Titel der Dissertation
„DEAD-box protein facilitated folding of the group II intron ai5γ 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: ai5γ.................................................................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 ai5γ in vivo.....................56
  3.4 The role of ATPase activity as well as RNA binding and unwinding of...............
    Mss116p in folding of ai5γ.........................................................................................58
    3.4.1 General approach.................................................................................................58
    3.4.2 The intracellular structure of ai5γ.......................................................................60
Figure Index

Figure 1.1: Influence of ions on RNA folding ................................................................. 13
Figure 1.2: Graphical representation of ai5γ ................................................................. 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 ai5γ 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 ai5γ folding ......................... 38
Figure 1.11: Mss116p-induced conformational changes in ai5γ ..................................... 40
Figure 2.1: Overview of the 2-step method for introducting 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 ai5γ 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 ai5γ ................................................. 63
Figure 3.6: DMS accessibility of the 5' part of the k−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 ai5γ in the K158R strain

Figure 3.10: DMS accessibility around stem d3 in the K158R strain

Figure 3.11: DMS accessibility around stem c1 in the K158R strain

Figure 3.12: DMS accessibility of J2/3 and D3 in the K158R strain

Figure 3.13: DMS accessibility of D4 in the K158R strain

Figure 3.14: DMS accessibility of D5 in the K158R strain

Figure 3.15: Differential summary map

Figure 3.16: DMS accessibility of the 5' part of the k–ζ element in the DXTD strain

Figure 3.17: DMS accessibility around the 3' part of the κ–ζ element in the ΔCTD strain

Figure 3.18: DMS accessibility around EBS2 and the coordination loop in the ΔCTD strain

Figure 3.19: DMS accessibility of the very 5' end of ai5γ in the ΔCTD strain

Figure 3.20: DMS accessibility around the d3 stem in the ΔCTD strain

Figure 3.21: DMS accessibility around the c1 stem in the ΔCTD strain

Figure 3.22: DMS accessibility of D5 in the ΔCTD strain

Figure 3.23: DMS accessibility of J2/3 and D3 in the ΔCTD strain

Figure 3.24: DMS accessibility of D4 in the ΔCTD strain

Figure 3.25: Differential summary map

Figure 3.26: DMS accessibility of the 5' part of the k–ζ element in the Q412A strain

Figure 3.27: DMS accessibility around the 3' part of the κ–ζ element in the Q412A strain

Figure 3.28: DMS accessibility around EBS2 in the Q412A strain

Figure 3.29: DMS accessibility of the very 5' end of ai5γ in the Q412A strain

Figure 3.30: DMS accessibility around the d3 stem in the Q412A strain

Figure 3.31: DMS accessibility around the c1 stem in the Q412A strain

Figure 3.32: DMS accessibility of J2/3 and D3 in the Q412A strain

Figure 3.33: DMS accessibility of D4 in the Q412A strain

Figure 3.34: DMS accessibility of D5 in the Q412A strain

Figure 3.35: Differential summary map
Figure 3.36: DMS accessibility of the exons.................................................................117
Figure 4.1: Comparison of DMS modification pattern of ai5γ in different genetic backgrounds...................................................................................................................121
Zusammenfassung


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 ai5γ remains largely unfolded if Mss116p is not present in vivo, but it is
enigmatic how Mss116p helps ai5γ 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 ai5γ in these different genetic backgrounds using an in
vivo chemical probing technique. The structure of ai5γ 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 ai5γ’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 configurations. 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 space 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. 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 regions. These RNA secondary structure elements are inherently stable which can lead to intermediate non-native structures interfering with native folding. 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 geometry\textsuperscript{7}. 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 nucleotides\textsuperscript{8}. The GNRA tetraloop is the most common and well studied type of closing loop and it plays an important role in tertiary structure stabilization\textsuperscript{9}. 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 kinking\textsuperscript{10}. 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\textsuperscript{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 helices\textsuperscript{13}. In three-way junctions, for example, it is proposed that coaxial stacking happens opposite of the longest stretch of junction nucleotides\textsuperscript{14}.

\subsection{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 ways\textsuperscript{15}. 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 \textit{in vitro} but tertiary interactions have to be established to stabilize this compact state\textsuperscript{16}.

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 structure\textsuperscript{17}. Coaxial stacking is mainly influenced by sequence but ionic conditions and junction topology also play a major role in its formation\textsuperscript{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 interactions\textsuperscript{19}. 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 adenosine\textsuperscript{20}. 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 structure\textsuperscript{21}. 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 kind\textsuperscript{22}. 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 GNRA\textsuperscript{23} and a receptor motif with different levels of structural complexity, tandem G-C pairs being a more simple example\textsuperscript{24}. 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 DV\textsuperscript{25}. 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 loop\textsuperscript{26}. Kissing loops are formed by Watson-Crick base pairing of two hairpins with complementary sequence in their closing loop\textsuperscript{27}. 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 formation\textsuperscript{28}. This ion atmosphere is created by monovalent ions, in particular K\textsuperscript{+}. 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\textsuperscript{+} are generally interpreted as diffuse ions and Mg\textsuperscript{2+} as to interact site specifically (Figure 1.1) but there are many different roles for different mono- and divalent ions in RNA structures\textsuperscript{29}. For example K\textsuperscript{+} was found to be also coordinated at tertiary structure level\textsuperscript{30} as well as divalent ions to be diffusely bound and exchanging rapidly\textsuperscript{29}. 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\textsuperscript{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 from\textsuperscript{28}.](image-url)
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 premRNA splicing by the spliceosome rely on their ribonucleic part for catalysis\textsuperscript{36,37}. Group II introns are some of the largest ribozymes in nature and found in bacteria\textsuperscript{38}, archaea\textsuperscript{39}, as well as the organelles of eukaryotic fungi and plants\textsuperscript{40} and even most recently in certain animals\textsuperscript{41}. Catalysis, structure and folding of this class of self-splicing ribozymes were studied intensely although mostly \textit{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 mechanism\textsuperscript{35}. As this class of ribozymes is found exclusively in organelles of eukaryotes they descend evolutionarily from endosymbionts\textsuperscript{42}. Fragmentation of the incorporated group II intron might have led to the spliceosome\textsuperscript{43}, components of the telomerase, LINE elements and most nuclear introns, thereby having a great influence on the human genome\textsuperscript{44}. Their ability to perform reverse splicing and therefore to act as mobile elements can accelerate genomic reorganizations\textsuperscript{45} and it was shown that environmental factors can influence group II intron mobility\textsuperscript{46}, hinting to an epigenetic influence of group II introns.

Group II introns are divided into three main families – group IIA, IIB and IIC\textsuperscript{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 families\textsuperscript{38}.

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 \textit{COXI} gene which is embedded in the mitochondrial genome of \textit{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 from 69.
intensely \textit{in vitro} rendering it a perfect candidate for studying \textit{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 \textit{in vivo}.

1.2.3 \textbf{Group II introns are capable of self-splicing}

Group II introns are able to promote forward as well as reverse splicing\textsuperscript{48} by catalyzing phosphodiester cleavage and ligation reactions in both RNA and DNA molecules\textsuperscript{49}. In this chemical mechanism a nucleophilic oxygen attacks the phosphate in the phosphodiester bond of the polynucleotide\textsuperscript{50}. During this process two metal ions coordinate both the 3'-OH leaving group and the attacking nucleophile in an intermediate state\textsuperscript{51}. 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 intron\textsuperscript{52}. Interestingly, the very same lariat structure as a result of intron splicing occurs in spliceosomal introns\textsuperscript{53}. Alternatively, the nucleophile can be the oxygen of a water molecule leading to a linear excised intron\textsuperscript{54}. This pathway has been observed in most group II introns that have been studied and is present both \textit{in vitro} and \textit{in vivo}\textsuperscript{54-56}. Both the lariat intron as well as the linear intron were shown to be able to perform reverse splicing\textsuperscript{57,58}, targeting not only RNA but also DNA molecules\textsuperscript{59}. In order to act as a mobile element a maturase, which is intron encoded, needs to be present\textsuperscript{60}. This protein typically consists of a RNA binding domain, a reverse transcriptase as well as an endonuclease domain\textsuperscript{65}. This maturase is not present in all group II introns, including the model RNA of this study ai5\gamma. Nevertheless it was shown that the excised ai5\gamma intron retains its structure \textit{in vivo}, a prerequisite of reverse splicing, underlining the importance of understanding group II intron folding\textsuperscript{61}. 
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 level\(^{47}\). 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 stacking\(^{62}\) 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 intron\(^{31}\). 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 ai5\(^g\) group II intron\(^{69}\).

**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 domains\(^{70}\). 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 D5\(^{71}\). D1 harbours two exon binding sites EBS1 and EBS2 which are complementary to the intron binding sites IBS1 and IBS2 in the 5'-exon\(^{72}\). In addition, EBS3 in D1 binds to the intron binding site IBS3 found in the 3'-exon\(^{73}\). Stabilized by the formation of the \(\beta–\beta'\) interaction\(^{38}\) 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 splicing\(^{74}\).

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 a\textsuperscript{75}. The T-loop is an often recurring five-nucleotide motif in noncoding RNA molecules that facilitates long-range tertiary interactions\textsuperscript{76}. The second essential folding step relates to the formation of the $\kappa$–$\zeta$ folding control element, which is a rate-limiting event in forming the native ai5$\gamma$ structure\textsuperscript{77,78}. The formation of the $\kappa$–$\zeta$ element might introduce a bend bringing the two lobes of D1 closer together, thereby allowing the establishment of the $\alpha$–$\alpha'$ and the $\beta$–$\beta'$ interactions\textsuperscript{55,77}. These two kissing loop interactions are important for the compaction of D1 and might also facilitate binding of the 5'-exonic substrate\textsuperscript{72}.

The catalytic center of ai5$\gamma$ is established by docking of D5 onto D1, both of which are essential and sufficient for the catalytic activity of group II introns\textsuperscript{71}. This docking is stabilized by the $\zeta$–$\zeta'$ and $\kappa$–$\kappa'$ tertiary contacts\textsuperscript{25,67}, whereas $\lambda$–$\lambda'$ and $\varepsilon$–$\varepsilon'$ are themselves required for catalysis, as these position D5 close to the 5'-exonic substrate\textsuperscript{64,79}.\textsuperscript{31} The tetraloop-receptor interaction 0–0' seems to add further stabilization to the native fold of group II introns\textsuperscript{23}. A motif most recently found in the crystal structure of the *Oceanobacillus iheyensis* group IIC intron, the so-called z-anchor, involves the $\lambda$–$\lambda'$ and $\varepsilon$–$\varepsilon'$ tertiary interactions, the internal loop of stem i, D5 and the 5' end of the intron\textsuperscript{75}. 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 activity\textsuperscript{72} though the basal stem with its important 0–0' 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 D3\textsuperscript{75}. 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 site\textsuperscript{65}. The second tertiary contact \( \eta-\eta' \) lies in the heart of D2\textsuperscript{23}. It is assumed to play an important role in establishing a silent conformation of D6\textsuperscript{55}.

**Domain III**

Domain III (D3) was shown to be a catalytic effector enhancing the rate of catalysis dramatically\textsuperscript{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 conservation\textsuperscript{83}. 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 ai5\textsuperscript{63}. This is in line with the earlier observation that this internal loop plays a major role in enhancing catalytic activity\textsuperscript{84}. 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 stacking\textsuperscript{55}. In group IIB introns D3 is extended forming the tertiary contact \( \mu-\mu' \) between the D3a terminal loop and D5, thereby fine tuning the \( \kappa-\kappa' \) interaction\textsuperscript{66}. Additionally, the single-stranded region between D3a and D3b was proposed to be part of the \( \theta-\theta' \) interaction site\textsuperscript{65}. In group IIB introns the conserved internal loop is close to the \( \varepsilon-\varepsilon' \) 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 \( \varepsilon-\varepsilon' \textsuperscript{85} \).

**Domain IV**

Domain IV (D4) is pointing away from the catalytic core in the group IIC intron of *Oceanobacillus iheyensis*\textsuperscript{31}, making it perfect for harboring open reading frames that encode a maturase\textsuperscript{86}. 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 domains and is positioned at the heart of group II introns. It is anchored in D1 by the very important tertiary contacts ζ–ζ' and κ–κ'25,67. The overall structure of D5 tremendously resembles that of the U6 snRNA of the spliceosome, suggesting little change in its evolution from bacteria to human (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 Mg2+ ions coordinate at this site, suggesting that these two metal ions participate in the catalytic reaction of group II introns50,55. 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 core 68. 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 splicing.92–94 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 site24,95. 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 D2. 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 needed. However, crosslinking studies showed that D5 and D6 are positioned close to each other before, during and after splicing. Thus, it has been proposed that the η-η' interaction might protect D6 from degradation or allows regulation of different splicing pathways.

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 RNA$^{103}$. 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-domain$^{104}$ 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 divers, 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 protected$^{63}$. The regions of protection appear all at the same Mg$^{2+}$ concentration, indicating cooperative folding$^{107}$. The level of needed counterions was very high and no regions of protections appeared before-hand. In analytical ultracentrifugation experiments the same Mg$^{2+}$ concentration that led to hydroxyl radical protection gave rise to compaction of D135$^{105}$. 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 happens$^{70}$. 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 folding\textsuperscript{108}. This could be explained by the appearance of kinetically trapped conformations that have to be resolved before native structure formation as seen for the \textit{Tetrahymena} group I intron\textsuperscript{100}. 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\textsuperscript{63,105}. Also no areas of internalization were observed at levels of Mg\textsuperscript{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\textsuperscript{70,105}. Strikingly also \textit{in vivo} there are no indications of misfolding when monitored by DMS chemical probing\textsuperscript{109}.

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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.4.png}
\caption{In vitro folding pathway of the D135 ribozyme. After addition of K\textsuperscript{+} secondary structure elements are formed (U). Addition of Mg\textsuperscript{2+} triggers the collapse of D1 (Ic) which leads to fast formation of the native structure (N). Figure was adapted from\textsuperscript{108}.}
\end{figure}

\textbf{Formation of the \kappa–\zeta folding control element is crucial for the collapse of ai5\gamma}

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 intron\textsuperscript{70}. 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\textsuperscript{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 Mg2+ in vitro and it was shown to form a divalent ion binding pocket78. Different concentrations of Mg2+ 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 ai5γ 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 ai5γ at near-physiological conditions**

Efficient folding of the group II intron ai5γ 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 Mg2+ native structure formation is triggered. Subsequent dilution of Mg2+ to a low concentration shifts natively folded ribozymes back to the near-native state, meaning that high Mg2+ concentration is needed for stabilization of the native state in vitro. Since the intracellular concentration of Mg2+ 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. Also the flanking sequence context differs *in vivo* and was shown to have an influence on the folding rate constants of RNA molecules 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 sequence.

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 Mg\(^{2+}\) transporter was shown to be important for folding of group II introns.

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. 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 activity, 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. For example the page T4 *td* intron core is stabilized in its native state by the Cyt18 splicing factor, contributing to the compactness of this group I intron. Cbp2 assists in native state formation of the bI5 intron by capturing the collapsed folding intermediate.

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*.
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\textsuperscript{130,131}. Defects in their function have been linked to different diseases as cancer, developmental defects or neurodegenerative diseases\textsuperscript{132,133}. Phylogenetic studies and comparative functional and structural analysis resulted in the classification of 6 superfamilies (SF) of helicases\textsuperscript{134}. All RNA helicases bind and hydrolyze ATP. Most but not all of them do also perform ATP dependent unwinding on appropriate substrates \textit{in vitro}\textsuperscript{135}, but these findings do not necessarily imply that this unwinding activity is part of the mechanistics of a helicase in its intracellular function.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mss116.png}
\caption{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 3I5X\textsuperscript{46}.}
\end{figure}
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 RecA. 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 superfamily.

The largest family of RNA helicases is part of the superfamily 2 with 37 members found alone in humans up until now. 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.

The two recA folds that build the catalytivc core remain flexible concerning their relative positions in the absence of bound ligands. Upon binding ATP and its target RNA the domains come into close proximity, forming two clefts on opposite sites of the helical core domains (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). 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. 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 lengths.
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\textsuperscript{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\textsuperscript{151,152}. In the case of Mss116p a truncated variant was used harboring part of the C-terminal domain\textsuperscript{146}. 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\textsuperscript{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\textsuperscript{155,156}. Most studies concerning biochemical features of DEAD-box proteins were done \textit{in vitro} and, if at all, than only with a small fraction of their intracellular complex cofactors\textsuperscript{138}.

**ATPase activity**

All members of the DEAD-box protein family that have been biochemically characterized \textit{in vitro} show ATP binding and hydrolysis activity\textsuperscript{157-161}. The two RecA-like domains are connected with a flexible linker and can move relative to each other if ATP and RNA are not bound\textsuperscript{154}. This is not case with processive helicases in which additional domains keep the helicase core domains in place\textsuperscript{154}. 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 molecules\textsuperscript{162}. 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\textsuperscript{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 triphosphate\textsuperscript{162}. These findings suggest that DI establishes the ATP-binding site in DEAD-box proteins\textsuperscript{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\textsuperscript{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\textsuperscript{146,147}. That the RNA substrate binds tightly to both DEAD-box helicase/ATP-analog complexes correlates nicely with these findings\textsuperscript{169}. Considering the very low affinity to RNA with bound ADP both structures most likely represent the protein after RNA strand separation\textsuperscript{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 activity\textsuperscript{169}. 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\textsuperscript{165,169,171,172}.

**RNA binding**

A double-stranded RNA molecule shows the same $K_d$ for DII and the DI/DII complex\textsuperscript{162}. 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 strand\textsuperscript{162}. 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 hypothesis^{146}. Ded1p and Mss116p are both able to load on an internal region of a RNA duplex and perform strand separation^{174} 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 4DB2^{162}. 

28
Unwinding activity

Although not all RNA helicases actually show unwinding activity\textsuperscript{175}, 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 RNA\textsuperscript{154}. Translocation and unwinding in these proteins depends on ATP binding and hydrolysis, whereby the phosphate release provides the energy for duplex unwinding\textsuperscript{130,154,176}. In contrast to those translocating SF1 or SF2 RNA helicases DEAD-box proteins that unwind RNA duplexes \textit{in vitro} do not show processivity\textsuperscript{174,177}. Consistently, unwinding by DEAD-box proteins is restricted to helices less than 25 nucleotides long\textsuperscript{159,178}. Additionally, the efficiency of unwinding decreases with increasing stability of the helix\textsuperscript{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 separation\textsuperscript{181}. 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 bound\textsuperscript{165}. 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 \textit{in vitro} with the bound non-hydrolyzable ATP analog ADP-BeFx with an efficiency modestly lower to that in the presence of ATP\textsuperscript{169}, showing that unwinding activity \textit{per se} is not dependent on hydrolysis of ATP. It was shown that the number of ATP molecules hydrolyzed by the DEAD-box protein \textit{Cyt19} depends on Mg\textsuperscript{2+} concentration and duplex length\textsuperscript{182}. At 2mM Mg\textsuperscript{2+} a 6 base pair long RNA duplex needed approximately half an ATP molecule for strand separation. For longer duplexes and/or higher [Mg\textsuperscript{2+}] 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 Mg\textsuperscript{2+} concentration and RNA duplex length with ATP molecules consumed was also tested for Mss116p and Ded1p, showing the same interdependence\textsuperscript{182}. 


The recently published crystal structure of Mss116p-DII bound to dsRNA gave insight into the unwinding mechanistics of DEAD-box proteins (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. 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*\(^1\text{83,184}\). Notably, the two DEAD-box proteins Ded1p and Mss116p are among the most potent strand annealers\(^1\text{83,185}\). 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 ssRNA\(^1\text{62}\). It is striking that those two proteins also share a distinct basic tail, distinguishing them from other members of the DEAD-box protein family\(^1\text{52}\).
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 proteins. Subsequently this protein was related to the RNA helicase family due to sequence homology. Mss116p is involved in splicing of all mitochondrial group I and group II introns as well as in mitochondrial translation and RNA processing. 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 ai5γ group II intron strain is not distinguishable from background noise in a Δmss116 whereas approximately 70-80% of the population is spliced in a wt strain (Figure 1.8).

Mss116p was shown to induce splicing of ai5γ in vitro under near-physiological conditions namely at 30°C and low Mg²⁺ concentration. The catalytic activity of ai5γ 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 ai5γ 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 from.
**Biochemical properties of Mss116p**

Mss116p binds group I and II introns non-specifically and with a similar $K_d$ ranging from 1 to 11 nM\(^{183}\). 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\(^{-1}\)\(^{188}\) and 56.7±13 min\(^{-1}\)\(^{187}\) to 112±6 min\(^{-1}\)\(^{183}\) and up to 168±10 min\(^{-1}\)\(^{189}\). 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 ATP\(^{183}\). In the presence of ATP Mss116p and Ded1p seem to promote a steady state between these two activities\(^{185}\).

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 RNA\(^{187}\) (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 retained\(^{187}\), however ATP binding was not tested\(^{188}\). This K158A mutation was not able to rescue the phenotype of a *mss116*-knock out strain, which cannot grow on non-fermentable carbon sources\(^{188}\). 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 separation. The conserved threonine 307 found in this motif interacts with a conserved histidine in motif VI 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. 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 duplex. 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 activity. 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 efficiency this was blamed on the long incubation time by the other and contrasted by an 8- to 28-fold reduction in splicing. Yeast cells harboring the very same double mutation were not able to grow on a non-fermentable carbon source, 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 experiments. 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 α5γ, 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.
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 substrate \(^{188}\) (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 ai5\(\gamma\) splicing resembles that of the wt protein \textit{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.

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 ai5\(\gamma\) splicing \(^{188}\). 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 ATP\textsuperscript{188}.

\textit{CTD}

Splicing of ai5\(\gamma\) can be efficiently rescued in a \(\Delta mss116\) strain by two other DEAD-box proteins namely the \textit{N. crassa} mitochondrial helicase Cyt19 and the \textit{S. cerevisiae} non-mitochondrial DEAD-box protein Ded1p\textsuperscript{183,187}. Structural comparison of Mss116p, Cyt19 and Ded1p show that only the first two proteins share the \(\alpha\)-helical C-terminal extension, whereas all three DEAD-box proteins have a characteristic basic tail at the very C-terminus\textsuperscript{152}. This tail was proposed to interact with structured RNA tethering the DEAD-box protein to its target molecule\textsuperscript{193}. Deletion of amino acids 569-664, which include the basic tail of Mss116p, do not interfere with respiratory growth of \textit{S. cerevisiae}\textsuperscript{152}. Splicing efficiency of ai5\(\gamma\) was reduced about 2 fold and the ATPase activity resembled that of wt protein \textit{in vitro}\textsuperscript{152}. In contrast, both unwinding and annealing activities were significantly reduced in the \(\Delta 569-664\) strain. The \(K_d\) of the truncated Ms116p to ai5\(\gamma\) 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 ai5γ 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).

![Figure 1.10: The two paradigms of Mss116p facilitated ai5γ folding.](image)

(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 from\(^{188}\).
Although it is generally assumed that Mss116p stimulates folding of ai5γ 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 ai5γ folding and can also be triggered by a [Mg\textsuperscript{2+}] of 100mM\textsuperscript{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 steps\textsuperscript{78}. 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 [Mg\textsuperscript{2+}] or a cofactor suggests that electrostatic properties play a crucial role in the folding pathway\textsuperscript{196,197}. Interestingly, compaction was efficient even without the addition of ATP\textsuperscript{195}, which in turn was found to be essential for duplex unwinding and protein recycling\textsuperscript{169}. 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 [Mg\textsuperscript{2+}] despite the presence of Mss116p\textsuperscript{183,195}. However, this is supposed to be exactly the opposite for kinetically trapped RNAs\textsuperscript{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 [Mg\textsuperscript{2+}]\textsuperscript{200}. Interestingly, Mss116p does not unfold the compact state of D1 RNA at low [Mg\textsuperscript{2+}]\textsuperscript{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 dependent\textsuperscript{201}. 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\textsuperscript{77,78}. Importantly, nucleotide analog interference mapping showed that Mss116p stabilizes this κ–ζ element, thereby
inducing a compact near-native fold of the ribozyme\textsuperscript{202}. In this near-native state the folding scaffold is maintained but docking of the other domains onto D1 is not stable and the active site center not established\textsuperscript{106,195,201}. Native structure formation seems to depend on ATP hydrolysis by Mss116p and is stabilized by binding of the exons to the intron binding sites.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Mss116p-induced conformational changes in ai5γ. The intron structure was mapped by in vivo structural chemical probing in a wild type and mss116-knockout strain. Differences in DMS accessibilities in both strains were plotted on a secondary structure map. Filled squares show enhanced and open squares reduced DMS accessibility in the knockout strain. Differences over two-fold are shown in red, differences 1.5-2 fold in gray\textsuperscript{109}.}
\end{figure}
To determine the effect of Mss116p on the structure of ai5γ in yeast cells we established DMS chemical probing in vivo\textsuperscript{109,203}. In the wt yeast strain ai5γ adopts a structure that highly resembles the in vitro fold of the D123456 ribozyme. In contrast, ai5γ could not adopt its tertiary fold without its cofactor although most but not all secondary structure elements were formed in vivo\textsuperscript{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 i stem which connects D1 to the central wheel in group II introns, is not formed in the mss116 knock out strain. As ai5γ is largely unfolded in the absence of the cofactor these findings suggest that Mss116p assists ai5γ 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 ai5γ folding\textsuperscript{183,189,190,204}. Ded1p and Cyt19 as mentioned above are able to stimulate splicing of ai5γ in vitro and in vivo\textsuperscript{70,183,187,204}, both of which were shown to destabilize misfolded structures of the Tetrahymena group I intron\textsuperscript{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\textsuperscript{152,193}. This domain was shown to play an important role in ai5γ binding\textsuperscript{152}. In contrast to the results stated above\textsuperscript{195} 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 ratio\textsuperscript{190}. 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 ai5γ folding. Additionally, the Mss116 S305A/T307A mutant shows a reduced ability to promote splicing of ai5γ\textsuperscript{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\textsuperscript{187,194}. By adding Mss116p to the pre-folded D135 with and without ATP it was shown that this ribozyme looses its splicing activity over time\textsuperscript{190}. 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. Interestingly, exonic sequences change the effect of Mss116p on ai5γ folding in vitro. Mss116p promotes splicing of ai5γ with long and short exons with a stronger stimulating effect with long exons. Additionally, the unwinding deficient Mss116p S305A/T307A mutant has a greatly reduced ability to induce splicing of ai5γ with long exons in vitro. Maybe the unwinding activity of Mss116p is needed to resolve exonic substructures that are detrimental for ai5γ native structure formation as observed for the Tetrahymena group I intron. 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 folding.

Although the ai5γ 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 ai5γ folding but also plays an important role in mitochondrial translation and RNA processing. 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 ai5γ 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>
<thead>
<tr>
<th>Table 1: Mutations introduced in MSS116.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>domain</strong></td>
</tr>
<tr>
<td>Motif I</td>
</tr>
<tr>
<td>Motif I</td>
</tr>
<tr>
<td>Motif Ib</td>
</tr>
<tr>
<td>QxxR</td>
</tr>
<tr>
<td>ΔCTD</td>
</tr>
</tbody>
</table>

MSS116 sequence

5’AGAGTTGCACGTTAGGCTGATAATGGTTGACCTCTATATTGATAAAAGGAGCTCGCACACTGTTCCTTGCAAGAGGAGAA
TTTATTAGCAAGCTACGTTGAGGACATATGCAATAATCTTTCATGAAAGGAGGAATTGCTGCTAAGGAAGGTTG
GCAAAGGTCTAGAGAAGACGATGATGAAGTGCACTTTGATAAGACCACTTTTTCCAAACTAATTCACGTCCCAAAAGA
AGATAATAGTAAGAGTAACTCCGCTGACTCTTACCTGGAGGAAAGGAGTCCCTTGAATAAGAATACATATAGACCGTACC
AGGAATTTGCCGTTTAAATCTCCTGACAGCAAGTTTTGCAAATTGGTGTTCCATCTGAGTTGGCTAATTA
TATTCATAGAATTGGTAGAACAGCCAGAAGTGGGAAAGAAGGATCTTCTGTGCTATTCATTTGTAAAGATGAATTGCC
ATTTGTTAGAAGTTGGAAAGTGGCAAAAAATATGCTAGTCCACAAAAACAAATGAGGACCGAGATTAGTT
AGTGGCTACCTCTGCCGCTGACAGGTAAAAGATGAGCGAAGTGGAGTTGGAAGATGCAAAAAATATAGTCATTGCCAAACA
AGAAAAATATGAGCCCAGTGAAGAAATAAACTGAGGTCCTAGAGGCAGTGACCGAAGAACCGGAAGACATATCAGATATTGTAA
AATCTGAGGTCCTAGAGGAGTTTGGGAAAAAATATTACATAATAATAAAAACAAATGGCAAGATGATGATGATTCAGAAATTTC
TTTCAGAGGCAACAAAAATATGATAATGGAGCTCCTTCTCCCTCCTCCCTCCTCCCTCCCTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCT
53
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.

<table>
<thead>
<tr>
<th>name</th>
<th>characteristics</th>
<th>Parent strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN16</td>
<td>GII-0 ai5γ, Matα adel1 lys1 ura3-51</td>
<td>a161</td>
</tr>
<tr>
<td>HRH189</td>
<td>MSS116Δ::kan</td>
<td>a161</td>
</tr>
<tr>
<td>yCW08</td>
<td>MSS116-3'UTR::CloNat</td>
<td>TN16</td>
</tr>
<tr>
<td>yCW01</td>
<td>MSS116::K158R, CloNat</td>
<td>TN16</td>
</tr>
<tr>
<td>yCW05</td>
<td>MSS116::Q412A, CloNat</td>
<td>TN16</td>
</tr>
<tr>
<td>yCW10</td>
<td>MSS116::stop codon at AA569/570, CloNat</td>
<td>TN16</td>
</tr>
<tr>
<td>yCW11</td>
<td>MSS116::R245E, CloNat</td>
<td>TN16</td>
</tr>
<tr>
<td>yCW12</td>
<td>MSS116::K158A, CloNat</td>
<td>TN16</td>
</tr>
</tbody>
</table>

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

Table 3: Primers used in MSS116 mutagenesis.

<table>
<thead>
<tr>
<th>name</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>yCW08F_1</td>
<td>TAACAAATTCCTCTATTAAGACGCTTGTATTTTAGCCCAATCTTTCCTGATTTCAATCTTT</td>
</tr>
<tr>
<td></td>
<td>CACAGCCTGGCTTGGCCGCCGG</td>
</tr>
<tr>
<td>yCW08R_1</td>
<td>AAAAGTAAGAATGGAGGTACCGGTCGTAAACTGACGATAAATATACCAGT</td>
</tr>
<tr>
<td></td>
<td>GGATCGACACTGGATGGCGGCGT</td>
</tr>
<tr>
<td>yCW01F</td>
<td>TATCGCAAGAGCTAAAACCGGTACAGGTAGCACATTGGCTTCTTGATTCGAAAATTCTGCAAAATAATATAAAGAACAAGTTTAG</td>
</tr>
<tr>
<td></td>
<td>TTAAAGATTC</td>
</tr>
<tr>
<td>yCW01R</td>
<td>CTTTGAGAGGAGGAAGAGCGAAGGGAATTGAGGAGGTACCGGTCGTAAATAAAGCATT</td>
</tr>
<tr>
<td></td>
<td>GAC</td>
</tr>
<tr>
<td>yCW05F</td>
<td>CCAAATTTTGGAATTTTCTGGTAAAAATTACTGCCAAATAAATAATAAAGAACAAGTTTAG</td>
</tr>
<tr>
<td></td>
<td>TTAAAGATTC</td>
</tr>
<tr>
<td>yCW10F</td>
<td>CCGAAATTTGCTCCCTATATGGTATGTTAAATGCTCTCAATTAGTAAAGGAGGAGGTACCGGTCGTAAATAAAGCATT</td>
</tr>
<tr>
<td></td>
<td>GATCCAAAGGG</td>
</tr>
<tr>
<td>yCW11F</td>
<td>CCAAACATTGGTAAATGGTCACCGCAGCGGAATTGATTTGGTGTTGGGAAATAATCGCC</td>
</tr>
<tr>
<td></td>
<td>AACAAATTTTTTATGTTCG</td>
</tr>
<tr>
<td>yCW12F</td>
<td>TATCGCAAGAGCTAAAACCGGTACAGGTACGGACATTTGCCTTTGATTCCAAATTTT</td>
</tr>
<tr>
<td></td>
<td>CCA</td>
</tr>
</tbody>
</table>

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 Schneiter\textsuperscript{207}. 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.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.1.png}
\caption{Overview of the 2-step method for introducing of point mutations in the yeast protein Mss116p.}
\end{figure}
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, 1µM 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.**

<table>
<thead>
<tr>
<th>name</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW01F</td>
<td>ACGCGTCGACATGTGACCTCTATATTGATAAAAG</td>
</tr>
<tr>
<td>CW01R</td>
<td>CGGAATTCTAATATATGTTGCTGTTTCTACTG</td>
</tr>
<tr>
<td>CW61F</td>
<td>ACCTCAGTTACTCACAGATG</td>
</tr>
<tr>
<td>CW61R</td>
<td>ATTGACGAATAATATACGACTTAG</td>
</tr>
<tr>
<td>yCW01-seq1</td>
<td>CCAGTGAAGACCATGATTTATCGC</td>
</tr>
<tr>
<td>yCW01-seq3</td>
<td>CCTAGAGGCGATGACCAGAAACCAG</td>
</tr>
<tr>
<td>CW01M</td>
<td>GACACAGTAGATAAGATGACCTGACATG</td>
</tr>
<tr>
<td>GCW01F</td>
<td>CTGGCGCAGAGGATGCTGATCG</td>
</tr>
<tr>
<td>GCW01R</td>
<td>GGTATTATACGATGAAAGTATTGAATGATGGGTC</td>
</tr>
</tbody>
</table>

**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 OD$_{600}$ 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 cells that 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.

<table>
<thead>
<tr>
<th>name</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sc-111</td>
<td>GCCCCAATGAAAAATTAATCAG</td>
</tr>
<tr>
<td>Sc-6</td>
<td>GAAATAGCCACCATTGATAATAC</td>
</tr>
<tr>
<td>Sc6</td>
<td>AAATATTATTTATGATAACTTTTCAGA</td>
</tr>
<tr>
<td>Sc82</td>
<td>TGTTACCATTATATACACCTTG</td>
</tr>
<tr>
<td>Sc160</td>
<td>GTTCTTCATCTTTTTTTTATATA</td>
</tr>
<tr>
<td>Sc234</td>
<td>GTTTTCAATTAGTGTTGTAAG</td>
</tr>
<tr>
<td>Sc312</td>
<td>TTCCAATACATAACATCAACC</td>
</tr>
<tr>
<td>Sc375</td>
<td>GTAATATCTAACTTAGCTCTC</td>
</tr>
<tr>
<td>Sc440</td>
<td>TTTCTTATCTCTAAAATTTTATA</td>
</tr>
<tr>
<td>Sc520</td>
<td>TTTCTTATCTCTAAAATTTTATA</td>
</tr>
<tr>
<td>Sc588</td>
<td>GCATTAGCTTTTTTATCAATC</td>
</tr>
<tr>
<td>Sc656</td>
<td>CAACCTATAGTAAAAAGTTACAG</td>
</tr>
<tr>
<td>Sc754</td>
<td>ATTTATCAGTATAATAAACCTCC</td>
</tr>
<tr>
<td>Sc805</td>
<td>CGGCTCAGCAATATC</td>
</tr>
<tr>
<td>Sc866</td>
<td>CCCGATAGGTAGACCTTTAC</td>
</tr>
</tbody>
</table>

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

\[ c \text{ (mol/L)} = \frac{A_{260}}{(\varepsilon_{\text{total}} \times \text{pathlength (cm)}) \times \text{dilution}} \]

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

**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 \( \gamma \)P\(^{32}\)-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 P\(^{32}\) 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 MgCl\(_2\), 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 use 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 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
pH 8.0, 45 µl TEMED, 450 µl 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 µl 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 phosphoimager screen (GE healthcare) for 1 to 3 days depending on the intensity of the signal and finally scanned on a Storm 820 phosphoimager (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 ai5γ 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 desribed before. 30 ml of YPD was inoculated with this preculture to an OD_{600} of 0.1. Cells were grown at 30 °C shaking with 180 rpm until the suspension reached an OD_{600} between 0.4 and 0.6; importantly the OD of all different strains has to be the same at time of harvest.
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 yeast\textsuperscript{128}. Its activities, ATP binding and hydrolysis as well as RNA unwinding and binding, were subject to numerous studies\textsuperscript{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 ai5γ we introduced different mutations into the \textit{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 recombination\textsuperscript{207}. The integrating PCR product harbors the desired mutation as well as a marker gene at the 3'-UTR of \textit{mss116}, in our case the cloNat resistance gene NAT1. By directly mutating the chromosomal copy of \textit{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 \textit{in vitro} studies cannot be neglected and high levels of the protein can for example lead to ai5γ splicing even in absence of an ATP substrate \textit{in vitro}\textsuperscript{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 \textit{in vitro} and \textit{in vivo} studies\textsuperscript{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 \textit{mss116}, named ΔCTD strain in which the distinct basic tail has been deleted\textsuperscript{152}. 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 \textit{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 \textit{nat1}
is integrated at the correct position in the 3'UTR of \textit{mss116} (Figure 3.1B). Expression and mitochondrial localization of the strains that were subject to \textit{in vivo} chemical probing (K158R, Q412A and ΔCTD) 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.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3_1.png}
\caption{Verification of correct location of integration cassette harboring the desired mutation of \textit{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.}
\end{figure}
3.2 Conserved motifs of Mss116p and their role in respiratory growth of yeast

The ai5γ intron is a perfect candidate for in vivo studies of RNA folding: First, it has been studied extensively in vitro. 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.](image)
poisoned primer assay *in vivo*. Last but not least only if ai5γ 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. The K158R mutation interferes with both leading to a loss of Mss116p's ability to assist in ai5γ splicing. 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 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. This mutation also disrupts ATP binding and hydrolysis. All three mutants were also shown to be unable to unwind a duplex RNA substrate.

The ΔCTD 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 activity. RNA binding was 10-fold decreased compared to wild type Mss116p. In our respiratory growth assay this yeast strain shows some aerobic 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 Δ569-664 strain displayed wild type-like growth. 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\textsuperscript{146,188}. In contrast to the previously mentioned mutations Q412A displays wild-type ATP binding and hydrolysis\textsuperscript{188}. 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 \textit{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 ai5γ \textit{in vivo}.

To monitor ai5γ splicing \textit{in vivo} we established a poisoned primer assay. Here the dNTP mix for reverse transcription of the \textit{cox1} group II intron ai5γ 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 ai5γ 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 ai5γ can be determined.
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 experiments\textsuperscript{109}. 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 levels\textsuperscript{152}. 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).

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

<table>
<thead>
<tr>
<th>Mss116p variant</th>
<th>Motif</th>
<th>Impaired activities</th>
<th>Relative splicing (1/k&lt;sub&gt;rel&lt;/sub&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>-</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Knockout</td>
<td>-</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>K158R</td>
<td>Motif I</td>
<td>ATP binding and hydrolysis, RNA unwinding</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>K158A</td>
<td>Motif I</td>
<td>ATP hydrolysis, RNA unwinding</td>
<td>0.02</td>
</tr>
<tr>
<td>R245E</td>
<td>Motif Ib</td>
<td>ATP binding and hydrolysis, RNA unwinding</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ΔCTD</td>
<td>Basic tail</td>
<td>RNA binding, RNA unwinding</td>
<td>0.1</td>
</tr>
<tr>
<td>Q412A</td>
<td>QxxR</td>
<td>RNA unwinding</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Relative splicing levels are determined by 1/k<sub>rel</sub>. k<sub>rel</sub> 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.
In order to obtain information on the structure of ai5γ \textit{in vivo} we established chemical probing of RNA molecules in yeast cells using dimethyl sulfate\textsuperscript{203}. 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\textsuperscript{69,75}.

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.
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 ai5γ 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 ai5γ

We have already shown that the DMS accessibility of nucleotides in ai5γ 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\textsuperscript{31,47,109}. The DMS pattern is also almost identical to that of the D123456 ribozyme in vitro\textsuperscript{109}. To be able to interpret the DMS modification pattern of ai5γ 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 ai5γ group II intron confirms previous results\textsuperscript{109}. Ai5γ 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 ai5γ in a wild type genetic background are in line with a fully catalytically active ribozyme. Since ai5γ splices efficiently in wt yeast, the chemical probing data reflects mostly already spliced intron molecules.
Figure 3.5: In vivo DMS modification pattern of ai5γ. Secondary structure map of ai5γ with tertiary interaction sites indicated by greek letters. Domains are labeled with roman numbers. Breaks are shown by black dots. DMS modification intensities are represented by squares from light (weak), medium (moderate) to dark (strong) yellow-green.
3.4.3 *Mss116p ATPase activity and ai5γ folding*

Disruption of the ATPase activity of Mss116p consequently leads to defective group II intron splicing\(^\text{187,188}\). 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 ATP\(^\text{195}\). 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\(^\text{146,169,170}\). 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 ATP\(^\text{201}\). 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 chain\(^\text{188}\).

The $\kappa$–$\zeta$ 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 $\kappa$–$\zeta$ element\(^\text{77}\). 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 catalysis\(^\text{31}\). 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 mutation\(^\text{187}\). The stems d, d' and d" that are part of the $\kappa$–$\zeta$ 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 $\alpha$–$\alpha'$ 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 $\beta$–$\beta'$ 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\textsuperscript{38,55,72,77}. Folding of the \( \kappa-\zeta \) 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 \( \kappa-\kappa' \) and \( \zeta-\zeta' \) interactions, play a fundamental role in establishing the active site\textsuperscript{25,67}. Enhanced DMS-accessibilities of A214 and A215 as well as A204 (Figure 3.6) show that the \( \kappa-\kappa' \) 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 \( \zeta-\zeta' \) tertiary contact is not established. These findings show that D5 is not docked into D1, an event which is triggered by D1 folding\textsuperscript{70} and absolutely crucial for catalysis.
Figure 3.6: DMS accessibility of the 5' part of the $\kappa-\zeta$ 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 $\beta$, $\zeta$, and $\kappa$. 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 backbone. Nevertheless several helices of D1 are not established in the presence of K⁺ ions alone *in vitro* but do only form after Mg²⁺ is added. The very same helices are also not formed *in vivo*, if Mss116p is not present to assist in ai5γ folding. 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 strain\textsuperscript{109} although this stem is formed \textit{in vitro} with only monovalent ions present\textsuperscript{78}. 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 \textit{in vitro}\textsuperscript{78}, were disrupted in the knockout strain\textsuperscript{109}. 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 ai5γ with its cofactor rendered unable to bind and hydrolyze ATP shows more disruptions of helical regions \textit{in vivo} than the D135 riboyme \textit{in vitro} without divalent ions. This might be due to the very high levels of K\textsuperscript{+} 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 ai5γ 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 elements\textsuperscript{61}. 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\textsuperscript{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\textsuperscript{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 pair\textsuperscript{15} at these sites in the native structure of ai5\gamma, 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\textsuperscript{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\gamma, showing that this DEAD-box protein is important for both secondary as well as tertiary structure formation of ai5\gamma\textsuperscript{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 \(\alpha-\alpha'\) and \(\beta-\beta'\), which are important for D1 collapse, are not in place in the K158R strain (Figure 3.6, Figure 3.8). The \(\theta-\theta'\) 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 elements\textsuperscript{75}, is formed.

Formation of the \(\kappa-\zeta\) 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 \(\zeta\) tetraloop receptor\textsuperscript{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 U381\textsuperscript{31}, is modified intensely (Figure 3.6). The \(\zeta-\zeta'\) interaction is crucial for docking of D5\textsuperscript{25} 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 \(\zeta'\) does not contact its receptor motif \(\zeta\) (Figure 3.14). A214-A215 as well as A204 are part of the \(\kappa-\kappa'\) tertiary interaction, which is the second crucial contact for docking of D5\textsuperscript{67}.
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.
These adenosines are modified intensely in the K158R strain, indicating that the $\kappa-\kappa'$ interaction does not form. Thus, D5 is not docked onto D1, an event that lays the foundation for active site formation$^{43}$. In the active site the so-called z-anchor region is
then established\textsuperscript{31}, in which $\lambda-\lambda'$ and $\varepsilon-\varepsilon'$ 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 $\lambda-\lambda'$ interaction that can be mapped with DMS, is forming a base triple with the C825-G836 pair\textsuperscript{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 $\lambda-\lambda'$ tertiary contact is not in place. C117 is part of the $\varepsilon-\varepsilon'$ 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\textgamma. 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\textsuperscript{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 helix\textsuperscript{43} is not established. A586 is enhanced in the K158R strain. In the native conformation of ai5\textgamma it should be protected, since it is contacting the backbone of G6 positioning the 5' end of the intron close to the dinucleotide bulge\textsuperscript{69}. Enhancement of A587 on the other hand shows that the $\gamma-\gamma'$ 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\textsuperscript{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 independent\textsuperscript{61}. 
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 ai5γ were only found in the effector D3\(^{61}\). Although in general dispensable for the catalytic activity of the ribozyme \textit{in vitro}\(^{71}\), D3 enhances the splicing activity tremendously\(^{65}\). Additionally, it contributes to the stability of the active site\(^{85}\). The \(\mu-\mu'\) 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 other\(^{66}\). 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 ai5γ \textit{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 ribose\(^{69}\), 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 \textit{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 nucleotides\(^{61}\). 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 \textit{in vitro} situation\(^{31}\). 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 \textit{in vivo}, especially since the J2/3 region again contacts D5 and is of utmost importance for the active site formation of the intron.
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 3.14: 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...
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 $\zeta-\zeta'$ and $\kappa-\kappa'$ seem to be far more effected than others as $\lambda-\lambda'$ and $\theta-\theta'$. 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 strain\textsuperscript{109}. \textit{In vitro} the ATPase activity of Mss116p was shown to be important for protein turnover but not for the collapse of ai5$\gamma$\textsuperscript{109,195,201}. The expected structure of ai5$\gamma$, 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 missing\textsuperscript{195}. The \textit{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 \textit{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 \textit{in vivo} Mss116p molecules might get trapped by nonproductive binding events\textsuperscript{195}.

**Mss116p cannot assist in ai5$\gamma$ folding without ATP binding and hydrolysis**

In a genetic background, in which Mss116p can neither bind nor hydrolyze ATP, the group II intron ai5$\gamma$ is largely unfolded \textit{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 $\kappa-\zeta$ element, whose formation was shown to be crucial for the collapse of the intron \textit{in vitro}\textsuperscript{78}. Modifications of the nucleotides involved in the intradomain tertiary interactions $\alpha-\alpha'$ and $\beta-\beta'$ are further evidence for an extended D1. Enhancements in $\zeta-\zeta'$ and $\kappa-\kappa'$
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).
Figure 3.15: Differential summary map. Differences in DMS accessibility of ai5γ in a wild type strain compared to the ATPase-deficient strain K158R. Legend as in Figure 3.6.
3.4.4 The basic tail of Mss116p and ai5γfolding

Some DEAD-box protein, namely Mss116p, Ded1p and Cyt19, show an abundance of basic amino acids in their C-terminal domain\textsuperscript{151,152}. Interestingly, Ded1p and Cyt19 are also able to rescue ai5γ folding \textit{in vitro}\textsuperscript{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 separation\textsuperscript{140}. 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 ΔCTD 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/Δ551-664, in which most of the α-helix 2 as well as the basic tail are deleted looses its ability to grow by respiration as well as to assist in ai5γ splicing \textit{in vitro}\textsuperscript{152}. In contrast, truncation of amino acids 569-664 removes the basic tail but in this mutant strain ai5γ splices at wild type level and is able to grow aerobically\textsuperscript{152}. In our hands, however, the Mss116p/Δ569-664 variant does not support ai5γ splicing \textit{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 plasmid\textsuperscript{152}. Since the distinct basic tail of Mss116p, Ded1p and Cyt19 is likely to be of importance for ai5γ folding, the intron structure was analyzed in the mss116/Δ569-664 (ΔCTD) 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 ΔCTD 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 ΔCTD 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 ΔCTD 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\textsuperscript{38,55,72,77} and the enhanced DMS accessibilities at these regions indicate that the intron is not collapsed in the ΔCTD strain.
Figure 3.17: DMS accessibility around the 3’ part of the $\kappa$–$\zeta$ element in the $\Delta$CTD strain. (A) Representative gel covering the 3’ part of the $\kappa$–$\zeta$ 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 ΔCTD 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 ΔCTD 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 codes as in Figure 3.6.
Several secondary structure elements of ai5γ are not formed in the ΔCTD 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 ΔCTD 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 ΔCTD 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 Δmss116 strain seem to be also affected in the ΔCTD 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 Δmss116 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 ΔCTD 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 ΔCTD strain (Figure 3.24).
Figure 3.20: DMS accessibility around the d3 stem in the ΔCTD 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 3.21: DMS accessibility around the c1 stem in the ΔCTD 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 ai5γ

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 ΔCTD 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 ΔCTD 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 D2 requires, which 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 elements.

The active site constituents ε–ε' and λ–λ' are involved in a complex structure, the so-called z-anchor, close to the dinucleotide bulge of D5, which is important for the formation of the active site. A117 which is part of the z-anchor is more accessible to DMS in the ΔCTD 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 ΔCTD strain. This complex structure, that brings the 5'exon close to the dinucleotide bulge, is not formed in the ΔCTD strain. The J2/3 nucleotides A586-587 and A589 show moderate enhancements (Figure 3.23). This indicates that only in a fraction of ai5γ molecules the contact between A586 and the backbone of G5 that positions the 5' exon and the ε–ε' 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 ΔCTD strain.
Figure 3.22: DMS accessibility of D5 in the ΔCTD 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 ΔCTD 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 ΔCTD 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 ai5γ 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 ΔCTD 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 ΔCTD strain rendering binding of the exons very unlikely. All in all the changes in DMS accessibility of ai5γ 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 $\varepsilon$ and $\lambda$ 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 ai5γ folding**

Formation of the $\kappa$–$\zeta$ 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 $\kappa$–$\zeta$ folding control element D1 is not in its collapsed state indicated by enhancements in the two D1 intradomain tertiary interactions $\beta$–$\beta'$ and $\alpha$–$\alpha'$. As a consequence D5 is not docked into D1 as shown by the enhancements of the tertiary interactions $\zeta$–$\zeta'$ and $\kappa$–$\kappa'$. As such, the active site is not formed, as indicated by the modification pattern around the z-anchor with its important tertiary interactions $\varepsilon$ and $\lambda$. Compared to the ATPase-deficient and the knockout strain, the structural consequences of this Mss116p variant on ai5γ seem to be less severe (Figure 3.1), suggesting that a subpopulation of ai5γ 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 ΔCTD 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, ai5γ 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 ai3γ in a wild type strain compared to the RNA binding-deficient strain ΔCTD. Legend as in Figure 3.6 using blue instead of red.
**3.4.5 Helicase activity of Mss116p and ai5γ folding**

An important activity of DEAD-box helicases lies in its name giving feature of being able to disrupt short helices of RNA. It is subject of ongoing discussion whether this helicase activity of Mss116p is its main contribution to ai5γ folding or, more generally speaking, if DEAD-box proteins act on RNA predominantly by unwinding kinetically trapped helical regions in a large RNA molecule. 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 ai5γ splicing. 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 severly affected. The T307A mutant though was neither able to rescue respiratory growth of yeast nor to stimulate splicing of ai5γ 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 ai5γ splicing. Hence, the ATPase activity of Mss116p was interpreted to be the essential part in its function as splicing cofactor of ai5γ. 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*. Therefore, the Q412A mutant is well suited to investigate the effects of helicase activity of the cofactor Mss116p on ai5γ folding since the SAT mutant is also affected in its ability to hydrolyze ATP. 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 ai5γ splicing.
An unwinding-deficient Mss116p variant interferes with the folding of the \( \kappa - \zeta \) element in a subpopulation of ai5γ

Formation of the \( \kappa - \zeta \) 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 folding\(^78\). Enhancements of nucleotides involved in the \( \kappa - \zeta \) folding control element can still be found at the involved nucleotides but they are significantly lower than those in the K158R and ΔCTD strains and almost all below threshold. A195-A196 which are adjacent to the \( \zeta \) 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 ai5γ molecules. A205 shows moderate enhancement (Figure 3.26). Hence the short helix connecting the \( \kappa \) and \( \zeta \) 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 ΔCTD strains. Actually, only A596 shows a weak enhancement (Figure 3.32).

Both D1 intradomain tertiary interactions \( \alpha-\alpha' \) and \( \beta-\beta' \) are intact at least in some of the RNA molecules. In case of \( \alpha-\alpha' \) 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 \( \beta-\beta' \) 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 ΔCTD strains.
Figure 3.26: DMS accessibility of the 5' part of the $\kappa$–$\zeta$ 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 $\beta$, $\zeta$ and $\kappa$. 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 ΔCTD 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 ΔCTD 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 ΔCTD 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).

Protection at the terminal base pairs of stem c2 (A128) and the basal D2 stem (A420) are very similar to the ones seen in ΔCTD 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 a' element. (B) (C) (D) Symbol and color code as in Figure 3.6.
Several tertiary interactions within ai5γ 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, ΔCTD 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 ai5γ molecules. The protection found at A671 is also not as strong as in the ΔCTD and K158R strains (Figure 3.33).
Figure 3.31: DMS accessibility around the c1 stem in the Q412A strain. (A) Representative gel covering the c1 stem, λ and ε. (B) (C) (D) Symbol and color code as in Figure 3.6.
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 triad, ζ′ tetraloop and the dinucleotide bulge. (B) (C) (D) Symbol and color code as in Figure 3.6.
The active site constituents $\varepsilon$ and $\lambda$ 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 DCTD 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 ai57 molecules in the unwinding deficient strain. A617-A618 are weakly enhanced compared to wild type hence the $\mu-\mu'$ 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 ai57 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 ai57 folding\textsuperscript{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 ai57 molecules**

Without its helicase activity Mss116p still retains its ability to assist in secondary structure formation of ai57 (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, DCTD or knock-out strains (Figure 3.1). Formation of the $\kappa-\zeta$ folding control element is only slightly disturbed, which also correlates with the minor enhancements in the D1 intradomain contacts $\alpha-\alpha'$ and $\beta-\beta'$. Hence, formation of the $\kappa-\zeta$ element seems to be independent from unwinding in a substantial fraction of the ai57 population.

Interestingly, the strongest enhancements are found at residues involved in the $\zeta-\zeta'$ interaction which is essential for docking of D5. Additionally, residues of the $\kappa$ 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 $\lambda-\lambda'$ 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 $ai5\gamma$ molecules. The very same is true for the $\mu-\mu'$ contact, which anchors the important effector D3 near the active site.
Figure 3.35: Differential summary map. Differences in DMS accessibility of ai5γ in a wild type strain compared to the helicase deficient strain Q412A. Legend as in Figure 3.6 using green instead of red.
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 ai5γ. 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, maybe structures involving exonic sequences have to be resolved in order to allow native structure formation of ai5γ. This would also nicely correlate with the finding that an unwinding deficiency in Mss116p strongly reduces splicing efficiency of ai5γ in vitro, if long exon sequences flank the intron.

Mapping the exonic regions of ai5γ 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 ai5γ 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 ai5γ 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 \textit{in vitro}\cite{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 structure\cite{115}. 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.

\textit{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 \textit{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 \textit{in vivo}, being influenced by the velocity and directionality of the RNA polymerase\cite{218–220}. Although metal ions play an important role in intracellular folding as well, their concentration is far below most used \textit{in vitro}. Instead many trans-acting co-factors are present \textit{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 yeast\cite{128}.

One of the major problems in studying RNA structure and folding \textit{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 ai5γ \textit{in vivo} at endogenous levels for the first time\cite{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 \textit{in vitro} or \textit{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 \textit{in vivo} and \textit{in vitro} as well as high consistency with the available crystal structure of the \textit{Oceanobacillus iheyensis} group IIC intron we showed the applicability of this method for this purpose.

In order to reach the native structure ai5γ depends on high amounts of [Mg$^{2+}$] and elevated temperature and although compaction of the ribozyme is also observed at near-physiological conditions, it is very ineffective\textsuperscript{106}. Mss116p is able to promote ai5γ splicing at near-physiological conditions, indicating that this DEAD-box protein substitutes for high [Mg$^{2+}$] used \textit{in vitro}\textsuperscript{183,187,189}. We showed that indeed ai5γ remains largely unfolded in a \textit{mss116}-knockout strain\textsuperscript{109}. The strongest effects could be seen in D1, which folds first and serves as a scaffold for the other domains during later stages of ai5γ 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 ai5γ were found at the κ–ζ folding control element, whose formation was shown to be crucial for the intron's collapse\textsuperscript{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 ai5γ does not misfold in the absence of Mss116p. Kinetically trapped RNA molecules are also prone to degradation\textsuperscript{221} which is not consistent with the strong RNA signal in the Δ\textit{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\textsuperscript{137,138}. These proteins display RNA binding, unwinding, annealing and ATPase activity\textsuperscript{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 ai5γ 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 ai5γ 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 ai5γ. 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 ai5γ 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 ai5γ 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), ΔCTD (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 ai5γ folding**

The ATPase activity of Mss116p is essential for ai5γ splicing both *in vitro* and *in vivo*\(^{28,187,189,190}\), although the precise effect of ATP binding and hydrolysis on the structure of ai5γ is still unclear. In a single molecule study it was shown that ATP is most crucial for the final step of ai5γ folding, the transition from a folded intermediate to the native state\(^{201}\). In fact, ATP is not needed at all to reach a near-native folding state of ai5γ\(^{195}\). These results are in line with the findings that ATP hydrolysis is needed for protein turnover\(^{169}\). Our results show that the mutation K158R, which abolishes binding and hydrolysis of ATP, renders Mss116p unable to assist in ai5γ 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 ai5γ 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 \(\kappa-\zeta\) folding control element, indicating that the stems embedding \(\kappa-\zeta\) 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 ai5γ 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 ai5γ 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 ai5γ 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 dissociation\(^{190}\). 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 ai5γ 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 ai5γ folding in vivo. DMS enhancements and protections in ai5γ 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 affected 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\textsuperscript{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\textsuperscript{162,165,183,208}. The drastic effect of the K158R mutation on ai5γ 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 ai5γ 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 ai5γ folding, namely the docking of the domains onto the folding scaffold D1, depend on ATP\textsuperscript{195,201}. On the other hand non-productive binding of Mss116p molecules would also be detrimental to folding of ai5γ 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 events\textsuperscript{195} 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 ai5γ is in excess to Mss116p.
In other words ATP binding and hydrolysis activity might not be needed for initial steps of ai5γ folding with an Mss116p to ai5γ ratio similar to the one used in vitro ranging from 1:1 to 20:1. 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 ai5γ splicing but also in RNA processing and mitochondrial translation.

**Non-specific binding by the C-terminal extension of Mss116p is important at early stages in ai5γ 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 ai5γ in vitro or in a Δmss116 strain. This domain is proposed to be of importance for binding to structured RNA tethering the DEAD-box protein to its target. Deletion of the basic tail in Mss116p reduced its affinity to RNA, but also its RNA unwinding and annealing activities in vitro. Specifically binding to the ai5γ 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 ai5γ 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 mss116Δ569-664 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 ai5γ was reduced significantly to 10% of the wild type splicing level. Thus, the truncated version of Mss116p causes a strong effect on ai5γ 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 ai5γ, which is about 10-fold increased compared to the wild type protein. This finding is further in line with the observed intracellular structure of ai5γ in the DCTD strain, as ai5γ is largely unfolded, adopting a
conformation comparable to the Δmss116Δl and ATPase-deficient strains. The DMS modification pattern is virtually identical in these three strains, except that the level of enhancements/protections is somewhat smaller in the ΔCTD strain, which might result from the fact that 10% of the ai5γ 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 ai5γ 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 ai5γ structure in some intron molecules, as Mss116p releases the RNA substrate after near-native folding.

The helicase activity of Mss116p plays an important role in the late steps of ai5γ 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. The K158R mutation abolishes both of those activities and it was proposed that in fact its ATP-dependent unwinding is necessary for the collapse of D1 and subsequent folding events. 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 ai5γ 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 ai5γ intron in the unwinding-deficient strain, which is in line with the model that ai5γ's folding follows a smooth energy landscape without kinetically trapped intermediates. Furthermore, ai5γ
splicing was shown to be efficient in unwinding-deficient strains\textsuperscript{187,188}. The Q412A mutation, which exhibits wild type-like ATP binding and hydrolysis, shows only very little to no helicase activity\textsuperscript{188}. In our \textit{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 ai5γ \textit{in vivo}, although to a far lower degree than ATP binding and hydrolysis or the basic tail of Mss116p. In case of the \textit{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 formation\textsuperscript{222}. If non-native stable structures would be resolved by Mss116p in the folding process of ai5γ, 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 \textit{Tetrahymena} group I intron premature formation of native tertiary interactions can also trap the intron in a non-native conformation\textsuperscript{100}. 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 \textit{in vitro}\textsuperscript{195,201}.

Resolving misfolded exonic substructures can be essential for self-splicing as shown before for group I introns\textsuperscript{206,223}. The correlation between length of exonic sequences and importance of the presence of ATP\textsuperscript{190} indicates that \textit{in vivo} with its exonic environment the ATPase or unwinding activity of Mss116p might be of higher relevance than \textit{in vitro} with a construct with no or short exons. The influence of exonic sequences on ATP dependence of ai5γ splicing \textit{in vitro}\textsuperscript{194,195} gives rise to the question if Mss116p unwinds
structures that involve exonic and/or intronic sequences and are detrimental for ai5γ 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 sequences190,194,202.

Mss116p is a multifunctional enzyme.
For a long time, it has been controversially discussed whether Mss116p facilitates ai5γ folding and splicing by resolving kinetically trapped structures or by stabilizing an on-pathway intermediate187–190,194. We showed that actually RNA-binding, ATP binding and hydrolysis and unwinding activity contribute to its mode of action in ai5γ 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 ai5γ, eventually resulting in intron splicing and the mature Cox1 mRNA transcript.
5 References


72. Qin, P. Z. & Pyle, A. M. The architectural organization and mechanistic function of

73. Michel, F. & Jacquier, A. Long-range intron-exon and intron-intron pairings

74. Hamill, S. & Pyle, A. M. The Receptor for Branch-Site Docking within a Group II

75. Toor, N. *et al.* Tertiary architecture of the Oceanobacillus iheyensis group II intron.

76. Nagaswamy, U. & Fox, G. E. Frequent occurrence of the T-loop RNA folding motif

77. Waldsich, C. & Pyle, A. M. A folding control element for tertiary collapse of a group

78. Waldsich, C. & Pyle, A. M. A Kinetic Intermediate that Regulates Proper Folding of

79. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5’ and 3’-terminal

80. Podar, M., Dib-Hajj, S. & Perlman, P. S. A UV-induced, Mg(2+)-dependent

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

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


Appendix

6.1 Publications

Chemical probing of RNA in living cells.

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


DEAD-box protein facilitated RNA folding in vivo.

Liebeg A, Mayer O, Waldsich C.


Probing RNA structure within living cells.

Liebeg A, Waldsich C.

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)
Professioneller Werdegang:

**Sozialbereich**

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

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

**Naturwissenschaftlicher Bereich**

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

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 (II/L1) in den Hertha Firnbergsschulen für Wirtschaft und Tourismus (hlw22) für Naturwissenschaften (Chemie, Biologie).