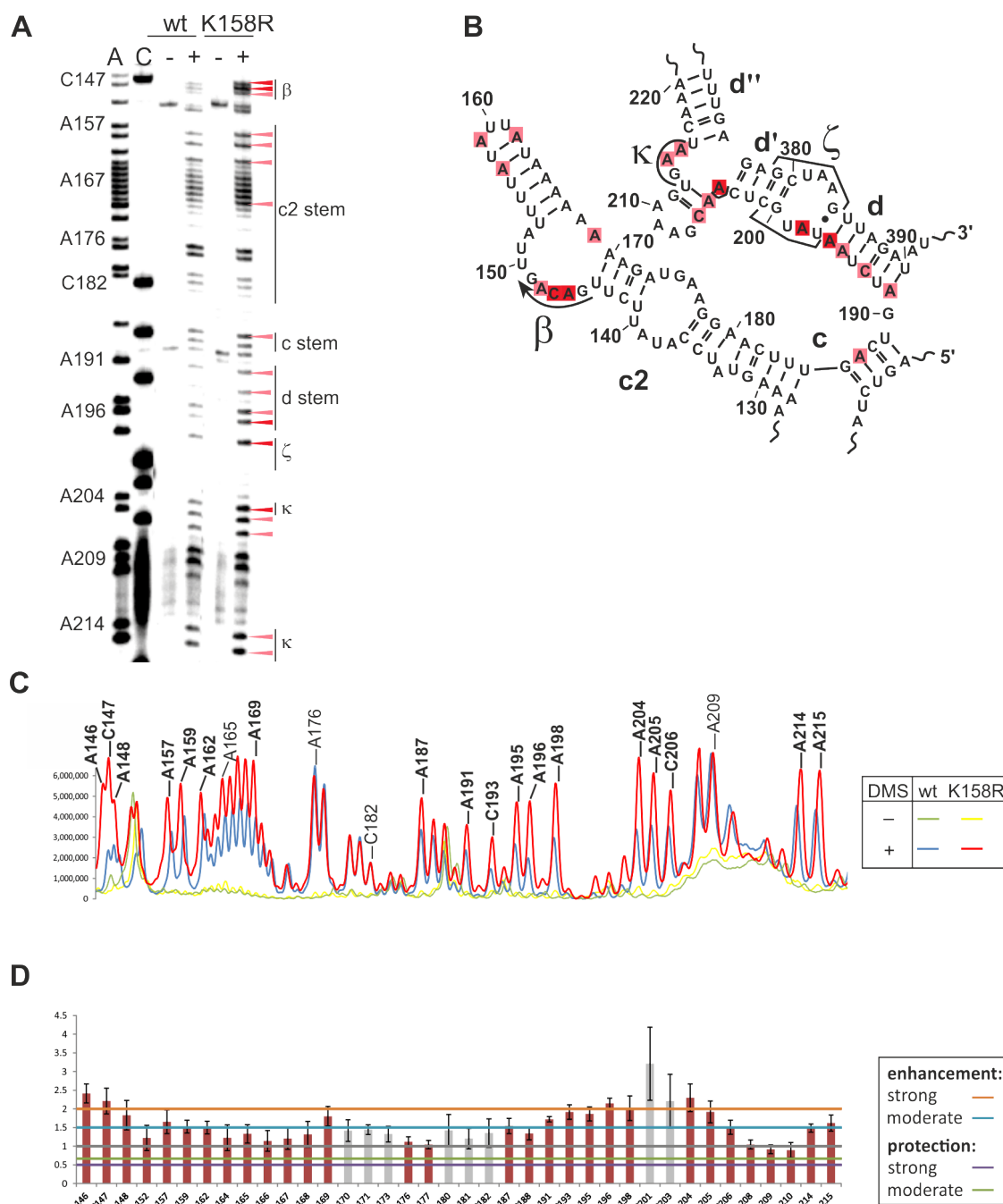


Figure 3.6: DMS accessibility of the 5' part of the κ - ζ element in the K158R strain.

(A) Representative gel covering the 3' part of the c2 stem as well as the 5' part of the c and d stem and tertiary interaction motifs β , ζ and κ . A and C lanes reveal adenosines and cytosines. - lanes indicate natural stops of the reverse transcriptase. In the + lanes the *in vivo* DMS pattern is shown. Comparing the - lane with the + lane of the respective strain reveals DMS induced stops of the reverse transcriptase and therefore accessible As and Cs.

Enhancements in and protections from DMS modification are revealed by comparing intensities of respective bands in + lanes. Filled arrowheads show strong while pink ones reflect minor enhancements. Open red arrowheads show strong and pink ones reflect minor protections. (B) Secondary structure map of the κ - ζ element, stem c and c2 with indicated enhancements and protections. Filled squares show enhancements and open squares protections. Color code as in A. Black dots indicate breaks. (C) Lane profile of the representative gel. Y axis shows the counts at a given position on the gel. (D) Average of counts of mutant to wild-type strain of a concerning nucleotide over at least three different experiments with standard deviation. The respect cutoffs used for evaluation are indicated with colored lines. The cutoff for strong changes in DMS-accessability was set at 2-fold and the cutoff for minor changes at 1.5 fold. 1 fold reflects an equal modification intensity in the different genetic backgrounds. Light gray bars refer to nucleotides that are neither modified in wild type nor mutant strain. Dark gray bars indicate natural stops of the reverse transcriptase.

Secondary structure formation partly depends on the ATPase activity of Mss116p

Secondary structures form *in vitro* in the presence of monovalent ions that shield the negative charge of the RNA backbone²⁸. Nevertheless several helices of D1 are not established in the presence of K⁺ ions alone *in vitro* but do only form after Mg²⁺ is added⁷⁸. The very same helices are also not formed *in vivo*, if Mss116p is not present to assist in ai5 γ folding¹⁰⁹. The DMS modification pattern in the yeast strain that harbors the K158R mutation in Mss116p highly correlates with these previous results. Stems d, d' and d'', which are not formed in the KCl state of the D135 ribozyme, are not formed *in vivo* without Mss116p as reflected by the enhancements in A191, A193, A195-A196, A205 and C206 (Figure 3.6), C217-A220 (Figure 3.8) as well as A378, A387 and A389 (Figure 3.7). Additionally, enhancements of the nucleotides A233, C234 and A365-366 (Figure 3.7) indicate the d''' stem as being unstable. A50-A52 are accessible to DMS (Figure 3.9) showing that stem a is not formed properly. Enhancements were also observed at A66-A67 and A69 that are involved in the formation of stem b, but

modification intensities in wild type yeast cells were elevated too, thus reducing the observable effect. A187, which is found in stem c, being part of the five-way junction of D1, shows enhanced DMS accessibility (Figure 3.6).

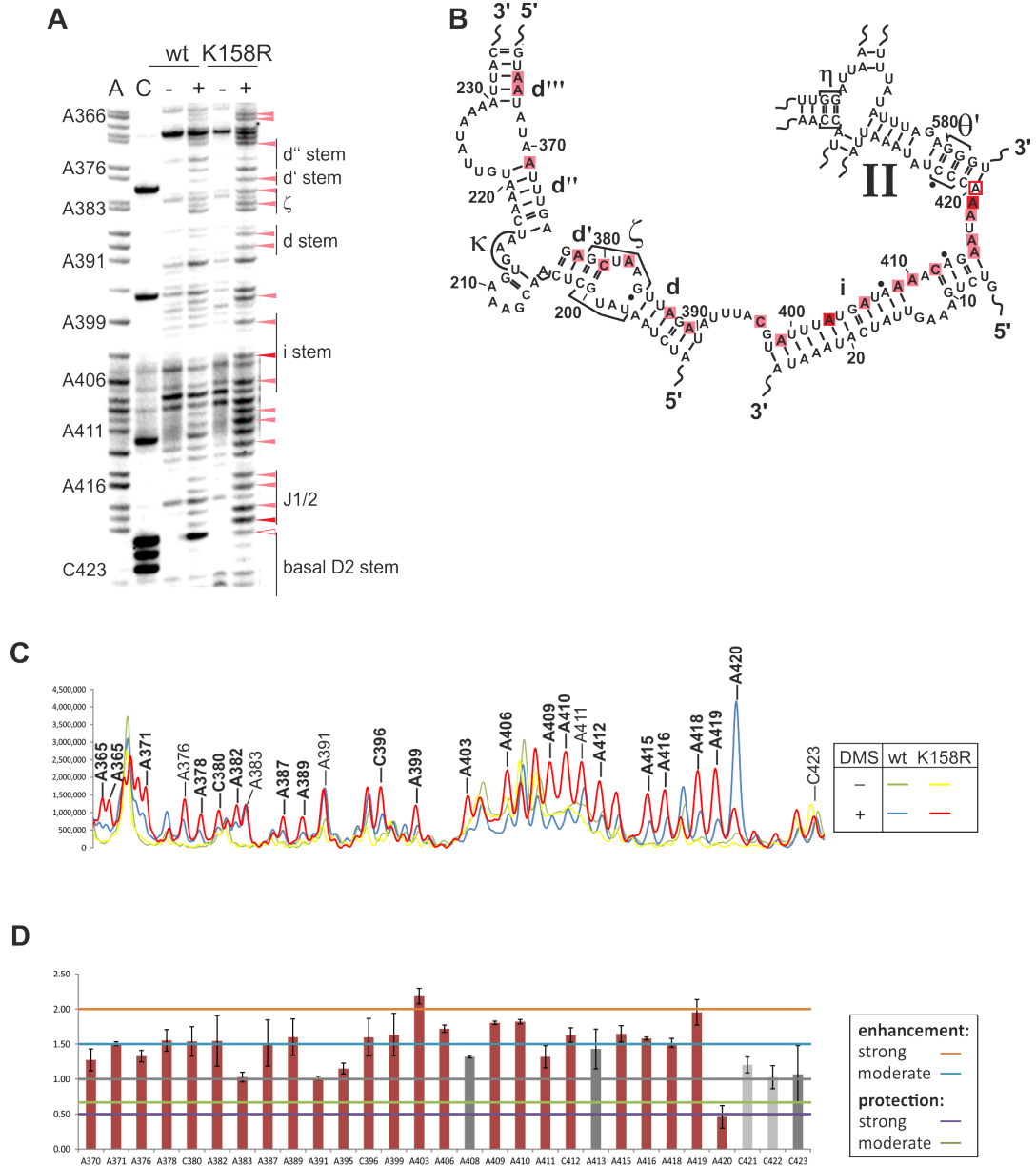


Figure 3.7: DMS accessibility around the 3' part of the κ - ζ element in the K158R strain. (A) Representative gel covering the 3' part of the κ - ζ element and the *i* stem as well as J1/2 and the basal D2 stem. (B) (C) (D) Symbol and color code as in Figure 3.6

Interestingly, also stem i, which links D1 to the central wheel of group II introns and harbours an important asymmetric internal loop, is disrupted in the knockout strain¹⁰⁹ although this stem is formed *in vitro* with only monovalent ions present⁷⁸. With the assistance of an ATPase-deficient cofactor A20, A22-A24 and A26 (Figure 3.9) are more accessible to DMS as well as C412, A406, A403 and A399 (Figure 3.7) showing that this stem is also not formed in the K158R strain.

Furthermore, stems in D3, which were shown to be formed in the KCl state *in vitro*⁷⁸, were disrupted in the knockout strain¹⁰⁹. Enhancements at the nucleotides A594 and A596-A597 (Figure 3.12) show that the basal stem of D3 is not properly folded in the ATPase-deficient strain. Also, stem b and to a minor degree stem c of D3 are unstable, as indicated by a higher reactivity to DMS of the nucleotides A632, A633, C642 and C656, respectively. Hence most interestingly ai5 γ with its cofactor rendered unable to bind and hydrolyze ATP shows more disruptions of helical regions *in vivo* than the D135 ribozyme *in vitro* without divalent ions. This might be due to the very high levels of K⁺ ions (f.c. 0.5 M) used in the mentioned study, but might also hint to a more broader and complex function of Mss116p in ai5 γ folding.

Nevertheless, our results also show that the DMS accessibility of nucleotides in stems c1, c2, d2a, d2b and d3 of D1 as well as stems in D2, D4 and D5 is not altered in the K158R genetic background (Figure 3.6, Figure 3.8, Figure 3.13, Figure 3.14, data for D2 not shown). The very same result could be seen in the knockout strain, indicating that Mss116p is not necessary for the establishment of the mentioned secondary structure elements⁶¹. Intriguingly, the disrupted stems are mostly found around the κ - ζ element and the very important five-way junction in D1. Thus, our results show that the ATPase activity of Mss116p plays a fundamental role in secondary structure formation around these two regions that are crucial for the structural organization of D1 and therefore for the collapse of the entire intron^{31,77,78}.

As one proposed mode of action of Mss116p is to disrupt misfolded regions of the intron and therefore allowing native structure formation^{189,190}, we also carefully looked for nucleotides that show less accessibility to DMS in the K158R strain than in the wild type genetic background. Due to the fact that we almost exclusively observed DMS

enhancements at helical regions, the existence of stable kinetically trapped intermediates is most unlikely. The scarce protections that can be seen in helical regions are very strong though and are confined to A128 and A420 found at the very end of the concerning stem (Figure 3.7, Figure 3.11). The respective stems are not found in group IIC introns, thus there is no high resolution structure available. An explanation for these sites of protection could be found in the establishment of a non-canonical reverse Hoogsteen A·U base pair¹⁵ at these sites in the native structure of ai5γ, which would result in a highly accessible N1 of the respective adenosine. Additionally, there is a strong enhancement at A671 (Figure 3.13) found in the single-stranded J3/4 region.

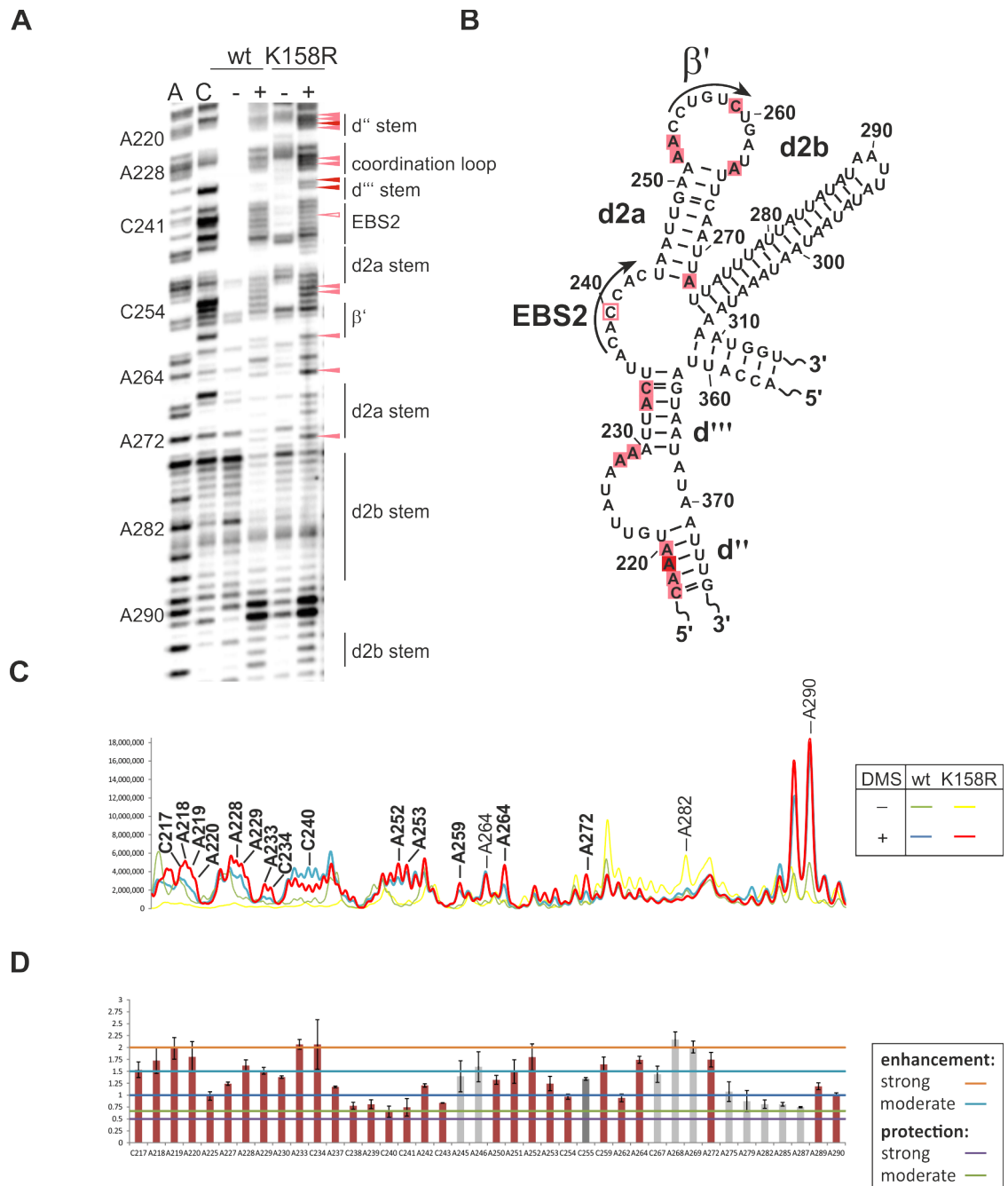


Figure 3.8: DMS accessibility around EBS2 and coordination loop in the K158R strain. (A) Representative gel covering the coordination loop, EBS2, the 5' part of stems d'' and d''' and b. (B) (C) (D) Symbol and color code as in Figure 3.6

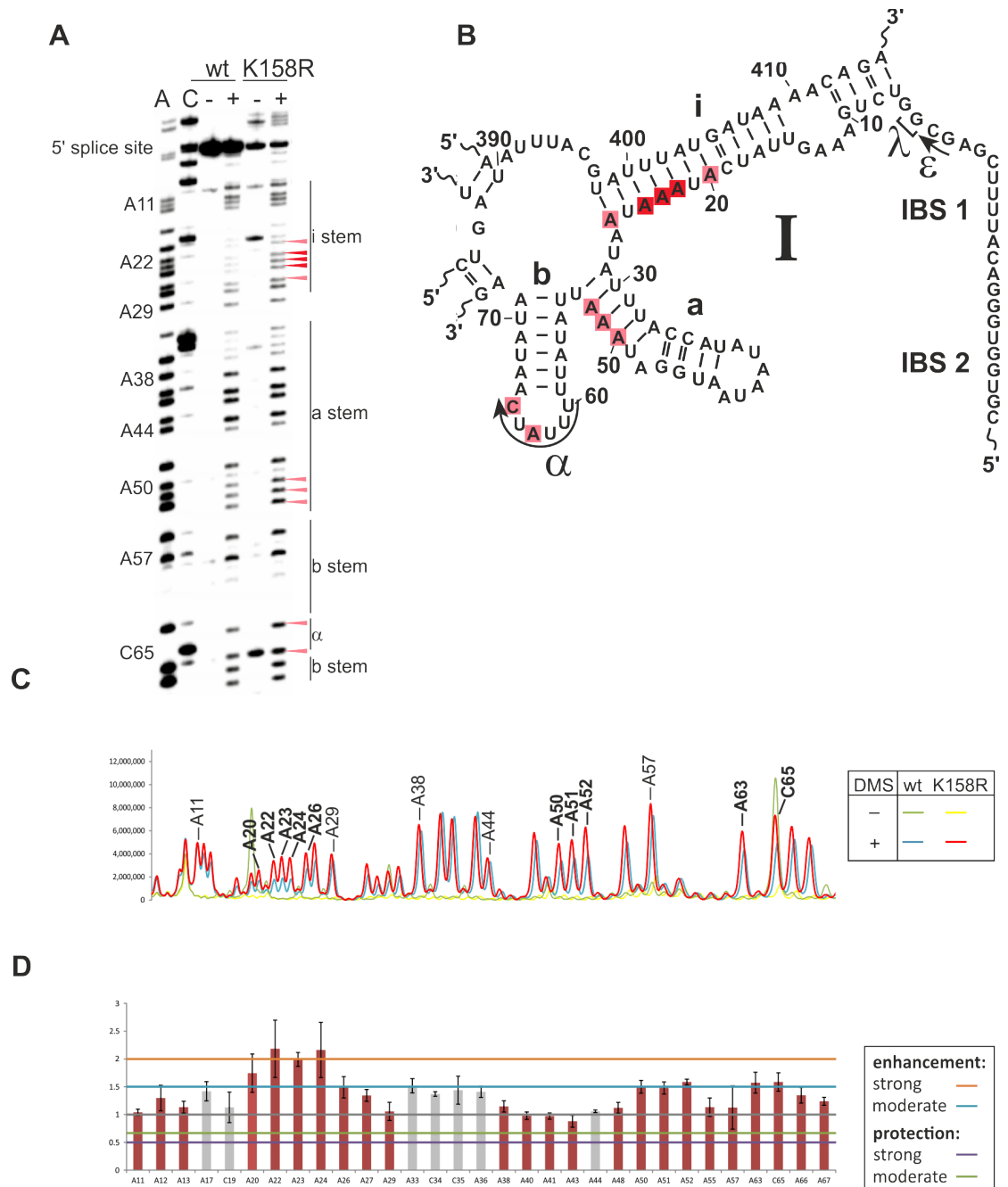


Figure 3.9: DMS accessibility of the very 5' end of *ai5γ* in the K158R strain. (A) Representative gel covering the 5' part of stem *i*, stem *a* and *b* as well as the α element. (B) (C) (D) symbol and color code as in Figure 3.6.

Formation of the active site depends on the ATPase activity of Mss116p

Group II introns harbor several long-range tertiary interactions that are fundamentally important for compaction and for establishing of the active site^{38,64,77,79}. Most of these tertiary interactions are formed either with or within D1, whose secondary structure formation is partly dependent on the ATPase activity of Mss116p (Figure 3.6, Figure 3.7, Figure 3.9). In the absence of Mss116p the tertiary interactions found are not established in ai5 γ , showing that this DEAD-box protein is important for both secondary as well as tertiary structure formation of ai5 γ ¹⁰⁹. We found that this is also mostly true for its ATPase-deficient variant of Mss116p. As stated in the previous chapter both D1 intradomain interactions α - α' and β - β' , which are important for D1 collapse, are not in place in the K158R strain (Figure 3.6, Figure 3.8). The θ - θ' interaction that connects the closing tetraloop of stem c1 with the basal stem of D2 seems to be affected only in a minor way displayed by the small differences in DMS accessibility of A100-A101 (Figure 3.11). This implies that in some molecules this tertiary interaction, which enables helical stacking of regions of D1 and D3 and in turn building a shell around the active site elements⁷⁵, is formed.

Formation of the κ - ζ element is not only important for the intron's collapse but also for establishing its active site. In a wild type yeast strain A382-383 show a distinct modification pattern (Figure 3.7), in which A382 becomes protected, whereas A383 still remains accessible to DMS, since they form an A·A platform as part of the ζ tetraloop receptor^{17,31,61}. In the K158R strain this conformation is not established, since both nucleotides show moderate modification intensities. Additionally, A198, which is protected in the native conformation of the intron, as it is involved in a reverse Hoogsteen-pair with U381³¹, is modified intensely (Figure 3.6). The ζ - ζ' interaction is crucial for docking of D5²⁵ and, if established, A830-A832 of D5 are not accessible to DMS. However, these residues are in fact methylated in the ATPase-deficient strain, indicating that the tetraloop ζ' does not contact its receptor motif ζ (Figure 3.14). A214-A215 as well as A204 are part of the κ - κ' tertiary interaction, which is the second crucial contact for docking of D5⁶⁷.

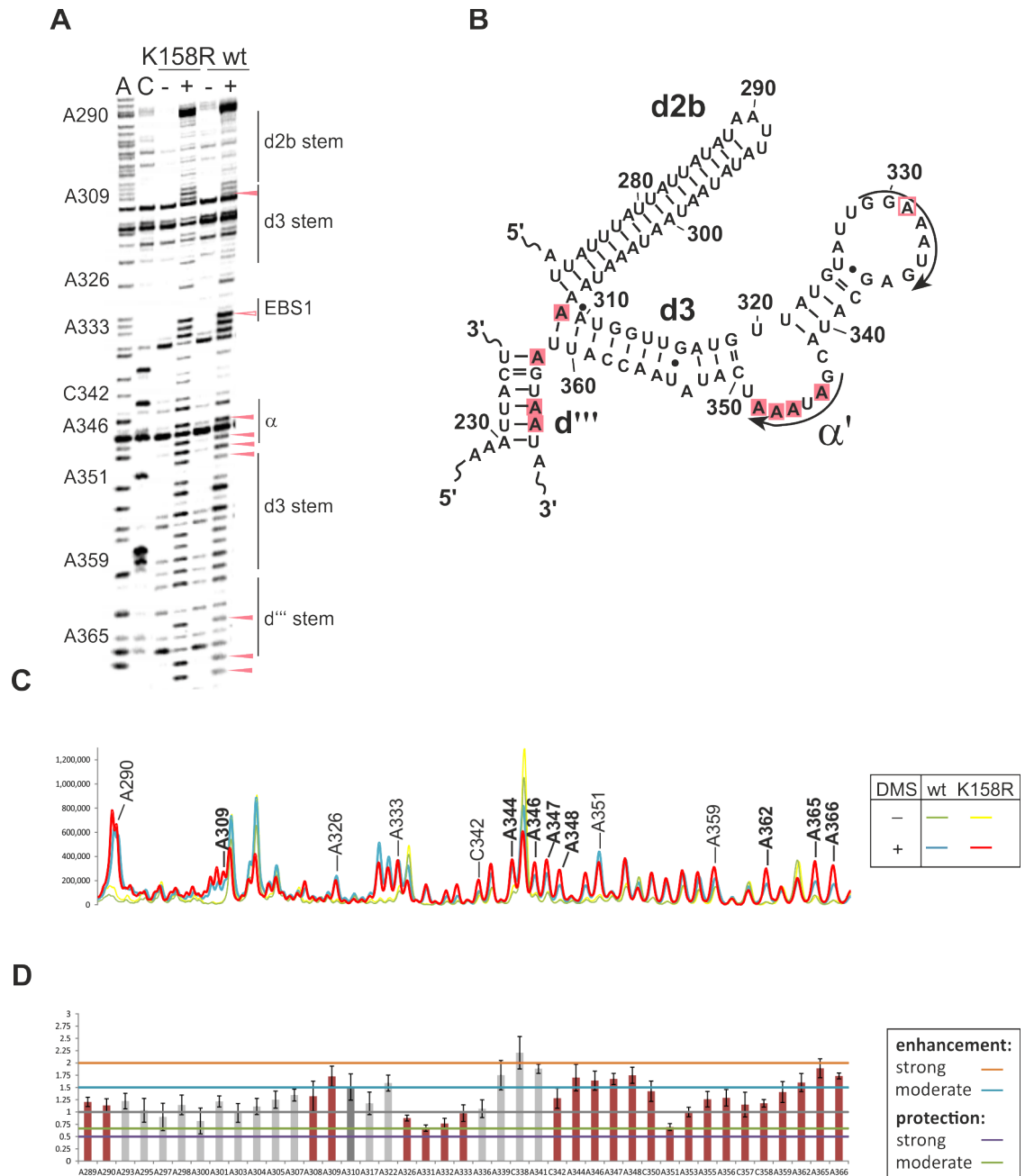


Figure 3.10: DMS accessibility around stem d3 in the K158R strain. (A) Representative gel covering the stems d2b, d3 and d''' as well as EBS1 and α' element. (B) (C) (D) Symbol and color code as in Figure 3.6.

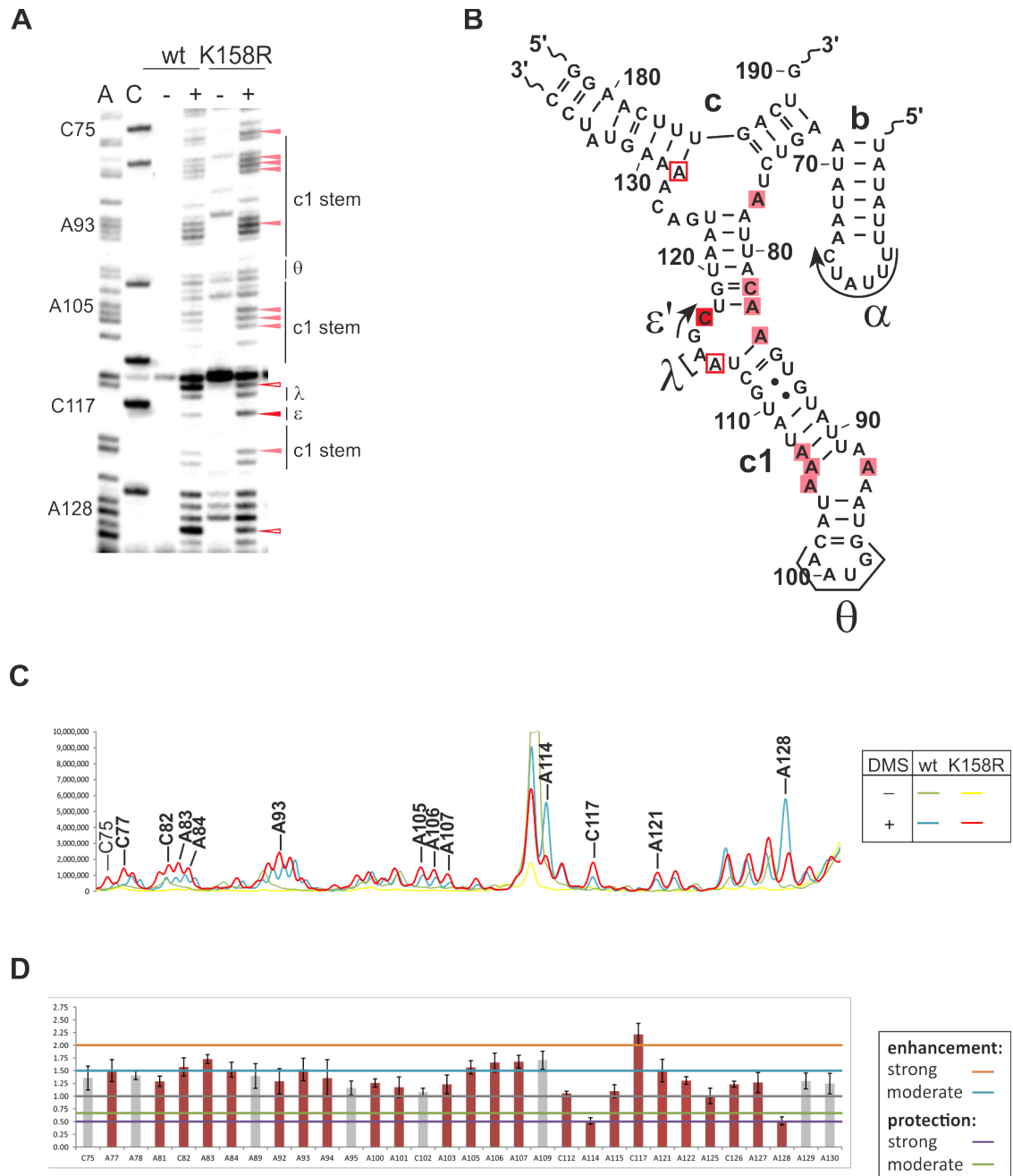


Figure 3.11: DMS accessibility around stem *c1* in the K158R strain. (A) Representative gel covering the *c1* stem, λ and ϵ element. (B) (C) (D) Symbol and colore code as in Figure 3.6.

These adenosines are modified intensely in the K158R strain, indicating that the κ - κ' interaction does not form. Thus, D5 is not docked onto D1, an event that lays the foundation for active site formation⁴³. In the active site the so-called z-anchor region is

then established³¹, in which λ - λ' and ε - ε' are close to the dinucleotide bulge of D5, which gets twisted creating a high negative potential at this site. A115, which is the only nucleotide of the λ - λ' interaction that can be mapped with DMS, is forming a base triple with the C825-G836 pair^{31,69}. In this conformation the A115-N1 is not part of the hydrogen bond network, but is weakly accessible to DMS (Figure 3.11). The adjacent A114 is pushed out to the solvent in the native conformation of this region. Though the accessibility of A115 is enhanced in the K158R strain, the difference is below our threshold, suggesting that the base triple is formed in some of the intron molecules (Figure 3.11). In contrast the neighboring A114 becomes strongly protected in the ATPase-deficient strain, suggesting that the λ - λ' tertiary contact is not in place. C117 is part of the ε - ε' interaction, forming a Watson-Crick base pair with G3 and positioning the 5' exon close to the active site center in the native structure of ai5 γ . In the K158R strain C117 is significantly modified by DMS, showing that this active site constituent is not formed. J2/3 forms a triple helix with the major groove of the lower stem in D5^{31,65}. The overall strongly increased DMS accessibility of J2/3 and especially of A589 (Figure 3.12) indicates that this triple helix is not formed, if Mss116p is not able to bind and hydrolyze ATP. In more detail, enhancement of A589 shows that the A-A non-canonical base pair with A816, which is important for the formation of the triple helix⁴³ is not established. A586 is enhanced in the K158R strain. In the native conformation of ai5 γ it should be protected, since it is contacting the backbone of G6 positioning the 5' end of the intron close to the dinucleotide bulge⁶⁹. Enhancement of A587 on the other hand shows that the γ - γ' tertiary interaction with U887 is not formed. The intradomain non-canonical contact between C839 and C818 seems to be in place at least in some of the molecules, since C839 is only weakly modified by DMS (Figure 3.14). The strong modification of A838 in both the wild type and the K158R strain suggests that the conformation of the dinucleotide bulge^{31,69} is established in both genetic backgrounds, although in a more unstable way in the ATPase-deficient strain, which could result from the lack of the stabilizing triple helix. That the intradomain contact between C839 and C818 seems to be partly formed correlates with the findings that secondary structure formation of D5 is Mss116p independent⁶¹.

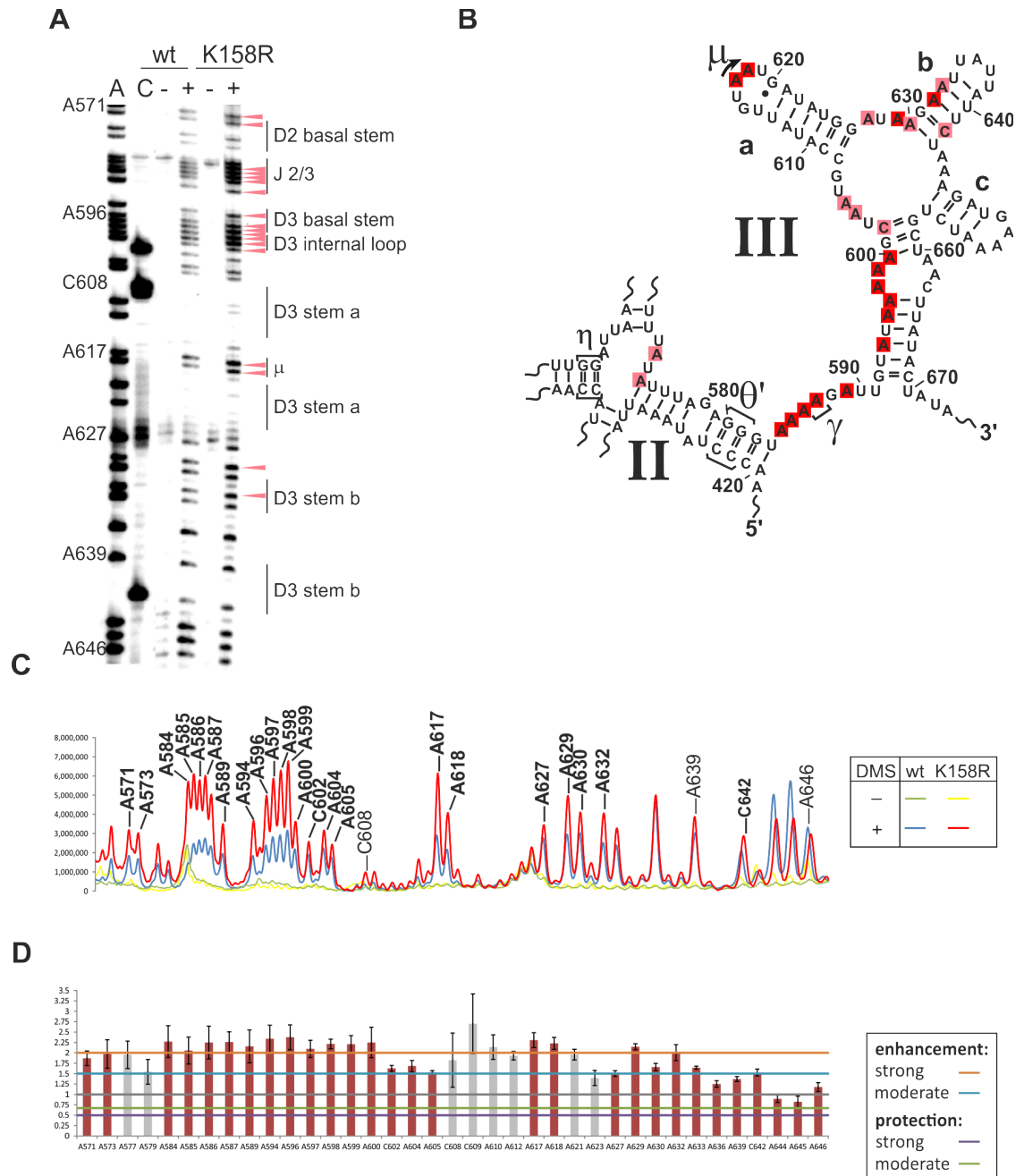


Figure 3.12: DMS accessibility of J2/3 and D3 in the K158R strain. (A) Representative gel covering the 3' part of the D2 basal stem, J2/3, D3 basal stem and internal loop, stems a and b and the μ element. (B) (C) (D) Symbol and color code as in Figure 3.6.

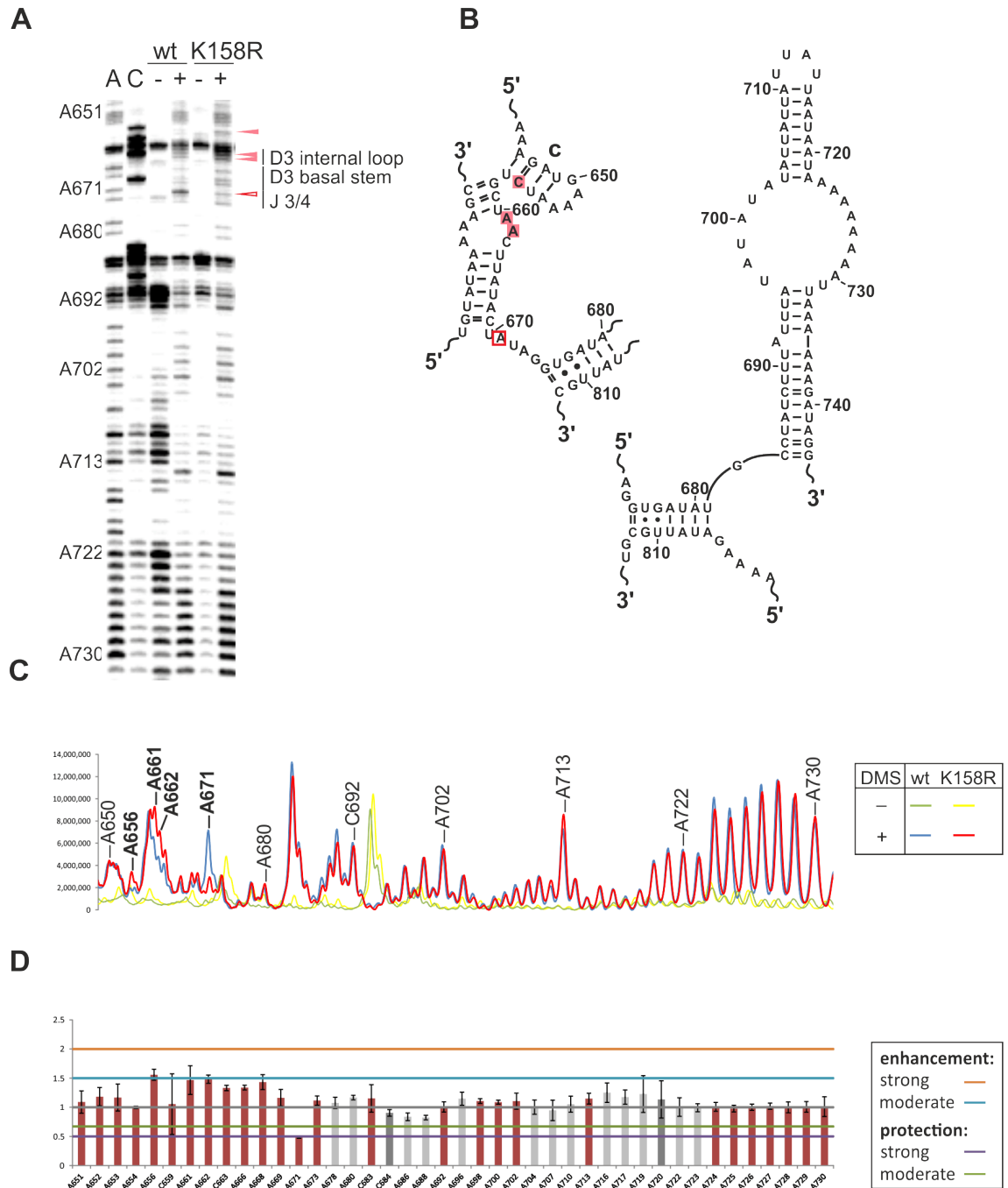


Figure 3.13: DMS accessibility of D4 in the K158R strain. (A) Representative gel covering the 3' part of the D3 basal stem and loop, J3/4 and part of D4. (B) (C) (D) Symbol and color code as in Figure 3.6.

Apart from D1, changes in DMS accessibility of secondary structure elements ai5 γ were only found in the effector D3⁶¹. Although in general dispensable for the catalytic activity of the ribozyme *in vitro*⁷¹, D3 enhances the splicing activity tremendously⁶⁵. Additionally, it contributes to the stability of the active site⁸⁵. The μ - μ' tertiary interaction, which forms between A617-N3, A618-N3 and G844-2'OH is very important for the orientation of D3 and D5 to each other⁶⁶. Enhancements in both A617 and A618 show that these contacts between D3 and D5 are not established in the K158R strain (Figure 3.12). The internal loop of D3 is highly conserved amongst group IIB introns. Two consecutive sheared A·A base-pairs were shown to play an essential role in enhancing the rate of catalysis of ai5 γ *in vitro*⁶⁶. In this A-rich asymmetric internal loop the Hoogsteen edge of A599 and A662 establish a sheared A·A base pair with the minor groove edge of A661 and A598, respectively. In this conformation A418 stacks on C662 and its ribose interacts with the A598 ribose⁶⁹, while C662 contacts A416 of the J1/2 region^{31,69}. Additionally, A418 and A419 form non-canonical base pairs with A584 and A585 respectively, and A420 with U583 a Watson-Crick pair *in vitro*, establishing a complex structure that involves J1/2, J2/3 and the internal asymmetric loop in D3. Enhancements at A416, A418, A419, A584 and A585 indicate that these contacts are not formed in the K158R strain (Figure 3.7, Figure 3.12). Furthermore, A420 becomes protected in the ATPase-deficient strain showing that the Watson-Crick base pair with U583 is not formed in the wild type strain, a fact that hints to a reverse Hoogsteen conformation of those two nucleotides⁶¹. The nucleotides of the internal loop of D3 show similar accessibility to DMS in the wild type strain, indicating that the conformation of this loop slightly differs at least for the spliced intron from the *in vitro* situation³¹. Nevertheless enhancements at A594, A596-A600 as well as in A661-A662 show that the basal stem as well as the internal loop are open compared to the wild type (Figure 3.12, Figure 3.13). Interestingly, the basal stem is the only one affected by the K158R mutation aside from D1 helical regions. The complex structure which is build by the internal loop, J1/2 and J2/3 seems to highlight the importance of this region *in vivo*, especially since the J2/3 region again contacts D5 and is of utmost importance for the active site formation of the intron.

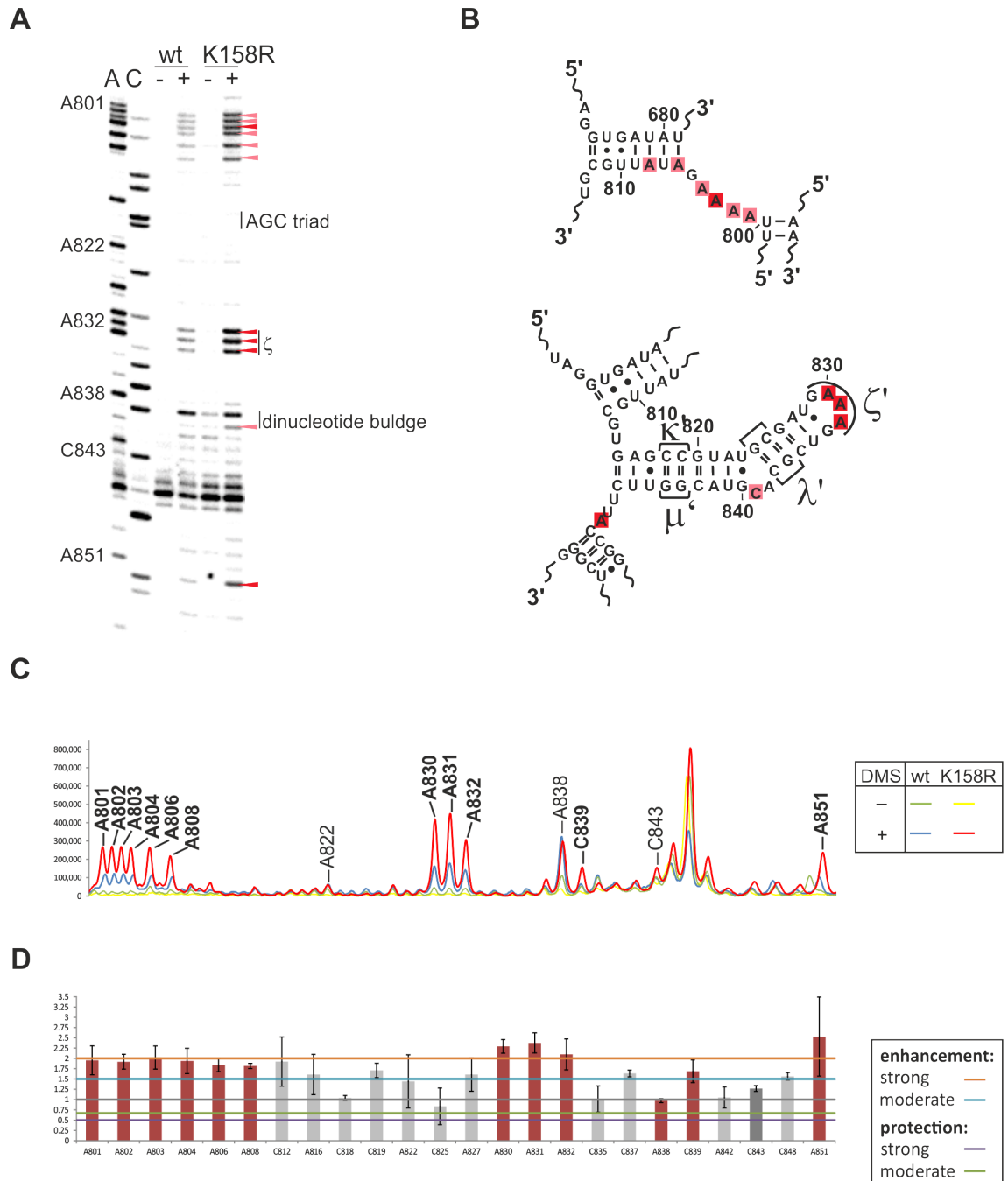


Figure 3.14: DMS accessibility of D5 in the K158R strain. (A) Representative gel covering the 3' part of the D4 basal stem and D5 with its AGC triad, ζ' tetraloop and the dinucleotide bulge. (B) (C) (D) Symbol and color code as in Figure 3.6.

The exon binding sites EBS1 and EBS2 are modified in the *ATPase-deficient* strain, but in a lesser degree than seen with the wild type genetic background (Figure 3.8, Figure

3.10). The protections in these regions, namely of the nucleotides C240 and C241 in EBS2 indicate that the geometry of this region in the 4-way-junction is affected by an ATPase-deficiency of Mss116p. In the wild type strain the EBS2 region gets pushed out to the solvent in order to bind IBS2. This important conformation is not established in the K158R strain. Although all of the nucleotides involved in tertiary interactions are more accessible to DMS in the ATPase-deficient strain, some like ζ - ζ' and κ - κ' seem to be far more effected than others as λ - λ' and θ - θ' . Intriguingly, tertiary interactions that are formed between different domains are particularly affected by the K158R mutation of Mss116p. Especially nucleotides in the regions, which are involved in active site formation, are generally more open and thus accessible to DMS, as seen before in the knockout strain¹⁰⁹. *In vitro* the ATPase activity of Mss116p was shown to be important for protein turnover but not for the collapse of ai5 γ ^{169,195,201}. The expected structure of ai5 γ , assisted in its folding by an ATPase-deficient strain, would resemble a near-native state, in which most of the folding process has taken place and only the last steps for native structure formation are missing¹⁹⁵. The *in vivo* DMS pattern of the K158R strain does not reflect this situation. Possibly, limiting amounts of Mss116p due to abolished protein turn-over lead to this knock-out like situation. If *in vivo* the protein to RNA ratio is very low, all the cofactors are trapped rendering the effects of ATPase independent folding invisible to chemical probing. Additionally, Mss116p binds RNA in an unspecific way and hence even at a higher protein to RNA ratio *in vivo* Mss116p molecules might get trapped by nonproductive binding events¹⁹⁵.

Mss116p cannot assist in ai5 γ folding without ATP binding and hydrolysis

In a genetic background, in which Mss116p can neither bind nor hydrolyze ATP, the group II intron ai5 γ is largely unfolded *in vivo* (Figure 3.15). Its structure is affected at both the secondary and the tertiary level. Several stems in D1, namely the stems i, a, c, d, d', d'', d''' and the basal stem of D3 are severely disrupted. Interestingly, most enhanced nucleotides that are part of helices are clustered around the κ - ζ element, whose formation was shown to be crucial for the collapse of the intron *in vitro*⁷⁸. Modifications of the nucleotides involved in the intradomain tertiary interactions α - α' and β - β' are further evidence for an extended D1. Enhancements in ζ - ζ' and κ - κ'

indicate that D5 is not docked onto D1, as a consequence the active site is not established, as revealed by the modification pattern of the z-anchor, the triple helical scaffold and the hydrogen bonding network of the D3 internal loop. The DMS modification pattern highly resembles the one found in the knockout strain, highlighting the importance of ATP binding and hydrolyzes for the function of Mss116p (Figure 3.1).

3.4.4 The basic tail of Mss116p and ai5 γ folding

Some DEAD-box protein, namely Mss116p, Ded1p and Cyt19, show an abundance of basic amino acids in their C-terminal domain^{151,152}. Interestingly, Ded1p and Cyt19 are also able to rescue ai5 γ folding *in vitro*^{183,187}. Cyt19 and Mss116p additionally show an α -helical region in their C-terminal domain preceding the basic tail. In the crystal structure of a truncated version of Mss116p, in which the so-called C-terminal extension (CTE) is present, but the basic tail was removed, was crystallized. The x-ray structure revealed that the CTE contacts the RNA substrate and introduces a second bend in the RNA, which was proposed to be essential for RNA strand separation¹⁴⁶. This second bend is caused by the alpha-helical regions of the CTE that pack against the RNA substrate.

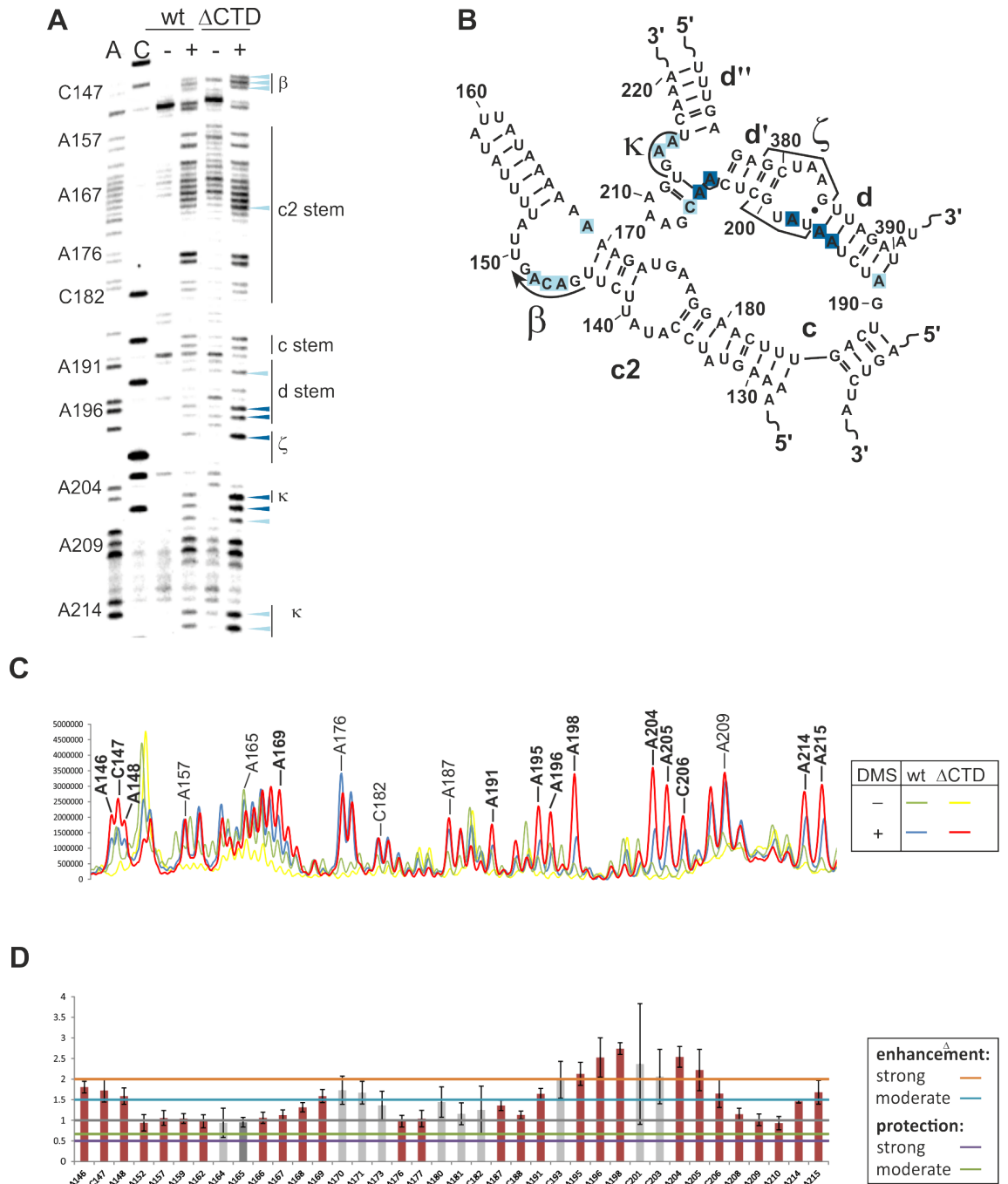


Figure 3.16: DMS accessibility of the 5' part of the κ - ζ element in the Δ CTD strain. (A) Representative gel covering the 3' part of the c2 stem as well as the 5' part of the c and d stems and tertiary interaction motifs β , ζ and κ . Dark blue color refers to strong, light blue color to moderate differences in DMS-accessability. This is also true for figures 3.17-3.24 concerning the DCTD strain. (B), (C) and (D) Symbol and color code as in Figure 3.6.

It was shown that Mss116p/ Δ 551-664, in which most of the α -helix 2 as well as the basic tail are deleted loses its ability to grow by respiration as well as to assist in ai5 γ splicing *in vitro*¹⁵². In contrast, truncation of amino acids 569-664 removes the basic tail but in this mutant strain ai5 γ splices at wild type level and is able to grow aerobically¹⁵². In our hands, however, the Mss116p/ Δ 569-664 variant does not support ai5 γ splicing *in vivo* and the respective strain shows only limited growth on non-fermentable carbon sources. While we integrated the truncated Mss116p variant into its genomic locus on chromosome IV, substituting the wt Mss116 gene, Lambowitz and coworkers expressed the truncated version of Mss116p from a plasmid¹⁵². Since the distinct basic tail of Mss116p, Ded1p and Cyt19 is likely to be of importance for ai5 γ folding, the intron structure was analyzed in the mss116/ Δ 569-664 (Δ CTD) strain.

The basic tail of Mss116p plays an important role in the formation of the κ - ζ element

Similarly to the K158R and the knockout strain most of the strongest DMS modifications are found at the κ - ζ folding control element. Enhancements in A218-A220 show that stem d'' is not formed in the Δ CTD strain (Figure 3.18). Modifications of A205 and C206 (Figure 3.16) indicate that this helix is not present as well and neither is stem d shown by enhancements of A191 and A195-196 as well as A387 and A389 respectively (Figure 3.17). A63, A65 (Figure 3.19) and A344 and A346-A348 (Figure 3.20) are more accessible to DMS than in the wild type strain, showing that the D1 tertiary interaction α - α' is not formed in the Δ CTD strain. Enhancements at A146, C147, A148 (Figure 3.16) indicate that the same is true for the β - β' contact, the second intradomain interaction that is important for compaction of D1. Notably, C255 and C259 are modified as well, but as these are equally methylated in the wild type and Δ CTD strain, these nucleotides are not included in the differential map (Figure 3.18). These two contacts play an important role in the collapse of the intron^{38,55,72,77} and the enhanced DMS accessibilities at these regions indicate that the intron is not collapsed in the Δ CTD strain.

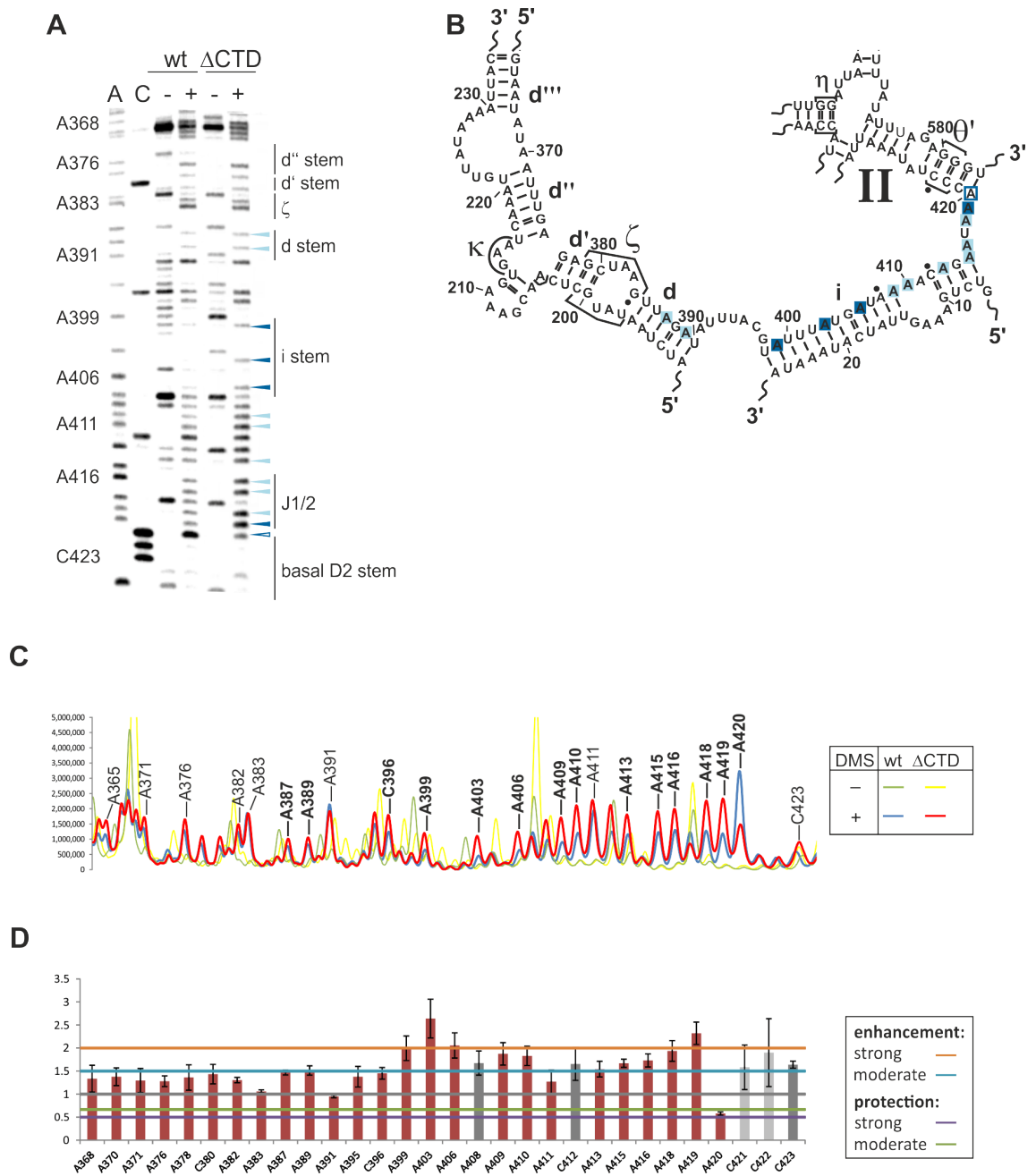


Figure 3.17: DMS accessibility around the 3' part of the κ - ζ element in the Δ CTD strain. (A) Representative gel covering the 3' part of the κ - ζ element and the i stem as well as J1/2 and the basal D2 stem. (B) (C) (D) Symbol and color code as in Figure 3.6.

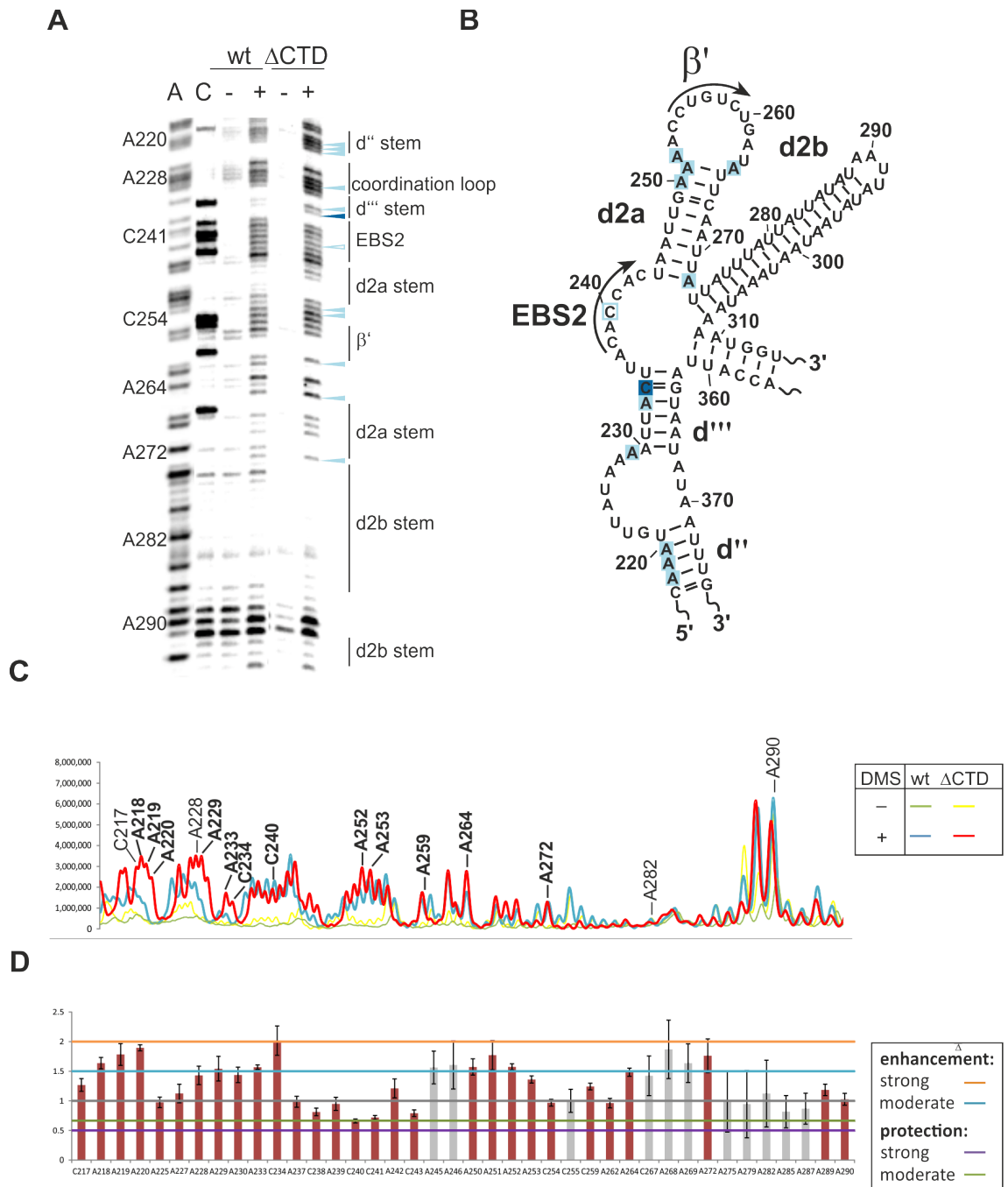


Figure 3.18: DMS accessibility around EBS2 and the coordination loop in the Δ CTD strain. (A) Representative gel covering the coordination loop, EBS2, the 5' part of stems d'' and d''' and β . (B) (C) (D) Symbol and color code as in Figure 3.6.

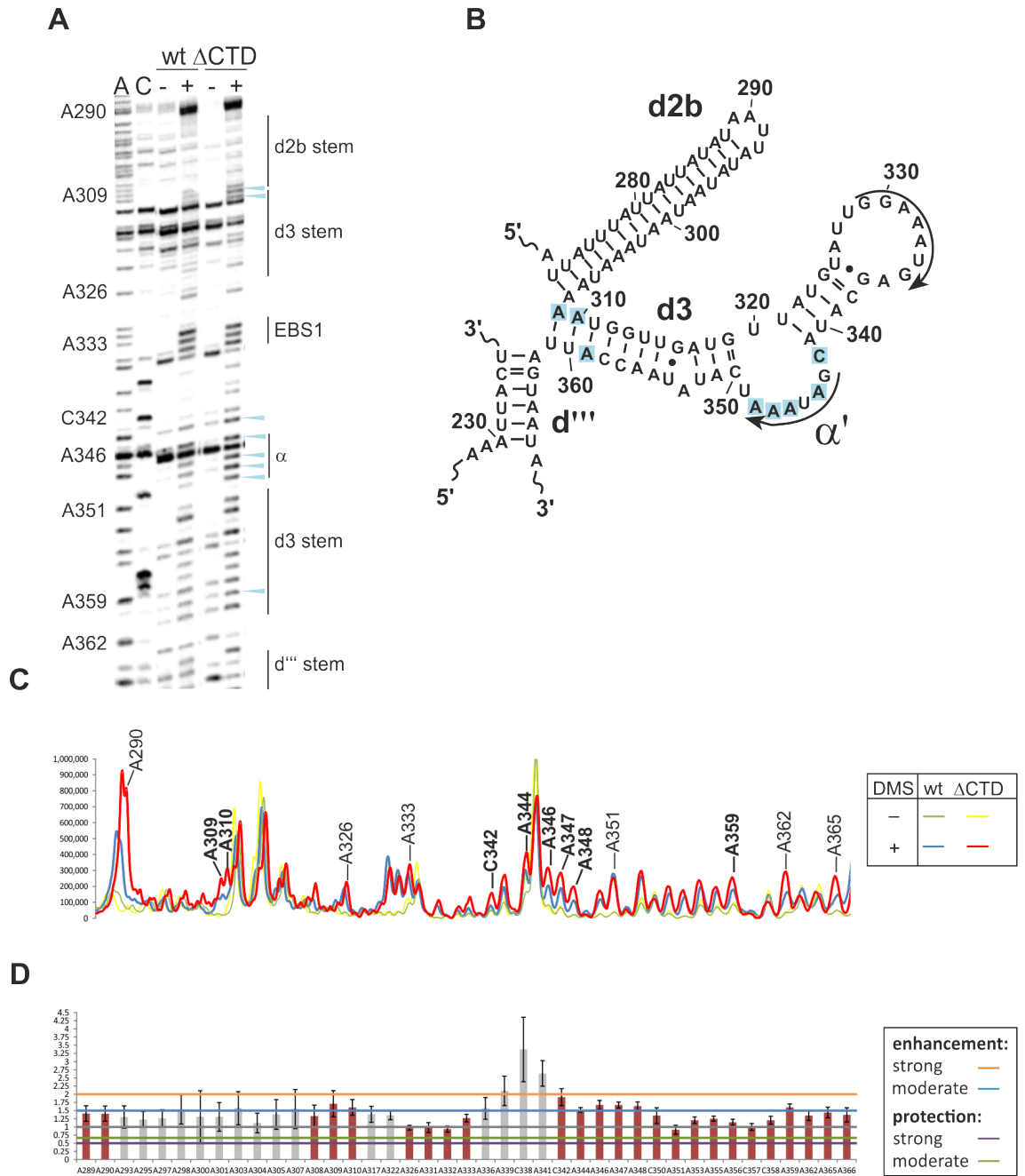


Figure 3.19: DMS accessibility of the very 5' end of *ai5γ* in the Δ CTD strain. (A) Representative gel covering the 5' part of stem *i*, stem *a* and stem *b* as well as α element. (B) (C) (D) Symbol and color code as in Figure 3.6.

Several secondary structure elements of *ai5γ* are not formed in the Δ CTD strain

Aside from the important helical regions that are part of the κ - ζ element several, other stems are affected by the deletion of the basic tail of Mss116p. Modifications of A22-A24 and A26 (Figure 3.19) as well as A399, A403, A406 and A413 (Figure 3.17) show that stem i that links D1 to the central wheel in group II introns is not formed in the Δ CTD strain. The respective enhancements are somewhat higher than in the ATPase-deficient strain. Also, stem a seems to be disrupted as indicated by enhancements at position A50-A52 (Figure 3.19). Nucleotides in stem b are open in both the wild type strain and the Δ CTD strain displaying similar DMS-accessibilities. Stem c that is part of the D1 five-way junction is more open as shown by enhancements of A187 although slightly below threshold and in a lesser degree as seen in the K158R mutant (Figure 3.16). The stem d''' adjacent to the coordination loop shows elevated DMS-accessibilities in its nucleotides A233-C234 (Figure 3.18) and A365-A366 (Figure 3.17). Overall all stems in D1 that were not folded properly in the ATPase-deficient as well as the Δ mss116 strain seem to be also affected in the Δ CTD strain. The basal stem of D3 although affected by the missing basic tail seems to be more stable than in the ATPase-deficient strain. Actually although there are enhancements in A666, A668 and C669 only the one seen in A594 and A596-A597 are above the threshold (Figure 3.20, Figure 3.24). Furthermore, modifications of A632-A633 hint to an instable formation of stem b in D3. Secondary structure elements found in D2, D4 and D5 seem to be unaffected by the deletion of the basic tail in Mss116p (Figure 3.24, Figure 3.22, D2 data not shown).

As in the K158R and the Δ mss116 strain protections were seen in A128 and A420 found at the base of the c2 and the D2 basal stems. A420 becomes strongly protected in both mutant strains whereas the effect on A128 is not as strong in the Δ CTD strain (Figure 3.21, Figure 3.17). The strong protection of A671 that was seen in the ATPase-deficient as well as the knockout strain is also found in the Δ CTD strain (Figure 3.24).

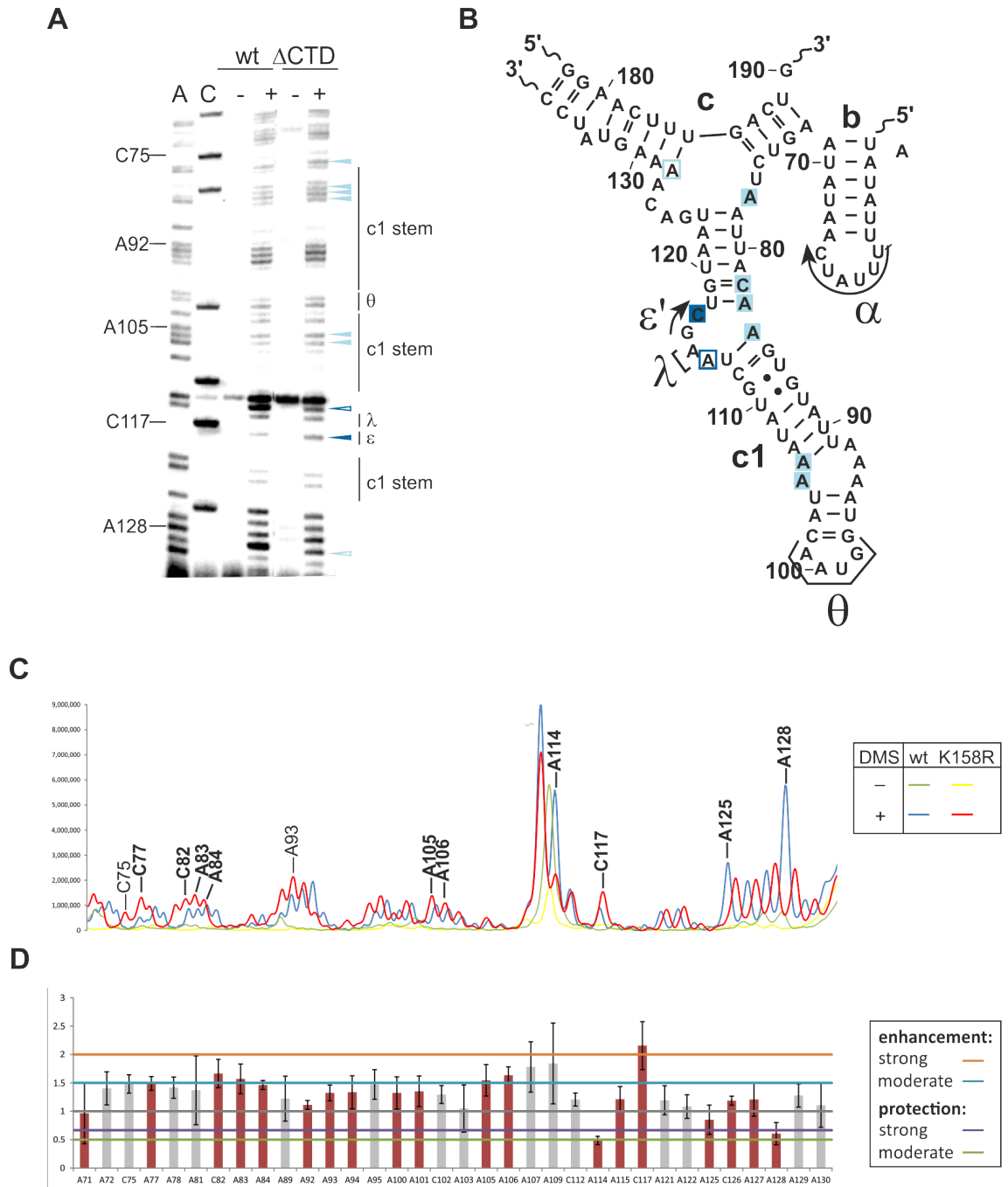


Figure 3.21: DMS accessibility around the *c1* stem in the Δ CTD strain. (A) Representative gel covering the *c1* stem, λ and ϵ . (B) (C) (D) symbol and color code as in Figure 3.6.

The C-terminal domain of Mss116p is important for formation of the active site within ai5 γ

The κ - κ' and ζ - ζ' interactions that are established by docking of D5 onto D1 are not in place, as indicated by enhancements of A214-215 and A204 (κ) as well as A198 (ζ) in D1 (Figure 3.16). Additionally, A830-A832 of the ζ' tetraloop motif are also strongly modified in the Δ CTD strain, confirming that D5 is not anchored in D1 (Figure 3.22). The enhancements of nucleotides involved in κ - κ' and ζ - ζ' are very similar to the ones found in the ATPase-deficient strain with the exception of A382 which is not more accessible to DMS in the Δ CTD strain (Figure 3.17). The θ - θ' interaction formed between the tetraloop at the end of stem c1 in D1 and the basal stem of D2²³ seems to be in place at least in some of the molecules as the modification intensity of A100-A101 is weak and resembles the one found in wild type cells (Figure 3.21). This region plays an important role in native structure formation, as it is involved in helical stacking with D1 and D3 forming a shell around the active site elements⁷⁵.

The active site constituents ε - ε' and λ - λ' are involved in a complex structure, the so called z-anchor³¹, close to the dinucleotide bulge of D5, which is important for the formation of the active site⁷⁵. A117 which is part of the z-anchor is more accessible to DMS in the Δ CTD strain, showing that the active site center is not established (Figure 3.21). λ - λ' does not show enhancement as the A115-N1 remains accessible upon forming the triple base interaction with C825-G836 base pair in D5. The adjacent A114, which is pushed out to the solvent in the native conformation of the z-anchor, becomes strongly protected in the Δ CTD strain. This complex structure, that brings the 5'exon close to the dinucleotide bulge, is not formed in the Δ CTD strain. The J2/3 nucleotides A586-587 and A589 show moderate enhancements (Figure 3.23). This indicates that only in a fraction of ai5 γ molecules the contact between A586 and the backbone of G5 that positions the 5' exon and the γ - γ' tertiary interaction between A587 and U887 are formed. Furthermore, the base triple involving A589 and A816 is not established, showing that the triple helix between J2/3 and the major groove of the lower D5 stem is not formed in the Δ CTD strain.

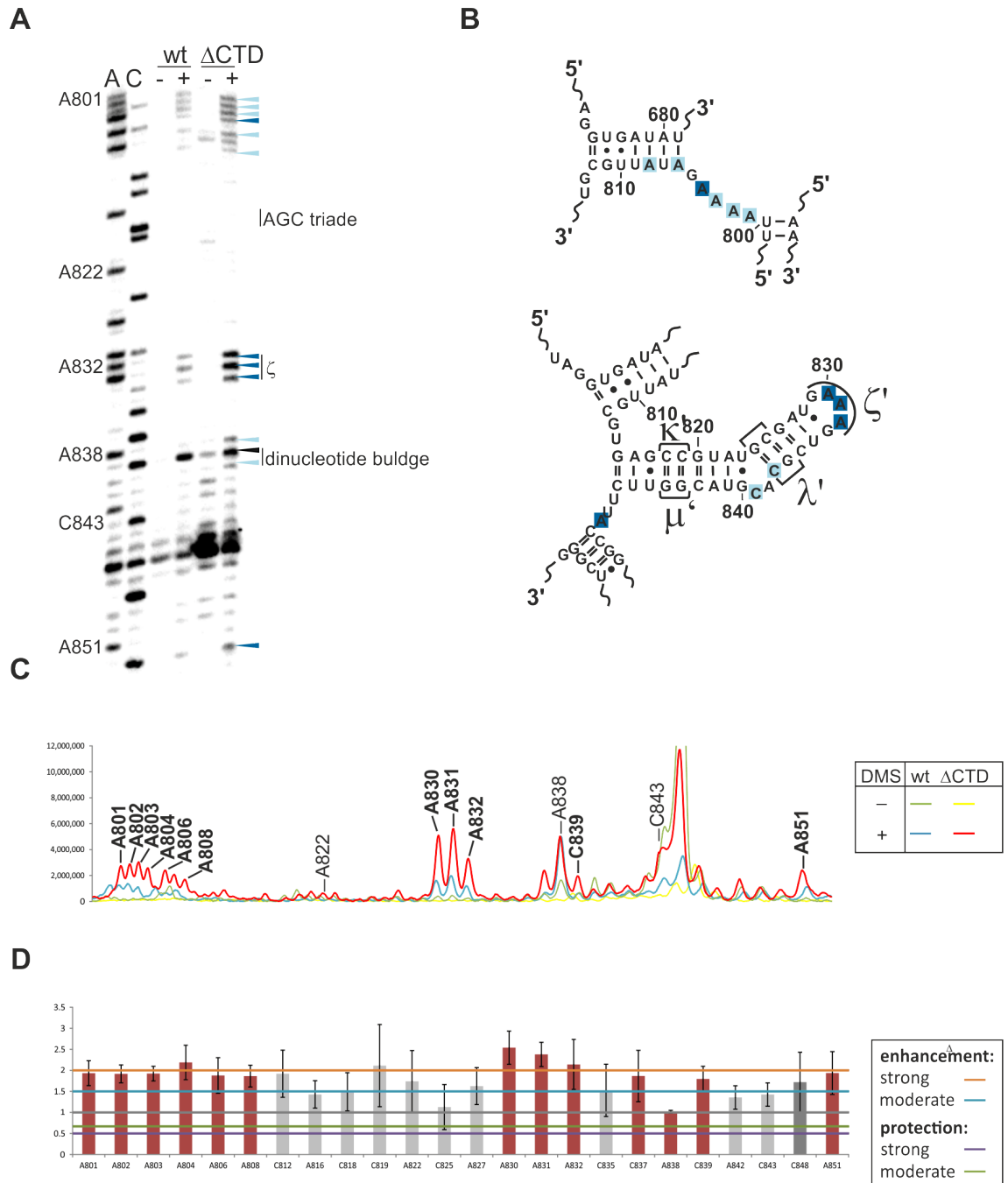


Figure 3.22: DMS accessibility of D5 in the Δ CTD strain. (A) Representative gel covering the 3' part of the D4 basal stem and D5 with its AGC triad, ζ' element and the dinucleotide bulge. (B) (C) (D) Symbol and color code as in Figure 3.6.

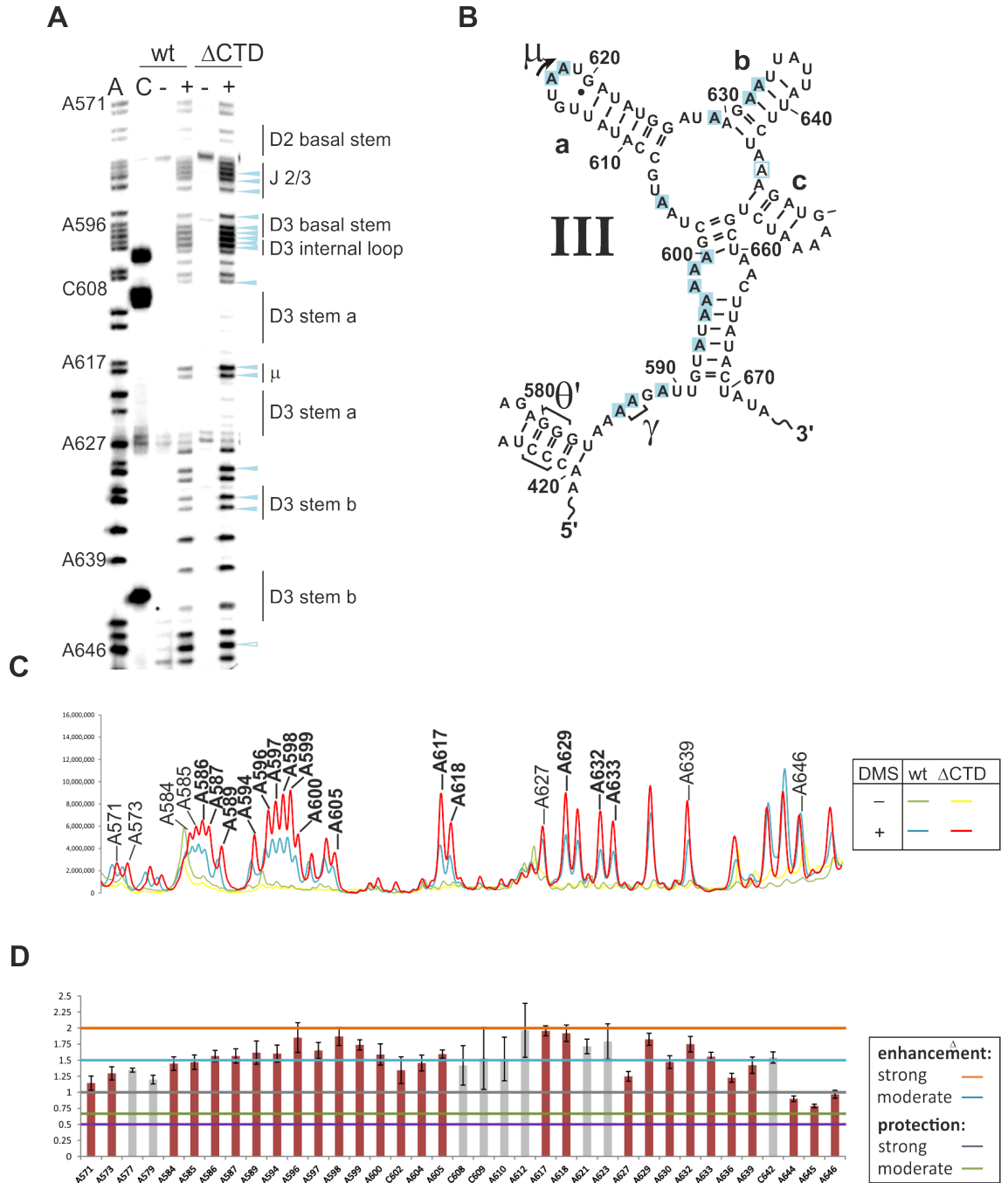


Figure 3.23: DMS accessibility of J2/3 and D3 in the Δ CTD strain. (A) Representative gel covering the 3' part of the D2 basal stem, J2/3, D3 basal stem and internal loop, stems a and b and the μ element. (B) (C) (D) Symbol and color code as in Figure 3.6.

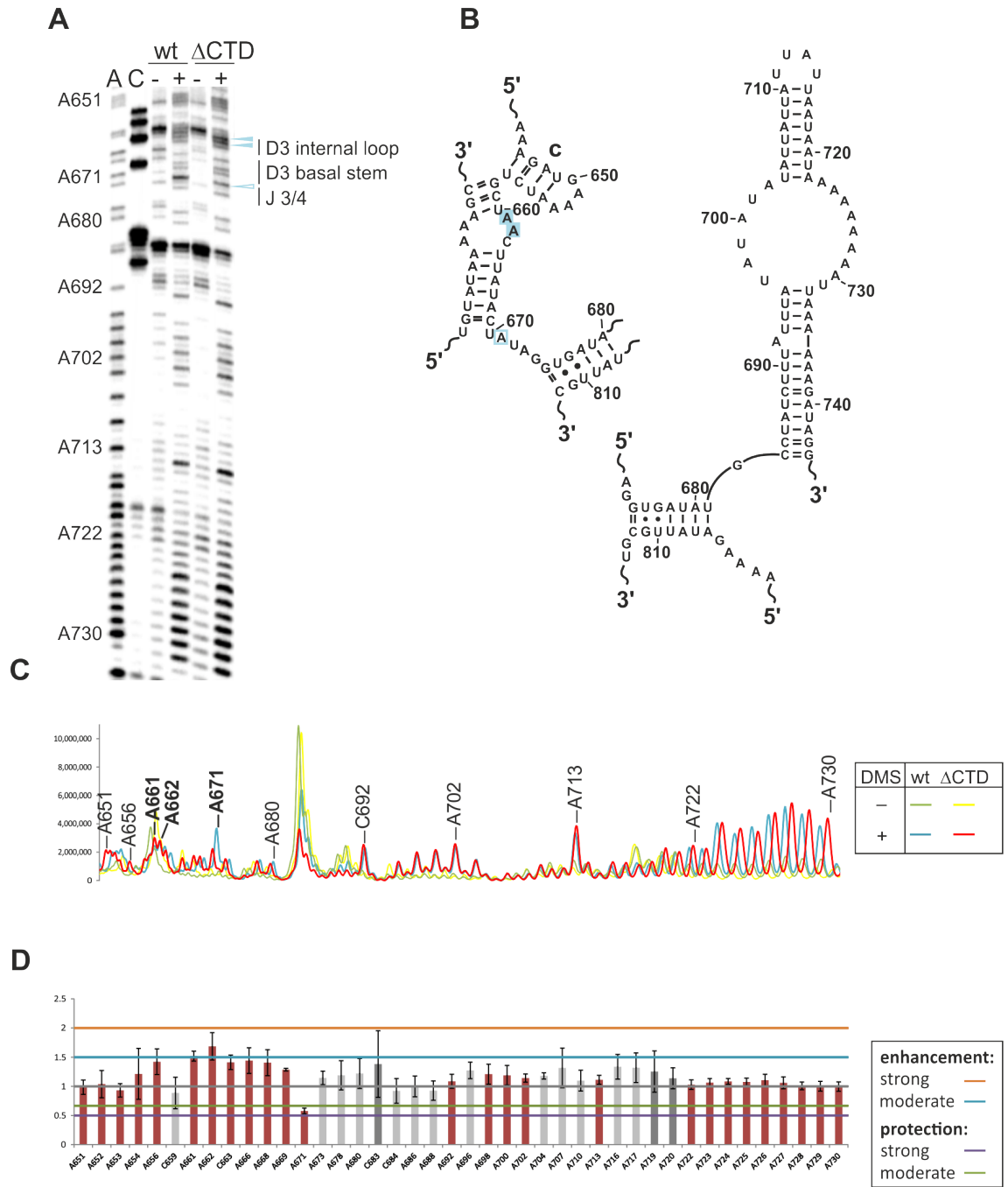


Figure 3.24: DMS accessibility of D4 in the Δ CTD strain. (A) Representative gel covering the 3' part of the D3 basal stem and loop, J3/4 and part of D4. (B) (C) (D) Symbol and color code as in Figure 3.6.

The modification pattern at the dinucleotide bulge with a strongly modified A838 and a moderately modified C839 indicates the non-canonical interaction between C839 and the base pair C818-G845 may form in a small subpopulation of ai5 γ molecules (Figure 3.22). The enhancement in C839 might result from the fact that the triple helix between J2/3 and the lower stem of D5 that stabilizes the twisted conformation of the dinucleotide bulge is not formed³¹.

In D3 the two nucleotides A617-A618 are both more accessible to DMS, showing that the D3-D5 tertiary interaction μ - μ' is not in place (Figure 3.23). Also the internal asymmetric loop of the D3 basal stem is more open than in the wild type strain, indicating that this important region is not in its native conformation. This is shown by enhancements of A594, A596 and A598 (Figure 3.23). Nevertheless all other nucleotides in this region show enhancements below threshold indicating a minor effect of a missing basic tail on the structure of this region compared to the ATPase-deficient strain. The asymmetrical loop contacts nucleotides in J2/3, which interacts with bases in J1/2. Since A416 and A418 are moderately and A419 is intensely modified in the Δ CTD strain neither the J1/2-J2/3 contacts nor the interactions of the asymmetric loop with J2/3 are established in this genetic background (Figure 3.17). As J2/3 itself is involved in a triple helix with the basal D5 stem, modifications of the nucleotides mentioned above show that the complex network of interactions connecting the asymmetric loop of D3 with the catalytic center D5 are not in place, when Mss116p is missing its basic tail. All nucleotides found in the exon binding sites EBS1 and EBS2 show strong or moderate DMS accessibility with the exception of A333 which is most likely a result of a different conformation in this terminal loop (Figure 3.8, Figure 3.10). This finding is in line with the fact that native structure formation is strongly disturbed in the Δ CTD strain rendering binding of the exons very unlikely. All in all the changes in DMS accessibility of ai5 γ if assisted in folding by Mss116p lacking its basic tail are formed at the very same regions as in a genetic background in which Mss116p is not able to bind and hydrolyze ATP. These changes are especially strong around the κ - ζ folding control element and often found in regions of high complexity as for example the z-anchor

region around tertiary interactions ε and λ or the triple helix involving J2/3 and the catalytic triad found in the lower stem of D5.

The basic tail of Mss116p plays an important role in ai5 γ folding

Formation of the κ - ζ folding control element and in turn D1 compaction depends on the presence of the basic C-terminal domain of Mss116p. The stems d, d'' and d''' are disrupted in this genetic background (Figure 3.25). The same is valid for stem i that is connecting D1 to the central wheel of the intron. As expected for an unstructured κ - ζ folding control element D1 is not in its collapsed state indicated by enhancements in the two D1 intradomain tertiary interactions β - β' and α - α' . As a consequence D5 is not docked into D1 as shown by the enhancements of the tertiary interactions ζ - ζ' and κ - κ' . As such, the active site is not formed, as indicated by the modification pattern around the z-anchor with its important tertiary interactions ε and λ . Compared to the ATPase-deficient and the knockout strain, the structural consequences of this Mss116p variant on ai5 γ seem to be less severe (Figure 3.1), suggesting that a subpopulation of ai5 γ molecules is able to fold into the native, splicing-competent state. This correlates well with the results of our respiratory growth assay, in which the Δ CTD strain was able to grow on a non-fermentable carbon source although with strongly reduced efficiency compared to the wild type strain (Figure 3.2). Furthermore, ai5 γ still showed residual splicing of about 10% when assisted by the basic tail missing variant of Mss116p (Figure 3.2).

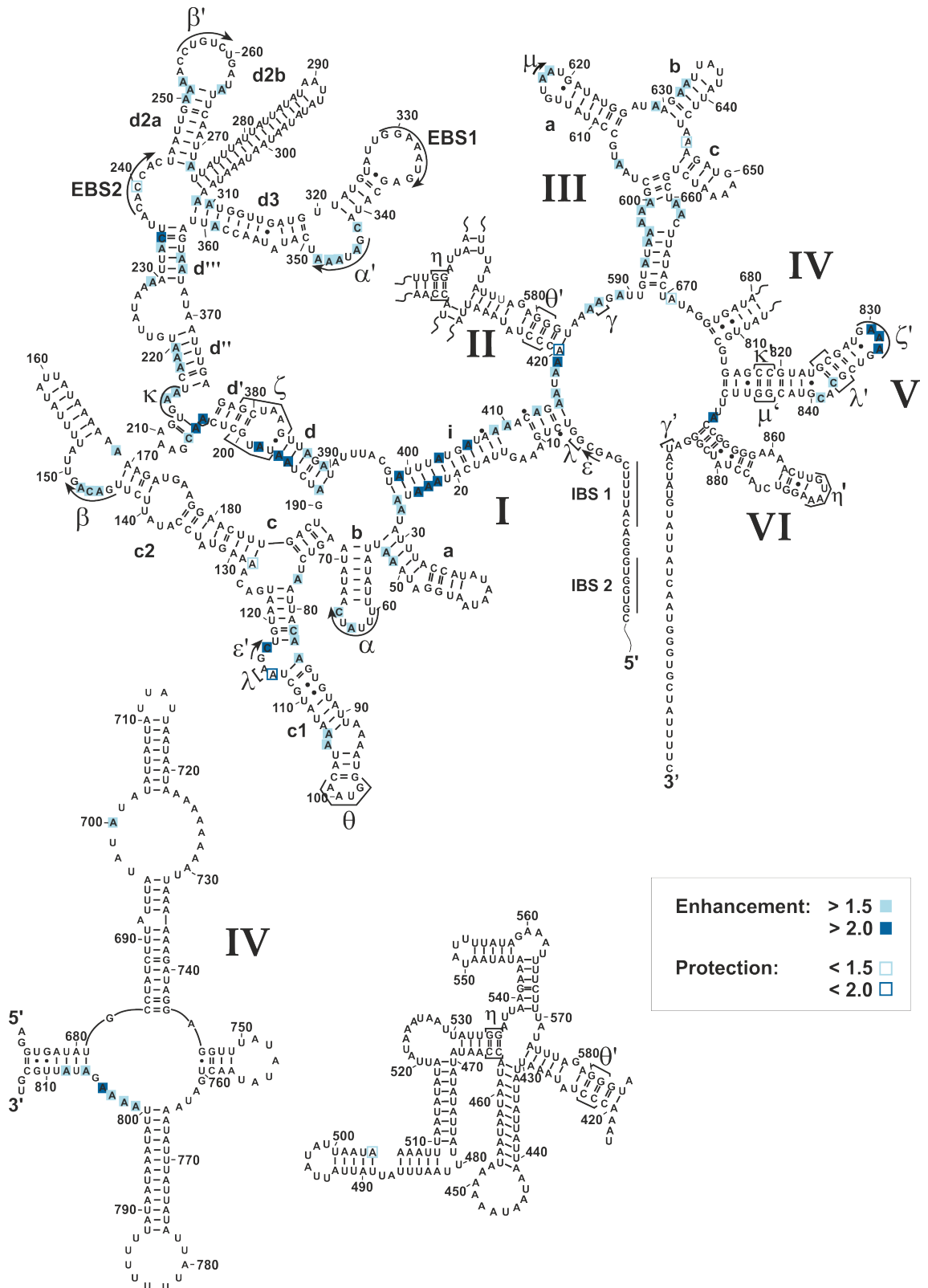


Figure 3.25: Differential summary map. Differences in DMS accessibility of *ai5γ* in a wild type strain compared to the RNA binding-deficient strain Δ CTD. Legend as in Figure 3.6 using blue instead of red.

3.4.5 Helicase activity of Mss116p and ai5 γ folding

An important activity of DEAD-box helicases lies in its name giving feature of being able to disrupt short helices of RNA^{137,138,143,154,162,171,172,177,179,181–183}. It is subject of ongoing discussion whether this helicase activity of Mss116p is its main contribution to ai5 γ folding or, more generally speaking, if DEAD-box proteins act on RNA predominantly by unwinding kinetically trapped helical regions in a large RNA molecule^{183,187,189,190,194,201,204,208}. The so-called Mss116p/SAT mutant was used in a remarkable study to show that a mutant, which is incapable of RNA unwinding, still retains its ability to assist in ai5 γ splicing¹⁸⁷. In an ensuing study it was shown that the SAT mutant can be dissected into a S305A mutant, which does not show any major differences to the wt protein in its ATPase or helicase activities, and a T307A mutant, in which both activities are severely affected¹⁸⁸. The T307A mutant though was neither able to rescue respiratory growth of yeast nor to stimulate splicing of ai5 γ *in vivo*. In the very same study several other variants of Mss116p were tested. Interestingly, all of the mutants that had defects in their ATPase activity did neither grow on a non-fermentable carbon source nor stimulate ai5 γ splicing. Hence, the ATPase activity of Mss116p was interpreted to be the essential part in its function as splicing cofactor of ai5 γ . One of the most interesting mutants used in this study, was Q412A. This mutant displays ATPase activity comparable to the wt protein but its helicase activity is significantly reduced *in vitro*¹⁸⁸. Therefore, the Q412A mutant is well suited to investigate the effects of helicase activity of the cofactor Mss116p on ai5 γ folding since the SAT mutant is also affected in its ability to hydrolyze ATP^{187,188,192}. Furthermore in contrast to the T307A as well as the S305A/T307A mutant, the Q412A variant is still able to rescue both respiratory growth and ai5 γ splicing^{188,192}.

An unwinding-deficient Mss116p variant interferes with the folding of the κ - ζ element in a subpopulation of ai5 γ

Formation of the κ - ζ element was shown to be crucial for both the collapse of the intron in early stages of folding as well as the establishment of the active site in the last steps of folding⁷⁸. Enhancements of nucleotides involved in the κ - ζ folding control element can still be found at the involved nucleotides but they are significantly lower than those in the K158R and Δ CTD strains and almost all below threshold. A195-A196 which are adjacent to the ζ receptor motif, are weakly enhanced (Figure 3.26). Similarly, nucleotides A218-A220 show increased reactivity to DMS (Figure 3.28). This indicates that stems d and d' are unstable at least in a fraction of ai5 γ molecules. A205 shows moderate enhancement (Figure 3.26). Hence the short helix connecting the κ and ζ element is not properly formed in at least some of the intron molecules in the helicase-deficient strain. Nucleotides at the basal stem of D3 are not as accessible to DMS as in the K158R and Δ CTD strains. Actually, only A596 shows a weak enhancement (Figure 3.32).

Both D1 intradomain tertiary interactions α - α' and β - β' are intact at least in some of the RNA molecules. In case of α - α' A63, C65, A344 and A346-348 are enhanced below threshold (Figure 3.29, Figure 3.30). The very same is true for the nucleotides A146, C147 and C259, which are found at the β - β' interaction (Figure 3.29, Figure 3.28). Only A148 is above threshold for weak enhancement. Nevertheless it must be stated that both elements are showing increased reactivity to DMS in the Q412A strain but in a lesser degree as seen with the K158R and Δ CTD strains.

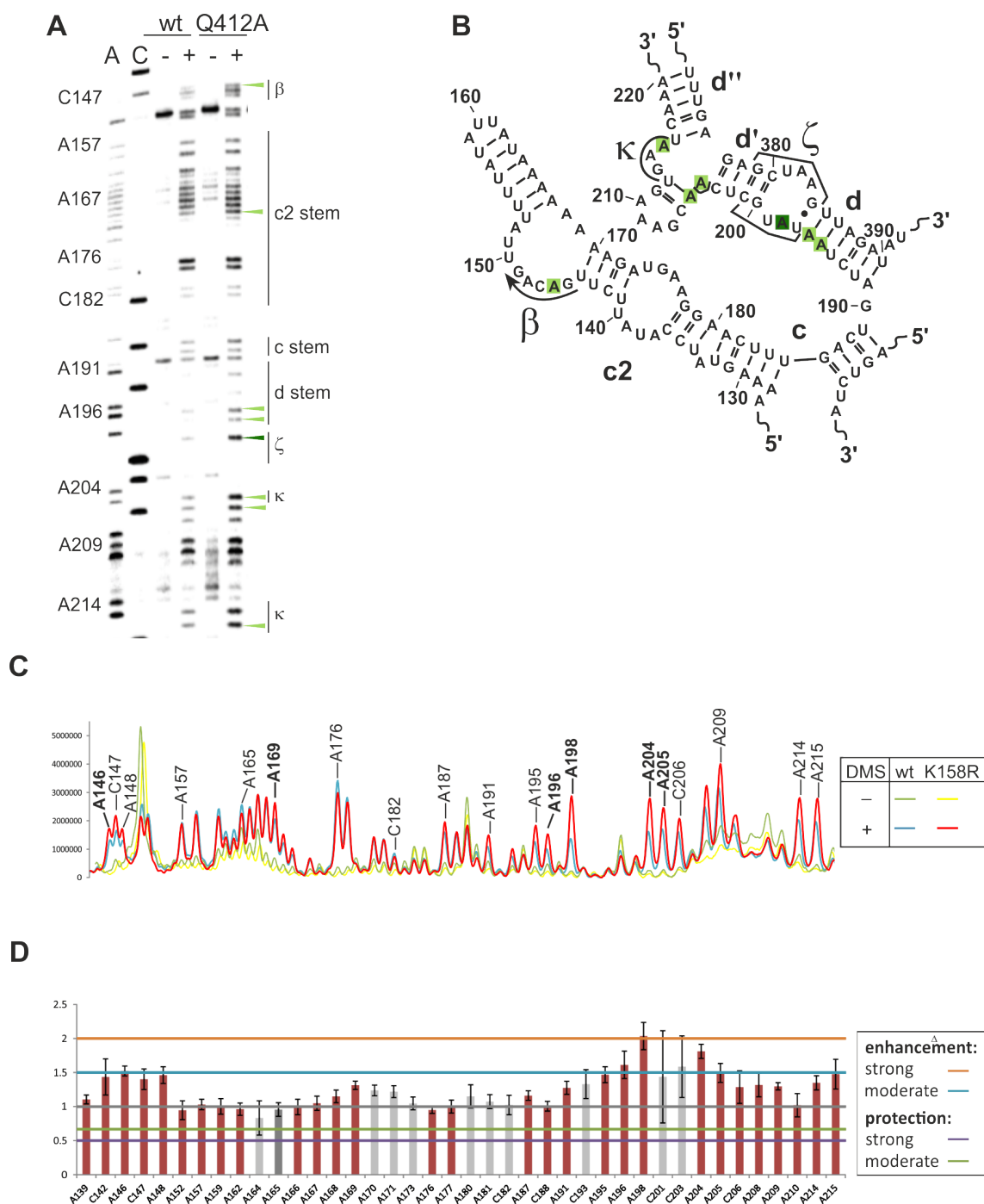


Figure 3.26: DMS accessibility of the 5' part of the κ - ζ element in the Q412A strain. (A) Representative gel covering the 3' part of the c2 stem as well as the 5' part of the c and d stem and tertiary interactions β , ζ and κ . Dark green color refers to strong light green color to moderate modification differences. (B), (C) and (D) Symbol and color code as in Figure 3.6.

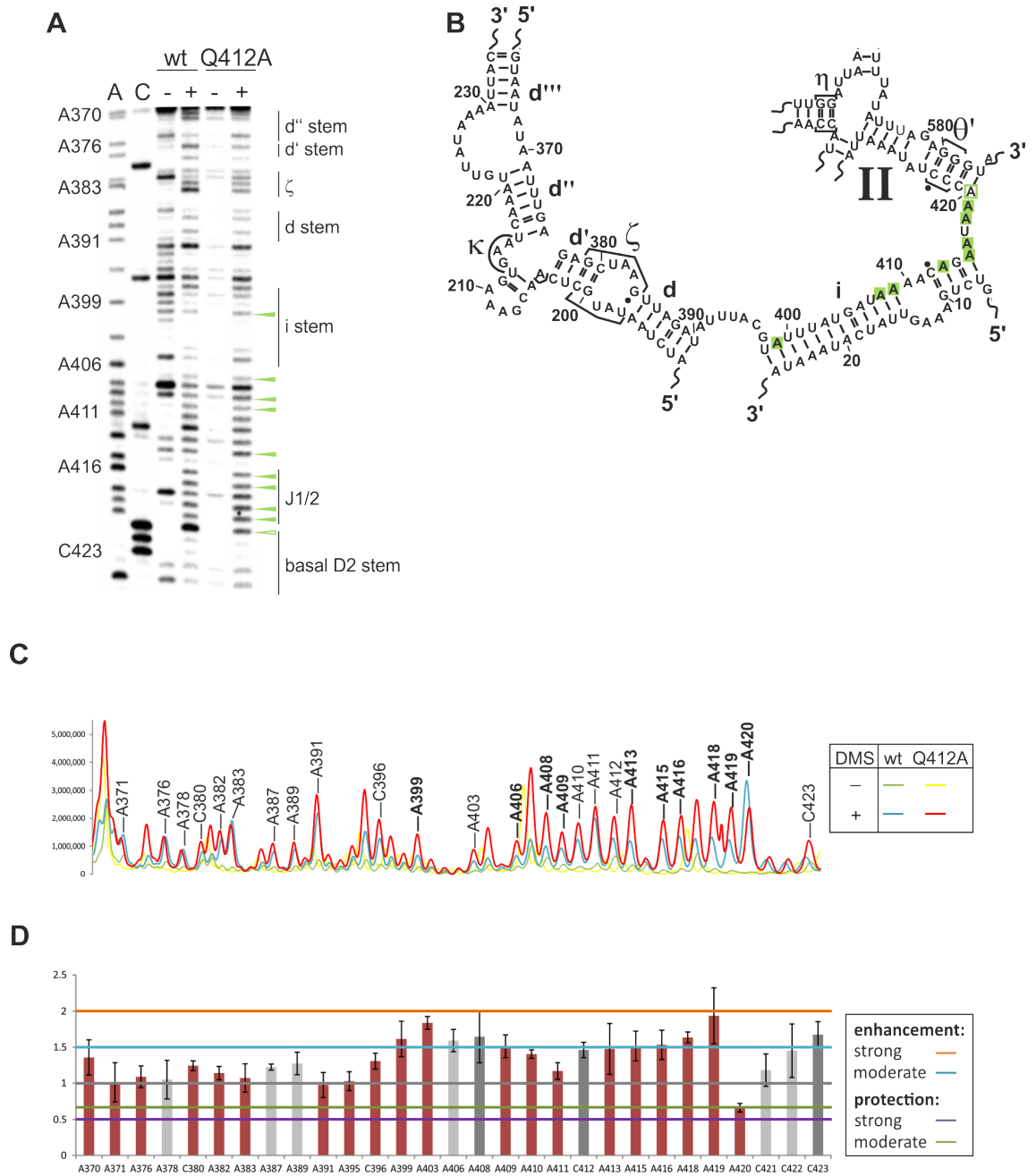


Figure 3.27: DMS accessibility around the 3' part of the κ - ζ element in the Q412A strain. (A) Representative gel covering the 3' part of the κ - ζ element and the i stem as well as J1/2 and the basal D2 stem. (B) (C) (D) Symbol and color code as in Figure 3.6.

Strand separation activity of Mss116p plays a role in secondary structure formation

One could assume that the disruption of Mss116p's ability to separate RNA strands would mostly affect the secondary structure formation of a given RNA molecule. In this respect we would intuitively expect the increase of stable secondary structures in *ai5γ*, since they are no longer resolved by Mss116p. But although secondary structure elements are not as severely disrupted as in the K158R or Δ CTD strain, we still observed enhancements rather than protections at the respective nucleotides.

A22-A24 show moderate enhancements compared to the wild type strain, whereas A20 and A26 are below threshold (Figure 3.29). Enhancements can also be seen for residues A399, A403, A406 and C412 (Figure 3.27). Thus stem i, which connects D1 to the central wheel of group II introns, cannot form in all intron molecules in the Q412A strain. Interestingly, the respective enhancements for nucleotides in stem i are significantly more pronounced in the K158R and Δ CTD strains. In stem a of D1 only A52 is above the threshold for moderate enhancement, but the DMS accessibility of A50-A52 is relatively similar to that observed in the strains harboring the Δ CTD or K158R mutation in *mss116* (Figure 3.29). Modification events at A75, A187 and C188 are as rare as in the wild type strain indicating that stem c is formed in most *ai5γ* molecules (Figure 3.26, Figure 3.29, Figure 3.31). Stem d''' seems to be disrupted in some intron molecules as A233 and C234 show weak DMS modification in the Q412A strain (Figure 3.28).

Protections at the terminal base pairs of stem c2 (A128) and the basal D2 stem (A420) are very similar to the ones seen in Δ CTD being above threshold in the case of A128 and of a substantially lower degree as seen in the ATPase-deficient strain in A420 (Figure 3.31, Figure 3.27).

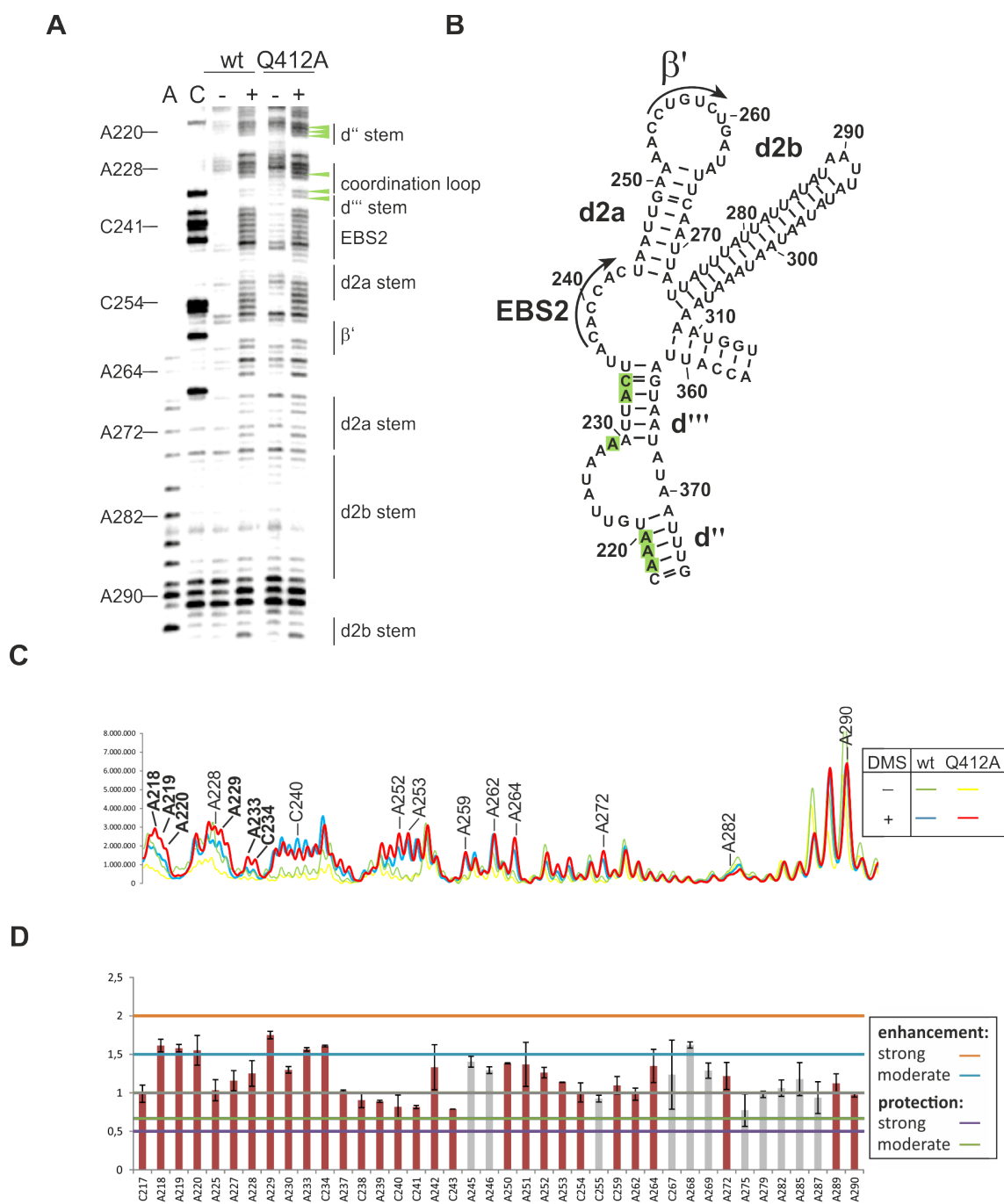


Figure 3.28: DMS accessibility around EBS2 in the Q412A strain. (A) Representative gel covering the coordination loop, EBS2, the 5' part of stems d'' and d''' and β . (B) (C) (D) Symbol and color code as Figure 3.6.

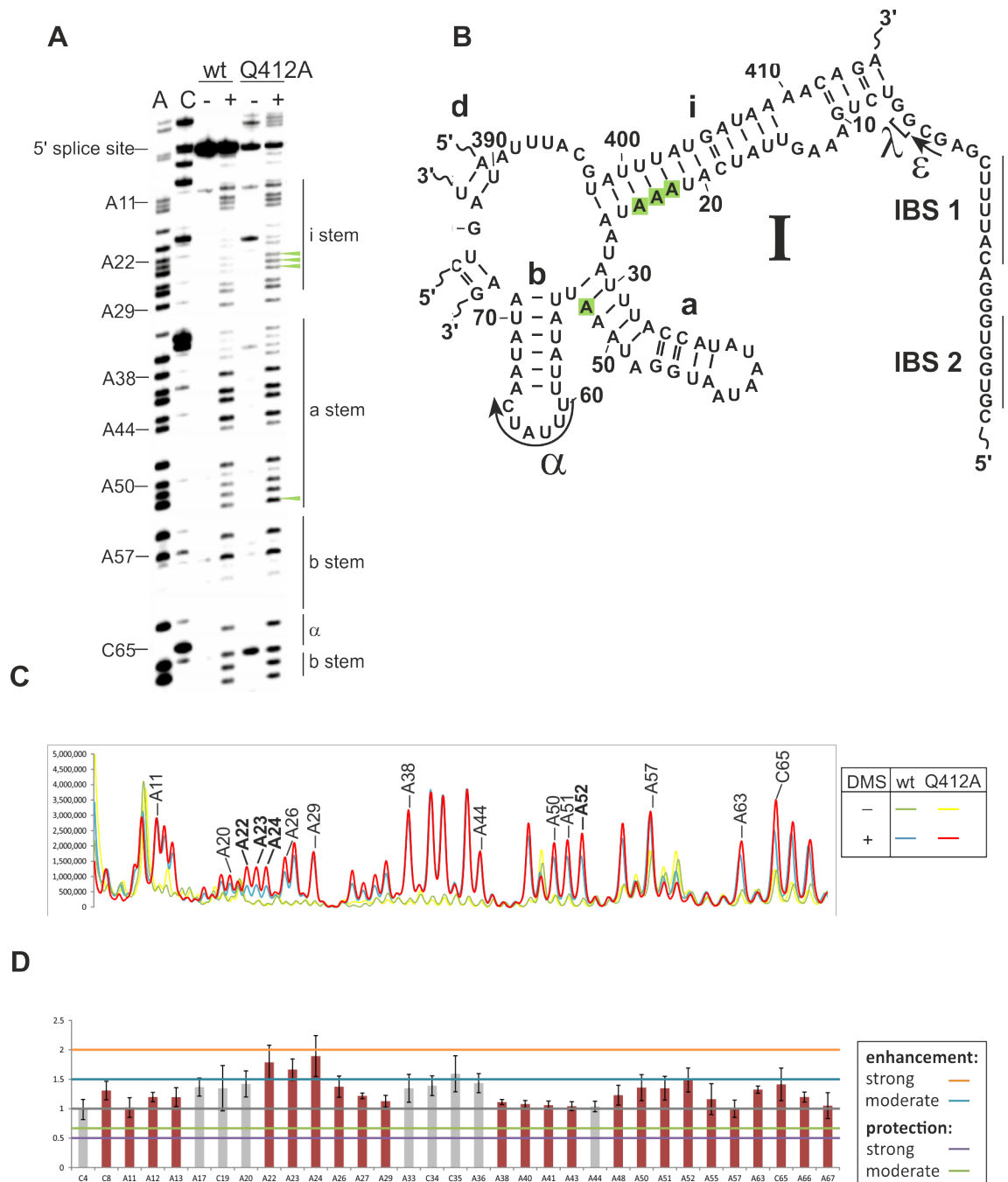


Figure 3.29: DMS accessibility of the very 5' end of *ai5γ* in the *Q412A* strain. (A) Representative gel covering the 5' part of stem *i*, stem *a* and stem *b* as well as the α element. (B) (C) (D) Symbol and color code as in Figure 3.6.

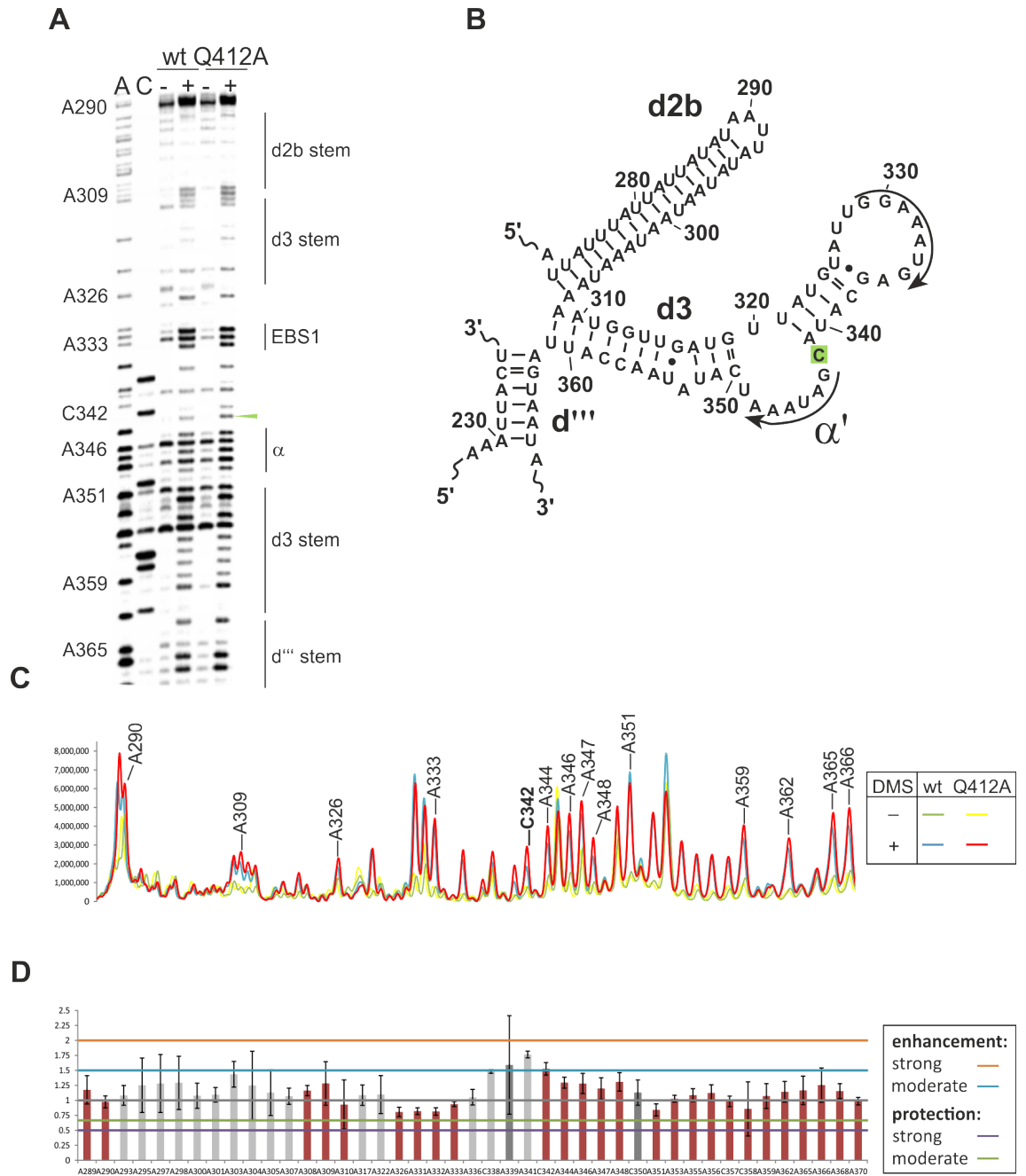


Figure 3.30: DMS accessibility around the d3 stem in the Q412A strain. (A) Representative gel covering the stems d2b, d3 and d''' as well as EBS1 and the α' element. (B) (C) (D) Symbol and color code as in Figure 3.6.

Several tertiary interactions within ai5 γ can form in the Q412A strain

Interestingly, in the presence of an unwinding-deficient Mss116p variant, the strongest enhancements are located at the tertiary interactions that anchor D5 on the D1 scaffold (Figure 3.26). A198 and A830-A831 show strong enhancements and A832 weak one in the Q412A genetic background. Enhancements in A382 (Figure 3.27) are of minor degree. Nucleotides found in the κ element are also more accessible to DMS in the unwinding-deficient strain. Although being slightly below threshold in A214, weak enhancements are observed in A204 and A215 (Figure 3.26). Hence the two tertiary interactions κ - κ' and especially ζ - ζ' seem to be in place in only a small fraction of the molecules. So although we see an effect of the disrupted helicase activity on folding of the κ - ζ folding control element, its degree is substantially lower than in case of an abolished ATPase activity or missing basic tail. Docking of D5 is affected at a comparable degree compared to the K158R, Δ CTD and knock-out strains. The architecture of D5 itself is as seen with the previously discussed strains not differing much from the wild type strain. The region around the dinucleotide bulge seems to be properly folded although C839 is weakly accessible to DMS, indicating that in some molecules the interaction with C818 is not formed (Figure 3.34). The J2/3 nucleotides A584-587 and A589 are equally modified in the wild type and Q412A strains (Figure 3.32). In contrast the J1/2 nucleotides A416 and A418-419 and residues of the D3 internal loop and D3 basal stem A596, A598, A599 and A600 show moderate enhancements (Figure 3.27, Figure 3.32) indicating that the intricate hydrogen network involving J1/2 and the D3 internal loop is not properly formed in a fraction of ai5 γ molecules. The protection found at A671 is also not as strong as in the Δ CTD and K158R strains (Figure 3.33).

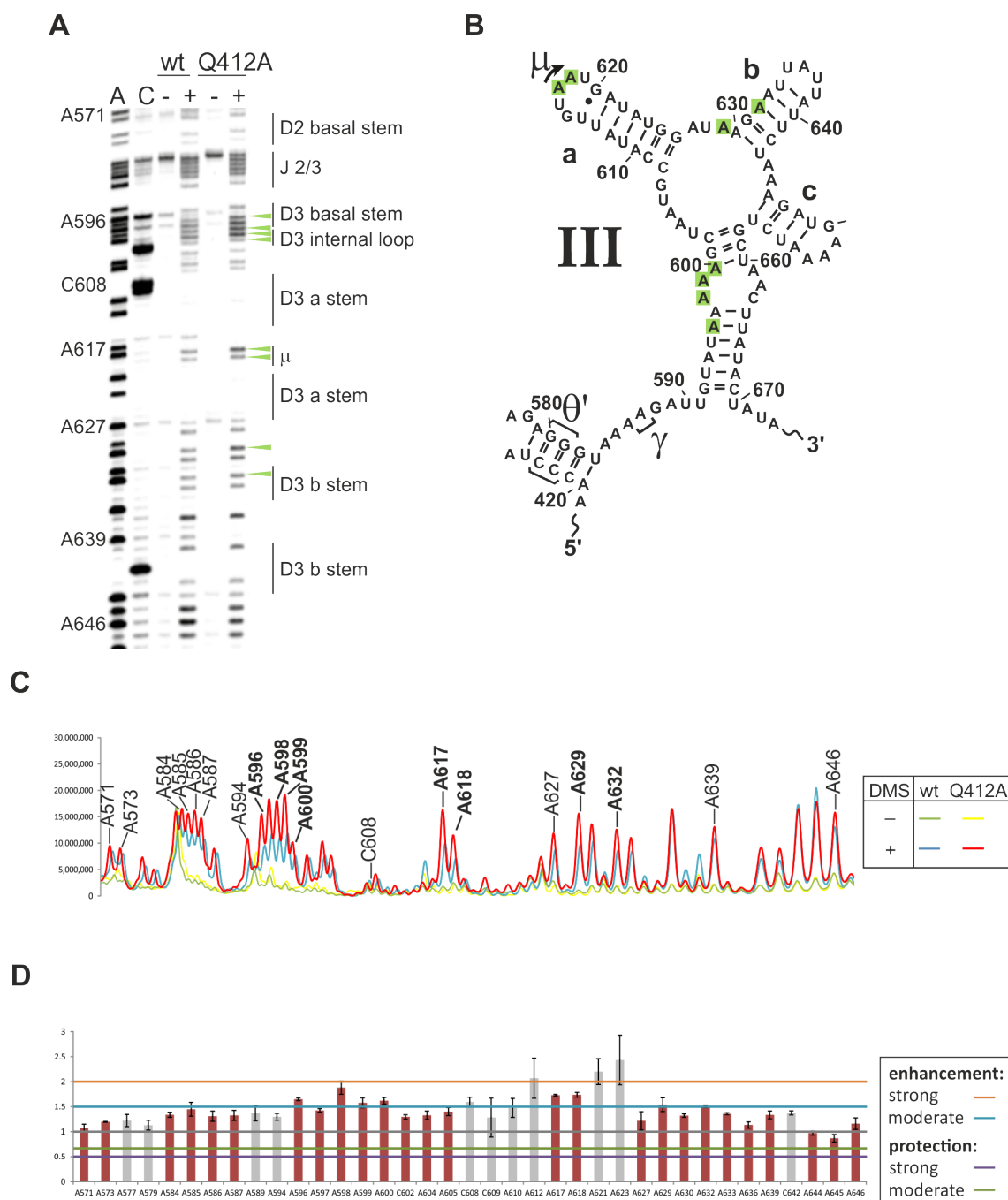


Figure 3.32: DMS accessibility of J2/3 and D3 in the Q412A strain. (A) Representative gel covering the 3' part of the D2 basal stem, J2/3, D3 basal stem and internal loop, stems a and b and the μ element. (B) (C) (D) Symbol and color code as in Figure 3.6.

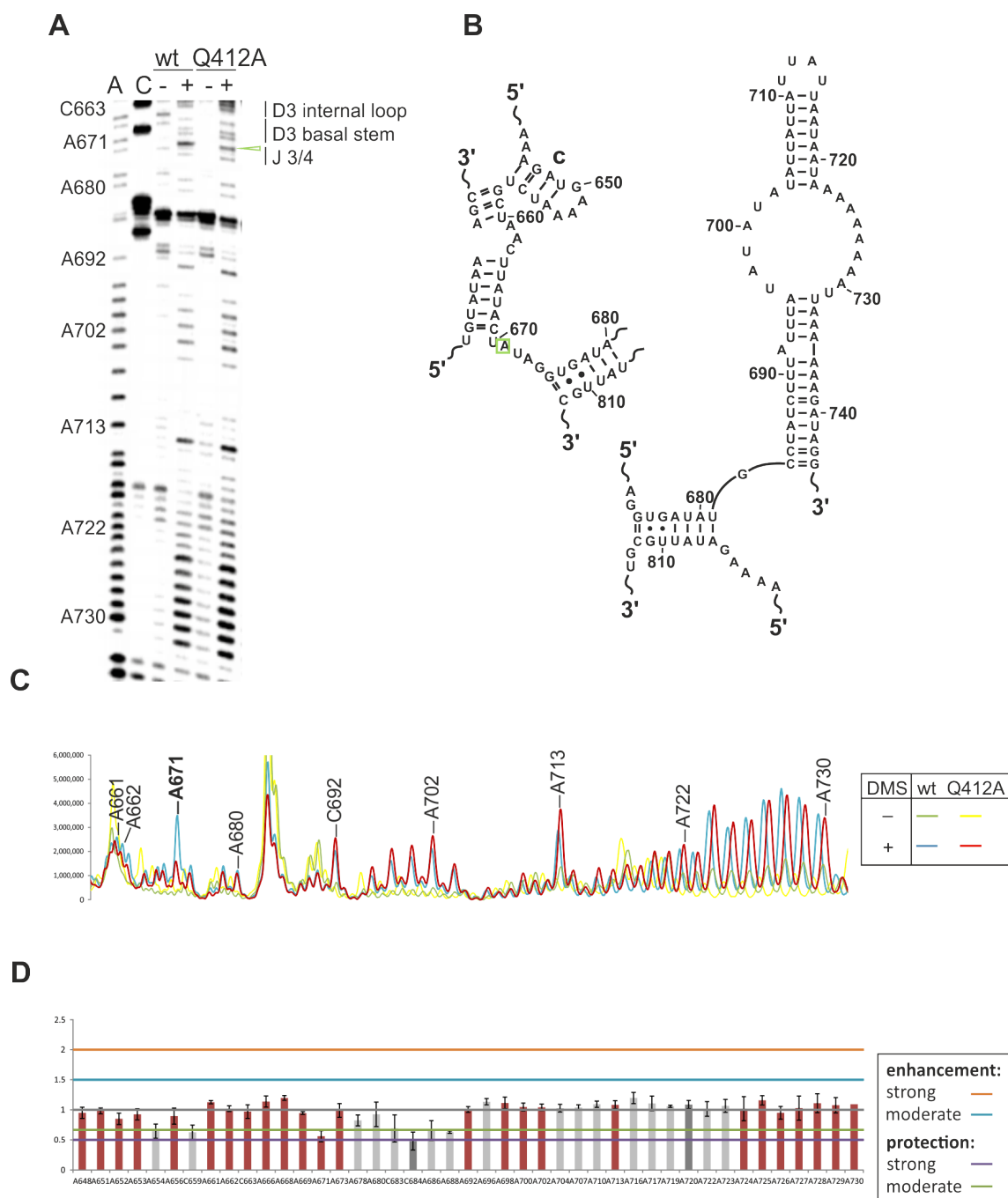


Figure 3.33: DMS accessibility of D4 in the Q412A strain. (A) Representative gel covering the 3' part of the D3 basal stem and loop, J3/4 and part of D4. (B) (C) (D) Symbol and color code as in Figure 3.6.

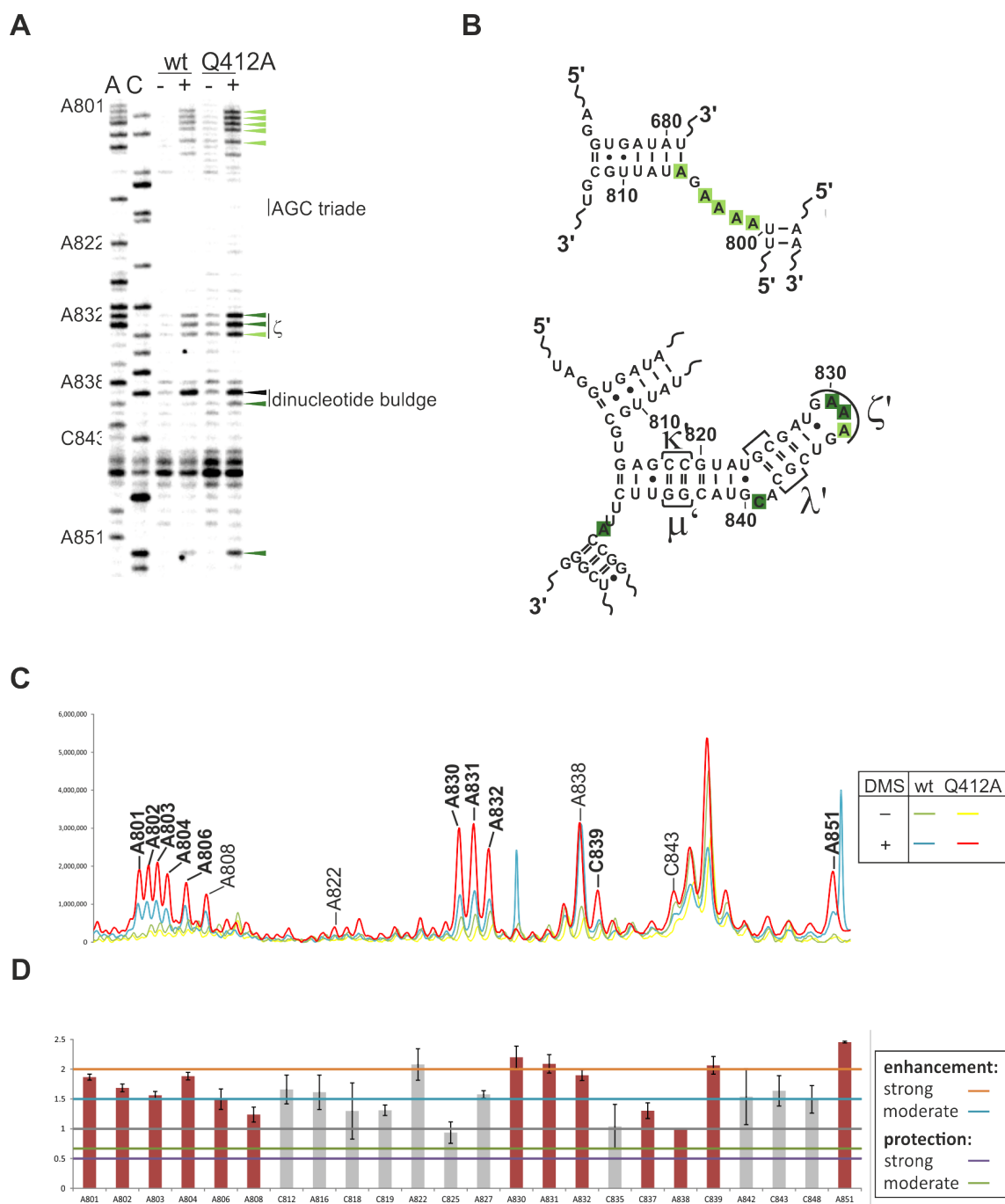


Figure 3.34: DMS accessibility of D5 in the Q412A strain. (A) Representative gel covering the 3' part of the D4 basal stem and D5 with its AGC triade, ζ' tetraloop and the dinucleotide bulge. (B) (C) (D) Symbol and color code as in Figure 3.6.

The active site constituents ε and λ seem to be in place in part of the intron population. C117 is only weakly enhanced in contrast to the strong enhancements found in the Δ CTD or K158R genetic background (Figure 3.31). Also, A114 is only moderately protected showing that the z-anchor region is established at least in a substantial fraction of ai5 γ molecules in the unwinding deficient strain. A617-A618 are weakly enhanced compared to wild type hence the μ - μ' tertiary contact is not established in all intron molecules (Figure 3.32).

Nucleotides found in EBS1 and EBS2 are all showing modification levels similar to the wild type genetic background, suggesting that in a majority of the RNA molecules the exons are not bound to the intron (Figure 3.28, Figure 3.31).

In this strain ai5 γ splices with an efficiency of 1/3 compared to the wild type strain. Thus, the effects on docking of D5 and on the active site might reflect the fact that disrupting the unwinding activity in Mss116p has a pronounced effect in the last stages of ai5 γ folding^{16,42}. Undoubtedly, the active site of the intron is not formed in all of the molecules.

Helicase activity is essential for docking of D5 and establishment of the active site in a substantial fraction of ai5 γ molecules

Without its helicase activity Mss116p still retains its ability to assist in secondary structure formation of ai5 γ (Figure 3.35). Disruptions of stems d, d'', d''', a and i are still detectable but to a much lesser degree than in the ATPase-deficient, Δ CTD or knock-out strains (Figure 3.1). Formation of the κ - ζ folding control element is only slightly disturbed, which also correlates with the minor enhancements in the D1 intradomain contacts α - α' and β - β' . Hence, formation of the κ - ζ element seems to be independent from unwinding in a substantial fraction of the ai5 γ population.

Interestingly, the strongest enhancements are found at residues involved in the ζ - ζ' interaction which is essential for docking of D5. Additionally, residues of the κ element, the second D5 anchoring interaction in D1, show elevated DMS reactivity. This can only be interpreted with disrupted docking of D5. The active site constituents λ - λ' and

$\varepsilon-\varepsilon'$ are formed in only a fraction of the intron molecules and the overall modification pattern around the important z-anchor suggests that this is also true for the entire region. Additionally, the complex interactions involving J1/2 and the D3 internal loop are not established in a fraction of ai5 γ molecules. The very same is true for the $\mu-\mu'$ contact, which anchors the important effector D3 near the active site.

3.4.6 Exons are accessible to DMS in all different strains

Disrupting the unwinding activity of Mss116p had a substantial effect on both the splicing level and structure of ai5 γ . Additionally, in neither of the mutant strains misfolded intron molecules could be revealed by *in vivo* DMS probing. As shown before for group I introns *in vitro*²⁰⁶, maybe structures involving exonic sequences have to be resolved in order to allow native structure formation of ai5 γ . This would also nicely correlate with the finding that an unwinding deficiency in Mss116p strongly reduces splicing efficiency of ai5 γ *in vitro*, if long exon sequences flank the intron^{190,194}. Mapping the exonic regions of ai5 γ close to the 5'- and 3'-splice site poses a challenge due to the different splicing levels in the different genetic backgrounds used in our studies. Since most of the ribozyme population self-splices in the wild type yeast strain it is not possible to map the exons. Furthermore to map the 5' exon different primers had to be used for the wild type and the mutant strains for the very same reason. By using a primer that actually binds in the downstream exon of ai5 γ in the wild type strain, it is possible to map all nucleotides being accessible to DMS in a similar manner (Figure 3.36B). A-6 and C-7, which are part of IBS1, as well as A-8, C-17, A-18, C-20, A-21, C-24 and A-25 are all strongly modified at comparable intensity in the wild type strain. For mapping of the upstream exon in the mutant strains we used a primer binding in the intron. There are no differences in DMS accessibility in neither of the mutant strains, indicating that the upstream exon does not form any stable structures neither in the mRNA (wild type strain) nor in the pre-RNA context (mutant strains).

The 3' exon can be mapped by using the same primer in all genetic backgrounds (Figure 3.36A). The modification pattern is virtually identical, when mapping the terminal cox1 exon in the individual strains. All involved nucleotides show mostly strong DMS accessibility, suggesting that also the downstream exon is not part of any stable structural elements in either of the genetic backgrounds.

Apparently, Mss116p does not resolve structures that involve exonic sequences as shown by the high DMS reactivity of almost all exonic nucleotides.

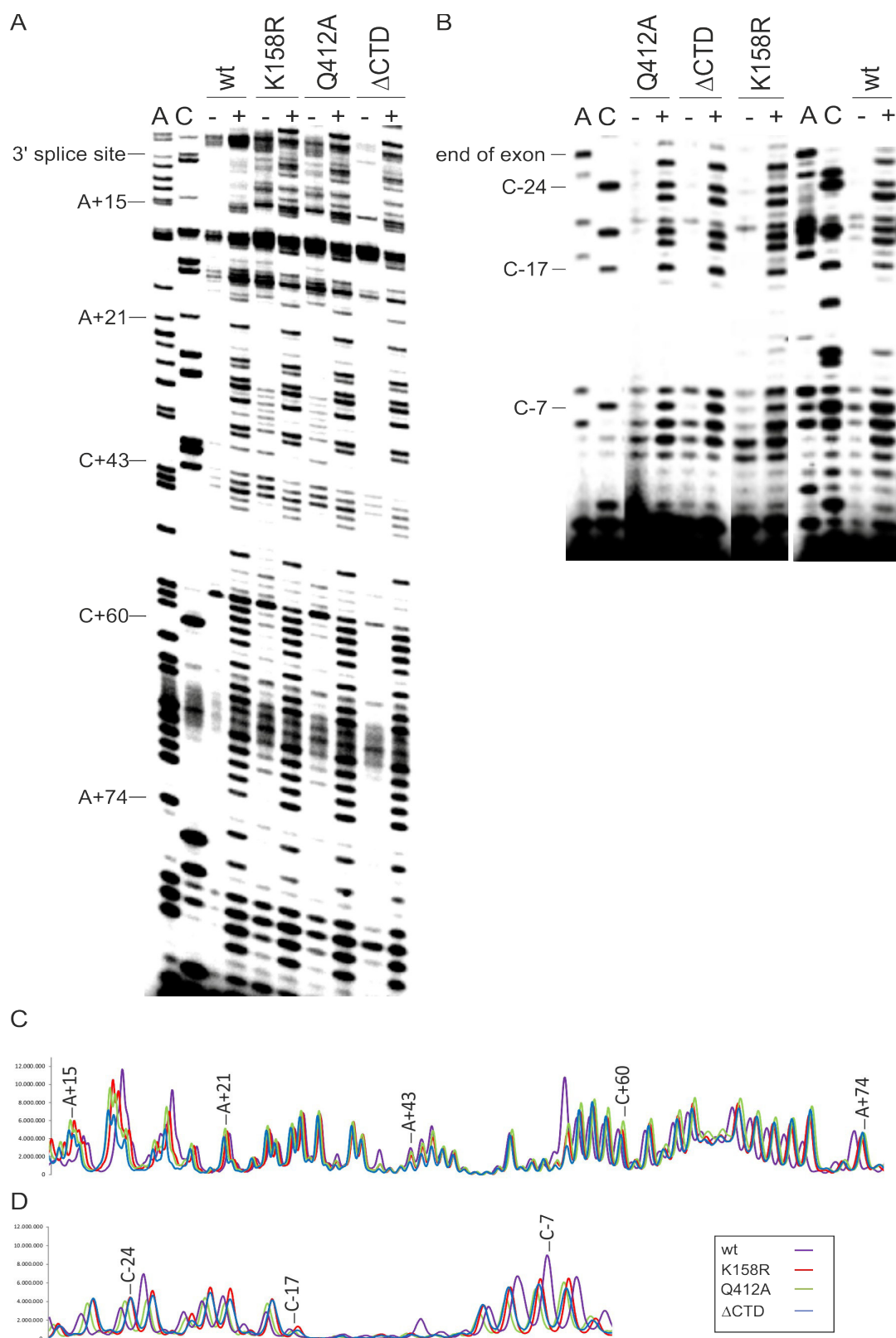


Figure 3.36: DMS accessibility of the exons. Representative gels showing (A) the 3'exon and (B) the 5'exon of *coxI* flanking the *ai5 γ* intron.

4 Discussion

By adopting its native structure a ribozyme can fulfill its function. The number of available high resolution structures of RNA molecules has tremendously increased in the last years giving us substantial insight into the relation of structure and function of catalytic RNA molecules. In order to reach its native catalytically active structure the given RNA molecule has to fold correctly. This process has been predominantly studied *in vitro*^{28,108,209–217} resulting in several folding paradigms for large RNAs. There are two major problems different RNA molecules encounter in adopting their native structure¹¹⁵. The folding energy landscape of RNA molecules is often rugged leading to misfolded species that have to be resolved in order to allow native structure formation. On the other hand folding can also follow a smooth energy landscape, in which an intermediate structure has to be stabilized to enable folding to the native state.

In vitro studies have been and will be of utmost importance for the understanding of RNA folding. Nevertheless, there are fundamental differences to the cellular environment. Most *in vitro* folding studies are carried out at non-physiological conditions, namely high metal ion concentrations and elevated temperature starting with the full length transcript after denaturation. In contrast, folding occurs cotranscriptionally *in vivo*, being influenced by the velocity and directionality of the RNA polymerase^{218–220}. Although metal ions play an important role in intracellular folding as well, their concentration is far below most used *in vitro*. Instead many trans-acting co-factors are present *in vivo* guiding a given RNA molecule towards its native structure. One of those is Mss116p, a DEAD-box helicase that assists in splicing of all mitochondrial group I and group II introns in yeast¹²⁸.

One of the major problems in studying RNA structure and folding *in vivo* is the small number of suitable methods, chemical probing being one of the few available. By applying DMS chemical probing in yeast cells we previously described the structure of the group II intron ai5 γ *in vivo* at endogenous levels for the first time^{109,203}. Of course this method does not provide any high resolution structural data as NMR or X-ray studies, but combined with available structures and models of the respective RNA, it can be of great help to analyze if its *in vitro* or *in silico* determined structure correlates

with the given native one in the cell. By finding a quite astonishing degree of correlation between the DMS modification pattern *in vivo* and *in vitro* as well as high consistency with the available crystal structure of the *Oceanobacillus iheyensis* group IIC intron we showed the applicability of this method for this purpose.

In order to reach the native structure ai5 γ depends on high amounts of [Mg²⁺] and elevated temperature and although compaction of the ribozyme is also observed at near-physiological conditions, it is very ineffective¹⁰⁶. Mss116p is able to promote ai5 γ splicing at near-physiological conditions, indicating that this DEAD-box protein substitutes for high [Mg²⁺] used *in vitro*^{183,187,189}. We showed that indeed ai5 γ remains largely unfolded in a *mss116*-knockout strain¹⁰⁹. The strongest effects could be seen in D1, which folds first and serves as a scaffold for the other domains during later stages of ai5 γ folding. This indicates that Mss116p plays a crucial role in folding of D1, which is transcribed first. Most interestingly, in the absence of Mss116p the majority of DMS enhancements in ai5 γ were found at the κ - ζ folding control element, whose formation was shown to be crucial for the intron's collapse^{77,78}. At this region also secondary structure elements were severely disrupted highlighting the importance of Mss116p in folding of this important region. Consistently, the active site was not properly formed in the knockout strain and neither D3 nor D5 were docked on D1. In the event of misfolding, RNA species with sites of DMS protection would arise. We did not observe such regions of protection indicating that indeed ai5 γ does not misfold in the absence of Mss116p. Kinetically trapped RNA molecules are also prone to degradation²²¹ which is not consistent with the strong RNA signal in the Δ *mss116* strain. Rather than unwinding misfolded regions of the intron to allow refolding to the native state, our data implicated that Mss116p stabilizes intermediate structures allowing native structure formation.

DEAD-box proteins are multifunctional enzymes that are important in many aspects of RNA metabolisms such as RNA decay, mRNA transport, ribosome biogenesis, RNA splicing and even RNP disassembly^{137,138}. These proteins display RNA binding, unwinding, annealing and ATPase activity^{137,138,154,191}. After our initial results we were interested in how these different activities contribute to folding of large and very complex RNAs, such as the ai5 γ group II intron.

Introducing different variants of Mss116p into *Saccharomyces cerevisiae*'s genome gave us an insight into the effects on ATPase, unwinding and non-specific RNA binding activities of a DEAD-box protein on the splicing efficiency and the structure of a large RNA molecule *in vivo*. DMS enhancements and protections in ai5 γ found in an ATP binding and hydrolysis deficient strain correlate strongly with the pattern we previously revealed in the knockout strain (Figure 4.1). Both secondary as well as tertiary structure elements are effected and there are no signs of non-native base-pairing. As with the knockout strain the strongest enhancements are found at the κ - ζ folding control element, suggesting that without ATP binding and hydrolysis most of the RNA molecules are handicapped at an early stage of folding. Disruption of unspecific RNA binding by deleting the C-terminal basic tail has a similar effect on the structure of ai5 γ . Nevertheless the degree of enhancements is not as strong as seen with the K158R strain. This is also reflected in the limited but observable ability to both grow on a non-fermentable carbon source as well as induce ai5 γ splicing. In contrast, most structural elements seem to be formed in the unwinding-deficient strain Q412A. Strong enhancements are found in interactions that are formed during late stages of ai5 γ folding.

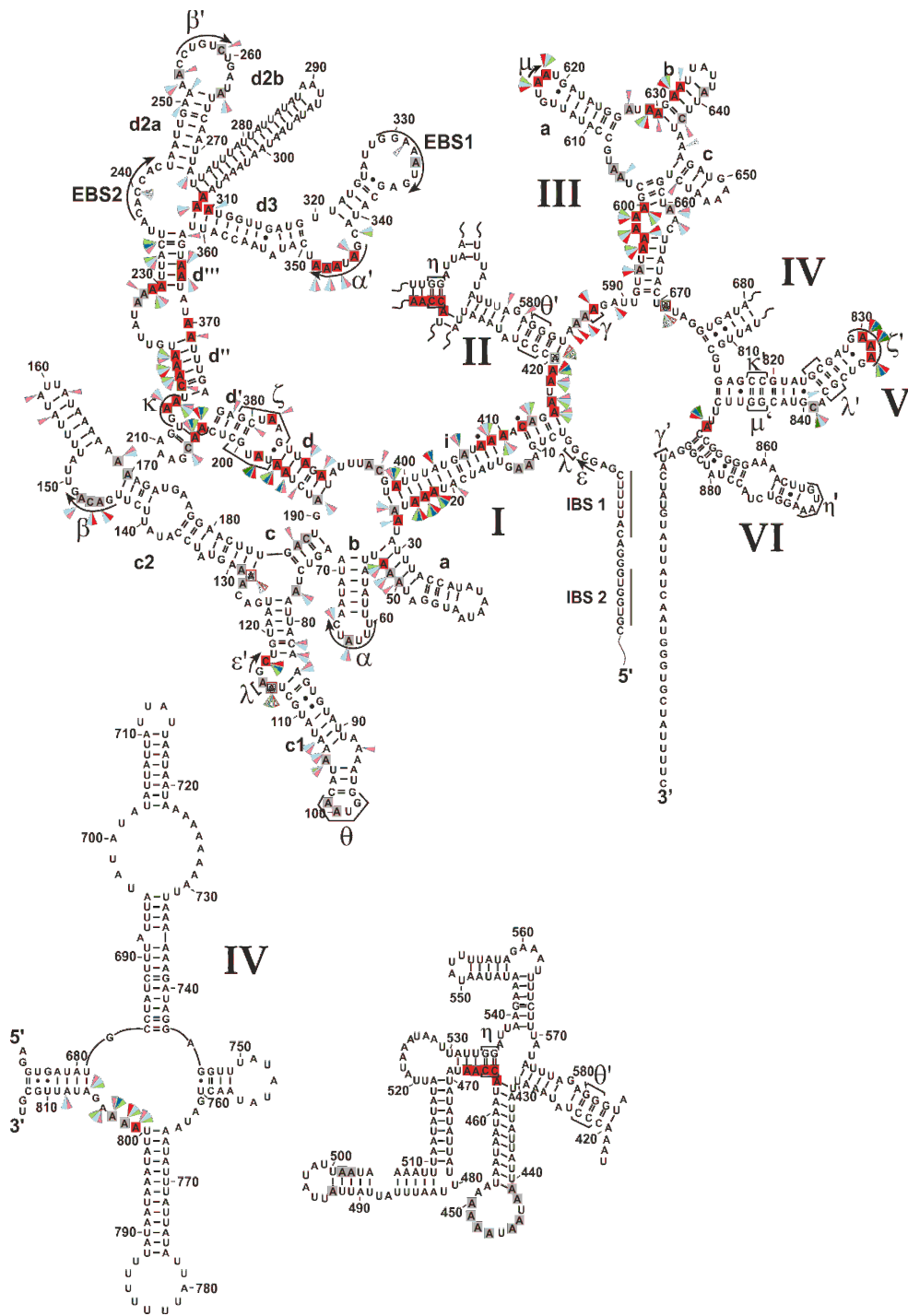


Figure 4.1: Comparison of DMS modification pattern of *ai5γ* in different genetic backgrounds. Filled symbols refer to enhancements, open ones to protections in the different strains *K158R* (red), Δ *CTD* (blue), *Q412A* (green). Dark colors show strong, light colors weak differences in DMS accessibilities compared to the wild type strain. Protections and enhancements in the knockout strain are shown by open and filled squares respectively. Red squares indicate strong and gray weak changes.

ATP binding and hydrolysis of Mss116p is crucial at an early stage of ai5 γ folding

The ATPase activity of Mss116p is essential for ai5 γ splicing both *in vitro* and *in vivo*^{128,187,189,190}, although the precise effect of ATP binding and hydrolysis on the structure of ai5 γ is still unclear. In a single molecule study it was shown that ATP is most crucial for the final step of ai5 γ folding, the transition from a folded intermediate to the native state²⁰¹. In fact, ATP is not needed at all to reach a near-native folding state of ai5 γ ¹⁹⁵. These results are in line with the findings that ATP hydrolysis is needed for protein turnover¹⁶⁹. Our results show that the mutation K158R, which abolishes binding and hydrolysis of ATP, renders Mss116p unable to assist in ai5 γ folding. The DMS pattern strongly correlates with the one found in the $\Delta mss116$ strain. It is therefore not in agreement with an intron being trapped in a compact state, in which only the final stages of ai5 γ folding, that depend on the ATPase activity of Mss116p^{195,201} are missing. As seen in the knockout strain, the strongest enhancements are again found in the κ - ζ folding control element, indicating that the stems embedding κ - ζ are not formed and in turn D5 cannot dock on D1. It was shown *in vitro* that establishing this important substructure is most crucial for the subsequent steps of ai5 γ folding^{77,78}. In fact neither is D1 collapsed nor the active site established without binding and hydrolysis of ATP by Mss116p, as shown by the enhanced DMS accessibility of the respective tertiary interactions within ai5 γ in the K158R strain. *in vitro* compaction of D1 as well as D135 and D1356 was shown to be independent on ATP at near-physiological conditions, as the rate constant for compaction of the ai5 γ ribozyme was barely affected^{195,201}. After initial ribozyme compaction proteolysis of Mss116p could trigger native structure formation without ATP, suggesting that indeed ATP binding and hydrolysis is mainly needed for protein dissociation¹⁹⁰. Trapped DEAD-box protein molecules, especially if these are located at crucial interaction sites of the intron, would interfere with tertiary folding. For example docking of D5 onto D1 could depend on dissociation of Mss116p. Constructs that harbored long exonic sequences though were dependent on ATP hydrolysis by Mss116p for promoting ai5 γ splicing *in vitro*¹⁹⁰. It was suggested that

structures involving exonic and intronic sequences have to be resolved using ATP hydrolysis in order to allow native structure formation.

In contrast to the *in vitro* findings our results suggest that ATP binding and hydrolysis is already needed at an early stage of ai5 γ folding *in vivo*. DMS enhancements and protections in ai5 γ found in the K158R strain correlate strongly with the pattern we previously revealed in the knockout strain (Figure 4.1). Both secondary as well as tertiary structure elements are effected and there are no signs of non-native base-pairing. As with the knockout strain the strongest enhancements are found at the κ - ζ folding control element, suggesting that without ATP binding and hydrolysis most of the RNA molecules are handicapped at an early stage of folding. Mss116p was proposed to bind its RNA substrate non-specifically^{152,183}, which could result in non-productive binding events. This would underline the importance of protein turnover, as Mss116p would need to dissociate again for productive binding. Non-specific binding does not need ATP *in vitro*, although the binding efficiency is increased with ATP, ADP-Pi or an unhydrolyzable ATP analog bound to Mss116p^{162,165,183,208}. The drastic effect of the K158R mutation on ai5 γ folding cannot be explained solely by its small decrease in RNA binding affinity, although it might contribute. As we work on endogenous levels of both protein and ribozyme, the ratio of Mss116p to ai5 γ molecules is not known. An access of intron over DEAD-box protein molecules could on one hand lead to productively bound Mss116p, which lacks the ability to dissociate from the RNA due to the K158R mutation. And indeed it was shown that only the last steps of ai5 γ folding, namely the docking of the domains onto the folding scaffold D1, depend on ATP^{195,201}. On the other hand non-productive binding of Mss116p molecules would also be detrimental to folding of ai5 γ in two ways. First, bound Mss116p might be in the way of certain folding events and secondly at low levels of protein these common non-productive binding events¹⁹⁵ might lead to protein depletion if ATP dependent protein recycling does not function. The lack of specificity in binding would make those binding events invisible to DMS probing. Hence our results would be in line with the proposed role of ATP hydrolysis in allowing Mss116p to repeatedly bind to and dissociate from its RNA substrate, if in the mitochondria ai5 γ is in excess to Mss116p.

In other words ATP binding and hydrolysis activity might not be needed for initial steps of ai5 γ folding with an Mss116p to ai5 γ ratio similar to the one used *in vitro* ranging from 1:1 to 20:1^{190,195,201}. But at endogenous levels recycling of Mss116p might be crucial due to low protein levels and partially non-productive binding events *in vivo*. Unfortunately, the protein to RNA ratio in the cell or more precisely in the mitochondria cannot be assessed, even more so as Mss116p is not only involved in ai5 γ splicing but also in RNA processing and mitochondrial translation¹²⁸.

Non-specific binding by the C-terminal extension of Mss116p is important at early stages in ai5 γ folding

The basic tail of Mss116p is a feature it shares with two other DEAD-box proteins, namely Ded1p and Cyt-19, which are both able to rescue splicing of ai5 γ *in vitro* or in a $\Delta mss116$ strain^{128,151,152}. This domain is proposed to be of importance for binding to structured RNA tethering the DEAD-box protein to its target¹⁹³. Deletion of the basic tail in Mss116p reduced its affinity to RNA, but also its RNA unwinding and annealing activities *in vitro*¹⁵². Specifically binding to the ai5 γ substrate was 10 times weaker compared to the wild type Mss116p protein. In contrast, binding to a short duplex RNA was unaltered. Surprisingly, in the same study both ai5 γ splicing and respiratory growth were unaffected by deletion of the basic tail in Mss116p *in vivo*. Our results do not correlate with these findings. Introducing *mss116*/ Δ 569-664 into *Saccharomyces cerevisiae*'s genome is expected to result in endogenous levels of Mss116p with a truncated C-terminal domain. This truncation of Mss116p leads to a strong reduction in cell growth on medium with a non-fermentable carbon source. In line with this observation, the *in vivo* splicing assay revealed that splicing of ai5 γ was reduced significantly to 10 % of the wild type splicing level. Thus, the truncated version of Mss116p causes a strong effect on ai5 γ splicing and in turn on respiratory growth. Interestingly, the degree of both effects correlates strongly with the effect of the truncation on Mss116p's K_D to ai5 γ , which is about 10-fold increased compared to the wild type protein¹⁵². This finding is further in line with the observed intracellular structure of ai5 γ in the DCTD strain, as ai5 γ is largely unfolded, adopting a

conformation comparable to the $\Delta mss116$ ⁶¹ and ATPase-deficient strains. The DMS modification pattern is virtually identical in these three strains, except that the level of enhancements/protectations is somewhat smaller in the Δ CTD strain, which might result from the fact that 10% of the ai5 γ population is able to reach the splicing-competent state despite the reduced binding affinity of truncated Mss116p lacking the basic tail. As with the ATPase-deficient strain a reduction in productive binding events of Mss116p to ai5 γ could explain the strong effect on the intron's structure. In the case of Mss116p/K158R this reduction results from the inability to dissociate from the RNA molecule once bound due to abolished ATPase activity necessary for protein recycling. On the other hand the truncated Mss116p variant is still able to recycle allowing rare binding events that result in the formation of native ai5 γ structure in some intron molecules, as Mss116p releases the RNA substrate after near-native folding^{195,201}.

The helicase activity of Mss116p plays an important role in the late steps of ai5 γ folding

The helicase activity of Mss116p was shown to depend on ATP binding but not ATP hydrolysis, although the latter is needed for recycling of the protein^{169,182}. The K158R mutation abolishes both of those activities¹⁸⁷ and it was proposed that in fact its ATP-dependent unwinding is necessary for the collapse of D1 and subsequent folding events^{181,190}. Following this line of arguments an unwinding-deficient strain should show the same phenotype as a strain in which Mss116p cannot hydrolyze ATP and is therefore also unable to perform RNA strand separation. Interestingly, the DMS modification pattern of ai5 γ in the unwinding-deficient strain Q412A differs vastly from the one seen in the ATPase-deficient strain K158R. Enhancements are mostly found at nucleotides that are involved in tertiary interactions that are formed during later stages of intron folding. Strong DMS modifications of nucleotides involved in the ζ - ζ' tertiary interactions indicate that docking of D5 does not occur in the Q412A strain. Strikingly, we also did not observe misfolding within the ai5 γ intron in the unwinding-deficient strain, which is in line with the model that ai5 γ 's folding follows a smooth energy landscape without kinetically trapped intermediates^{63,70,105,107–109}. Furthermore, ai5 γ

splicing was shown to be efficient in unwinding-deficient strains^{187,188}. The Q412A mutation, which exhibits wild type-like ATP binding and hydrolysis, shows only very little to no helicase activity¹⁸⁸. In our *in vivo* assay its splicing efficiency was reduced to one third relative to wild type Mss116p providing evidence that strand separation is actually important for folding of ai5 γ *in vivo*, although to a far lower degree than ATP binding and hydrolysis or the basic tail of Mss116p. In case of the *Tetrahymena* group I intron most of the RNA population ends up in a kinetically trapped structure, which has to be resolved to allow native structure formation²²². If non-native stable structures would be resolved by Mss116p in the folding process of ai5 γ , we should observe sites of protection in a unwinding-deficient strain. However, we could only see protections at single nucleotides, which rather suggests different nucleotide conformation than intron misfolding. Interestingly, it was shown that not only mispaired nonnative structures can lead to kinetically trapped intermediate structures. In the *Tetrahymena* group I intron premature formation of native tertiary interactions can also trap the intron in a non-native conformation¹⁰⁰. Such a kinetic trap cannot be detected with DMS chemical probing. The overall DMS pattern of D1 in the unwinding-deficient strain Q412A indicates that the κ - ζ element is formed and D1 is collapsed in a considerable fraction of molecules. Since D5 is not docked it is probable that native tertiary interactions are established too early, potentially blocking the docking site of D5 on D1. Based on our data it could be assumed that D3 has docked prematurely or that the z-anchor is in place before the ζ - ζ' contact could be established. Such kinetic traps that are caused by native contacts would have to be resolved by Mss116p in order to allow docking of D5. This could also explain as to why Mss116p enables formation of the introns near-native state even without ATP when enough protein molecules are present eliminating the importance of protein turn-over *in vitro*^{195,201}.

Resolving misfolded exonic substructures can be essential for self-splicing as shown before for group I introns^{206,223}. The correlation between length of exonic sequences and importance of the presence of ATP¹⁹⁰ indicates that *in vivo* with its exonic environment the ATPase or unwinding activity of Mss116p might be of higher relevance than *in vitro* with a construct with no or short exons. The influence of exonic sequences on ATP dependence of ai5 γ splicing *in vitro*^{194,195} gives rise to the question if Mss116p unwinds

structures that involve exonic and/or intronic sequences and are detrimental for ai5 γ folding. It can be assumed that after splicing the exons remain mostly unstructured which we indeed could confirm *in vivo* for the wild type strain. Most interestingly and also unexpectedly the very same is true for the exons in the ATPase- and unwinding-deficient strain excluding the proposed role of Mss116p in resolving kinetic traps that involve exonic sequences^{190,194,202}.

Mss116p is a multifunctional enzyme.

For a long time, it has been controversially discussed whether Mss116p facilitates ai5 γ folding and splicing by resolving kinetically trapped structures or by stabilizing an on-pathway intermediate^{187–190,194}. We showed that actually RNA-binding, ATP binding and hydrolysis and unwinding activity contribute to its mode of action in ai5 γ splicing. Following our results we propose that Mss116p binds its RNA substrate with the assistance of its basic tail. Frequent non-productive binding events are resolved by protein turnover, which is triggered by ATP hydrolysis. As a result of productive binding Mss116p stabilizes the κ – ζ folding control element triggering the D1 collapse. After D1 compaction Mss116p ensures with its unwinding activity that native structure elements are not prematurely formed. Subsequent to near-native folding Mss116p has to dissociate using its ATPase activity in order to allow the late steps in native structure formation of ai5 γ , eventually resulting in intron splicing and the mature Cox1 mRNA transcript.

5 References

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
2. Kim, S. H. & Rich, A. Single crystals of transfer RNA: an x-ray diffraction study. *Science* (80-.). **162**, 1381–1384 (1968).
3. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982).
4. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
5. Brion, P. & Westhof, E. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137 (1997).
6. Butcher, S. E. & Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **44**, 1302–11 (2011).
7. Broyde, S. & Hingerty, B. “A” forms of RNAs in single strands, duplexes and RNA-DNA hybrids. *Nucleic Acids Res.* **5**, 2729–41 (1978).
8. Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.* **38**, 221–243 (2005).
9. Correll, C. C. & Swinger, K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA* **9**, 355–363 (2003).
10. Vallurupalli, P. & Moore, P. B. The solution structure of the loop E region of the 5 S rRNA from spinach chloroplasts. *J. Mol. Biol.* **325**, 843–856 (2003).
11. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
12. Strobel, S. A., Adams, P. L., Stahley, M. R. & Wang, J. RNA kink turns to the left and to the right. *RNA* **10**, 1852–1854 (2004).

13. Hohng, S. *et al.* Conformational flexibility of four-way junctions in RNA. *J. Mol. Biol.* **336**, 69–79 (2004).
14. Lescoute, A., Leontis, N. B., Massire, C. & Westhof, E. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* **33**, 2395–2409 (2005).
15. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
16. Russell, R. *et al.* Rapid compaction during RNA folding. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4266–4271 (2002).
17. Cate, J. H. *et al.* Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
18. Bailor, M. H., Sun, X. & Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**, 202–206 (2010).
19. Almakarem, A. S. A. *et al.* Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.* **40**, 1407–23 (2012).
20. Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**, 339–343 (2001).
21. Bevilacqua, P. C. & Turner, D. H. Comparison of binding of mixed ribose-deoxyribose analogues of CUCU to a ribozyme and to GGAGAA by equilibrium dialysis: evidence for ribozyme specific interactions with 2' OH groups. *Biochemistry* **30**, 10632–10640 (1991).
22. Tamura, M. & Holbrook, S. R. Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.* **320**, 455–474 (2002).
23. Costa, M. & Michel, F. Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.* **16**, 3289–3302 (1997).

24. Pley, H. W., Flaherty, K. M. & McKay, D. B. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111–113 (1994).
25. Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285 (1995).
26. Shen, L. X. & Tinoco, I. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247**, 963–978 (1995).
27. Chang, K. Y. & Tinoco, I. Characterization of a “kissing” hairpin complex derived from the human immunodeficiency virus genome. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 8705–8709 (1994).
28. Draper, D. E., Grilley, D. & Soto, A. M. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 221–243 (2005).
29. Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335–343 (2004).
30. Conn, G. L., Gittis, A. G., Lattman, E. E., Misra, V. K. & Draper, D. E. A compact RNA tertiary structure contains a buried backbone-K⁺ complex. *J. Mol. Biol.* **318**, 963–973 (2002).
31. 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).
32. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. Metal ion catalysis in the Tetrahymena ribozyme reaction. *Nature* **361**, 85–88 (1993).
33. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6498–6502 (1993).
34. Sigel, R. K., Vaidya, A. & Pyle, A. M. Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**, 1111–6 (2000).
35. Gordon, P. M., Sontheimer, E. J. & Piccirilli, J. A. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA* **6**, 199–205 (2000).

36. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
37. Valadkhan, S. The spliceosome: a ribozyme at heart? *Biol. Chem.* **388**, 693–697 (2007).
38. Ferat, J. L. & Michel, F. Group II self-splicing introns in bacteria. *Nature* **364**, 358–361 (1993).
39. Toro, N. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environ. Microbiol.* **5**, 143–151 (2003).
40. Bonen, L. Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* **8**, 26–34 (2008).
41. Vallès, Y., Halanych, K. M. & Boore, J. L. Group II introns break new boundaries: Presence in a bilaterian's genome. *PLoS One* **3**, (2008).
42. Koonin, E. V. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct* **1**, 22 (2006).
43. Keating, K. S., Toor, N., Perlman, P. S. & Pyle, A. M. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**, 1–9 (2010).
44. Boeke, J. D. The unusual phylogenetic distribution of retrotransposons: a hypothesis. *Genome Res.* **13**, 1975–1983 (2003).
45. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38**, 1–35 (2004).
46. Coros, C. J. *et al.* Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol. Microbiol.* **56**, 509–524 (2005).
47. Michel, F., Kazuhiko, U. & Haruo, O. Comparative and functional anatomy of group II catalytic introns — a review. *Gene* **82**, 5–30 (1989).

48. Fedorova, O. & Zingler, N. Group II introns: structure, folding and splicing mechanism. *Biol. Chem.* **388**, 665–678 (2007).
49. Aizawa, Y., Xiang, Q., Lambowitz, A. M. & Pyle, A. M. The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell* **11**, 795–805 (2003).
50. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507 (2012).
51. Gordon, P. M., Fong, R. & Piccirilli, J. A. A second divalent metal ion in the group II intron reaction center. *Chem. Biol.* **14**, 607–612 (2007).
52. Lambowitz, A. M. & Belfort, M. Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62**, 587–622 (1993).
53. Sharp, P. A. Splicing of messenger RNA precursors. *Science* **235**, 766–771 (1987).
54. Van der Veen, R., Kwakman, J. H. & Grivell, L. A. Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *EMBO J.* **6**, 3827–3831 (1987).
55. Pyle, A. M. The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.* **45**, 215–232 (2010).
56. Podar, M., Chu, V. T., Pyle, A. M. & Perlman, P. S. Group II intron splicing *in vivo* by first-step hydrolysis. *Nature* **391**, 915–918 (1998).
57. Augustin, S., Müller, M. W. & Schweyen, R. J. Reverse self-splicing of group II intron RNAs *in vitro*. *Nature* **343**, 383–386 (1990).
58. Roitzsch, M. & Pyle, A. M. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA* **15**, 473–482 (2009).
59. Morl, M., Niemer, I. & Schmelzer, C. New reactions catalyzed by a group II intron ribozyme with RNA and DNA substrates. *Cell* **70**, 803–810 (1992).

60. Matsuura, M., Noah, J. W. & Lambowitz, A. M. Mechanism of maturase-promoted group II intron splicing. *EMBO J.* **20**, 7259–7270 (2001).
61. Liebeg, A., Mayer, O. & Waldsich, C. DEAD-box protein facilitated RNA folding in vivo. *RNA Biol.* **7**, 803–11
62. Cech, T. R. Conserved sequences and structures of group I introns: building an active site for RNA catalysis--a review. *Gene* **73**, 259–271 (1988).
63. Swisher, J., Duarte, C. M., Su, L. J. & Pyle, A. M. Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.* **20**, 2051–2061 (2001).
64. Boudvillain, M., de Lencastre, A. & Pyle, A. M. A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature* **406**, 315–318 (2000).
65. Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330**, 197–209 (2003).
66. Fedorova, O. & Pyle, A. M. Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3. *EMBO J.* **24**, 3906–3916 (2005).
67. Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091–7104 (1998).
68. Konforti, B. B., Liu, Q. & Pyle, A. M. A map of the binding site for catalytic domain 5 in the core of a group II intron ribozyme. *EMBO J.* **17**, 7105–7117 (1998).
69. Somarowthu, S., Legiewicz, M., Keating, K. S. & Pyle, A. M. Visualizing the ai5 γ group IIB intron. *Nucleic Acids Res.* 1–12 (2013). doi:10.1093/nar/gkt1051
70. Su, L. J., Waldsich, C. & Pyle, A. M. An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res.* **33**, 6674–6687 (2005).
71. Koch, J. L., Boulanger, S. C., Dib-Hajj, S. D., Hebbar, S. K. & Perlman, P. S. Group II introns deleted for multiple substructures retain self-splicing activity. *Mol. Cell. Biol.* **12**, 1950–1958 (1992).

72. Qin, P. Z. & Pyle, A. M. The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.* **8**, 301–308 (1998).
73. Michel, F. & Jacquier, A. Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 201–12 (1987).
74. Hamill, S. & Pyle, A. M. The Receptor for Branch-Site Docking within a Group II Intron Active Site. *Mol. Cell* **23**, 831–840 (2006).
75. Toor, N. *et al.* Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. *RNA* **16**, 57–69 (2010).
76. Nagaswamy, U. & Fox, G. E. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **8**, 1112–9 (2002).
77. Waldsich, C. & Pyle, A. M. A folding control element for tertiary collapse of a group II intron ribozyme. *Nat. Struct. Mol. Biol.* **14**, 37–44 (2007).
78. Waldsich, C. & Pyle, A. M. A Kinetic Intermediate that Regulates Proper Folding of a Group II Intron RNA. *J. Mol. Biol.* **375**, 572–580 (2008).
79. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213**, 437–447 (1990).
80. Podar, M., Dib-Hajj, S. & Perlman, P. S. A UV-induced, Mg(2+)-dependent crosslink traps an active form of domain 3 of a self-splicing group II intron. *RNA* **1**, 828–40 (1995).
81. Griffin, E. A., Qin, Z., Michels, W. J. & Pyle, A. M. Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem. Biol.* **2**, 761–70 (1995).
82. Xiang, Q., Qin, P. Z., Michels, W. J., Freeland, K. & Pyle, A. M. Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37**, 3839–49 (1998).

83. Toor, N., Hausner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**, 1142–52 (2001).
84. Jestin, J. L., Dème, E. & Jacquier, A. Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J.* **16**, 2945–2954 (1997).
85. Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. *RNA* **14**, 1048–1056 (2008).
86. Watanabe, K. & Lambowitz, A. M. High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. *RNA* **10**, 1433–1443 (2004).
87. Schmidt, U., Podar, M., Stahl, U. & Perlman, P. S. Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing *in vitro* and in vivo: phenotypes and suppressor mutations. *RNA* **2**, 1161–1172 (1996).
88. Peebles, C. L., Zhang, M., Perlman, P. S. & Franzen, J. S. Catalytically critical nucleotide in domain 5 of a group II intron. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 4422–6 (1995).
89. Podar, M., Dib-Hajj, S. & Perlman, P. S. A UV-induced, Mg(2+)-dependent crosslink traps an active form of domain 3 of a self-splicing group II intron. *RNA* **1**, 828–840 (1995).
90. Griffin, E. A., Qin Zhifeng, Michels, W. J. & Pyle, A. M. Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem. Biol.* **2**, 761–770 (1995).
91. Xiang, Q., Qin, P. Z., Michels, W. J., Freeland, K. & Pyle, A. M. Sequence specificity of a group II intron ribozyme: Multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37**, 3839–3849 (1998).
92. Peebles, C. L. *et al.* A self-splicing RNA excises an intron lariat. *Cell* **44**, 213–223 (1986).

93. Van der Veen, R. *et al.* Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell* **44**, 225–234 (1986).
94. Schmelzer, C. & Schweyen, R. J. Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell* **46**, 557–565 (1986).
95. De Lencastre, A., Hamill, S. & Pyle, A. M. A single active-site region for a group II intron. *Nat. Struct. Mol. Biol.* **12**, 626–627 (2005).
96. Podar, M. & Perlman, P. S. Photocrosslinking of 4-thio uracil-containing RNAs supports a side-by-side arrangement of domains 5 and 6 of a group II intron. *RNA* **5**, 318–329 (1999).
97. Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15**, 3466–3476 (1996).
98. Batey, R., Rambo, R. & Doudna, J. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.* **38**, 2326–2343 (1999).
99. Pan, J., Thirumalai, D. & Woodson, S. A. Folding of RNA involves parallel pathways. *J. Mol. Biol.* **273**, 7–13 (1997).
100. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. Kinetic intermediates trapped by native interactions in RNA folding. *Science* **279**, 1943–1946 (1998).
101. Lagerbauer, B., Murphy, F. L. & Cech, T. R. Two major tertiary folding transitions of the Tetrahymena catalytic RNA. *EMBO J.* **13**, 2669–76 (1994).
102. Russell, R. & Herschlag, D. New pathways in folding of the Tetrahymena group I RNA enzyme. *J. Mol. Biol.* **291**, 1155–67 (1999).
103. Pan, T., Fang, X. & Sosnick, T. Pathway modulation, circular permutation and rapid RNA folding under kinetic control. *J. Mol. Biol.* **286**, 721–31 (1999).
104. Fang, X.-W., Thiyagarajan, P., Sosnick, T. R. & Pan, T. The rate-limiting step in the folding of a large ribozyme without kinetic traps. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8518–8523 (2002).

105. Su, L. J., Brenowitz, M. & Pyle, A. M. An alternative route for the folding of large RNAs: Apparent two-state folding by a group II intron ribozyme. *J. Mol. Biol.* **334**, 639–652 (2003).
106. Fedorova, O., Waldsich, C. & Pyle, A. M. Group II Intron Folding under Near-physiological Conditions: Collapsing to the Near-native State. *J. Mol. Biol.* **366**, 1099–1114 (2007).
107. Swisher, J. F., Su, L. J., Brenowitz, M., Anderson, V. E. & Pyle, A. M. Productive folding to the native state by a group II intron ribozyme. *J. Mol. Biol.* **315**, 297–310 (2002).
108. Pyle, A. M., Fedorova, O. & Waldsich, C. Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.* **32**, 138–145 (2007).
109. Liebeg, A., Mayer, O. & Waldsich, C. DEAD-box protein facilitated RNA folding in vivo. *RNA Biol.* **7**, 803–811 (2010).
110. Schroeder, R., Grossberger, R., Pichler, A. & Waldsich, C. RNA folding in vivo. *Curr. Opin. Struct. Biol.* **12**, 296–300 (2002).
111. Zemora, G. & Waldsich, C. RNA folding in living cells. *RNA Biol.* **7**, 634–641 (2010).
112. Cao, Y. & Woodson, S. A. Refolding of rRNA exons enhances dissociation of the Tetrahymena intron. *RNA* **6**, 1248–1256 (2000).
113. 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**, 557–567 (1999).
114. Gregan, J., Kolisek, M. & Schweyen, R. J. Mitochondrial Mg(2+) homeostasis is critical for group II intron splicing in vivo. *Genes Dev.* **15**, 2229–2237 (2001).
115. Herschlag, D. RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).
116. Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **4**, 118–130 (2007).

117. Waldsich, C., Grossberger, R. & Schroeder, R. RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo. *Genes Dev.* **16**, 2300–2312 (2002).
118. Mayer, O., Rajkowitsch, L., Lorenz, C., Konrat, R. & Schroeder, R. RNA chaperone activity and RNA-binding properties of the E. coli protein StpA. *Nucleic Acids Res.* **35**, 1257–1269 (2007).
119. Adilakshmi, T., Bellur, D. L. & Woodson, S. A. Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature* **455**, 1268–72 (2008).
120. 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).
121. Caprara, M. G., Lehnert, V., Lambowitz, A. M. & Westhof, E. A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. *Cell* **87**, 1135–45 (1996).
122. Dai, L. *et al.* A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol. Cell* **30**, 472–85 (2008).
123. 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).
124. Talkington, M. W. T., Siuzdak, G. & Williamson, J. R. An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628–32 (2005).
125. Webb, A. E. & Weeks, K. M. A collapsed state functions to self-chaperone RNA folding into a native ribonucleoprotein complex. *Nat. Struct. Biol.* **8**, 135–40 (2001).
126. Mohr, G., Zhang, A., Gianelos, J. A., Belfort, M. & Lambowitz, A. M. The *Neurospora* CYT-18 protein suppresses defects in the phage T4 td intron by

- stabilizing the catalytically active structure of the intron core. *Cell* **69**, 483–494 (1992).
127. Webb, A. E. & Weeks, K. M. A collapsed state functions to self-chaperone RNA folding into a native ribonucleoprotein complex. *Nat. Struct. Biol.* **8**, 135–140 (2001).
 128. Huang, H.-R. *et al.* The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 163–168 (2005).
 129. Séraphin, B., Boulet, A., Simon, M. & Faye, G. Construction of a yeast strain devoid of mitochondrial introns and its use to screen nuclear genes involved in mitochondrial splicing. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6810–6814 (1987).
 130. Jankowsky, E. RNA helicases at work: Binding and rearranging. *Trends Biochem. Sci.* **36**, 19–29 (2011).
 131. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 helicases: Family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
 132. Abdelhaleem, M. Do human RNA helicases have a role in cancer? *Biochim. Biophys. Acta - Rev. Cancer* **1704**, 37–46 (2004).
 133. Hanada, K. & Hickson, I. D. Molecular genetics of RecQ helicase disorders. *Cell. Mol. Life Sci.* **64**, 2306–2322 (2007).
 134. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
 135. Jankowsky, E. & Fairman, M. E. Duplex unwinding and RNP remodeling with RNA helicases. *Methods Mol. Biol.* **488**, 343–355 (2008).
 136. Caruthers, J. M. & McKay, D. B. Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123–133 (2002).
 137. Linder, P. Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res.* **34**, 4168–4180 (2006).

138. Linder, P. & Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**, 505–516 (2011).
139. Caruthers, J. M., Johnson, E. R. & McKay, D. B. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13080–13085 (2000).
140. Story, R. M., Li, H. & Abelson, J. N. Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1465–1470 (2001).
141. Shi, H., Cordin, O., Minder, C. M., Linder, P. & Xu, R.-M. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17628–17633 (2004).
142. Von Moeller, H., Basquin, C. & Conti, E. The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat. Struct. Mol. Biol.* **16**, 247–254 (2009).
143. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural Basis for RNA Unwinding by the DEAD-Box Protein *Drosophila* Vasa. *Cell* **125**, 287–300 (2006).
144. Andersen, C. B. F. *et al.* Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313**, 1968–1972 (2006).
145. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The Crystal Structure of the Exon Junction Complex Reveals How It Maintains a Stable Grip on mRNA. *Cell* **126**, 713–725 (2006).
146. Del Campo, M. & Lambowitz, A. M. Structure of the Yeast DEAD Box Protein Mss116p Reveals Two Wedges that Crimp RNA. *Mol. Cell* **35**, 598–609 (2009).
147. Nielsen, K. H. *et al.* Mechanism of ATP turnover inhibition in the EJC. *RNA* **15**, 67–75 (2009).
148. Fan, J. S. *et al.* Solution and Crystal Structures of mRNA Exporter Dbp5p and Its Interaction with Nucleotides. *J. Mol. Biol.* **388**, 1–10 (2009).

149. Hardin, J. W., Hu, Y. X. & McKay, D. B. Structure of the RNA binding domain of a DEAD-box helicase bound to its ribosomal RNA target reveals a novel mode of recognition by an RNA recognition motif. *J. Mol. Biol.* **402**, 412–427 (2010).
150. Rudolph, M. G. & Klostermeier, D. The *Thermus thermophilus* DEAD box helicase Hera contains a modified RNA recognition motif domain loosely connected to the helicase core. *RNA* **15**, 1993–2001 (2009).
151. Grohman, J. K. *et al.* Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry* **46**, 3013–3022 (2007).
152. Mohr, G. *et al.* Function of the C-terminal Domain of the DEAD-box Protein Mss116p Analyzed *in vivo* and *in vitro*. *J. Mol. Biol.* **375**, 1344–1364 (2008).
153. Tanner, N. K. & Linder, P. DExD/H box RNA helicases: From generic motors to specific dissociation functions. *Mol. Cell* **8**, 251–262 (2001).
154. Pyle, A. M. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* **37**, 317–336 (2008).
155. Zhang, Z.-M. *et al.* Crystal Structure of Prp5p Reveals Interdomain Interactions that Impact Spliceosome Assembly. *Cell Rep.* **5**, 1269–1278 (2013).
156. Chen, H.-C., Tseng, C.-K., Tsai, R.-T., Chung, C.-S. & Cheng, S.-C. Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. *Mol. Cell. Biol.* **33**, 514–25 (2013).
157. Rocak, S., Emery, B., Tanner, N. K. & Linder, P. Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. *Nucleic Acids Res.* **33**, 999–1009 (2005).
158. Blum, S. *et al.* ATP hydrolysis by initiation factor 4A is required for translation initiation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 7664–7668 (1992).
159. Huang, Y. & Liu, Z.-R. The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. *J. Biol. Chem.* **277**, 12810–12815 (2002).

160. O'Day, C. L., Dalbadie-McFarland, G. & Abelson, J. The *Saccharomyces cerevisiae* Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J. Biol. Chem.* **271**, 33261–33267 (1996).
161. Oh, J. Y. & Kim, J. ATP hydrolysis activity of the DEAD box protein Rok1p is required for *in vivo* ROK1 function. *Nucleic Acids Res.* **27**, 2753–2759 (1999).
162. Mallam, A. L., Del Campo, M., Gilman, B., Sidote, D. J. & Lambowitz, A. M. Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* **490**, 121–125 (2012).
163. Rudolph, M. G., Heissmann, R., Wittmann, J. G. & Klostermeier, D. Crystal Structure and Nucleotide Binding of the *Thermus thermophilus* RNA Helicase Hera N-terminal Domain. *J. Mol. Biol.* **361**, 731–743 (2006).
164. Schütz, P. *et al.* Comparative structural analysis of human DEAD-Box RNA helicases. *PLoS One* **5**, 1–11 (2010).
165. Henn, A., Cao, W., Hackney, D. D. & De La Cruz, E. M. The ATPase Cycle Mechanism of the DEAD-box rRNA Helicase, DbpA. *J. Mol. Biol.* **377**, 193–205 (2008).
166. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* **37**, 2180–93 (1998).
167. Iost, I., Dreyfus, M. & Linder, P. Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. *J. Biol. Chem.* **274**, 17677–83 (1999).
168. Talavera, M. A. & De La Cruz, E. M. Equilibrium and kinetic analysis of nucleotide binding to the DEAD-box RNA helicase DbpA. *Biochemistry* **44**, 959–70 (2005).
169. Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20209–20214 (2008).

170. Iost, I., Dreyfus, M. & Linder, P. Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. *J. Biol. Chem.* **274**, 17677–17683 (1999).
171. Henn, A., Bradley, M. J. & De La Cruz, E. M. ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annu. Rev. Biophys.* **41**, 247–67 (2012).
172. Henn, A. *et al.* Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase, DbpA. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4046–50 (2010).
173. Collins, R. *et al.* The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J. Biol. Chem.* **284**, 10296–10300 (2009).
174. Yang, Q. & Jankowsky, E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol.* **13**, 981–986 (2006).
175. Dürr, H., Flaus, A., Owen-Hughes, T. & Hopfner, K.-P. Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Res.* **34**, 4160–4167 (2006).
176. Wang, Q., Arnold, J. J., Uchida, A., Raney, K. D. & Cameron, C. E. Phosphate release contributes to the rate-limiting step for unwinding by an RNA helicase. *Nucleic Acids Res.* **38**, 1312–1324 (2010).
177. Bizebard, T., Ferlenghi, I., Iost, I. & Dreyfus, M. Studies on three *E. coli* DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* **43**, 7857–7866 (2004).
178. Chamot, D., Colvin, K. R., Kujat-Choy, S. L. & Owtrim, G. W. RNA structural rearrangement via unwinding and annealing by the cyanobacterial RNA helicase, CrhR. *J. Biol. Chem.* **280**, 2036–2044 (2005).
179. Rogers, G. W., Richter, N. J. & Merrick, W. C. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.* **274**, 12236–12244 (1999).

180. Rogers, G. W., Lima, W. F. & Merrick, W. C. Further characterization of the helicase activity of eIF4A. Substrate specificity. *J. Biol. Chem.* **276**, 12598–12608 (2001).
181. Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. DEAD-Box Proteins Unwind Duplexes by Local Strand Separation. *Mol. Cell* **28**, 253–263 (2007).
182. Chen, Y. *et al.* DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20203–20208 (2008).
183. 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**, 835–855 (2007).
184. Uhlmann-Schiffler, H., Jalal, C. & Stahl, H. Ddx42p--a human DEAD box protein with RNA chaperone activities. *Nucleic Acids Res.* **34**, 10–22 (2006).
185. Yang, Q. & Jankowsky, E. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**, 13591–13601 (2005).
186. Séraphin, B., Simon, M., Boulet, A. & Faye, G. Mitochondrial splicing requires a protein from a novel helicase family. *Nature* **337**, 84–87 (1989).
187. Solem, A., Zingler, N. & Pyle, A. M. A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol. Cell* **24**, 611–617 (2006).
188. Bifano, A. L., Turk, E. M. & Caprara, M. G. Structure-guided mutational analysis of a yeast DEAD-Box protein involved in mitochondrial RNA splicing. *J. Mol. Biol.* **398**, 429–443 (2010).
189. Del Campo, M. *et al.* Do DEAD-Box Proteins Promote Group II Intron Splicing without Unwinding RNA? *Mol. Cell* **28**, 159–166 (2007).

190. Potratz, J. P., Del Campo, M., Wolf, R. Z., Lambowitz, A. M. & Russell, R. ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing *in vitro* and *in vivo*. *J. Mol. Biol.* **411**, 661–679 (2011).
191. Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. The DEAD-box protein family of RNA helicases. *Gene* **367**, 17–37 (2006).
192. Mohr, G. *et al.* High-throughput genetic identification of functionally important regions of the yeast DEAD-box protein Mss116p. *J. Mol. Biol.* **413**, 952–972 (2011).
193. Mallam, A. L. *et al.* Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12254–12259 (2011).
194. Zingler, N., Solem, A. & Pyle, A. M. Dual roles for the Mss116 cofactor during splicing of the ai5 γ group II intron. *Nucleic Acids Res.* **38**, 6602–6609 (2010).
195. Fedorova, O., Solem, A. & Pyle, A. M. Protein-Facilitated Folding of Group II Intron Ribozymes. *J. Mol. Biol.* **397**, 799–813 (2010).
196. Ramaswamy, P. & Woodson, S. A. Global Stabilization of rRNA Structure by Ribosomal Proteins S4, S17, and S20. *J. Mol. Biol.* **392**, 666–677 (2009).
197. Ramaswamy, P. & Woodson, S. A. S16 throws a conformational switch during assembly of 30S 5' domain. *Nat. Struct. Mol. Biol.* **16**, 438–445 (2009).
198. Pan, J., Thirumalai, D. & Woodson, S. A. Magnesium-dependent folding of self-splicing RNA: exploring the link between cooperativity, thermodynamics, and kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6149–6154 (1999).
199. Rook, M. S., Treiber, D. K. & Williamson, J. R. An optimal Mg(2+) concentration for kinetic folding of the tetrahymena ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12471–12476 (1999).
200. Bhaskaran, H. & Russell, R. Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* **449**, 1014–1018 (2007).

201. Karunatilaka, K. S., Solem, A., Pyle, A. M. & Rueda, D. Single-molecule analysis of Mss116-mediated group II intron folding. *Nature* **467**, 935–939 (2010).
202. Fedorova, O. & Pyle, A. M. The brace for a growing scaffold: Mss116 protein promotes RNA folding by stabilizing an early assembly intermediate. *J. Mol. Biol.* **422**, 347–365 (2012).
203. Liebeg, A. & Waldsich, C. Probing RNA structure within living cells. *Methods Enzymol.* **468**, 219–238 (2009).
204. Del Campo, M. *et al.* Unwinding by Local Strand Separation Is Critical for the Function of DEAD-Box Proteins as RNA Chaperones. *J. Mol. Biol.* **389**, 674–693 (2009).
205. Sachsenmaier, N. & Waldsich, C. Mss116p: a DEAD-box protein facilitates RNA folding. *RNA Biol.* **10**, 71–82 (2013).
206. Woodson, S. A. & Cech, T. R. Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the Tetrahymena intervening sequence RNA. *Biochemistry* **30**, 2042–2050 (1991).
207. Toulmay, A. & Schneiter, R. A two-step method for the introduction of single or multiple defined point mutations into the genome of *Saccharomyces cerevisiae*. *Yeast* **23**, 825–831 (2006).
208. Cao, W. *et al.* Mechanism of Mss116 ATPase reveals functional diversity of DEAD-Box proteins. *J. Mol. Biol.* **409**, 399–414 (2011).
209. Treiber, D. K. & Williamson, J. R. Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **11**, 309–314 (2001).
210. Sosnick, T. R. Kinetic barriers and the role of topology in protein and RNA folding. *Protein Sci.* **17**, 1308–1318 (2008).
211. Shcherbakova, I., Mitra, S., Laederach, A. & Brenowitz, M. Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr. Opin. Chem. Biol.* **12**, 655–666 (2008).

212. Baird, N. J., Fang, X.-W., Srividya, N., Pan, T. & Sosnick, T. R. Folding of a universal ribozyme: the ribonuclease P RNA. *Q. Rev. Biophys.* **40**, 113–161 (2007).
213. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5**, 908–919 (2004).
214. Sosnick, T. R. & Pan, T. RNA folding: Models and perspectives. *Curr. Opin. Struct. Biol.* **13**, 309–316 (2003).
215. Treiber, D. K. & Williamson, J. R. Exposing the kinetic traps in RNA folding
Daniel K Treiber* and James R Williamson~. *Curr. Opin. Struct. Biol.* **9**, 339–345 (1999).
216. Woodson, S. A. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* **57**, 796–808 (2000).
217. Woodson, S. A. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **39**, 61–77 (2010).
218. Brehm, S. L. & Cech, T. R. Fate of an intervening sequence ribonucleic acid: excision and cyclization of the Tetrahymena ribosomal ribonucleic acid intervening sequence in vivo. *Biochemistry* **22**, 2390–2397 (1983).
219. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 161–175 (2006).
220. Heilman-Miller, S. L. & Woodson, S. A. Effect of transcription on folding of the Tetrahymena ribozyme. *RNA* **9**, 722–733 (2003).
221. Jackson, S. A., Koduvayur, S. & Woodson, S. A. Self-splicing of a group I intron reveals partitioning of native and misfolded RNA populations in yeast. *RNA* **12**, 2149–2159 (2006).
222. Woodson, S. A. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* **57**, 796–808 (2000).

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

6 Appendix

6.1 Publications

Chemical probing of RNA in living cells.

Wildauer M, Zemora G, Liebeg A, Heisig V, Waldsich C.

Methods Mol Biol. 2014;1086:159-76. doi: 10.1007/978-1-62703-667-2_9.

DEAD-box protein facilitated RNA folding in vivo.

Liebeg A, Mayer O, Waldsich C.

RNA Biol. 2010 Nov-Dec;7(6):803-11. Epub 2010 Nov 1.

Probing RNA structure within living cells.

Liebeg A, Waldsich C.

Methods Enzymol. 2009;468:219-38. doi: 10.1016/S0076-6879(09)68011-3. Epub 2009 Nov 17.

6.2 Curriculum vitae

Personal Data

Name: Andreas Liebeg
Geburtsdatum: 21.12.1978
Geburtsort: Aarau (CH)
Staatsbürgerschaft: Österreich
Familienstand: geschieden
Kinder: Lilly Goldmann, 5.3.2007
Hannah Goldmann, 5.11.2008

Ausbildung:

1989-1997 BG/BRG Fürstenfeld, Matura mit Auszeichnung
1997-1998 Auslandsdienst im Rahmen eines Gedenkdienstes bei dem Verein „Niemals Vergessen“ in der Gedenkstätte für Euthanasie, Bernburg (D)
Okt.1999- Juli 2003 Studium der Heilpädagogik / Behindertenpädagogik an der FH Zittau/Görlitz (D), Abschluss: Diplom-Heilpädagoge (FH)
Okt.2003-Sep.2009 Studium der molekularen Biologie an der Uni Wien.
Okt.2009-Feb.2014 Doktoratsstudium im Department für Biochemie, MFPL, Universität Wien.

Professioneller Werdegang:

Sozialbereich

1.Juni 2004 – 31.Dez.2005

Anstellung als Behindertenbetreuer in einer Wohngemeinschaft für Erwachsene mit Mehrfachbehinderung bei HABIT

1.Juni 2006 – 31.Okt.2008

Anstellung als Betreuer von Menschen mit psychischen Erkrankungen im Bereich ambulante Betreuung des Vereins LOK

Naturwissenschaftlicher Bereich

1.Juli 2009 – 30.Sep. 2009

Anstellung als CTA im Labor Waldsich, MFPL, Department für Biochemie, Biozentrum Bohrgasse

1.Okt 2009 – 28.02.2014

Anstellung als wissenschaftlicher Projektmitarbeiter im Rahmen des START Wissenschaftspreis im Labor Waldsich, MFPL, Department für Biochemie, Biozentrum Bohrgasse

Pädagogischer Bereich

Seit 28.08.2014

Anstellung als Lehrer (IIL/L1) in den Hertha Firnbergschulen für Wirtschaft und Tourismus (hlw22) für Naturwissenschaften (Chemie, Biologie).