universität wien

DIPLOMARBEIT

DEAD-box protein facilitated folding of the group II intron ai5γ *in vivo*

angestrebter akademischer Grad

Magister/Magistra der Naturwissenschaften (Mag. rer.nat.)

Verfasserin / Verfasser:	Andreas Liebeg
Matrikel-Nummer:	9811770
Studienrichtung /Studienzweig (lt. Studienblatt):	Molekularbiologie (A490)
Betreuerin / Betreuer:	Dr. rer.nat. Christina Waldsich

Wien, 07.09.2009

Contents

Introduction

Group II introns – ai5γ

In the beginning of the 80s of the last century the finding that an intervening region in the 26S ribosomal RNA of *Tetrahymena thermophila* is able to catalyse its own excision from the precursor RNA (Kruger et al., 1982) and that the RNA moiety of ribonuclease P is the catalytic subunit of the RNP (Guerrier-Takada et al., 1983) triggered a paradigm shift concerning the functions of RNA. RNA was no longer only a short-term mobile information carrier but a molecule able to show complex catalytic activity.

The mentioned self-splicing intervening region turned out to be a member of a new class of ribozymes, the so-called "group I introns". Another class of self-splicing introns differing in their structure from group I introns have been found in organellar genomes of fungi and plants as well as in some bacterial genomes (Michel et al. 1989). Being able to carry out phosphotransesterification too, the splicing mechanism differs. Interestingly its mechanism resembles nuclear pre-mRNA splicing by the spliceosome, in which as in ribosomes (Nissen 2000) the ribonucleic part is carrying out the catalytic activity (Valadkhan, 2007). According to the later discovery this group of self-splicing introns were called "group II introns".

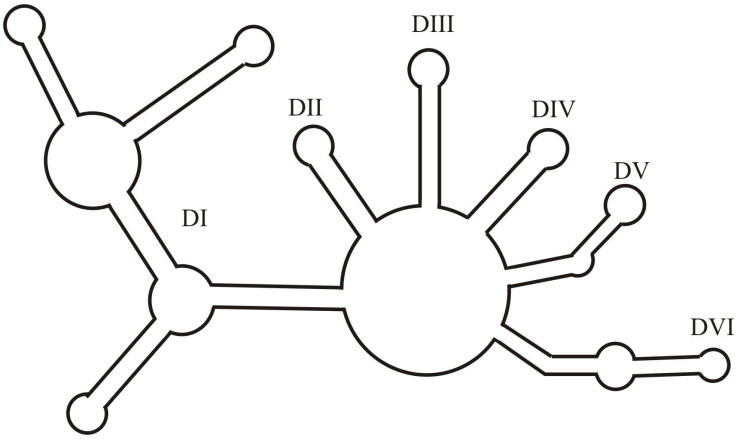

Fig1: *Conserved secondary structure of group II introns*. Domains are labelled in roman numbers I-VI.

Group II introns are among the largest known catalytic RNA molecules ranging from 500 to 1000 nucleotides. They are generally poorly conserved in their primary sequence,

but the secondary structure shows high conservation (Michel, 1989). The secondary structure map of group II introns shows some very interesting characteristics. It consists of 6 domains protruding from a central wheel (Fig. 1).

In contrast to group I introns there are functional structure elements established by base-pairing or coaxial stacking involving different domains (Qin and Pyle 1998). On the other hand a ribozyme derived from the group II intron ai5γ folds into a compact structure despite this lack of long-range secondary structure elements indicating that long-range tertiary interactions are present (Swisher et al. 2001). Ai5γ is found in the Cox1 gene in the mitochondrial genome. Therefore its efficient splicing is crucial for a functional respiratory chain in yeast. Most studies on ai5γ were carried out *in vitro* under high salt conditions and elevated temperature. The collapse of this group II intron requires high amounts of magnesium (Swisher et al. 2001).

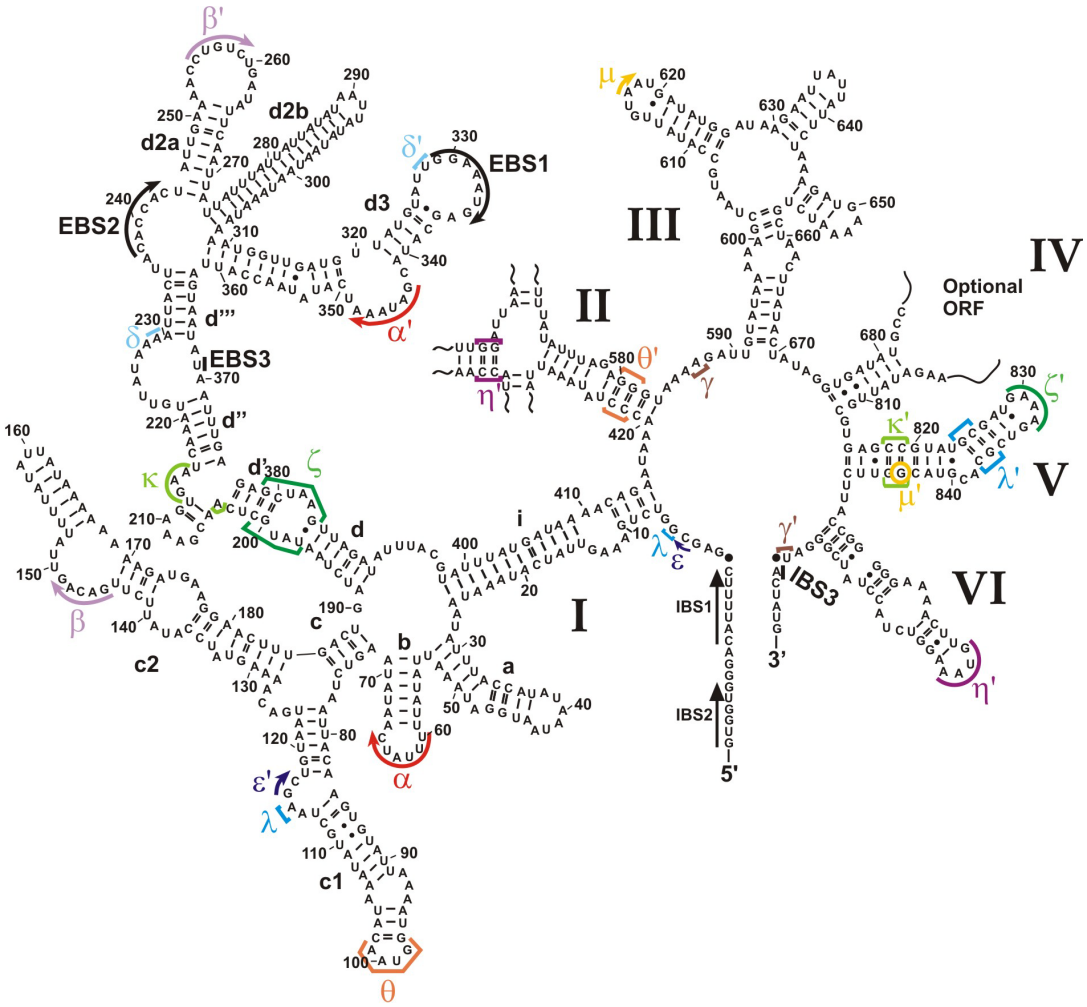

Fig.2: *Secondary structure of group II intron ai5γ.* Domains are indicated in Roman numbers I-VI. Tertiary interactions are shown in Greek letters. EBS (exon binding sites) and IBS (intron binding site) interact. DII and IV are only shown partially (adapted from .

Splicing mechanism

Group II introns most prominent catalytic activity is to excise from precursor mRNA forming a lasso like structure, the so-called lariat (Peebles 1986). Remarkably the same lariat spliced intron can be found in spliceosomal introns suggesting a common ancestor of both pathways (Seetharaman et al. 2006).

Most of the data available on splicing of group II introns were gained in studies (Fedorova and Zingler 2007) carried out under high salt conditions *in vitro*.

Forward splicing

The splicing activity of group II introns consists of two transesterification steps (Fedorova and Zingler 2007), which are carried out sequentially. In the first step the 2'OH of a highly conserved bulged Adenosine in D6 attacks the phosphate at the 5' splice site releasing the 5'-exon while forming a lariat like structure. The released 5'-exon remains bound to the intron by base pairing interactions and its free 3'OH is positioned correctly to attack the phosphate at the 3' splice site resulting in an excised intron and ligated exons.

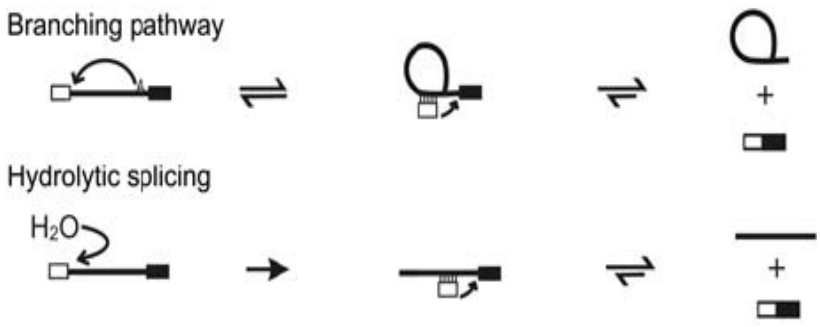

Fig. 3: *Branching and hydrolytic splicing pathway*. Adapted from Fedorova and Zingler 2009

Alternatively a H_2O molecule can act as the nucleophile attacking the 5' splice site resulting in a linear spliced intron. In vitro it was shown that under different salt conditions either the bracnching or the hydrolytic pathway is predominant (Daniels et al. 1996).

Reverse splicing

Spliced group II intron have the possibility to do reverse splicing in both the lariat (Augustin et al. 1990) and the linear form (Roitzsch and Pyle 2009). In fact both steps of forward splicing are reversible although considerably slower. It has also been shown that reverse splicing does not only work with RNA but also with DNA targets (Mörl et al. 1992). Reverse splicing therefore is the first step to intron mobility. In order to act as a mobile genetic element an intron-encoded protein (IEP), the so-called "maturase" is required (Matsuura et al. 2001). Maturases are encoded in many group II introns and typically consist of a reverse transcrptase, an endonuclease and an RNA binding domain (Lambowitz and Zimmerly 2004). Anyhow, ai5γ neither has an IEP nor is able to act as a mobile genetic element. Nevertheless the ability for reverse splicing further expands the importance of group II introns.

Recently it was shown, that also linear introns resulting from hydrolytic splicing are able to catalyse efficient reverse splicing (Roitzsch and Pyle 2009).

Domains

Domain 1:

Domain 1 is the largest of the six domains. Together with the finding that in the folding of ai5γ D1 folds first (Su et al. 2005) D1 was proposed to act as a folding scaffold allowing subsequent docking of other domains and forming the core structure. This is also consistent with folding of D1 being the rate-limiting step. It is inevitable for the catalytic activity of the ribozyme together with domain V (Koch et al. 1992). Harbouring two 5'-exonic substrate recognition sequences (EBS1 and EBS2) D1 is also responsible for binding the complementary regions at the 3'-end of the 5'-exon (Qin and Pyle 1998) defining both splice sites. Additionally, EBS1 and EBS2 in D1 are essential for targeting RNA or DNA substrates in reverse splicing. Folding of D1 was shown to be independent of other domains (Qin and Pyle 1997).

D1 forms several intra- and interdomain tertiary interactions (Fig. 2). The α–α' interaction are required for splicing *in vitro*. It plays an important role in folding of ai5γ

(Waldsich and Pyle 2007), facilitates binding of the exon splicing site (Qin and Pyle, 1998) and is helping in campaction of ai5γ (Waldsich and Pyle 2007). β–β' interaction is found in a subset of group II introns (Qin and Pyle 1998) although its function is still unclear and it is not necessary for splicing it was found to play a role in compaction (Waldsich and Pyle 2007). ε–ε' plays an important role in forward splicing and may help in defining the 5' exon splicing site (Jacquier and Michel 1990). Three tertiary interactions were identified between D1 and D5, the catalytic center of group II introns: ζ–ζ' and κ–κ' are important for docking of D5 in D1, while λ–λ' is taking part in catalysis and puts D5 close to the 5' splicing site making it an important factor of the formation of the active site together with the ε–ε' contact as involved residues are proximal to the dinucleotide bulge in D5 and J2/3, both key players in active site organisation (Fedorova and Pyle 08). θ–θ' is proposed to be responsible for structural stabilisation of the native structure (Costa et al. 1997). Additionally, D1 contains an asymmetrical internal loop, the "coordination loop", that is proposed to align all components necessary for splicing (DeLancestre et al. 2005) and even more importantly acts as the binding site for the bulged A880 that is responsible for the branching pathway in forward splicing (Hamill and Pyle 2006).

Domain 2

The function of D2 is still unclear. Its deletion hardly has any effect on cis-splicing activity (Qin and Pyle 1998) although the θ–θ' interactions where still present. However D2 forms long-range tertiary contacts to D1, θ–θ', and D6, η–η' (Costa et al. 1997).

Domain 3

Although shown to be dispensable for catalysis (Koch 1992) DIII is enhancing the rate of catalysis dramatically (Fedorova et al. 2003, Fedorova and Pyle 2005) making it a catalytic effector. It was shown that D3 is partly protected against hydroxyl radicals with one of the most highly protected regions of the ai5γ derived ribozyme D135 at the internal loop of the basal stem indicating contacts with the active site (Swisher et al. 2001, Fedorova and Pyle 2008). The μ-μ' tertiary interaction with D5, identified by

NAIS analysis (Fedorova and Pyle 2005), is emphasizing the important role of D3 at the active site center of group II introns.

Domain 4

D4 contains an open reading frame coding for the maturase protein in several but not all group II introns. The maturase is required for mobility (Lambowitz and Zimmerly 2004). It was shown that D4 of the LI.LtrB group II intron contains a high-affinity binding site for the maturase protein LtrA (Watanabe and Lambowitz 2004). Since D4 is located at the surface of the intron it might actually serve as a general binding platform (Fedorova and Zingler 2007). The ai5g group II intron does not encode for a maturase.

Domain 5

D5 is strictly required for catalysis. It is the most conserved domain and docks into D1 via two tetraloop-receptor interactions, ζ–ζ' and κ–κ' (Costa and Michel 1995, Boudvillain and Pyle 1998). D5 harbours two catalytically very important substructures: the dinucleotide bulge (Schmidt et al. 1996) and the catalytic triad (Boulanger et al. 1995), the latter also found in the U6 snRNA of the spliceosome (Valadkhan and Manley 2001). The backbones of this two substructures are brought closely together resulting in the formation of a negatively charged pocket that complexes two Mg^{2+}-ions accessible to the 5' and 3' splice sites (Toor et al. 2008). EBS1 as well as the 5' and 3' end of the intron are positioned closely to the bulge suggesting a two metal ion mechanism in catalysis (Steitz and Steitz 1993). D5 also interacts with the highly conserved nucleotides of the linker region between D2 and D3 (J2/3), which seem to play a crucial role in catalysis. The most highly conserved nucleotides in D5 are found in the catalytic triad AGC, where the G (G817 in ai5γ) is invariant. The functions of this invariant G are interesting as its major groove atoms play a crucial role in catalysis, while minor groove substituents facilitate ground-state binding of D5 to the ribozyme core (Konforti et al. 1998).

Domain 6

D6 contains the highly conserved bulged Adenosine, which serves as a branch point in the lariat forming splicing pathway (Fedorova and Zingler 2007).

Folding of ai5γ

In order to fulfil its catalytic function, an RNA molecule has to adopt its native structure. Especially for large RNAs, like ai5γ, folding is a very complex phenomenon. Considering the primary sequence, the number of building blocks is very limited compared to proteins (Herschlag 1995). Furthermore nucleotides show far less diversity in size and physiology than amino acids do. These attributes generally hinder RNA to adopt a certain structure. At the level of secondary structure the high thermodynamic stability of RNA helices compared to protein a-helices pose a problem in RNA folding (Herschlag 1995). Thus kinetic traps are likely to arise. Additionally problematic is the establishment of a tertiary fold mediated by 2'-
hydroxyls, phosphoryl groups, metal ions or by the formation of nonstandard base/base interactions, which can further stabilize misfolded species (Herschlag 1995).
It is impossible to derive an RNA secondary structure from its sequence for large RNAs (Schroeder et al. 2004). Considering on one hand the low conservation of primary sequence of group II introns that adopt the same secondary structure and on the other hand the existence of RNAs like riboswitches that are able to adopt different structures according to their environment, this might never be possible.

The folding of a RNA molecule is hierarchic (Schroeder et al. 2004). First secondary structure elements involving proximate regions are built. Afterwards, coaxial stacking of helices and long-range tertiary interactions are formed, eventually resulting in the final and native structure. The establishment of secondary and tertiary interactions lead to a more compact species. This Mg^{2+}-induced compaction happens mostly very fast in less than 50ms. But interestingly the compaction of ai5γ needs more than 30s (Pyle et al. 2007).

Global and local folding events in ai5γ

Early studies dealing with the solvent accessibility of group II introns using hydroxyl radical footprinting showed that 40% of the ai5γ D135 ribozyme was protected from cleavage (Swisher et al. 2001). Phylogenetically conserved and catalytically important regions were highly internalized and ai5γ identified as a tightly folded RNA. Hydroxyl footprinting revealed that all regions of D135 become protected at the same time and Mg^{2+} concentration (Swisher et al. 2001). Similar results for Mg^{2+} dependence and the timescale were found in global molecular compaction (Su et al. 2003 and 2005). This collapse of the unstructured RNA was further differentiated. First there is a partial collapse due to secondary structure formation in the presence of K^{+} called the "unfolded state" (Su et al. 2003). In the presence of high levels of Mg^{2+} the tertiary structure is obtained resulting in further compaction. Furthermore, D135 can do rounds of folding and unfolding without loosing its activity; the energy for unfolding and folding are the same and like folding unfolding is done in a cooperative manner (Su et al. 2003).

Kinetic traps

As mentioned before compaction is relatively slow in D135 compared to folding of orther large ribozymes. This could be explained by the arising of kinetic traps that have to be resolved in order to allow native structure formation as with the *Tetrahymena* group I intron (Treiber et al. 1998).

In that case subdenaturing concentrations of urea should reduce folding time (Pyle et al. 2007). However, neither catalysis nor compaction are accelerated by adding of urea (Su et al. 2003, Fedorova et al. 2007). Furthermore, stable misfolded species should appear on a native gel since its radius would differ from both the unfolded as well as the native structure (Pyle et al. 2007).

Folding control element

Another explanation for the slow collapse of ai5γ is the necessity of forming a small nucleation element of folding in an Mg^{2+} dependent manner. Compaction and formation of long-range tertiary interactions would depend on the formation of this substructure.

It was shown that D1 folds independently (Qin and Pyle 1997) and that its footprinting pattern resembles its counterpart in D135 (Su et al. 2005). Furthermore D1 and D135 need the same Mg2+ concentration for local folding and molecular compaction. Additionally all regions of D1 internalize at the same time (Su et al. 2005). D1 once folded can independently bind an exonic substrate and folding of D1, D13, D15 and D135 show the same rate constant implying that D1 folding is the rate limiting step in the folding pathway of the ai5γ group II intron (Su et al. 2005). Therefore D1 is a folding scaffold for the docking of other domains and contains an obligate intermediate along the folding pathway of D135 (Su et al. 2005). In D135 D3 and D5 dock independently into D1, which is further evidence for the importance of folding of D1 in the collapse of D135.

To narrow down the region being responsible for the slow collapse of D135, Nucleotide Analogue Interference Mapping (NAIM) was carried out (Waldsich and Pyle 2008). In this approach nucleotide analogues carrying single-atom modifications are incorporated randomly in the RNA. By adding 100mM Mg^{2+} tertiary folding is triggered. After ~$t_{1/2}$ for compaction the RNA is loaded on a native gel to separate folded from unfolded D135. Iodine sequencing then reveals important sites for D135 collapse. All NAIM signals were located in D1 underlining its crucial role in folding. The strongest NAIM effects were found within the D1 central stem around the so-called κ–ζ element. It was revealed that the formation of the κ–ζ element is Mg^{2+} dependent. Since the amplitude for compaction but not the rate constant is dependent on the Mg^{2+} concentration a conformational capture mechanism was proposed (Waldsich and Pyle 2008) meaning that the proper conformation of the κ–ζ element is captured by Mg^{2+} inducing the collapse of D1. Additionally it was shown, that the formation of certain helical elements in the central stem of D1 is dependent on Mg^{2+} indicating that not only tertiary contacts but also secondary structure formation plays an important role on the collapse of ai5γ (Waldsich and Pyle 2008). Tertiary contacts then stabilize the collapsed structure and subsequent docking of D3 and D5 results in the native structure of the ribozyme. With these findings it was possible to describe a folding pathway for ai5γ *in vitro* (Figure 4).

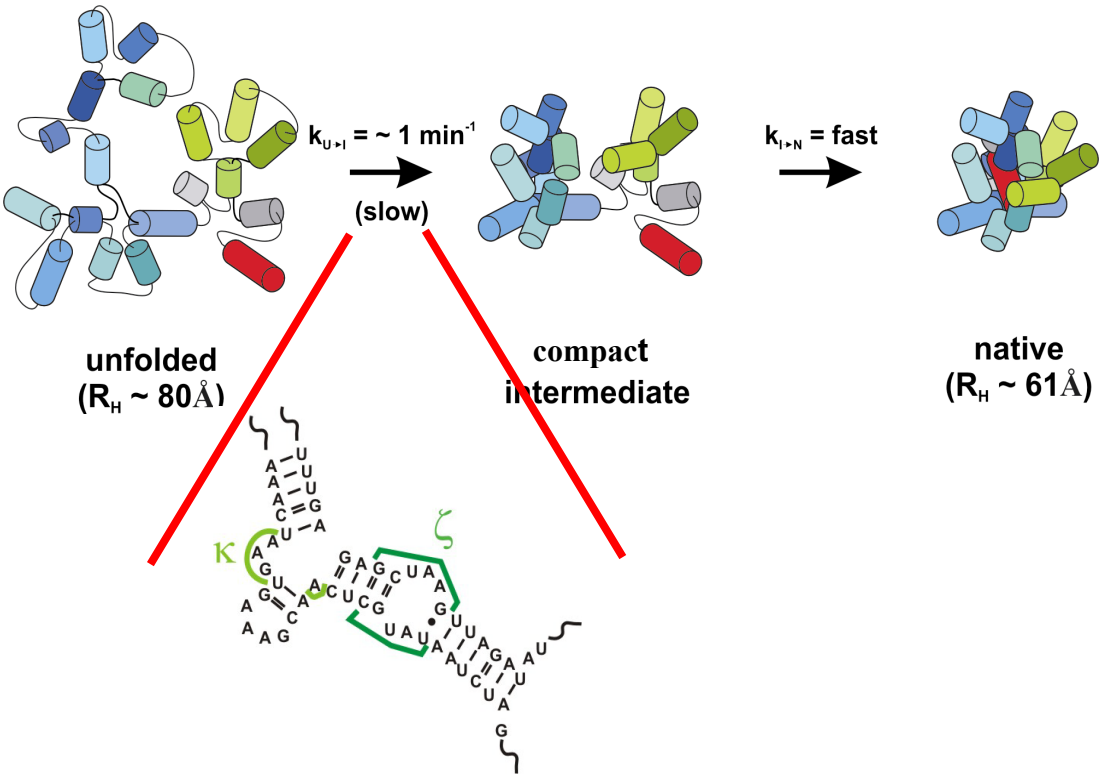

Figure 4: *The proposed folding pathway of ai5γ in vitro.* Here the electrostatically relaxed intron collapses slowly to a compact intermediate controlled by the formation of the k-z element. Subsequent docking and native structure formation then happens fast. Adapted from Su et al. 2005.

Folding in vivo

In the cellular context neither mono- nor divalent ions are present at a concentration required for folding of RNA *in vitro* (Schroeder et al. 2002). Folding and in turn splicing of ai5γ depends on the activity of the DEAD-box protein Mss116p (Seraphin et al. 1987).

Generally little is known about folding of RNA *in vivo*. Despite the difference in ion concentration, the flanking sequence context of RNA *in vivo* is not available *in vitro* and has significant influence in RNA folding rates (Cao and Woodson 2000). Additionally, observed misfolding of the *Tetrahymena* group I intron *in vitro* can be significantly reduced if the flanking sequence context is present *in vivo* (Nikolcheva and Woodson 1999). Furthermore folding of RNA *in vivo* happens co-transcriptionally, which impacts folding due to its directionality and velocity as well as pausing of the RNA Polymerase (Schroeder et al. 2002).

Maybe the biggest difference of RNA folding *in vitro* and *in vivo* results from the presence of proteins, which can assist in folding *in vivo*.

One class is formed by the so-called RNA chaperones (Herschlag 1995, Rajkowitsch et al. 2007). These proteins are found to act on RNA non-specifically by resolving misfolded intermediates or by impeding kinetic traps. They are ATP independent and bind RNA non-specifically. StpA, a RNA chaperone found in *E.coli,* for example shows enhanced chaperone activity if its RNA binding is weakened by a mutation (Mayer et al. 2007). Consistent with this weak binding RNA chaperones are also not needed to maintain the native structure of a RNA (Rajkowitsch et al. 2007). Nevertheless little is known about the mechanisms and relevance of RNA chaperones *in vivo*.

Another type of proteins assisting in RNA folding are specific binding cofactors that are stabilizing an on pathway intermediate or the native state itself. An example for this can be found in CPB2 that captures its target RNA the bI5 group I intron from the collapsed to the native state by binding (Webb and Weeks 2001). Cyt18, a group I intron splicing factor, binds to a defined structural element and thereby induces compaction of the intron and stabilizes its tertiary structure (Mohr et al. 1992). The action of these proteins is required to maintain the native structure (Rajkowitsch et al. 2007). Another point of action for proteins in RNA folding is ion homeostasis since especially Mg^{2+} plays a crucial role in folding of most large RNAs as shown for group II introns and the magnesium transporter Mrs2p (Gregan et al. 2001).

A class that assists in RNA folding and has recently drawn a lot of attention to it is the so-called DEAD-box helicase family (Cordin et al. 2006). Helicases are ATP dependent proteins that are able to unwind both RNA and DNA strands (Pyle 2008). Mss116p was proposed to act as a chaperone (Delcampo et al. 2007) using its unwinding activity to resolve kinetic traps and thereby allow native structure conformation. In contrary the same protein was found to be able to assist in folding of the group II intron ai5γ without unwinding but by stabilization (Solem et al. 2006).

DEAD–box proteins

The family of DEAD-box proteins is widely spread in all eukaryotes and most prokaryotes (Cordin et al. 2006). They are part of the helicase superfamily SF2 and share 9 conserved motifs (Figure 5) including the Walker B motif which is responsible for its name due to its amino acid sequence D-E-A-D (Pyle 2008). DEAD-box proteins are organized in Domain 1 and 2 which harbor the conserved motifs. These domains are structurally conserved and superimposable with a RecA fold (Pyle 2008). Some DEAD-box proteins have an additional alpha helical domain Dα linking the two RecA folds to a C-terminal domain, which is arginine rich and proposed to interact with the negatively charged RNA backbone (Solem et al. 2006).

A vast majority of processes involving RNA are on one or another hand dependent on the activity of DEAD-box proteins. These processes involve cellular functions as different as ribosomal biogenesis and RNA decay, where DEAD-box proteins are required for structural reorganisation (De la Cruz et al. 1999).

Figure 5: *Conserved motifs with consensus sequence for DEAD box proteins*. Showing >60 % homology. Adapted from Cordin et al. 2006

Motifs

The Q motif, which is found at the N-ternegative, is necessary for efficient binding of single stranded RNA and for conformational changes resulting from nucleotide binding and ATP hydrolysis. It might therefore play an important role in acting as a sensor for bound substrates triggering hydrolysis of ATP (Cordin et al. 2004).

Motif I, also called Walker A motif, is found in all helicases. It is inevitable for ATPase as well as unwinding activites. It is conserved in structure and sequence. The

importance of motif I is highlighted through the interactions with the Q motif, motif II and III as well as the phosphates of ATP (Figure 6, Cordin et al. 2006).

Motifs Ia and Ib are only poorly understood but crystal structures showed that they are involved in RNA binding together with motif IV and V (Cordin et al. 2006).

Motif II is also known as Walker B motif and necessary for the ATPase activity. It interacts with motif I, III and IV and in that possibly forms the ATP binding pocket (Cordin et al. 2006).

Motif III, called the SAT motif due to its sequence, is coupling ATPase and unwinding activities. While mutations lead to an abolishment of unwinding activity, ATP binding, hydrolysis and RNA binding are still present (Pause and Sonenberg 1992).

Motif IV is still very poorly understood and its functions are unknown.

Motif V was suggested to bind RNA together with motifs Ia, Ib and IV. It might also be involved in signalling RNA binding to the ATPase domain and thereby regulating ATP hydrolysis (Schneider et al. 2004).

Motif VI is located between domain 1 and 2 (Figure 6). It is important both for ATPase activity and RNA binding. Motif II and VI have different consensus sequences in DEAD and DExH-box proteins suggesting a covariation and potential communication between those two motifs (Cordin et al. 2006).

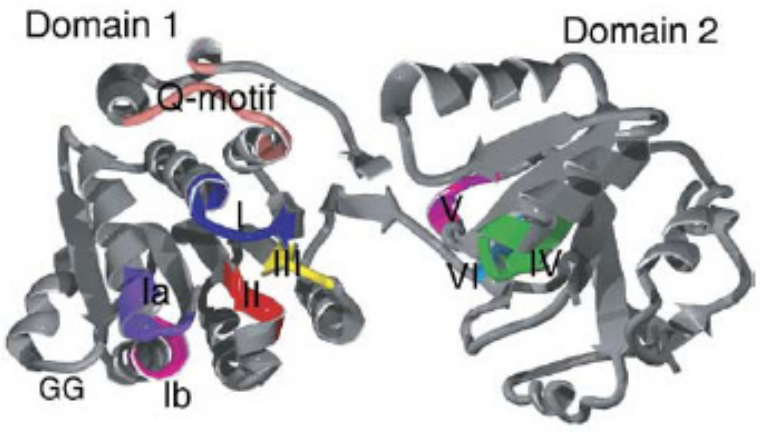

Figure 6: *Crystal structure of full length DEAD box protein eIF4A*. Showing positions of the conserved motifs. Adapted from Cordin et al. 2006.

Unwinding activity

Although SF2 including DEAD-box proteins has originally been defined as helicases they display a diversity of functions (Pyle 2008). Namely the DEAD-box proteins behave differently compared to other SF2 helicases. They have a different structure and are nonprocessive, meaning they are not translocating continuously and directionally on a filament like a RNA strand while for example seperating it (Pyle 2008). Furthermore, they have another strategy for binding and coupling their substrate and ATP hydrolysis.

ATP binding and hydrolysis happens between the two RecA folds domain 1 and 2. Compared to processive SF2 proteins the connection between the two RecA domains is more flexible and only rigidification by binding of its substrate RNA or other proteins allows ATP binding (Jankowsky and Fairman 2007). By binding RNA and ATP domain 1 and 2 function as a clamp changing their orientation (Figure 7). Another important difference between DEAD-box proteins and the rest of the SF2 proteins can be found in the structure of bound RNA. While RNA bound by processive SF2 proteins show an extended shape its backbone is severely bent when associated to DEAD-box protein, which was revealed by cocrystallisation of SF2 helicases and a RNA substrate (Jankowsky and Fairman 2007). Interestingly DEAD-box proteins bind the phosphate backbone exclusively without the involvement of the bases.

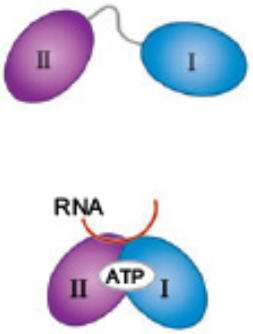

Figure 7: *Conformational changes of the DEAD box protein Vasa upon RNA and ATP binding*. Adapted from Pyle 2008.

The nonprocessive character of DEAD-box proteins does not allow extensive strand separation. In contrast to for example viral DexH proteins the length of unwinding is limited by 2 helical turns and is often even less.

The actual mode of unwinding by DEAD-box helicases is still unknown. As in all helicases ATP binding and hydrolysis is needed for its helicase activity. It is also known that substrate binding is greatly enhanced in the presence of ATP and reduced in its absence or the presence of ADP differing from processive helicases (Yang and Jankowsky 2005). For the DEAD-box protein Ded1 it was proposed, that during binding and hydrolysis of ATP a transient conformational state is destabilizing duplex and stabilizing single stranded RNA (Jankowsky and Fairman 2007). In contrast it was suggested for Vasa, that the unwinding might be a more active process (Sengoku et al. 2006). In this model the protein exerts force on the duplex RNA to actively wrench it apart as it kinks bound single stranded RNA in a way that does not allow duplex formation. Nevertheless this could also imply that Vasa acts by stabilising transient single strand RNA and thereby driving the reaction toward the unwound state (Pyle 2008).

Other functions

Although DEAD-box proteins are mostly referred to as helicases many of them do not unwind RNA (Cordin et al. 2006). Strikingly until now many other functions of DEAD-box proteins have been revealed as for example RNA stabilisation, protein displacement or transcriptional regulation (Pyle 2008).

A very interesting additional function revealed especially in DEAD-box proteins with a basic C-terminal protein as Ded1 or Mss116p is strand annealing activity (Yang and Jankowsky 2005). Ded1 stabilizes duplexes in the absence of ATP while establishing equilibrium between unwinding and annealing in its presence. This might be a key function in the assembly and disassembly of RNA structures.

Many different DEAD-box proteins are required for structural organisation or folding of RNA as with self-splicing group II introns. But although a lot of those do so by unwinding kinetically trapped substructures, other modes of actions have to be present. In the case of Mss116p it was proposed that not unwinding of a misfolded structure but stabilizing leads to the establishment of the native structure of the ribozyme (Solem et al. 2006).

Mss116p

Mss116p was identified in *Saccharomyces cerevisiae* in a screen for nuclear mutants that are unable to grow by respiration on nonfermentable carbon sources when their mitochondrial DNA contain introns, but not when their mtDNAs lack them (Seraphin et al. 1987). In *S. cerevisiae* splicing of all mitochondrial introns is facilitated by Mss116p *in vivo* (Huang et al. 2005). In the case of group II introns this effect was very severe with a reduction in splicing of 88-97%. It was shown in the same study that another DexH/D-box protein Cyt19 could partially rescue the phenotype of a mss116 knock out strain. Mss116p also plays a role in mitochondrial translation (Huang et al. 2005) and its overexpression can partly rescue defects of Δsuv3 strains, a helicase of the degradosom (Minczuk 2002). Therefore Mss116p seems to be a multifunctional enzyme involved in different RNA dependent processes.

Mss116p also stimulates splicing of the group II intron ai5γ under near-physiological conditions *in vitro* lowering the needed [Mg^{2+}] significantly (Solem et al. 2006, Halls et al. 2006). Mss116p hydrolyzes ATP after RNA binding and unwinds short dsRNA with a 3' or 5' single stranded overhang as well as with a blunt end (Halls et al. 2006). This previously led to the assumption that resolving of kinetic traps is the mode of action in DEAD-box proteins and that Mss116 acts as an ATP-dependent RNA chaperone (Huang et al. 2005).

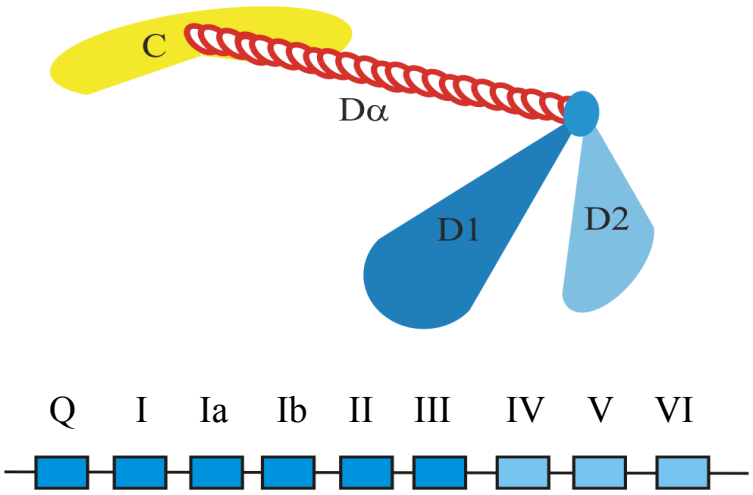

Figure 8: *A schematic of the architecture of Mss116.* Below the 9 conserved motifs are shown and their location.

To shed more light on the mechanism of Mss116p Solem et al. (2006) applied Mss116p with different mutations to their *in vitro* splicing and unwinding assays. Of special interest were the mutation of motif III (SAT → AAA) that couples ATPase and unwinding activity and the mutation of motif I (K158R) which is necessary for ATP binding and hydrolysis. The SAT→ AAA mutation did not influence RNA and ATP binding but reduced RNA stimulated ATPase activity by 80%. More importantly unwinding activity was eliminated with the substrate used (Solem et al. 2006) although with shorter substrate residual unwinding activity still could be observed (DelCampo et al. 2007). Nevertheless although splicing was reduced in a minor manner, which might result from reduced ATPase activity, it still resulted in about half the amount of mRNA compared to wild type Mss116. In contrast the K158R mutation, which abolishes ATP binding as well as ATPase and unwinding activities while maintaining RNA binding, shows no spliced product (Solem et al. 2006). These results show that ATP binding and hydrolysis is necessary for splicing of ai5γ whereas unwinding activity is not. This is further indicated by the elevated ratio between Mg^{2+} and ATP needed for efficient Mss116p-stimulated splicing of the group II intron ai5γ (1mM ATP versus 8mM Mg^{2+}), which was proposed to be optimal in nearly equimolar concentrations for unwinding activity (Iost et al. 1999). This large excess of Mg^{2+} additionally inhibits the helicase activity of Mss116p (Solem et al. 2006).

Consistent with the unimportance of unwinding activity for splicing of the ai5γ gintron is the fact that ai5γ folds properly *in vitro* without misfolding (Su et al. 2003 and 2005). A proposed mechanism of Mss116p is stabilising an on pathway intermediate instead of unwinding RNA. This would correlate with the finding of the folding control element κ–ζ which has to be formed in order to trigger the collapse of ai5γ (Waldsich and Pyle 2008). ATPase activity might be needed for conformational changes required for substrate release.

To understand the interchangeability of Mss116p and Cyt19, Solem et al. (2006) applied a variety of different DEAD-box proteins to their *in vitro* splicing assay and found out that additionally to Mss116p and Cyt19 also Ded1 could promote splicing, while NS3, Dbp8 and Dp8/Esf2 couldn't. All 6 contain the helicase core but only Mss116, Cyt19 and Ded1 have a distinct C-terminal domain harboring several arginine side chains. It was shown that truncation of this CTD abolished RNA annealing activity in Ded1 (Yang and jankowsky 2005) and promotes splicing in Mss116p (Huang 2004).

Besides adding of Mss116p lowers the amount of Mg^{2+} ai5γ needs for establishing its native structure significantly (Solem et al. 2006).

These findings indicate that there is a distinct subgroup of DEAD-box proteins that assist in folding of large RNAs by stabilizing on pathway intermediates. Alternatively Lambowitz and coworkers proposed Mss116p and related proteins to act as RNA chaperones by unwinding misfolded structures allowing native structure conformation (Halls et al. 2006, DelCampo et al. 2007).

Aims

Up to now a lot of progress has been made in elucidating RNA structure and their folding pathways *in vitro* (Woodson 2005, Williamson 2000, Pyle et al. 2007). It has been shown that folding of RNA molecules is diverse in its pathways and dependencies of extrinsic factors. While e.g. some group I introns where shown to need RNA chaperones to dissolve kinetic traps (Rajkowitch et al. 2007) other RNA molecules do not have a rugged free-energy landscape but fold directly into their native structure (Pyle et al. 2007). Although the effort in understanding the folding of large RNAs *in vitro* has been enourmous, *in vivo* only a few studies are available on that topic (Zhang et al. 1995, Waldsich et al. 2002, Semrad and Schroeder 1998).

An interesting extrinsic factor many large RNA molecules need to fold properly *in vitro* is an unusual concentration of Mg^{2+}. This non-physiological amount of divalent ions is not present in the actual cellular environment, which raises two questions: how does magnesium assist in folding and what cellular factor copes for that?

Magnesium has two possibilities to influence tertiary folding of RNA. First it can interact electrostatically with the phosphate backbone of the RNA molecule in its "diffused" form and second it can directly be bound by the RNA backbone in its "chelated" (Draper 2005). Since the crystal structure of the *O. iheyensis* group II intron (Toor et al. 2008) revealed several Mg binding pockets both attributes of magnesium might contribute to folding of group II introns. Since Mss116p lowers the required magnesium level to a near-physiological one (Solem et al. 2006) a linkage between the function of this DEAD box protein and Mg^{2+} is possible.

In our work we first wanted to get more insight into the intracellular structure of ai5γ comparing it with structural data obtained *in vitro* at high salt and elevated temperature conditions. To do that we used an *in vivo* DMS chemical probing assay. DMS penetrates the yeast cell and all its compartments quickly methylating the N1 of adenosines and the N3 of cytosines if they are solvent accessible meaning neither involved in hydrogen bonding nor protected by binding of a protein or other ligands. With this method it is possible to obtain structural information on both the secondary structure and the tertiary structure. Additionally footprints of proteins bound to the RNA can be found and therefore binding sites identified. After having information on the structure of ai5γ *in vivo* we wanted to determine the influence of Mss116p or, more accurate, of the lack of Mss116p on the structure of ai5γ. To do that we applied the same technique on a Δmss116 strain and compared the DMS pattern to the wild type strain RNA to reveal nucleotides with enhanced or protected N1 and N3.

Methods

Cell growth and DMS modification

Strains

During my work I used two *Saccharomyces cerevisiae* strains kindly provided by Philip Pearlman (Huang et al 2005). The first one was a wild type strain (*MATa ade1 lys1 ura3-52*) that has been used previously to create the second one, a mss116-knock out strain (MATa ade1 lys1 ura3-52 mss116Δ::kan^r).

Cell culture

Yeast strains were grown on YPD plates (1% yeast extract, 2% peptone, 2% glucose, 2% agar). Single colonies were picked, inoculated in 5 ml liquid YPD and grown shaking over night at 30°C. This preculture was transferred to 100 ml YPR with 2% raffinose instead of glucose and grown shaking at 30°C until an OD_{600} of about 1.

DMS modification

The protocol was adapted from Wells et al. (2000). Two times 30 ml of each cell suspension were centrifuged at 6000 rpm and 20°C. The cell pellets were resupended in prewarmed (30°C) YPD to a volume of 1 ml. One of the sample was treated with 50 mM end concentration of DMS (10.7 M) and vortexed briefly. Immediately after DMS treatment all samples were put on 30°C shaking at 300 rpm for 2 min. After this they were put on ice and DMS modification stopped with 0.7 M end concentration of β-mercaptoethanol (14.3 M) and 50 µl isoamylalcohol (98%) and the samples vortexed intensely. The cells were then pelleted at 6000 rpm at 4°C and the cell pellet was resuspended in 1 ml TM-buffer. Again β-mercaptoethanol was added as before to the DMS treated cells and all samples were pelleted again under same conditions. After

removing the supernatant cell pellets were frozen at –80°C for at least 30 min to maximum over night.

RNA preparation

RNA preparation with hot acid phenol extraction

Frozen cells were first resuspended in 600 µl AE buffer and 100 µl 10 % SDS. Then the resulting cell suspension is divided into two 1.5 ml tubes that contain each 700 µl preheated (65°C) phenol pH 4.5.

After vortexing briefly the tubes were put into liquid nitrogen until completely frozen. Then the samples are thawed at 65°C, vortexed for 30 seconds and shaken at 65°C for 4 min. This freezing-thawing step was done 3 times.

Then the samples were centrifuged 5 min at 13200 rpm and room temperature. The aqueous phase was transferred to tubes containing 700µl PCI. After strong vortexing for 30 seconds the samples were centrifuged as before and the aqueous phase transferred to 700 µl CI. The same procedure as with PCI extraction resulted in an aqueous phase containing RNA that was precipitated with 2.5x Volume ethanol / 0.3 M NaOAc for at least 1 hour at –20 °C. After discarding the supernatant and drying the pellet, the latter was resuspended in 100 µl ddH_2O. To get rid of residual DNA 3 µl DNAse I [2 U/µl], 0.4 µl RNAse Inhibitor [40 U/µl] and 12 µl 10 x DNAse I buffer was added and the reaction incubated on ice for 1 hour. After this PCI and CI extraction was carried out as before and the RNA precipitated with 2.5x Volume ethanol / 0.3 M NaOAc and 2 µl 0.5M EDTA pH 8.0 for at least 1 hour at –20 °C finally resuspending the RNA pellet in 10 µl ddH_2O.

Additionally this procedure was modified by a douncing step introduced after resuspending the frozen cells. In this douncing step the cell wall was to be actively disrupted by shear forces.

RNA preparation with glass beads

First 2 g of acid treated glass beads (0.2 mm) were put into a 15 ml Falcon tube. 1 ml of LETS buffer equilibrated phenol pH 4.5 was added. The frozen cell pellet was resuspended in 800 µl ice-cold LETS buffer and added to the glass beads/phenol mix. The solution was then vortexed strongly at 4°C for 3 minutes. After adding another 1.5 ml ice-cold LETS buffer and short vortexing the sample was centrifuged 5 min at 5000 rpm. The aqueous phase was extracted first twice with 1.5 ml PCI by vortexing and centrifuging as above and then once with 1.5 ml CI. After this the aqueous phase was precipitated with 2.5x Volume ethanol / 0.3 M NaOAc for 3 hours at –20°C, centrifuged as before and resuspended in 100 µl ddH_2O. DNA digestion and subsequent PCI and CI extraction were done as before and the pellet resuspended in 10 µl ddH_2O.
Additionally this procedure was modified by first treating the cells as in "RNA preparation with hot acid phenol extraction" until the freezing-thawing step was applied and after this switching to glass beads as described.

RNA preparation with NucleoSpin^{TM} RNA L Kit

RNA was prepared according to the provided protocol. After elution the RNA was precipitated with 2µl 0.5 M EDTA pH 8.0 and 2.5x Volume ethanol / 0.3 M NaOAc over night and centrifuged at 4°C and 16000 x g for 1 hour. After discarding the supernatant and drying the pellet was resuspended in 10 µl ddH_2O.

RNA preparation with Ribopure^{TM}-Yeast Kit

RNA was finally prepared using Ribopure^{TM}-Yeast Kit by Ambion according to the protocol. The main steps are cell breakage with Zirconia beads, phenol extraction of the lysate and glass-fiber filter purification. The vortexing (cell breakage) step was done at 4°C for 15 min and the elution volume was 100µl of elution solution. After DNA digestion the RNA was again precipitated with 2µl 0.5 M EDTA pH 8.0 to remove nucleotides and 2.5x Volume ethanol / 0.3 M NaOAc over night and centrifuged at 4°C and 16000 x g for 1 hour. After discarding the supernatant and drying the pellet was resuspended in 10 µl ddH_2O.

Determination of RNA concentration

0.5 µl of the prepared RNA were diluted in 500 µl dH_2O and A_{260}, A_{280} as well as A_{320} were measured. After checking the ratio of A_{260}:A_{280} to ensure proper removal of ethanol, phenol or proteins, the concentration of RNA was determined with

$$[RNA]\ \mu g/\mu l = (A_{260} - A_{320}) \cdot 40.$$

RNA quality controls

Quality control for RNA on agarose gel

To get a first glance at RNA quality 1 µg RNA was loaded on a native 1% agarose gel (in 1x TBE). The gel ran at 100 mA as limiting parameter.

RT-PCR for analysis of mitochondrial DNA contamination

To check for contamination of mitochondrial DNA in the RNA solution reverse transcription with subsequent PCR was applied.

For reverse transcription 4 µl of a 0.5 µg/µl RNA solution where mixed with 10 µl ddH_2O and 1µl RP_ai5g_R primer. The reaction was then incubated for 5 minutes at 70 °C and subsequently put on ice for 5 minutes. Then 1 µl MMLV reverse transcriptase (200 U/µl), 5 µl 10x MMLV-RT buffer, 1.3 µl 10 mM dNTPs, 1µl RNAse Inhibitor (40 U/µl) and 2.7 µl ddH_2O were added. The reaction was put on 42°C for 1 hour for primer extension and then 15 minutes on 70 °C for enzyme deactivation. After this 1 µl RNAse H (5 U/µl) was added and the RNA digested for 20 minutes at 37°C. Finally the RNAse H was inactivated by putting the sample on 65°C for 20 minutes.

Thereafter the PCR reaction was set up by mixing 10 µl 5x GoTaq™ Green buffer, 1 µl 10 mM dNTPs, 0.5 µl of each 100µM primer RP_ai5g_R, RP_ai5g_M, RP_ai5g_F, 0.25 µl GoTaq DNA Polymerase (5 U/µl) and 32.25 µl ddH_2O. To check if the desired RNA is present 5 µl of the cDNA sample from the previous RT reaction was added. For negative control 5 µl ddH_2O was added instead. For mitochondrial DNA check 2µg RNA was applied in a volume of 5 µl. The PCR was run with following parameters:

First the sample was denatured 5 minutes at 95°C. Then a denaturing step of 95°C for 45 seconds, an annealing step of 58°C for 45 seconds and an elongation step of 72°C for 2 minutes were applied for 30 cycles. The final extension was carried out at 72°C for 10 minutes and the reaction was kept on 4°C until loaded on a 1% agarose gel for analysis.

PCR for analysis of nuclear DNA contamination

To check for nuclear DNA contamination PCR using primers for nuclear encoded Mss116p was applied. 2 µg RNA in a volume of 5 µl were mixed with 10 µl 5x GoTaq^{TM} Green buffer, 1 µl 10 mM dNTPs, 0.5 µl of each 100µM primer CW01F, CW01R, 0.25 µl GoTaq DNA Polymerase (5 U/µl) and 33.25 µl ddH_2O. For positive control the plasmid CW01 carrying the Mss116 gene was used instead of RNA. For negative control 5 µl ddH_2O was added instead. The PCR was run with the same parameters as before.

Primer purification and kinasing

Gel purifying of oligonucleotides

Incoming oligonucleotides for reverse transcription were mixed 1:1 with 1x Loading Buffer and loaded on a 20% denaturing acrylamide gel. The gel was run with 20 W until the Xylene Cyanol marker which runs at 25 nucleotides length reached the second half of the gel. Oligonucleotide bands were then identified by UV shadowing. The gel pieces were transferred to perforated 0.5 ml PCR tubes, inserted in 1.5 ml Eppendorf tubes and centrifuged at top speed for 10 minutes for crushing. 1 ml Elution buffer was added and the tubes were shaking overnight at room temperature before 2.5 times the volume was added in ethanol (p. a.). The oligonucleotides were precipitated over night at –20 °C, centrifuged at top speed and 4°C and resuspended in dH_2O. The concentration was determined by using following Extinction coefficients:

$\varepsilon_{adenine} = 15346.6$

$\varepsilon_{guanine} = 11751.8$

$\varepsilon_{thymine} = 9929.7$

$\varepsilon_{cytosine} = 7623.2$

To account for neighbouring effects a factor of 0.9 was included in calculation. The final Extinction coefficient then was determined by

$$\varepsilon_{total} = 0.9 \cdot (\varepsilon_{adenine} \cdot n_{adenine} + \varepsilon_{guanine} \cdot n_{guanine} + \varepsilon_{thymine} \cdot n_{thymine} + \varepsilon_{cytosine} \cdot n_{cytosine})$$

The concentration of the related oligonucleotide was determined by measuring A_{260} and using the formula

$$\text{concentration (mol/L)} = A_{260}/(\varepsilon_{total} \cdot \text{pathlength (cm)}) \cdot \text{dilution}$$

were the pathlength of the cuvette was 1 cm.

Primer Kinase Reaction

Oligonucleotides for reverse transcription were diluted to a concentration of 10 µM. 1 µl of that dilution (10 pmol DNA) was mixed with 2 µl of 10x T4 polynucleotide kinase reaction buffer (NEB), 10µl dH_2O, 1 µl T4 Polynucleotide Kinase (NEB, 10 U/µl) and 6 µl of γ-P^{32}-ATP (6000 Ci/mmol). This solution was incubated for 30-45 minutes at 37 °C. The reaction was stopped by adding 1µl 0.5 M EDTA pH 8.0 (to remove excess ATP) and incubating at 95 °C for 1 min. After cooling down on ice for 2 minutes, 1 µl glycogen was added and the labelled oligonucleotide was precipitated with 2.5x Volume ethanol / 0.3 M NaOAc for 1h to overnight. After spinning for 1 hour on 4°C and top speed and air drying the pellet was resuspended in 30 µl dH_2O.

Reverse transcription and subsequent gel run

Reverse Transcription using AMV Reverse Transcriptase

For hybridisation 1 µl of 4.5x Hybridisation buffer was mixed with 2.5 µl RNA (40 µg) and 1 µl of 5'end P^{32}-labeled primer. This mixture was then incubated at 95°C for 1 minute and then slowly cooled down to 42°C. For extension a master mix was prepared using 0.6 µl Extension buffer, 0.3µl 2.5 mM dNTP mix, 1.0 µl ddH2O and 0.3 µl AMV Reverse Transcriptase [10 U/µl]. 2.2 µl of this master mix were added to the annealing reaction that was previously cooled down. To obtain sequencing lanes for adenine and cytosine 0.75 µl of 1 mM ddTTP (for adenine) or 1 mM ddGTP (for cytosine) was

added the respective reactions. The extension reaction was then incubated for 1 hour at 42°C. To degrade the RNA 1.5 µl 1 M NaOH was added to each reaction, which was then again incubated at 42°C for 1 hour. The pH was neutrilized by 1.5 µl 1 M HCl and the cDNA precipitated with 1µl 0.5 M EDTA pH 8.0 and 30 µl Ethanol / 0.3 M NaOAc pH 5.0 at –20°C for 1 hour to overnight. After this the samples were centrifuged for 1 hour at 13200 rpm. After discarding the supernatant the pellet was air dried for about 10 min and resuspended in 8 µl Loading buffer.

Reverse Transcription using ThermoScript™

Some regions were GC-rich and had to be mapped at a higher temperature with another enzyme in order to get a quantifiable signal. For this ThermoScript™ was used. For hybridisation 2 µl of 4.5x Hybridisation buffer was mixed with 2.5 µl RNA (40 µg), 2.5 µl ddH_2O and 2 µl of 5'end P^{32}-labeled primer. This mixture was then incubated at 95°C for 1 minute and then slowly cooled down to 55°C. For the extension 0.5 µl ThermoScript™ [15 U/µl], 2 µl 10 mM dNTPs, 1µl 0.1 M DTT, 4 µl 5x cDNA Synthesis buffer and 3.5 µl ddH_2O were added. To obtain sequencing lanes for adenine and cytosine 1.5 µl of 5 mM ddTTP (for adenine) or 5 mM ddGTP (for cytosine) was added the related reactions. The resulting reaction was incubated at 55°C for 45 minutes. Then the sample was put on 85°C for 5 minutes. After adding 4.5 µl 1 M NaOH the RNA was digested for 30 minutes. The pH was then neutralized with 4.5 µl 1 M HCl and the cDNA precipitated as before but with 2µl 0.5 M EDTA pH 8.0 and 60 µl Ethanol / 0.3 M NaOAc pH 5.0.

Poisoned Primer Assay

All steps are carried out as described in "Reverse Transcription using AMV Reverse Transcriptase". The only difference is that in the extension mix instead of dNTPs the poisoned mix containing ddCTP instead of dCTP is used.

Resolving the cDNA pool on a denaturing gel

For resolving the cDNA pool resulting from reverse transcription an 8 % denaturing polyacrylamid gel was prepared. This gel was set up with 1x TBE as running buffer. The pockets were rinsed with 1x TBE and the gel prerun 1 hour with 40 W. After the prerun half of the sample was loaded on the gel and under identical conditions run until the Bromphenol Blue dye reached the bottom. The gel was then dried for a minimum time of 1 hour and exposed in phosphoimager cassettes from overnight to 3 days, which were than scanned on a phosphoimager (GE Healthcare).

Analysis

For quantitative analysis of the phosphoimager scans ImageQuantTM (Molecular Dynamics, GE Healthcare) was used.

Analysis of wild type strain DMS modification

Line objects of same length were applied to the RT stop control lane and the DMS modification lane. The resulting graph was further analyzed with Excel. Sequencing lanes were used to assign nucleotides adenosine and cytosine to referring peaks. The RT stop control that was not treated with DMS revealed bands that correspond to natural stops of reverse transcription. Peaks related to this bands where treated as breaks instead of DMS modification. DMS modification intensity was characterized by building the sum of all peak heights and calculating the percentage of the individual peak intensities.

Analysis of knock out strain DMS modification

Line objects of same length were applied to the RT stop control lane and the DMS modification lane of both the wild type strain and the knock out strain. DMS modification intensities were defined as before. To reveal nucleotides that were enhanced in or protected from DMS modification in the knock out strain the ratio between DMS intensities of nucleotides carrying DMS modifications were determined

(Δmss116 : wild type). Ratios over 1.5 were assigned as medium, over 2.0 as strong enhancements. Ratios under 2/3 were assigned as medium, under 0.5 as strong protections.

Analysis of the poisoned primer assay

Rectangular objects of the same size were applied to the bands related to pre-RNA and mRNA. For background subtraction the same element was also applied to a band unrelated to both species. The volume of the objects were defined with ImageQuantTM and analyzed with Excel. There the background intensity was subtracted from both volumes related to one species. Then the percentage of each species was determined for the wild type and knock strain.

Materials

Primers

All sequences are written from 5’ to 3’.

Oligonucleotides for mapping ai5γ

Sc82:	GTTACCATTTTAATACACTTG
Sc160:	GTTCCTTCATCTTTTTTTTATAA
Sc234:	GTTTTCAATTAGTGGTGTAAG
Sc312:	TTTCCAATACATAACATCAACC
Sc375:	GTAAATATCTAACTTAGCTCTC
Sc440:	TTGGTATTATTATTATTTTTTATTATTA
Sc520:	TATTTCTTAATCCAATAATTATTTATA
Sc588.	GCTTTTTATAC
Sc656:	CACCTATAGTATAAGTTAGCAG
Sc682:	ATAATAATAATATATATATAAATAAAGATAGGC
Sc745:	ATTTATCAGTTATATATAAACCTCC
Sc805:	CGGCTCACGCAATATC
Sc866:	CCCGATAGGTAGACCTTTAC
Sc-6:	GAAAATAGCACCCATTGATAATAC

Primers for analysis of mitochondrial DNA contamination

RP_ai5g_1F:	CTTACTACGTGGGACATTTTC
RP_ai5g_1M:	GGTTTATTCTGTTTTATCATAAATACG
RP_ai5g_1R:	CATATAAAATATAGATAAATAAGAATAATG

Primers for analysis of nuclear DNA contamination

CW01F: ACGCGTCGACATGTTGACCTCTATATTGATAAAAG
CW01R: CGGAATTCCTAATATATGTTGCTGTTTCTACTG

Enzymes and Kits

Enzymes

Enzyme	concentration (U/µl)	Buffer (if used)	Company
AMV Reverse Transcriptase	10		Promega
MMLV Reverse Transcriptase	200	10x MMLV RT Buffer	Promega
GoTaq™ DNA Polymerase	5	5x Green GoTaq™ Buffer	Promega
ThermoScript Reverse Transcriptase	15	5x cDNA synthesis Buffer	Invitrogen
DNAseI	2	10x DNAse I Buffer	NEB
RNAse Inhibitor	40		NEB
RNAse H	5		NEB
T4 Polynucleotide Kinase	10	T4 Polynucleotide Kinase Reaction Buffer	NEB

Kits

RiboPure™-Yeast Kit (Ambion)
NucleoSpin™ RNA L Kit (Machery-Nagel)

Buffers, Solutions and Media

All buffers, solutions and medias had to be RNAse free. Therefore all stock solutions and all media had to autoclaved. Raffinose (10%) and glucose (10%) solutions were sterile filtrated.
If not mentioned otherwise the solution is prepared with ddH_2O

Cell culture:

YPD-plates:
1 % Yeast Extract
2 % Peptone
2 % Glucose
2 % Agar
55 mg/L Tyrosine
55 mg/L Adenine

YPD-media:
1 % Yeast Extract
2 % Peptone
2 % Glucose
55 mg/L Tyrosine
55 mg/L Adenine

YPR-media:
1 % Yeast Extract
2 % Peptone
2 % Raffinose
55 mg Tyrosine

55 mg Adenine

Raffinose and glucose were always added after autoclaving right before inoculating.

General solutions

1 M Tris pH 7.5 bought from AppliChem

0.5 M EDTA pH 8.0
pH adjusted with 5 M NaOH

2 M Tris pH 8.0 pH adjusted with HCl (concentrated)

1 M $MgCl_2$

3 M NaOAc pH 5.4
pH adjusted with acidic acid (concentrated)

Ethanol (p.a.) / 0.3 M NaOAc pH 5.4

10 % SDS (sodium dodecylsulphate, not sterile)

5 M LiCl

1x TBE: 0.089 M TRIS Base
0.089 M Boric Acid
2 mM EDTA pH 8.0

Loading buffer: 7 M Urea
25 % Sucrose
0.25 % Bromphenol Blue
0.25 % Xylene Cyanol

in 1x TBE

1 M K-Hepes pH 7.0
pH adjusted with KOH

3 M KCl

50 % Sucrose

DMS-modification

TM-buffer: 10 mM TRIS pH 7.5
1 mM $MgCl_2$

RNA preparation

AE-buffer: 50 mM NaOAc pH 5.4
10 mM EDTA pH 8.0

LETS-buffer: 0.1 M LiCl
10 mM EDTA
10 mM TRIS pH 7.5
0.2 % SDS

Acid-treated glass beads (diameter 0.2 mm)

Phenol pH 4.0, water saturated

CI: Chloroform : Isoamyl alcohol 24:1

PCI: Phenol : CI = 1:1

RNA quality control

Agarose gel (1% agarose in 1x TBE)

dNTP mix	10mM dATP
	10mM dCTP
	10mM dGTP
	10mM dTTP

Primer purification and kinasing

20 % Acrylamide Solution:

7 M Urea
20 % Acrylamide / Bisacrylamide 19:1
in 1x TBE

Elution buffer:	250 mM NaOAc pH 5.4
	10 mM TRIS pH 7.5
	2 mM EDTA pH 8.0
	0.1 % SDS

γ-P32-ATP: 6000 Ci/mmol

Reverse transcription and subsequent cDNA resolvation

4.5x Hybridisation Buffer:

225 mM K-Hepes pH 7.0
450 mM KCl

10x Extension Buffer:

1.3 TRIS pH 8

100 mM $MgCl_2$

100 mM DTT

dNTP solution: 2.5 mM dATP

2.5 mM dCTP

2.5 mM dGTP

2.5 mM dTTP

1 mM ddTTP

1 mM ddGTP

1 M NaOH

1 M HCl

Poisoned mix 0.88 mM dATP

0.88 mM dGTP

0.88 mM dTTP

3.68 mM ddCTP

8 % Acrylamide Solution 7 M Urea

8 % Acrylamide / Bisacrylamide 19:1

in 1x TBE

10 % Ammoniumpersulfate

TEMED

Results

General approach

To get structural data for the ai5γ intron *in vivo* we used dimethyl sulfate to chemically probe its structure (Figure 9). DMS penetrates the cell and its compartments quickly thereby methylating all accessible N1 of adenosines and N3 of cytosines. Hydrogen bonding or binding of a protein abolishes the accessibilities of these nitrogens. After RNA preparation modified adenosines and cytosines were identified by reverse transcription since attached methyl groups cause the reverse transcriptase to stop the extension. As the RNA pool is composed by an ensemble of RNA molecules the accessibility to DMS *in vivo* has to be interpreted as an average value. The cDNA pool then was resolved on a denaturing gel revealing all stops caused by DMS modification, as they don't show up in the RT stop control lane. The intensities of the related bands directly correlate to the solvent accessibility of the respective nucleotide and these results can then be plotted on the secondary structure map.

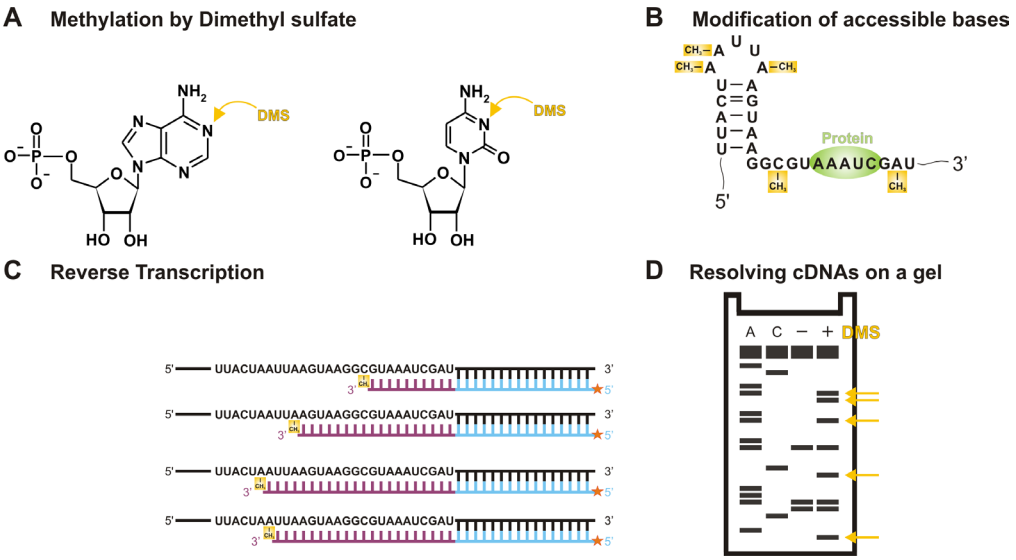

Figure 9: *The basic principles of the in vivo DMS chemical probing technique.* A) Methylation site of adenosine (left) and cytosine (right). B) Only solvent accessible N1 of As and N3 of Cs are methylated. C) After RNA preparation reverse transcriptase stops at modified nucleotides. D) The cDNA pool after reverse transcription is resolved on a denaturing gel. The first two lanes are sequencing lanes. The third lane is showing cDNA derived from cells that were not treated with DMS to reveal all natural stops of the RT. The forth lane is showing bands corresponding either to a natural stop, which would also be present in the RT stop control lane, or DMS modification

Screening for an adequate RNA preparation method

The biggest challenge in the method of *in vivo* DMS probing is working with RNA expressed at endogenous levels. There is no reasonable possibility to get rid of rRNA or tRNA, which are the most abundant RNAs in the cell (Figure 12). Furthermore ai5γ is located in the mitochondria additionally increasing the importance of a proper RNA preparation technique that efficiently lyses cells and their organelles. Besides that the quality of the RNA is crucial as the resolution of the final denaturing gel is at single nucleotide. Since in the end the intensity of the bands or in other words the band pattern has to be interpreted, breaks have to be kept at a minimum.
In short the RNA preparation had to fulfil following premises: Mitochondria have to be efficiently lysed to ensure the highest mtRNA versus cytoplasmatic and nuclear RNA ratio. The quality of the RNA has to be very high meaning that degradation and residual cellular components have to be kept at a minimum. DNA contamination has to be very low.
The latter is important to use approximately the same amount of ai5γ RNA in reverse transcription.
The different extraction procedures that were initially tested are described in the chapter "Methods".

Growth media

Although preculture yeast cells are typically cultivated in YPD, raffinose as carbohydrate source resulted in significantly higher amount of ai5γ RNA although the total RNA used was the same, since more mitochondria are produced when using this sugar (Figure 10).

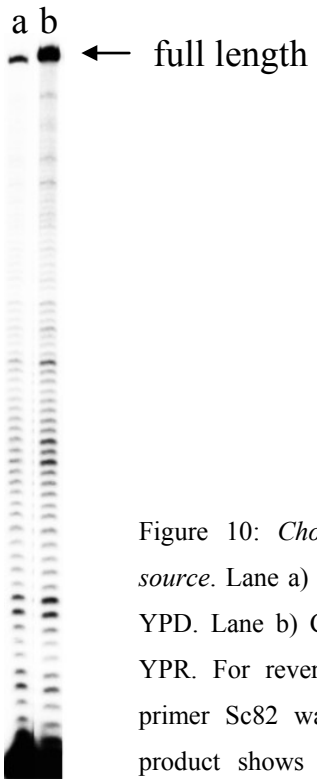

Figure 10: *Choosing carbohydrate source*. Lane a) Cells were grown in YPD. Lane b) Cells were grown in YPR. For reverse transcription the primer Sc82 was used. Full-length product shows significantly higher output in YPR as growth media.

Total RNA extracted

The different methods varied significantly in the amount of RNA extracted from a cell culture with 10^7 cells (~1ml at an OD_{600} of 1). Due to the endogenous level of ai5γ RNA, a huge amount of RNA was needed for mapping the entire intron.
The "hot acid phenol" method with an included douncing step to actively disrupt the cell wall yielded most RNA with 20 µg per ml cell culture. The least amount of RNA was obtained with the NucleoSpin™ Kit (see table 1).

DNA contamination

DNA contamination was very prominent in the two "homemade" protocols "Hot acid phenol" and "Glass beads". Actually only the Ribopure™ Yeast Kit (Ambion) barely showed nuclear or mitochondrial DNA contamination (Figure 11).

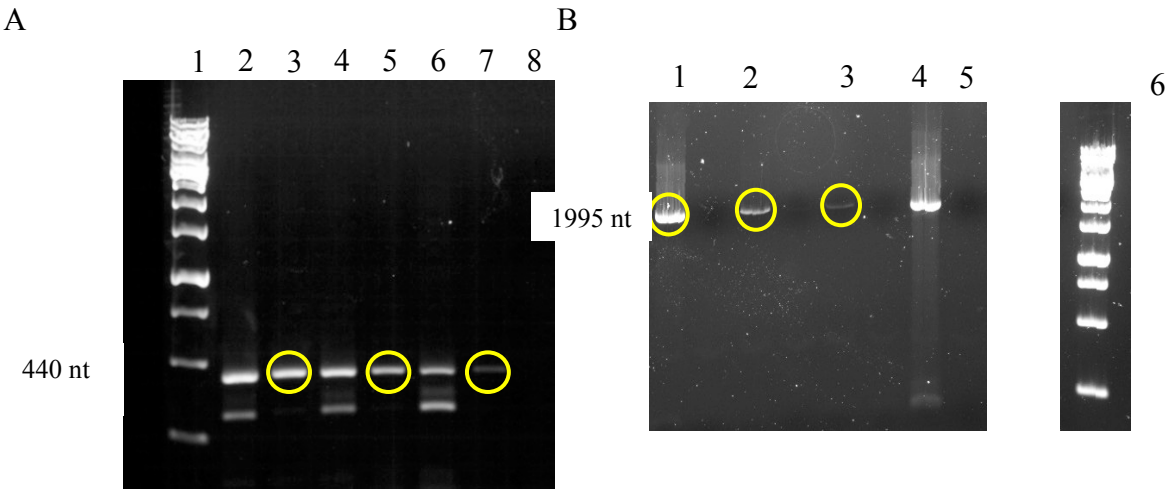

Figure 11: A) *Mitochondrial DNA contamination*; 1 – 1 kb DNA ladder (NEB), 2 – RT-PCR of RNA prepared with hot acid phenol method, 3 – PCR of RNA prepared with hot acid phenol method showing DNA contamination, 4 – RT-PCR of RNA prepared with NucleoSpin™, 5 – PCR of RNA prepared with NucleoSpin™, 6 - RT-PCR of RNA prepared with Ribopure™, 7 - PCR of RNA prepared with Ribopure™, 8 – negative control showing PCR without template. B) *Nuclear DNA contamination*; 1 – PCR of RNA prepared with hot acid phenol method, 2 – PCR of RNA prepared with NucleoSpin™, 3 – PCR of RNA prepared with Ribopure™, 4 - positive control with plasmid DNA as template, 5 - negative control showing PCR without template, 6 - 1 kb DNA ladder (NEB)

RNA quality

RNA quality (figure 12) and signal to noise ratio on the cDNA Polyacrylamid gel (figure 13) are directly related making this the most important aspect of RNA preparation. The strongest band referring to full-length cDNA using Sc82 and lowest hydrolysis background showed RNA prepared with Ribopure™ Yeast Kit. Least quality produced NucleoSpin™ (Figures 12 and 13).

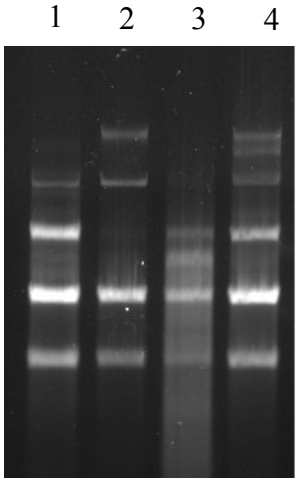

Figure 12: *Agarose gel showing RNA quality of the different RNA preparation methods*. 1 – hot phenol method, 2 – glass bead method, 3 - NucleoSpin™, 4 - Ribopure™

Figure 13: *Representative gel showing signal to noise ratio of different preparation methods*. Also mitochondrial lysis efficiency is shown by the amount of full length. RNA was prepared with 1 – Ribopure™, 2 – glass bead method, 3 – hot acid phenol method, 4 - NucleoSpin™

Efficiency of mitochondrial lysis

This could be seen in the amount of full-length cDNA after reverse transcription (Figure 13), since the same amount of RNA is used in reverse transcription. Here both kits used were significantly more efficient than the homemade protocols.

Summary

Comparing all the tested features the Ribopure^{TM} Yeast Kit was by far showing the best results (table 1).

Method	**RNA (µg) per 10^7 cells (average**	**DNA contamination**	**RNA quality**	**Efficiency of mitochondrial lysis**
Hot Acid Phenol	10	n.d	n.d	n.d
Hot Acid Phenol*	20	+++	++	+
Glass bead	10	++	++	+
Glass bead**	6	n.d	n.d	n.d
NucleoSpin (Machery-Nagel)	3.5	+	+	++
RiboPure Yeast (Ambion)	7	+	+++	+++

Table 1: *Summary of the different features of RNA that were important for the applied in vivo chemical probing assay*. * douncer step was included, ** hot acid phenol treatment in beginning, n.d. – not determined

Determining the optimal DMS concentration

Since the actual experiment in chemical probing of ai5γ *in vivo* is DMS modification, this step is crucial for the quality of data obtained in the end. Here the amount of DMS should lead to maximum one modification per RNA molecule. To find the optimal concentration of DMS we carried out a concentration series starting with 0 and ending with 150 mM DMS. Additionally we performed a stop control to check efficient quenching of DMS by β-mercaptoethanol (Figure 14).

Comparing lane 3 with lane 4-7 shows that 50 mM of DMS is optimal, as it shows a comparable same amount of full-length cDNA and still has a good signal to noise ratio. In the stop control lane no modification is seen indicating that the used amount of β-mercaptoethanol is sufficient to quench the DMS molecules.

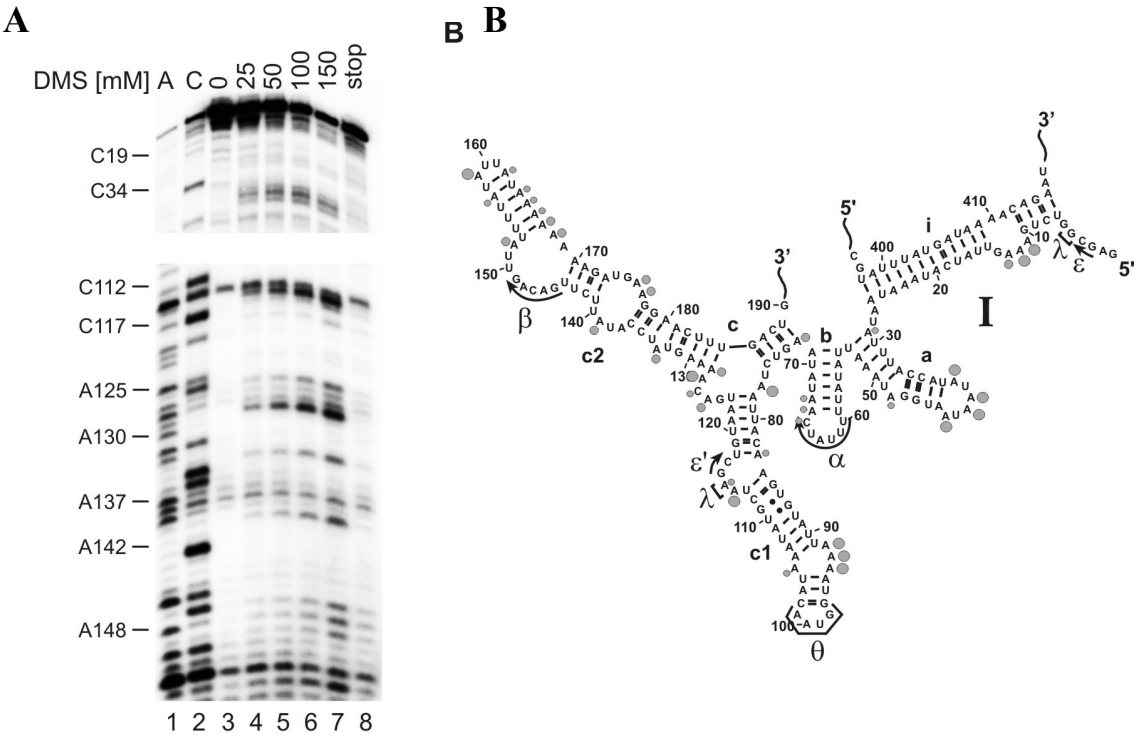

Figure 14: *Gel for determining DMS-concentration and proper quenching.* A) Lanes 1 and 2 show sequencing for adenosine and cytosine, in lane three cells were not treated with DMS, lanes 4 to 7 show cells treated with an increasing amount of DMS starting with 25 mM ending with 150 mM, in lane 8 the quenching reagent β-mercaptoethanol was added before DMS. Residues 46 to 108 are not shown. B) Secondary structure map showing the modification pattern represented by lane 5 (50 mM DMS). The size of the gray dots corresponds to the relative modification intensity.

Ai5γ splices efficiently *in vivo* in the presence of Mss116p

In the poisoned primer assay reverse transcriptase has only three nucleotides in their deoxy form available. Instead of dCTP ddCTP is added as the fourth nucleotide. This terminates the extension of the Sc-6 primer either after 8 nucleotides if the intron is still integrated (pre-RNA) or after 14 nucleotides if the intron is spliced (mRNA) (figure 14). With this *in vivo* splicing assay spliced mRNA and unspliced precursor RNA can be detected and the splicing efficiency of ai5γ in the respective strain can be determined. 70 % of the intron is spliced in the wild type strain. In the knock out strain, though, less than 0,5% of the intron has been spliced a signal that is not significantly more intense than noise (figure 15). Additionally DMS treatment does not interfere with the splicing efficiency as seen by comparing the splicing ratio of untreated and treated cells (data not shown).

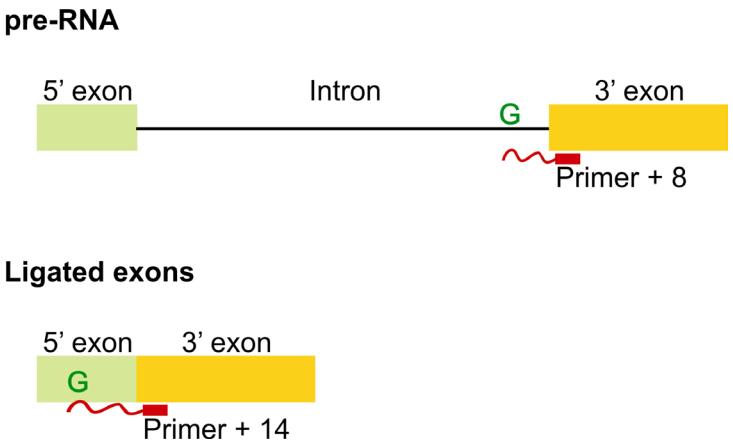

Figure 14: *Poisoned primer assay.* Pre-RNA and mRNA are shown schematically including the first G after primer binding site.

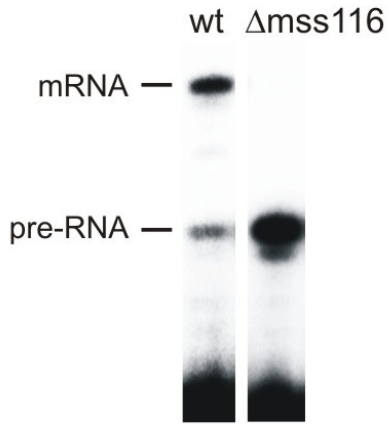

Figure 15: *In vivo splicing assay*. Spliced (mRNA) and unspliced (pre-RNA) products are shown. While ~ 75 % of the intron is spliced in the wild type strain splicing is completely abolished in the knock out strain

The intracellular structure of ai5γ

The secondary structure of group II introns has been predicted early by phylogenetic studies (Michel et al. 1989). There is no high resolution tertiary structure available for ai5γ but a 3D-Modell is partly available due to identified tertiary contacts and surface solvent accessibility data (Costa and Michel 1995, Costa et al. 1997, Boudvillain and Pyle 1998, Boudvillain et al. 2000, Fedorova and Pyle 2005, Michel and Ferat 1995, Swisher et al. 2001).

As described before ai5γ folds into its native structure at elevated temperature and high salt conditions (Su et al. 2003, Swisher et al. 2001). Both parameters are not available in the cellular mitochondria, were folding of ai5γ takes place. To get more insight into the intracellular structure of ai5γ we therefore applied the *in vivo* DMS chemical probing technique (Liebeg and Waldsich 2009). To map the whole intron 13 primers were used (see materials).

The most critical intron domains adopt their proper secondary structure in vivo

In this chapter I will discuss DMS modification data according to the secondary structure of ai5γ *in vivo*. Nucleotides that are not protected by taking part in secondary structure elements but by being involved in tertiary interactions will be discussed later.

The asymmetrical internal loop at the stem i of D1 is accessible to DMS showing strong modification of A11, A12, A13 (figure 16) and A411 (figure 17). The i-stem itself is formed since all N1 of adenosines and N3 of cytosines involved are protected from methylation by Watson Crick base pairing. Stem a is closed with the adenosines in the heptaloop accessible to DMS. Some N1 of adenosines are weakly modified in stem b (A66, A67 and A69; figure 16), suggesting that while the secondary structure in this region is established in some molecules of the ensemble it isn't. The c1 stem has adapted its native conformation due to the DMS pattern observed. As expected from the proposed secondary structure A92, A93, and A94 are accessible since located in an internal asymmetrical loop (figure 16). Unexpectedly A105, which should not be involved in base pairing, did only show weak modification below the threshold (figure 16). This points to a non-canonical base-pair interaction. The strong modification of

A114, which is part of a 4-nucleotide bulge, then again fits to the proposed secondary structure model (figure 16); A115 and A117 are involved in tertiary interactions and will therefore be discussed later. In the c-c1-c2 3-way junction A125, C126, A127 and A128 that were shown to be important for the collapse of the ai5γ ribozyme D135 *in vitro* (Waldsich and Pyle 2007) are accessible in ai5γ *in vivo* (figure 16). Stem c2 is consistent with the secondary structure in its DMS modification pattern except for A157, A162 and A164-166 that show weak to moderate modification intensities although proposed to be present in a helical conformation (figure 16). Maybe this helix

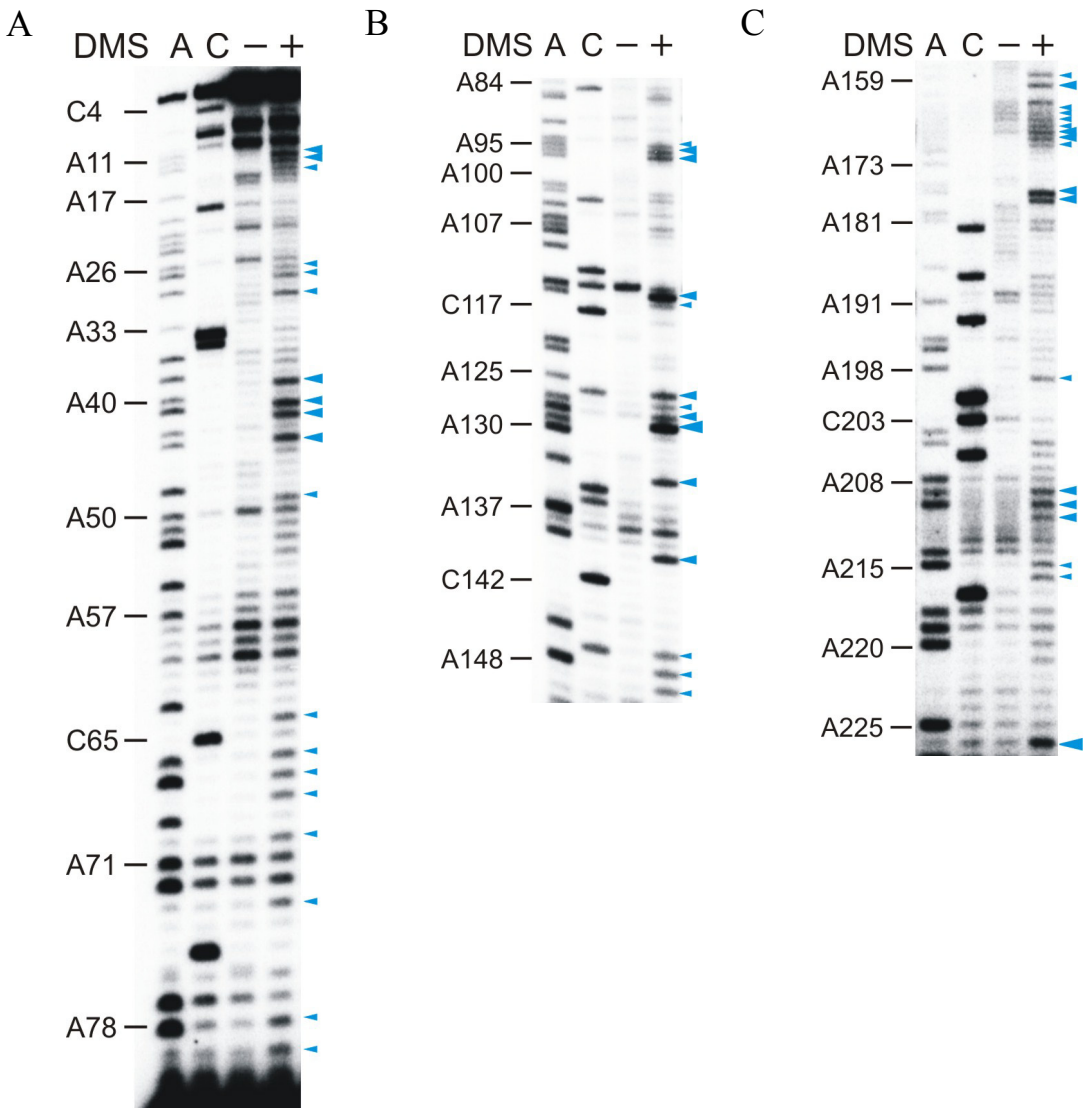

Figure 16: *Representive gels for DMS pattern of ai5γ in wild type cells.* Reference nucleotides are indicated on the left. Lane A and C are sequencing lanes for adenosine and cytosine. Lane – corresponds to wild type cells that were not treated with DMS, lane + to DMS treated wt cells. On the right modified nucleotides are shown. The size of arrows corresponds to intensity of modification (weak, moderate, strong). A) Nts C4 to A78. B) Nts A84 to A148. C) Nts A159 to A225.

is not very stable since it is only short and exclusively involving A-U base pairs (Tinoco et al 1971) suggesting that this helix is only formed in a fraction of the RNA pool. Nevertheless both internal loops show reasonable modification intensities to allow the assumption that the secondary structure of stems c c1 c2 is established. The same can be said for stem d, where no modification is taking place. A208,

A209 and A210 that are part of a tetraloop at the κ element show strong modification as expected (figure 16). Furthermore adenosines and cytosines located in d', d'', d''', d2a, d2b and d3 are not methylated by DMS. A225, A227, A229 (figure 17, strong modification) and A368 (figure 17, moderate modification) that are involved in asymmetrical internal loop between d'' and d''' are modified by DMS indicating the proper folding of this loop that binds the bulged A880 (Hamill and Pyle 2006) and is also called "coordination loop" as it was proposed to align all reaction components necessary for splicing (DeLancestre et al. 2005). Unmodified A228 and A370 might in this context be involved in tertiary interactions. Nucleotides of the 4-way junction

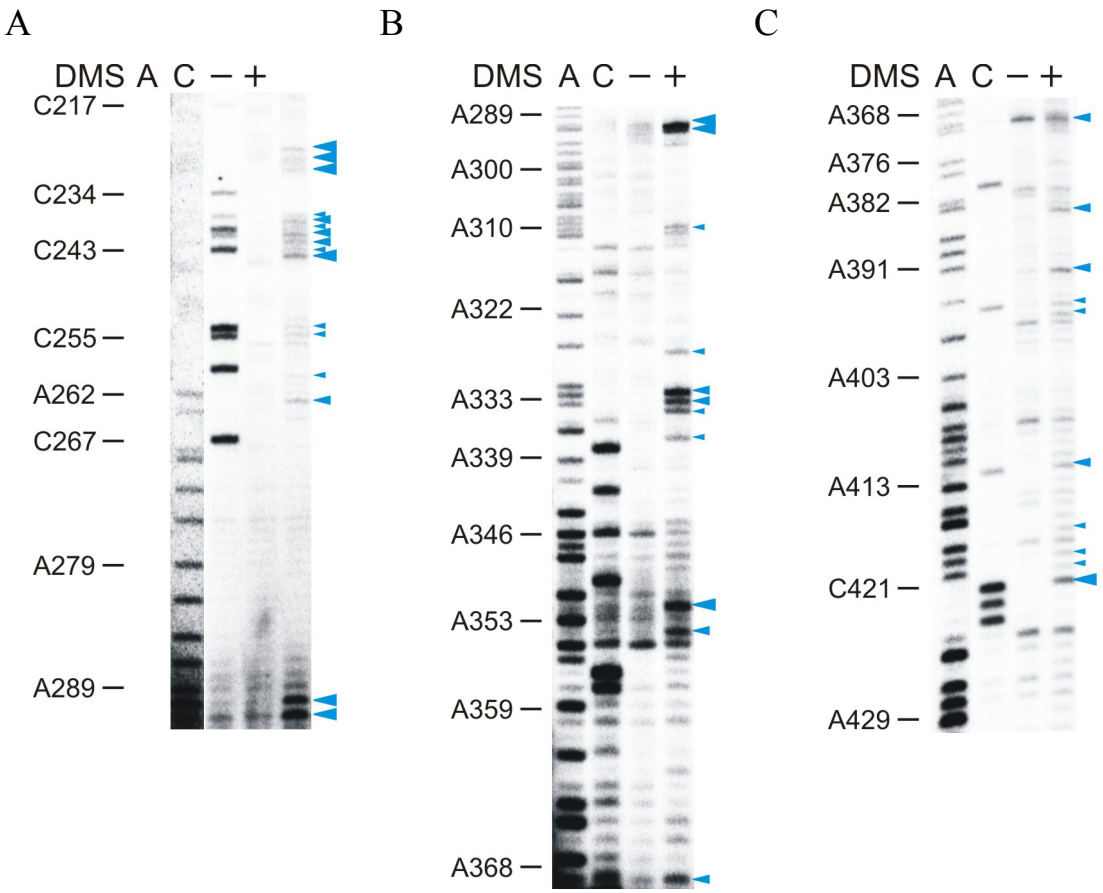

Figure 17: *Representive gels for DMS pattern of ai5γ in wild type cells.* Legend as Figure 16. A) Nts C217 to A290. B) Nts A289 to A368. C) Nts A368 to A429.

between d''', d2a, d2b and d3 are modified weakly (A237, A239 and A242), moderately (A238, C240 and C241) or strongly (C243) indicating that this loop has formed properly and that exons are not bound (figure 17). A289 and A290 that are part of the tetraloop in d2b are modified strongly (figure 17). A326, A331-333 and A336 being located in a terminal loop and the bulged A353 are modified (figure 17). Interestingly A351, which should be involved in Watson-Crick base pairing, is strongly modified indicating a non-canonical base-pair interaction at this nucleotide (figure 17).

In D3 nucleotides involved in a, b and c stems are protected from modification. As A636 in the tretraloop of stem b and A651-653 in the tetraloop of the c stem are moderately accessible to DMS N1 of this adenosines are not involved in any interactions (figure 18). The asymmetrical internal loop at the basal stem of D3 is present due to modifications of A597, A598, A599, A662 and C663 (figure 18 and 19).

In D5 the only significant high modifaction could be seen at A838 that is part of the

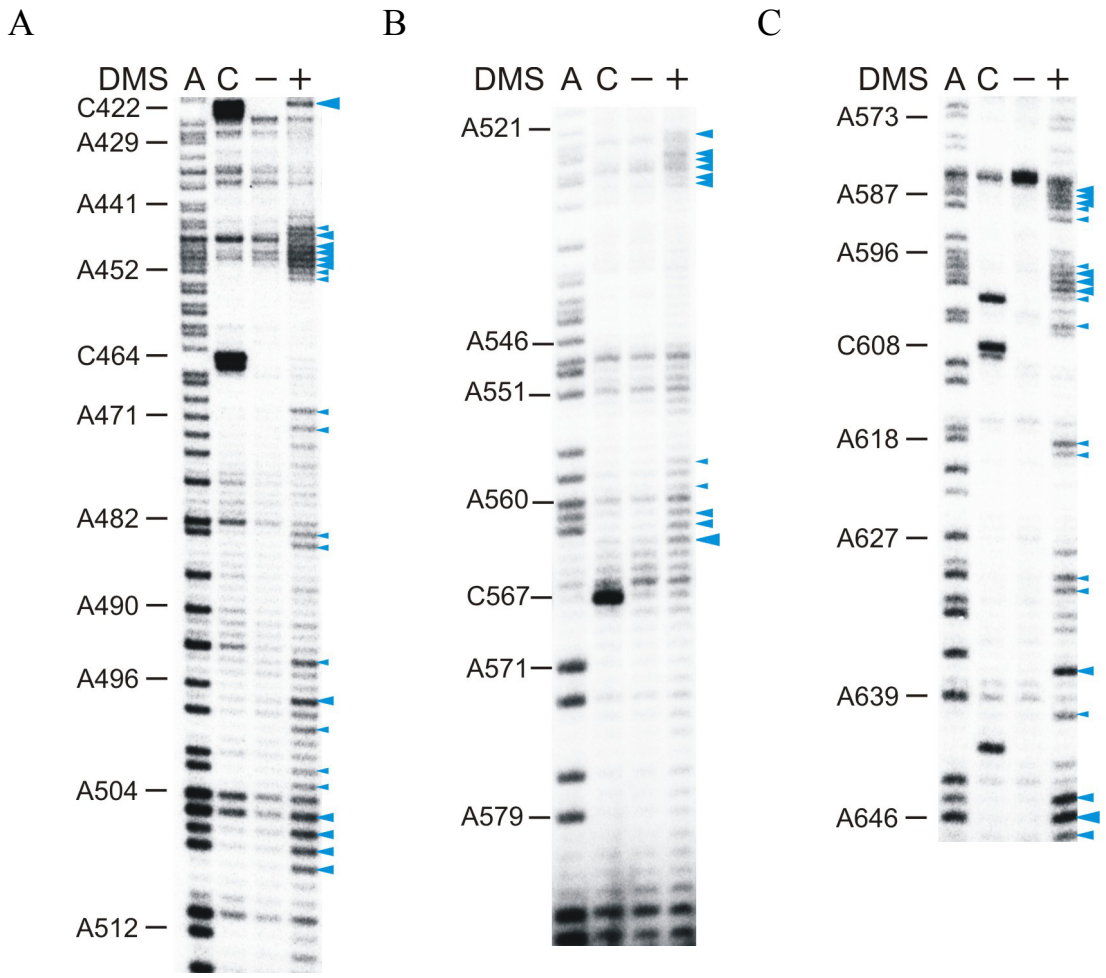

Figure 18: *Representive gels for DMS pattern of ai5γ in wild type cells.* Legend as Figure 16. A) Nts C422 to A512. B) Nts A521 to A579. C) Nts A573 to A646.

important dinucleotide bulge showing that D5 adopts its structure *in vivo* (figure 19). C839 is involved in a tertiary interaction and therefore discussed later.

An interesting phenomenon in general can be seen at some nucleotides at the ends of several stems, e.g. A128 (figure 16), A420 (figure 17), A308 (figure 17) or A630 (figure 18) since they show moderate or even strong modification intensities. This actually hints to a non-canonical base pair interaction like a reverse Hoogsteen A·U pair or might also be explained by breathing of terminal base pairs in duplexes.

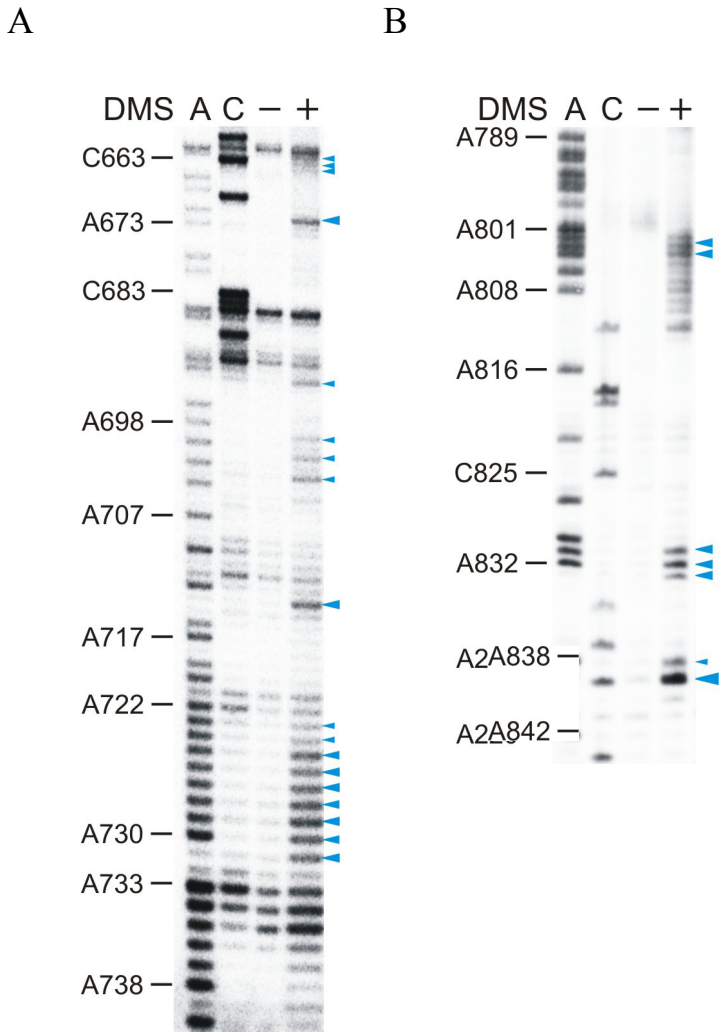

Figure 19: *Representive gels for DMS pattern of ai5γ in wild type cells.* Legend as Figure 16. A) Nts C663 to A738. B) Nts A789 to A842.

D2 adopts a different structure in vivo

In D2 the helix harbouring the important docking site η for D6 is formed properly. Furthermore all loops are present as shown by the modifications of A560-562, A526-528, A523-525, A521, A498, A496, A448-A453 and A444-445 (figure 18). Interestingly A501, A502 and A503-A507 are modified as well although they should be

involved in a helix according to the predicted secondary structure (figure 20). This could only be explained by a different structure in this region of D2.

In vivo structure of D4

The structure of D4 in ai5γ has not been determined until know. Therefore we applied our *in*
vivo DMS chemical probing assay to get further insights into the structure of D4. Doing this in combination with the RNA folding program mfold (Zuker 2003) we found a possible and consistent secondary structure for D4 presented in Figure 20. In this scheme the basal stem participates in a 4-way junction. The other protruding stems are closed by loops with nucleotides accessible to DMS (A713, A751, A753, A755) except for A780 that might be involved in tertiary or non-canonical interactions (figure 19, gel showing A751, A753 and A755 not shown). The largest stem has an additional internal asymmetrical loop (A698, A700, A702, A722-730) and a bulged A692 (figure 19, gel showing A22-730 not shown).
DMS modification data on D6 is not available since it is the first possible binding site for primers.
To conclude, the DMS modification data on ai5γ *in vivo* (figure 20) correlates well with structural probing data derived from ribozymes *in vitro* (Fedorova et al. 2007, Swisher et al. 2001, Su et al. 2005) and the phylogenetically derived secondary structure map (Michel et al. 1989). This indicates that ai5γ adopts its native structure *in vivo* and that this native structure strongly resembles the structure ai5γ adopts under high salt conditions and elevated temperature *in vitro*.

Long-range tertiary interactions are in place in vivo

As ai5γ is very compact in its native conformation (Swisher et al. 2001) the network of long-range tertiary interactions plays a crucial role in establishing the catalytically active structure of this ribozyme.
Concluding from the DMS pattern the D1 intra-domain α–α' contact, which is inevitable for exon binding and compaction (Su 2002, Waldsich and Pyle 2007) is established *in vivo*. A63 and C65 that are found in the α region are only weakly

modified (figure 16) and A344 as well as A346-348 in the α' element lack modification (figure 17). Likewise the intra-domain β–β' interaction is present *in vivo* although as in the α–α' contact A146, C147 and A148 in the β region show weak modification intensities (figure 16) whereas C255 and C259 in the β' region don't (figure 17). Both phenomenons might be a result of the fact that we are always looking at a population of RNA molecules with a significant amount of unspliced ai5γ of about 25%. Furthermore the β–β' interaction is not phylogenetically conserved (Michel and Ferat 1995) and dispensable for its catalytic activity *in vitro* (Su 2002).

Besides the D1 intra-domain long-range tertiary interactions are in place as well. A100 and A101 show no modification (figure 16) indicating that the θ–θ' interaction between the c1 tetraloop and the basal stem of D2 is established (Costa et al. 1997). This interaction brings stem c1 with its λ and ε' regions close to the essential joining element J2/3 (Fedorova et al. 2003). Furthermore establishment of the helix in D2 harboring the η region, as seen by lacking modifications of C464-465 and A466-467 (figure 20), suggests that the η–η' tertiary interactions between D2 and D6 is present bringing the branch point A880 close to the active site (Chanfreau and Jacquier 1996, Costa et al. 1997, Hamill and Pyle 2006). Also the coordination loop, which is the binding site for A880, is formed properly *in vivo,* due to its modification pattern (figure 20).

In short it seems that the network of long-range tertiary interactions necessary for establishing the compact native structure of ai5γ are in place.

The active site of the spliced ai5γ remains assembled

In a wt yeast strain the ai5γ group II intron splices efficiently with approximately 70 % of the entire population spliced. This means that in the wild type strain the DMS modification pattern observed mainly evolves from the structure of the spliced intron. Consistent with this are the modifications of nucleotides found at the exon binding sites EBS2 (A237, C238, A239, C240, C241, A242, and C243) and EBS1 (A331-A333) indicating that the exon substrates are no longer bound (figure 17).

Intriguingly the catalytic centre of the intron seems to remain established in the spliced intron. The ζ–ζ' interaction is in place indicated by the distinct modification intensities of A382 and A383 (figure 17) that are forming an A·A platform suggesting a canonical

form of this receptor (Cate et al. 1996). The weak modification of A198 (figure 16) pairing with U381 and of A830 and A832 (figure 19) suggests that most of the ai5γ molecules have their ζ–ζ' interaction between D1 and D5 established. Additionally A204 lacks modification and A214-215 are only weakly modified (figure 16) meaning that also the κ–κ' contact is present in the spliced intron *in vivo*. This indicates that the catalytic domain D5 is anchored in its receptor site in D1, which is a pre-requisite for active site formation. Further evidence for active site assembly could be seen in the established λ–λ' and ε–ε' interactions: C117 is not accessible to DMS (figure 16), which is in line with interactions of C117-N3 with G3-N3 (Boudvillain and Pyle 1998). Weak modification of A115 indicates that the base triple interaction named λ–λ' with the Watson-Crick base pair G825-G836 is present (figure 16). In analogy to the crystal structure of the *O. iheyensis* group II intron (Toor et al. 2008) C839, G588, A589 and the catalytic triad A816, G817 and C818 could be involved in a triple helix. The minor modification of A589 (figure 18) and C839 (figure19) hints to established base triples with A816-U847 and C818-G845, respectively. As A587 is also only modified in a minor way (figure 18) the interaction between this nucleotide in J2/3 and U887 at the very 3' end of the intron seems to be established. The catalytic effector D3 has two important regions considering tertiary interaction. A617-618 are only weakly modified (figure 18) implying an established μ–μ' contact with the 2'OH of G844 in D5 (Fedorova and Pyle 2005). The second important region is the internal loop at the basal stem of D3. There two consecutive A·A base pairs are needed for enhancing catalysis (Fedorova and Pyle 2005). The Hoogsteen face (N6, N7) of A599 and A662 is thereby H-bonding with the minor groove edge (C2, N3, 2'OH) of A661 and A598. Furthermore this region is contacting the ε–ε' region and energetically communicating with the dinucleotide bulge in D5 (Fedorova and Pyle 2005). Interestingly this conformation is not maintained in the spliced intron *in vivo* as A597, A598 and C663 were all equally modified by DMS at medium intensity (figure 18). Alternatively, in *vivo* interaction between D1 and D3 happens differently or is abolished after splicing.

Except from this small difference in the conformation of D3 the structure of ai5γ *in vivo* (figure 20) is very similar to the one observed *in vitro* (figure 27). Therefore these data suggests that the spliced intron maintains its native structure after splicing and that the active site remains assembled, which would be important for group II introns that are able to act as mobile genetic elements and perform reverse splicing (Pyle and Lambowitz 2006).

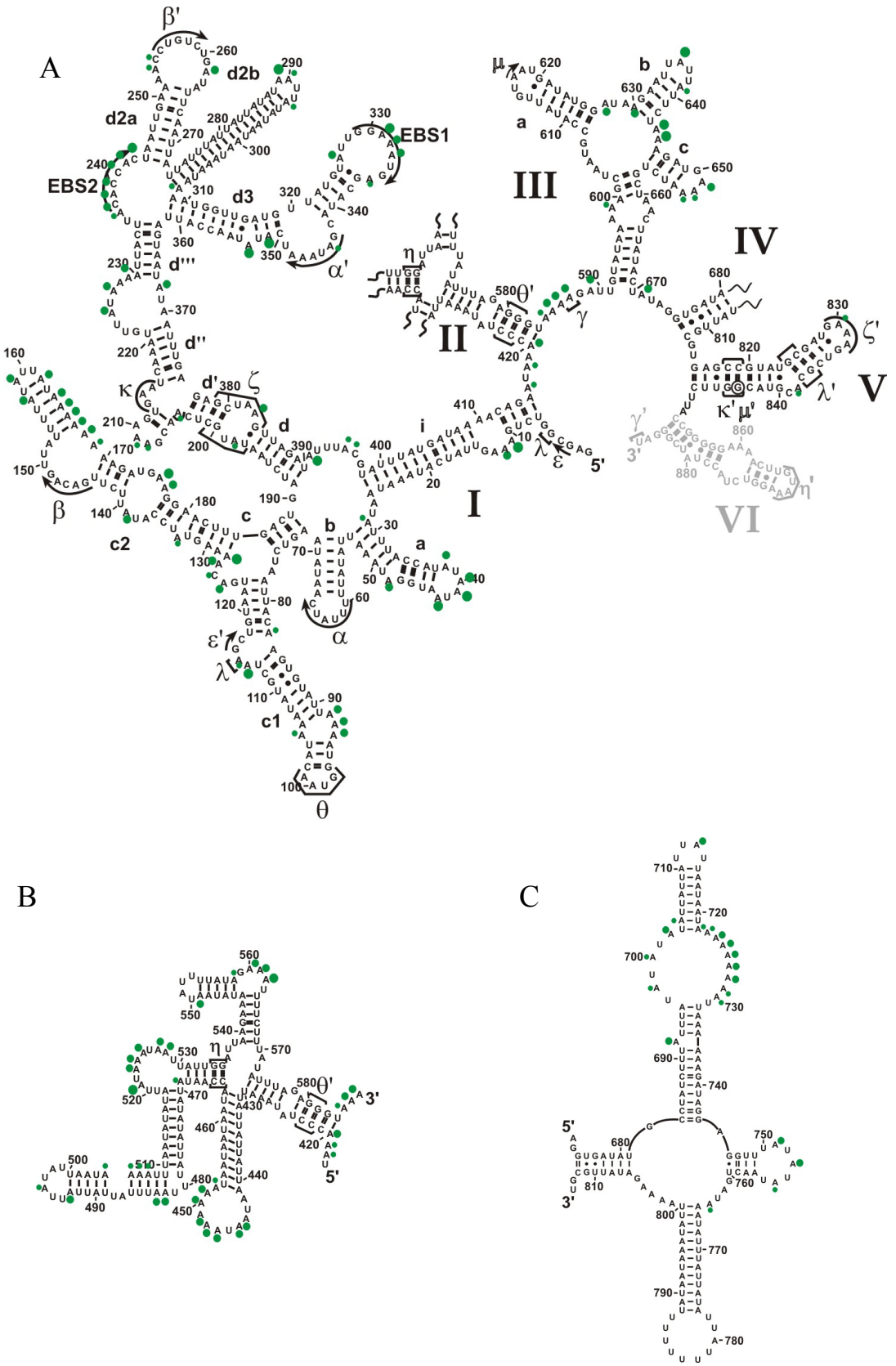

Figure 27: *DMS modifications of the ai5γ derived ribozyme D123456*. DMS modifications were plotted according to their intensities (weak, moderate, large) on the secondary structure map of ai5γ. A) D1, D3, D5 and D6 are shown. D6 is in grey since it couldn't be mapped with this method. B) D2 is shown. C) D4 is shown. Data kindly provided by Dr. Oliver Mayer.

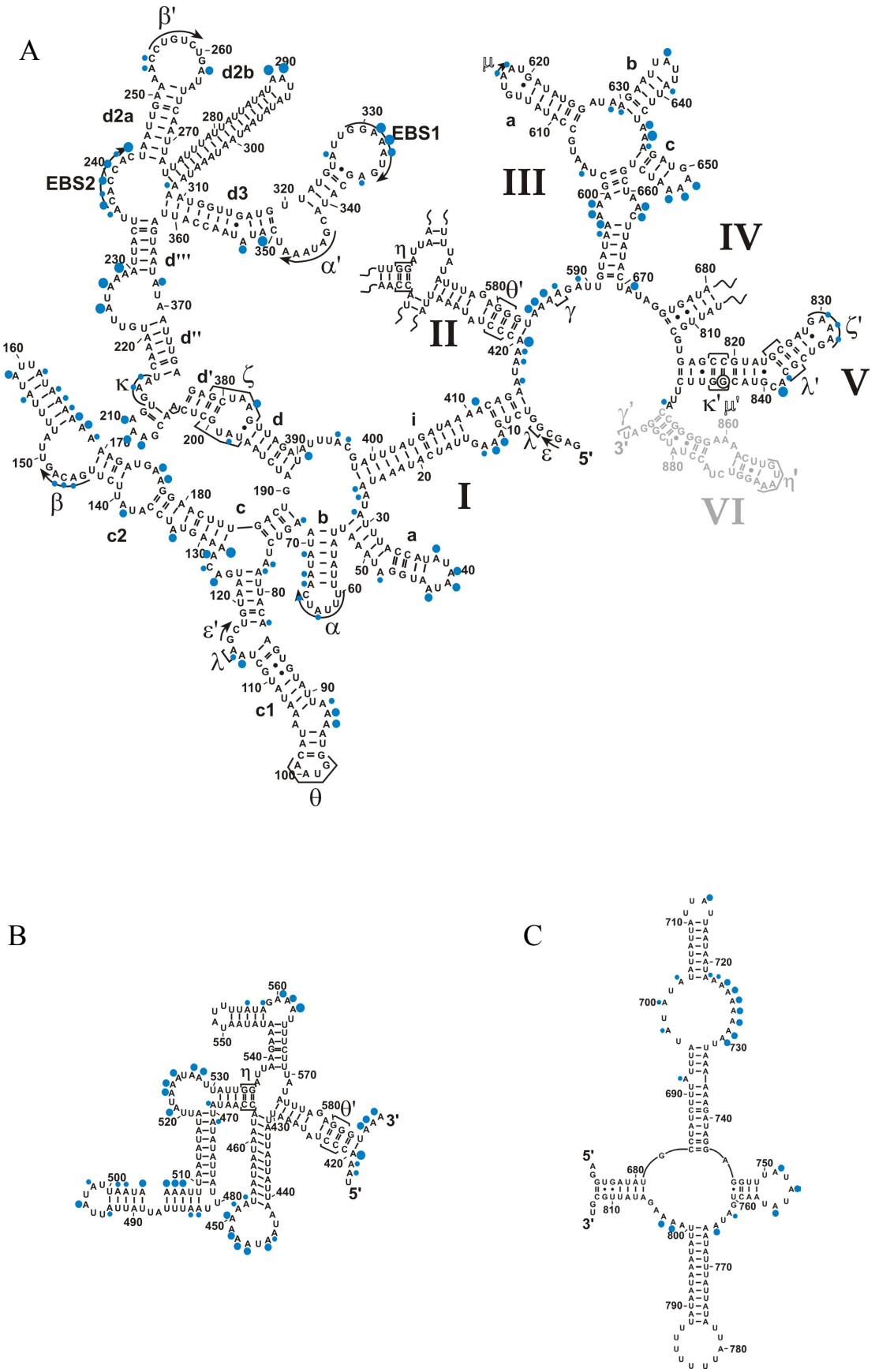

Figure 20: *DMS modifications of ai5γ in wild type cells*. DMS modifications were plotted according to their intensities (weak, moderate, large) on the secondary structure map of ai5γ. A) D1, D3, D5 and D6 are shown. D6 is in grey since it couldn't be mapped with this method. B) D2 is shown. C) D4 is shown

Mss116p is not bound to the spliced intron

DMS can modify N1 of adenosine and N3 of cytosine if they are solvent accessible. Binding of a protein would render a stretch of nucleotides inaccessible to DMS. By comparing the DMS pattern obtained *in vivo* and *in vitro* we could not find such a footprint for Mss116p on the most important domains D1, D3, D5 and D6. As other group II intron cofactors where shown to bind in D4 (Pyle and Lambowitz 2006) we focused on that region of ai5γ. But neither did we find a footprint on D4 nor on D2. This result implies that Mss116p does not remain bound to the spliced intron in its native conformation. Since with the method applied we can only survey N1 of adenosines and N3 cytosines it still remains possible that Mss116p binds to other functional groups or nucleotides. However, binding of a large protein should at least result in reduced accessibility of the respective nucleotides resulting in reduced intensities of DMS modifications. We were not able to make such an observation. A second alternative reason might be a high k_{off} of Mss116p resulting in only transient binding making it impossible to monitor. This is also very unlikely since *in vitro* Mss116p binds very tightly to ai5γ with a K_D around 1nM (Halls et al. 2006). As a consequence we conclude that Mss116p is not bound to ai5γ in its native state and hence not important for maintaining the native conformation of this group II intron.

The κ–ζ folding control element cannot adopt its structure in the absence of Mss116p

Formation of the κ–ζ element is crucial for the tertiary collapse of the ai5γ derived ribozyme D135 *in vitro* (Waldsich and Pyle 2008) and thereby the rate limiting step. Additionally it plays an important role for active site formation during later stages of folding.

The importance of this element drew our attention to specific structural differences of this region between the wild type and the knock out strain. As expected the κ–ζ element is not formed in absence of Mss116p. Most of the strongest differences in intensity of DMS modification are found in this region of the intron. *In vitro* the formation of d, d'

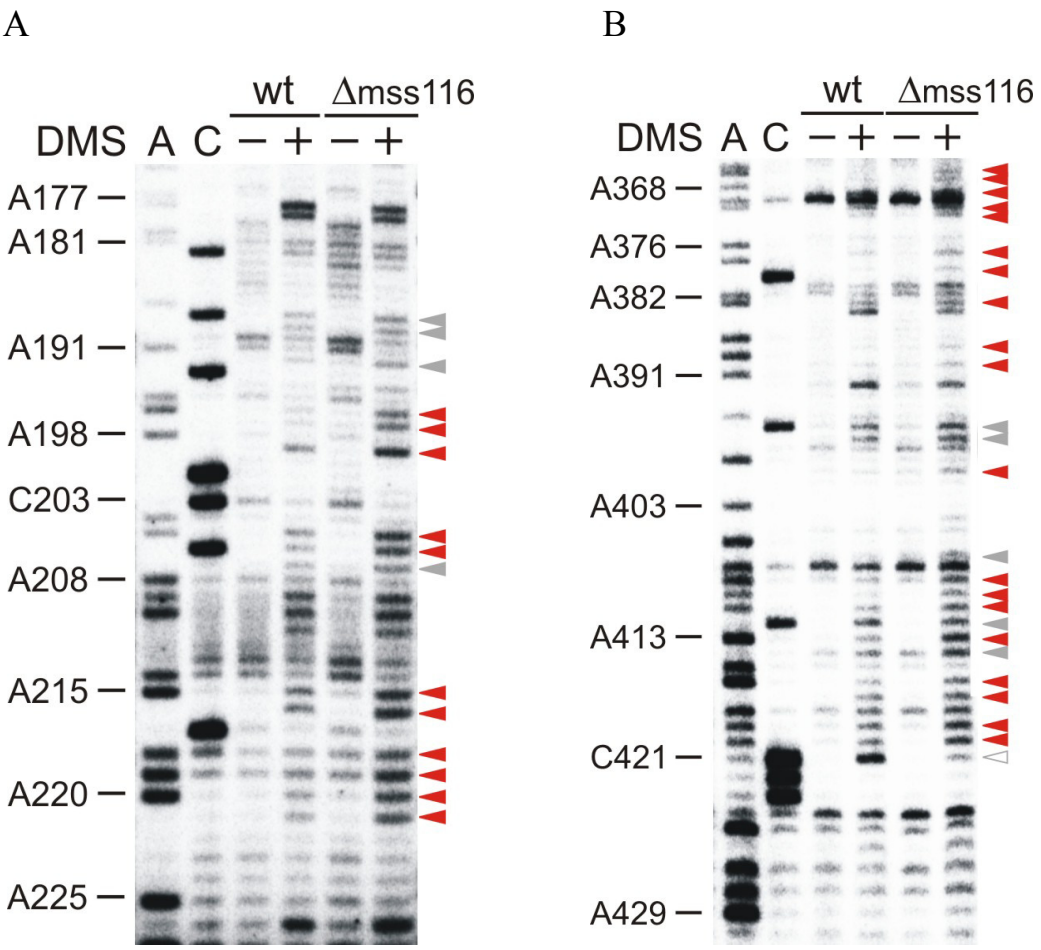

Figure 21: *Different DMS modifications intensities of ai5γ in wild type and ΔMss116 cells at the κ–ζ element*. Differences in DMS modification intensities are indicated according to their intensities. Red filled arrows show >2x intensity in the mutant strain compared to wild type, red open arrows <0.5x intensity. Grey filled arrows show >1.5x intensity in the mutant strain compared to wild type, grey open arrows <2/3x intensity. A) Nts from A177 to A225. B)Nts from A368 to 429.

and d'' are Mg^{2+} dependent. All three helices do not form in the knock out strain shown by DMS modification of A195, A196, A387 and 389 in d, of A378 in d'' and of C217, A218-A220 and A376 in d'' (figure 21). Enhanced modification of A204 and A214-A215 hints to lacking formation of the κ–κ' contact with D5 (figure 21). Furthermore A198 as well as A382 are significantly more modified in the knock out strain (figure 21) showing that also the ζ–ζ' interaction is not in place, which is strengthened by the enhanced modification of A830-A832 in the D5 tetraloop (figure 24). The explanation for these enhancements is that D1 is not docked onto D5 and because of that the active site is not established in the knock out strain. In accordance with that A115 and C117 are more accessible to DMS (figure 22) indicating that the active site constituents ε–ε' and λ–λ' are not formed. The reduced accessibility of A114 in the knock out strain (figure 22) is a result from the lacking λ–λ' contact that leads to a twist of A114 pushing its N1 outside. These observations clearly show that Mss116p is essential for the formation of the κ–ζ folding control element.

Mss116p acts on both secondary and tertiary structure formation

As the folding control element in D1 of ai5γ triggers a cascade of folding steps in the D135 ribozyme *in vitro* (Waldsich and Pyle 2007 and 2008) we were interested in the effects of the absent cofactor Mss116 on the secondary as well as the tertiary structure level.

Both D1 intradomain contacts α–α' and β–β' are not established as the involved observable nucleotides A63 (α), A344 and A346-348 (α'), A146, C147 and A148 (β) as well as C259 (β') are significantly more modified (figure 22 and 23) in the knock out strain. Interdomain tertiary contacts are affected to a similar extent. Enhanced modification of A100 and A101 (figure 22) indicates that the θ–θ' contact, that brings the c1 stem harbouring the ε' and λ regions close to the basal stem of D2 and the joining segment J2/3, is not in place. As the helix in D2 that includes the η element is not

formed in the knock out strain indicated by enhanced modification of C464-C465 and A466-A467 (figure 23) the η–η' contact that docks D6 onto D2 bringing the branch point adenosines in close proximity to the active site. In line with this, A868-A869 in the η'-tetraloop is accessible to DMS in the knock out strain (figure 24), where an exon binding primer can be used to map this region. Nucleotides (A227-A229, A386 and A370; figure 22 and 23) that are part of the coordination loop, which coordinates A880 (Hamill et al. 2006), are significantly more intensely modified showing again that the active site is not formed in the knock out strain. As A617 and A618 are more accessible to DMS in the D3 pentaloop the μ-μ' interaction that docks D3 in the

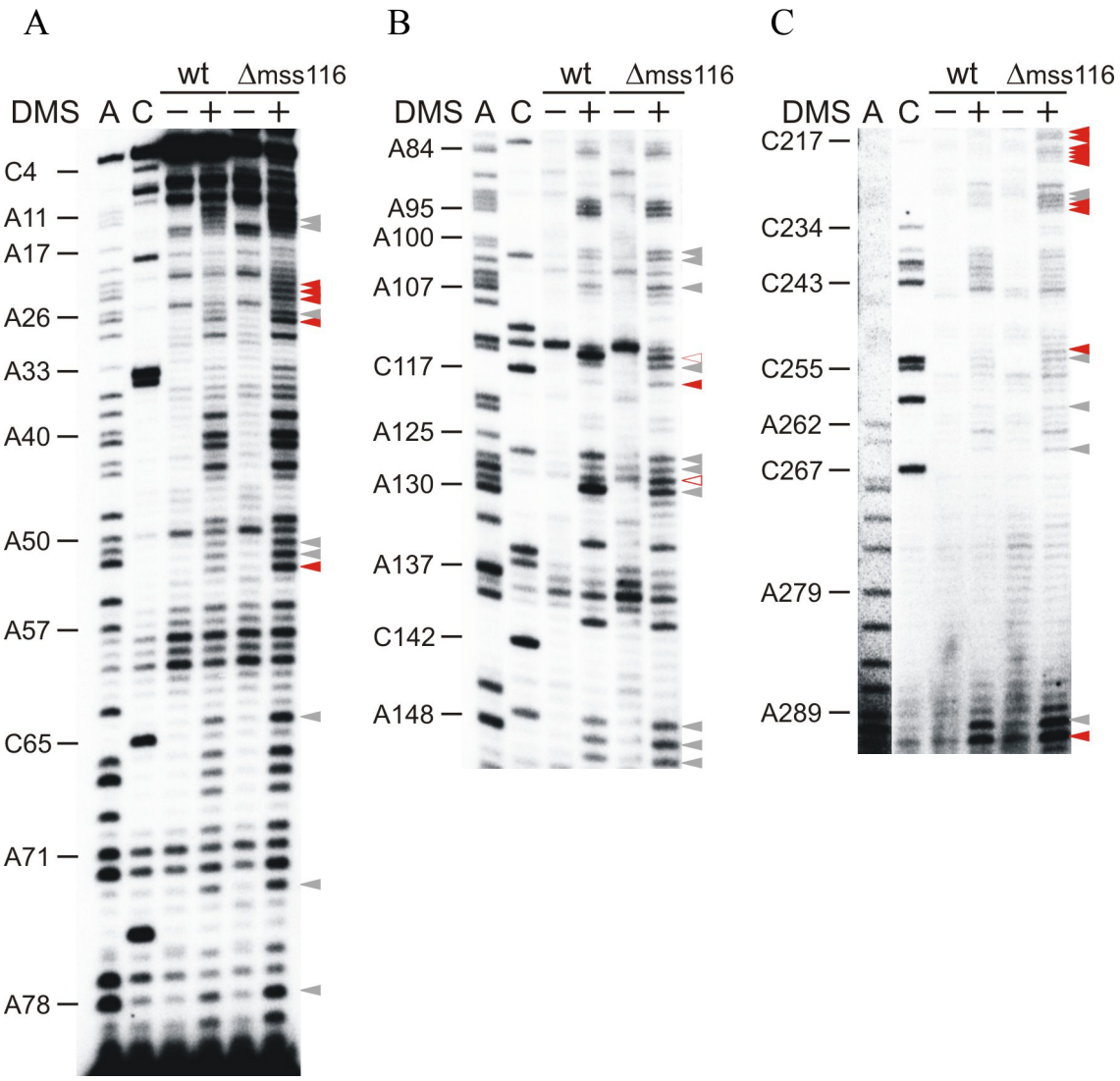

Figure 22: *Different DMS modifications intensities of ai5γ in wild type and ΔMss116 cells.* Legend as Figure 21. A) Nts C4 to A78. B) Nts A84 to A148. C) Nts A217 to A290.

active site is not established (figure 23). Finally J2/3 (A484-A486) that forms a triple helix with the catalytic triad (Toor et al. 2008) in D5 shows enhanced modification (figure 23). Taken together all tertiary contacts seem to be open in the knock out strain according to their DMS. These findings imply that neither the active site is formed nor the compact conformation of ai5γ obtained when Mss116p is not present as a cofactor.

The effects off absent Mss116p on the tertiary structure of ai5γ were expected although the extent wasn't. The consequences for the secondary structure of the group II intron *in vivo* were more surprising. Aside from the already mentioned stems d, d', d'' also the helices d''', a, c and i of D1 are not formed in the knock out strain revealed by enhanced modification of e.g. A22-24, C408, A411-A412 and A399 (i), A50-A52 (a), A72, A187

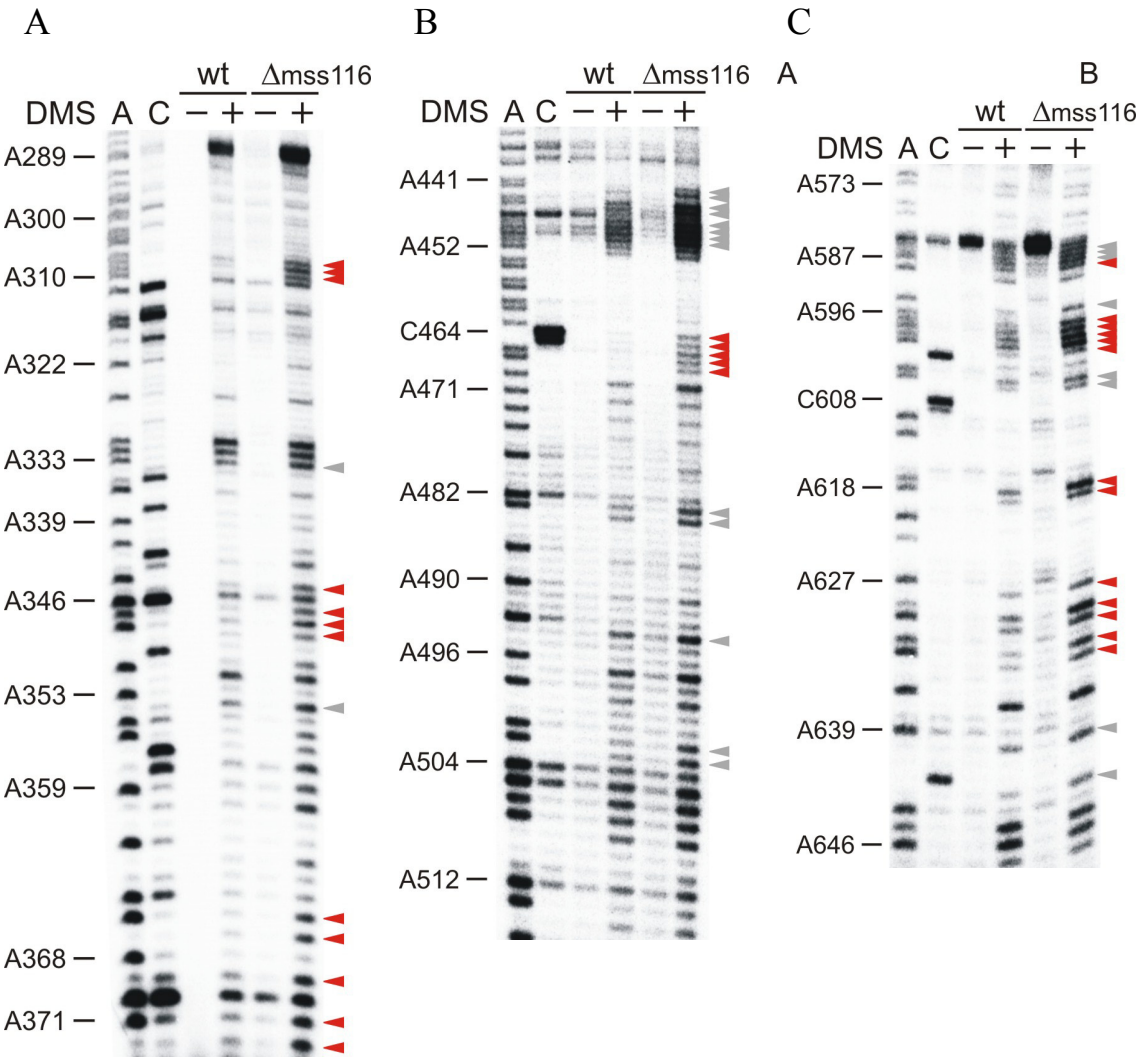

Figure 23: *Different DMS modifications intensities of ai5γ in wild type and ΔMss116 cells.* Legend as Figure 21. A) Nts A289 to A371 B) Nts A441 to A512. C) Nts A573 to A646.

and C188 (c) and A230 as well as A365-A366 (d''') (figure 21, 22 and 23). Interestingly except for stem i all these helices need Mg^{2+} *in vitro* to be formed (Waldsich and Pyle 2007) implying that the secondary structure of ai5γ in the knock out strain resembles the *in vitro* one in presence of K^{+} ions but without Mg^{2+} added (Waldsich and Pyle 2007). Additionally to D1 the secondary structure of D3 is affected in the Δmss116 strain. Here, the basal stem (A594 and A596-A597) and the internal loop (A598-A560) as well as stem b (A630, A632-A633, A639 and C642) are not formed properly due to enhanced accessibility of the mentioned nucleotides (figure 23). Actually the unfolded state *in vitro* contains more secondary structure elements than the unspliced intron *in vivo* in the absence of Mss116p (figure 25, Waldsich and Pyle 2008).

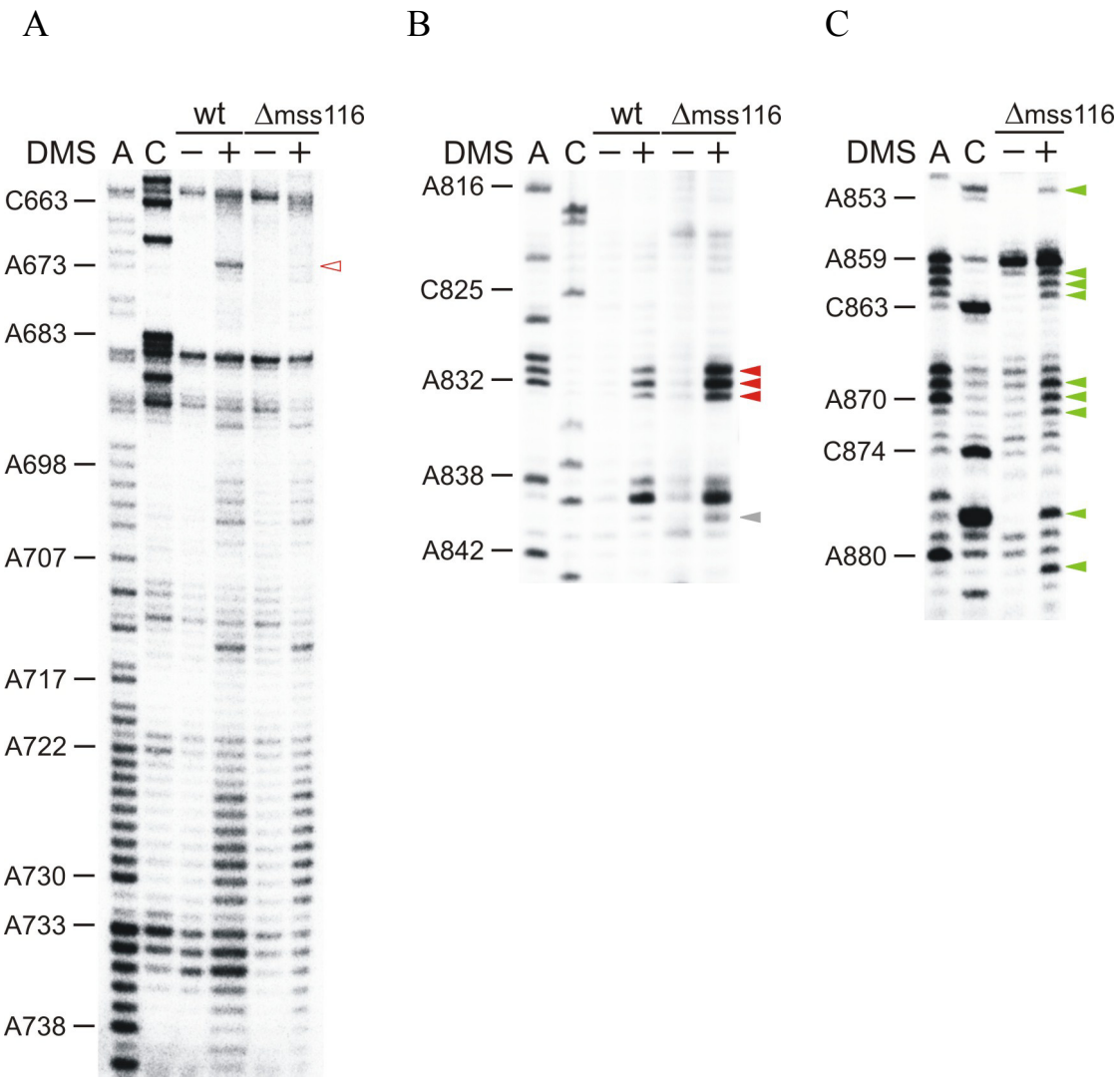

Figure 24: *Different DMS modifications intensities of ai5γ in wild type and ΔMss116 cells.* Legend as Figure 21. Green arrows in B show modifications in the mutant strains at positions that are not expected to be accessible in the wild type strain, which couldn't be primed due to splicing A) Nts C663 to A738.B) Nts A816 to A842. C) Nts A853 to A880

This could be explained by the nonphysiological concentration of monovalent ions applied to get the unfolded state of ai5γ where most of the secondary structure elements are present.

In D2 only the helix that contains the η element shows strong enhancement by the absence of Mss116p (figure 24). Likewise in D4 only A801-A803 show enhanced modification intensities in the Δmss116 strain (gel not shown). This suggests that Mss116p plays a minor role in the structural organisation of D2 and D4. Also secondary structure formation of D5 and D6 seems independent of Mss116p. Therefore Mss116p seems to act mostly on the D1 scaffold and the effector D3 as well as tertiary contacts. In the knock out strain these regions are unstable suggesting that Mss116p acts by stabilizing ai5γ. Since we did not see a stretch of residues being more structured in the knock out strain kinetic traps seem not to occur in the absence of Mss116p.

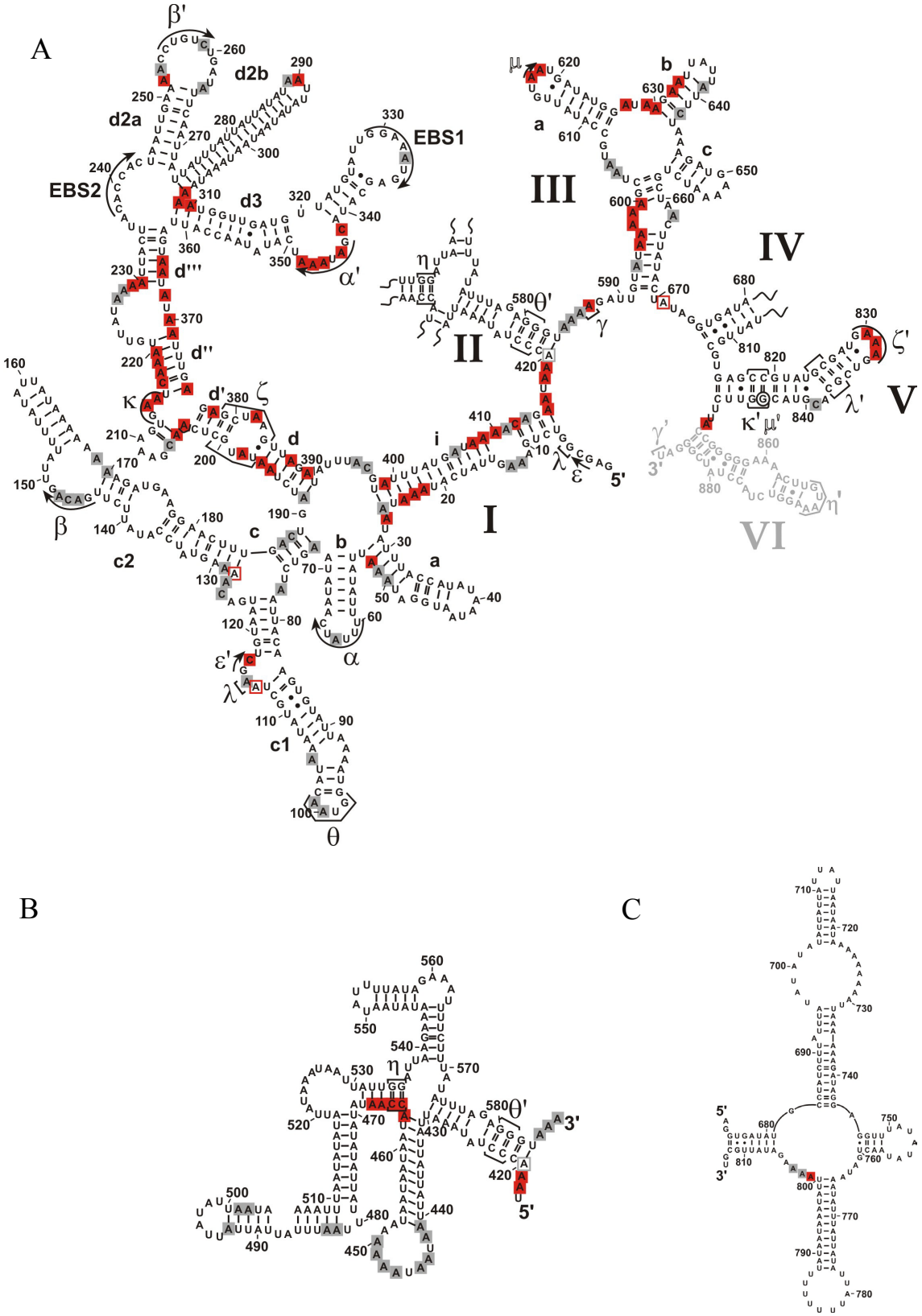

Figure 25: *Different DMS modifications intensities of ai5γ in wild type and ΔMss116 cells*. Differences in DMS modification intensities were plotted according to their intensities on the secondary structure map of ai5γ. Colour code as in figure 21. A) D1, D3, D5 and D6 are shown. D6 is in grey since it could only be mapped in the mutant strain. B) D2 is shown. C) D4 is shown

Discussion

After transcription catalytic RNA has to adopt a specific structure in order to fulfil its function. Folding into this structure is fundamental and in general a complex process. Folding has been studied intensely *in vitro* and little is known about this process *in vivo* where folding happens co-transcriptionally (Schroeder et al. 2002). Ribozymes derived from the group II intron ai5γ in *S. cerevisiae* have proven to be a good model system for the folding of large RNAs (Pyle et al. 2007). But there have been only few attempts to compare folding of RNA *in vitro* and *in vivo*.

The intracellular structure of ai5γ

During my diploma research I first had a look at the intracellular structure of ai5γ to compare it with data gained in *in vitro* studies.
To do that we applied the described *in vivo* DMS chemical probing assay (see Methods) to obtain structural information both on secondary and tertiary structure. We found out that the intracellular structure vastly resembles the structure of D123456 *in vitro* despite of the very different folding environment. *In vivo* ai5γ adopts its secondary structure. Also all inter- and intradomain tertiary contacts are in place and most importantly the active site is properly assembled. Indeed most of the ai5γ RNA molecules are spliced. Interestingly the intron keeps its native conformation after splicing. Recently it was shown that the ai5γ group II intron is able to perform reverse splicing (Roitzsch and pyle 2009). The data of this thesis suggests that this group II intron does not have to undergo major structural rearrangements for its ability to reverse splice. This might be important for other group II introns that are in contrast to ai5*g* not only able to reverse splice but also known to act as mobile genetic elements in assistance of a self-encoded maturase (Pyle and Lambowitz 2006).

The function of Mss116p in ai5γ folding

Folding of ai5γ-derived ribozymes has been studied intensely *in vitro* both at non-physiological conditions with elevated temperature and high K^+ and Mg^{2+} levels (Su et al. 2003, Swisher et al. 2001, Waldsich and Pyle 2007) and at near-physiological levels (Fedorova et al. 2007). Interestingly ai5γ collapses to the compact intermediate *in vitro* even at near-physiological conditions but complete compaction takes 26 h, which is tremendous considering the needs of a living cell (Fedorova et al. 2007). Furthermore the native state can only be approached *in vivo* at physiological temperatures when [Mg^{2+}] is raised to 100mM. However adding of the DEAD box protein Mss116p *in vitro* lowers the needed [Mg^{2+}] to a near-physiological level for native state conformation (Halls et al. 2006, DelCampo et al. 2007, Solem et al. 2006). Furthermore it was shown that Mss116p is essential for splicing of ai5γ *in vivo* (Huang et al 2005). It was proposed that Mss116p acts as a RNA chaperone that actively disrupts kinetically trapped molecules allowing them folding into the native state (DelCampo et al. 2007). However previous studies showed that the unwinding activity of Mss116p is not essential for splicing activity of ai5γ (Solem et al. 06). Additionally to ai5γ efficient splicing of all mitochondrial group I and II introns in yeast depends on Mss116p (Huang et al. 2005). Furthermore the family of DEAD box proteins is found in almost all cellular processes involving RNA structural organisation (Cordin et al. 2006). Although there has been quite an effort and success in understanding the functions of DEAD-box proteins, it is still enigmatic how folding of large RNAs can be facilitated by non-processive DEAD-box helicases. We wanted to know how one of those proteins, Mss116p assists in folding of ai5γ in vivo.

To do that we applied the mentioned *in vivo* chemical probing assay additionally on a strain where mss116 was knocked out. The first striking result was the extent of impact Mss116p has on the structure of ai5γ. Without its assistance not only that none of the intra- or interdomain interactions are in place but even more intriguingly the entire intron is largely unfolded. This indicates that Mss116p acts at a very early stage of folding since the secondary structure is severely perturbed. In context with that it is noteworthy that only a few nucleotides showed protection from methylation in the knock out strain compared to the wild type strain and that these nucleotides where all standing alone hinting to a different conformation of this one nucleotide. A stretch of

protection indicating misfolding of ai5γ in the knock out strain was not found. Instead especially the scaffold D1 seems to be inherently unstable without Mss116p assisting. This is also consistent with the almost total loss of splicing activity in the knock out strain (<1%) as the native structure cannot be obtained and the active site not assembled. Therefore we propose that Mss116p acts on ai5γ folding by stabilizing on pathway intermediates instead of resolving misfolded structures. This is also in good agreement with the unimportance of the unwinding activity of Mss116p for splicing of ai5γ *in vitro* (Solem et al. 2006). Since D1 also folds first *in vitro* and was proposed as a folding scaffold (Su et al. 2005) acting of Mss116p on this structurally most relevant domain makes absolute sense. Additionally *in vivo* folding happens co-transcriptionally. By having Mss116p acting on D1 folding can immediately take place, which of course is favourable in an evolutionary point of view.

In vivo folding pathway

With this data present and having the knowledge from previous experiments dealing with the folding pathway of ai5γ *in vitro* (Pyle et al. 2007, Steiner et al. 2008) it is possible to describe an *in vivo* folding pathway of this group II intron. Pyle and co-workers revealed that the ai5γ group II ribozyme D135 is following a direct path to its native structure where the formation of the scaffold D1 is the prerequisite for the collapse of the intron into a compact intermediate (Su et al. 2005). This first step of tertiary folding is unusually slow with a rate constant of ~1 min^{-1}. It is noteworthy that the starting point of this experiment was an electrostatically relaxed state of the molecule by adding monovalent ions. The relaxed molecule however did not show any tertiary fold although significant compaction indicated formation of secondary structure elements (Su 05). The collapse itself is controlled by the Mg^{2+} dependent formation of a small substructure within D1 the κ–ζ folding control element (Waldsich and Pyle 2007 and 2008). Interestingly this same element plays a crucial role in active site assembly later on (Boudvillain and Pyle 1998, Costa et al. 1995). Introducing a component of the active site as a control element for folding is a sophisticated way to ensure proper folding of the core region of ai5γ. Folding from the intermediate state to its native conformation then happens fast. Another astonishing feature of the folding pathway of ai5γ *in vitro* is the lack of kinetic traps (Swisher et al. 2002, Steiner et al 2008).

Furthermore it has been shown that substrate binding stabilizes the compact intermediate (Steiner et al. 2008).

The biological relevance of this folding pathway for ai5γ in its actual intracellular environment was still to be confirmed as needed [Mg^{2+}] for folding *in vitro* are not present *in vivo*. In our study we first of all saw that the intracellular cofactor, known to be essential for splicing of all group II introns, Mss116p was not bound to the intron in its native conformation, as it did not show any footprint. In light of the *in vitro* folding pathway and together with the largely unfolded structure in the knock out strain this indicates that Mss116p acts at an early stage of ai5γ folding and is not required to maintain the active native structure of the intron. As in the knock out strain the strongest enhancements of modification are found in the folding scaffold D1, Mss116p seems to assist in the transition of the unfolded intron to the compact intermediate where D1 has adapted its structure. Since the folding control element is severely disturbed in its structure without Mss116p present, the cofactor might act on this small element. Subsequent docking of the other domains for active site assembly then needs Mss116p to release the intron to allow tertiary interactions. Interestingly formation of the k-z element is dependent on Mg^{2+}, which proposes possible similarities in functions of Mss116p and magnesium. Therefore Mss116p might act by stabilizing RNA structure and in our case by actually stabilizing the proper structure of the κ–ζ folding control element. Furthermore it was shown that in contrast to unwinding activity ATPase activity of Mss116p is absolutely required for splicing activity of ai5γ both *in vitro* (Solem et al. 2006) and *in vivo* (Mohr et al. 2006). As ATPase activity causes structural changes in DEAD box proteins, this might allow the release of the cofactor rendering docking of the catalytic center D5 and of other domains possible. To summarize the data from this diploma thesis suggests the following folding pathway *in vivo* (figure 26):

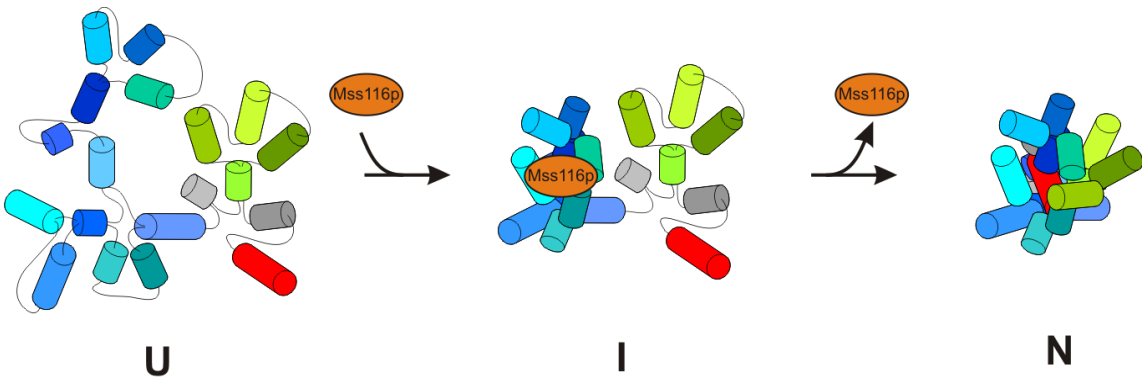

Figure 26: *Schematic of proposed role of Mss116p in ai5γ folding*. Mss116p binding allows collapse of ai5γ. In order to allow folding into native structure Mss116p has to be released.

Mss116p perhaps stabilizes the formation of the κ–ζ element found in D1. After D1 has properly formed ATPase activity ensures release of the cofactor. Exon binding stabilizes the formed compact intermediate maintaining the structure without further assistance by Mss116p. With the folding control element and D1 present subsequent docking of the other domains and active site assembly in short native structure formation is fast. The resulting native structure is stable *in vivo* without binding of Mss116p rendering ai5γ splicing-competent.

Mss116p is a multifunctional enzyme

Previously it was suggested that Mss116p and DEAD box proteins in general act as RNA chaperones that actively disrupt misfolded structures allowing native structure formation (Mohr et al. 2006, Halls et al. 2007, Delcampo et al. 2007). It was shown that Mss116p shows unwinding activity *in vitro* (Solem et al. 2006, Delcampo et al. 2007). Furthermore Mss116p functions in RNA processing and translation in yeast mitochondria (Huang 2004). As the latter ones only hint to more functions of the same protein the unwinding and stabilizing activity are somehow contradictory. Intriguingly only a subset of DEAD box proteins where shown to be able to rescue splicing activity of ai5γ *in vitro* (Solem et al 2006) and *in vivo* (Mohr et al. 2006). The speciality of these proteins is a distinct C-terminal domain that is arginine rich. Therefore the stabilizing function of helicases might be evolutionary younger than the unwinding activity and have evolved at this distinct C-terminal domain. Mss116p also acts on all mitochondrial group I and II introns suggesting different mode of action as a cofactor. For example the folding intermediate of the bI5 group I intron is stabilized by CBP2 *in vitro* (Weeks and Cech 1996). Thus Mss116p might act differently on this RNA.
Having a multifunctional enzyme present in mitochondria is definitely a worthy strategy from an evolutionary point of view since a lot of energy can be saved in transcription, translation and mitochondrial import.

Perspectives

The findings of this diploma thesis evoke a lot of further questions. Seeing that Mss116p acts by stabilizing RNA structure makes one wonder how it actually does that. Our findings that the κ–ζ folding control element is not formed in absence of Mss116p would suggest a focus on possible interactions of Mss116p with this region of the intron. Since Mss116p has different activities, namely unwinding, ATPase and RNA binding, it would also be interesting what importance those activities have considering the structure of ai5γ.

As ATPase activity is essential for splicing of ai5γ *in vivo* DMS modification of a strain carrying the K158R mutation that was shown to reduce ATPase activity tremendously (Solem et al. 2006) would be a logical next step. If the proposed mechanism of Mss116p were correct the resulting DMS pattern would resemble the one obtained with a wild type strain. Differences that relate to undocked domains and an immature active site centre would strengthen our theory. An occurring footprint would be consistent with our suggestion that the ATPase activity is needed for cofactor release. Structural probing of strains with differently truncated C-terminal domains might shed more light onto the function of this distinct and highly positively charged domain. Furthermore DMS modification of a strain with the SAT -> AAA mutation that uncouples ATPase and unwinding activity would elucidate the role of unwinding activity in native structure formation of ai5γ. Finally a strain with a disrupted κ–ζ folding control element would help to confirm the importance of this region for the tertiary collapse of ai5γ.

Literature

1. Augustin, S., Müller, M. W. & Schweyen, R. J. Reverse self-splicing of group II intron RNAs *in vitro*. Nature 343, 383-386 (1990).

2. Boudvillain, M., Delencastre, A. & Pyle, A. M. A new RNA tertiary interaction that links active-site domains of a group II intron and anchors them at the site of catalysis. Nature 406, 315-318 (2000).

3. Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. EMBO J. 17, 7091-7104 (1998).

4. Boulanger, S. C. et al. Studies of point mutants define three essential paired nucleotides in the domain 5 substructure of a group II intron. Mol. Cell. Biol. 15, 4479-4488 (1995).

5. Cao, Y. & Woodson, S. A. Refolding of rRNA exons enhances dissociation of the Tetrahymena intron. RNA 6, 1248-56. (2000).

6. Cate, J. H. et al. RNA tertiary structure mediation by adenosine platforms. Science 273, 1696-1699 (1996).

7. Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. EMBO 15, 3466-3476 (1996).

8. Cordin, O., Tanner, N. K., Doere, M., Linder, P. & Banroques, J. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. EMBO J 23, 2478-87 (2004).

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

10. Costa, M., Deme, E., Jacquier, A. & Michel, F. Multiple tertiary interactions involving domain II of group II self- splicing introns. J. Mol. Biol. 267, 520-36 (1997).

11. Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. EMBO J. 14, 1276-1285 (1995).

12. Daniels, D. L., Michels, W. J., Jr. & Pyle, A. M. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. J. Mol. Biol. 256, 31-49 (1996).

13. de la Cruz, J., Kressler, D. & Linder, P. Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem Sci 24, 192-8 (1999).

14. Del Campo, M. et al. Do DEAD-box proteins promote group II intron splicing without unwinding RNA? Mol. Cell 28, 159-166 (2007).

15. DeLencastre, A., Hamill, S. & Pyle, A. M. A single active-site region for a group II intron. Nat. Struct. Mol. Biol. 12, 626-627 (2005).

16. Draper, D. E., Grilley, D. & Soto, A. M. Ions and RNA folding. Annu Rev Biophys Biomol Struct 34, 221-43 (2005).

17. Fedorova, O., Boudvillain, M., Kawaoka, J. & Pyle, A. M. Nucleotide analog interference mapping and suppression: specific applications in studies of RNA tertiary structure, dynamic helicase mechanism and RNA–protein interactions. (eds. Hartmann, R. K., Bindereif, A., Schoen, A. & Westhof, E.) (Wiley-VCH, Weinheim, 2005).

18. Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 interact to form critical elements of the group II intron active site. J. Mol. Biol. 330, 197-209 (2003).

19. Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. RNA 14, 1048-56 (2008).

20. Fedorova, O., Waldsich, C. & Pyle, A. M. Group II intron folding under near-physiological conditions: collapsing to the near-native state. J. Mol. Biol. 366, 1099-1114 (2007).

21. Fedorova, O. & Zingler, N. Group II introns: structure, folding and splicing mechanism. Biol Chem 388, 665-78 (2007).

22. Gregan, J., Kolisek, M. & Schweyen, R. J. Mitochondrial Mg(2+) homeostasis is critical for group II intron splicing in vivo. Genes Dev 15, 2229-37. (2001).

23. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35, 849-57 (1983).

24. Halls, C. et al. Involvement of DEAD-box Proteins in Group I and Group II Intron Splicing. Biochemical Characterization of Mss116p, ATP Hydrolysis-dependent and -independent Mechanisms, and General RNA Chaperone Activity. J. Mol. Biol. 365, 835-855 (2006).

25. Hamill, S. & Pyle, A. M. The receptor for branch-site docking within a group II intron active site. Mol Cell 23, 831-40 (2006).

26. Herschlag, D. RNA chaperones and the RNA folding problem. J. Biol. Chem. 270, 20871-20874 (1995).

27. Huang, H. R. et al. The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. Proc Natl Acad Sci U S A 102, 163-8 (2005).

28. Huang, H.R. Functional studies of intron- and nuclear-encoded splicing factors in the mitochondria of Saccharomyces cerevisiae, PhD thesis (Dallas, Texas: The University of Texas Southwestern Medical Center).

29. Iost, I., Dreyfus, M. & Linder, P. Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. J Biol Chem 274, 17677-83 (1999).

30. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. J. Mol. Biol. 213, 437-447 (1990).

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

32. Koch, J. L., Boulanger, S. C., Dib-Hajj, S. D., Hebbar, S. K. & Perlman, P. S. Group II Introns deleted for multiple substructures retain self-splicing activity. Mol. Cell Biol. 12, 1950-1958 (1992).

33. Konforti, B. B. et al. Ribozyme catalysis from the major groove of group II intron domain 5. Mol. Cell 1, 433-441 (1998).

34. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell 31, 147-157 (1982).

35. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. Annu Rev Genet 38, 1-35 (2004).

36. Liebeg, A. & Waldsich, C. Probing RNA structure within living cells. Invited for publishing in Methods Enzymol (2009)

37. Matsuura, M., Noah, J. W. & Lambowitz, A. M. Mechanism of maturase-promoted group II intron splicing. EMBO J 20, 7259-70 (2001).

38. Michel, F. & Ferat, J.-L. Structure and activities of group II introns. Annu. Rev. Biochem. 64, 435-461 (1995).

39. Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns--a review. Gene 82, 5-30 (1989).

40. Minczuk, M., Dmochowska, A., Palczewska, M. & Stepien, P. P. Overexpressed yeast mitochondrial putative RNA helicase Mss116 partially restores proper mtRNA metabolism in strains lacking the Suv3 mtRNA helicase. Yeast 19, 1285-93 (2002).

41. Mohr, G., Zhang, A., Gianelos, J. A., Belfort, M. & Lambowitz, A. M. The neurospora CYT-18 protein suppresses defects in the phage T4 td intron by stabilizing the catalytically active structure of the intron core. Cell 69, 483-94 (1992).

42. Mörl, M., Niemer, I. & Schmelzer, C. New reactions catalyzed by a Group II intron ribozyme with RNA and DNA substrates. Cell 70, 803-810 (1992).

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

44. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. Science 289, 920-30 (2000).

45. Pause, A. & Sonenberg, N. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J 11, 2643-54 (1992).

46. Peebles, C. L. et al. A self-splicing RNA excises an intron lariat. Cell 44, 213-223 (1986).

47. Pyle, A. M., Fedorova, O. & Waldsich, C. Folding of group II introns: a model system for large, multidomain RNAs? Trends Biochem. Sci. 32, 138-145 (2007).

48. Pyle, A. M. & Lambowitz, A. M. in The RNA World. (eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 469-534 (Cold Spring Harbor Laborartory Press, Cold Spring Harbor, New York, 2006).

49. Pyle, A. M. Translocation and unwinding mechanisms of RNA and DNA helicases. Annu Rev Biophys 37, 317-36 (2008).

50. Qin, P. Z. & Pyle, A. M. Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg^{2+} ions in the tertiary structure. Biochemistry 36, 4718-4730 (1997).

51. Qin, P. Z. & Pyle, A. M. The architectural organization and mechanistic function of group II intron structural elements. Curr. Opin. Struct. Biol. 8, 301-308 (1998).

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

53. Roitzsch, M. & Pyle, A. M. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. RNA 15, 473-482 (2009).

54. Schmidt, U., Podar, M., Stahl, U. & Perlman, P. S. Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing in vitro and in vivo: Phenotypes and suppressor mutations. RNA 2, 1161-1172 (1996).

55. Schneider, S., Campodonico, E. & Schwer, B. Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. J Biol Chem 279, 8617-26 (2004).

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

57. Schroeder, R., Grossberger, R., Pichler, A. & Waldsich, C. RNA Folding in vivo. Curr. Opin. Struct. Biol. 12, 296-300 (2002).

58. Seetharaman, M., Eldho, N. V., Padgett, R. A. & Dayie, K. T. Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. RNA 12, 235-47 (2006).

59. Semrad, K. & Schroeder, R. A ribosomal function is necessary for efficient splicing of the T4 phage *thymidylate synthase* intron *in vivo*. Genes Dev. 12, 1327-1337 (1998).

60. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell 125, 287-300 (2006).

61. Seraphin, B., Boulet, A., Simon, M. & Faye, G. Construction of a yeast strain devoid of mitochondrial introns and its use to screen nuclear genes involved in mitochondrial splicing. Proc Natl Acad Sci U S A 84, 6810-4 (1987).

62. Solem, A., Zingler, N. & Pyle, A. M. A DEAD Protein that Activates Intron Self-Splicing without Unwinding RNA. Mol. Cell 24, 611-617 (2006).

63. Steiner, M., Karunatilaka, K. S., Sigel, R. K. & Rueda, D. Single-molecule studies of group II intron ribozymes. Proc Natl Acad Sci U S A 105, 13853-8 (2008).

64. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA 90, 6498-6502 (1993).

65. Su, L. in Biochemistry and Molecular Biophysics 187 (Columbia University, New York, 2002).

66. Su, L. J., Brenowitz, M. & Pyle, A. M. An alteRNAtive route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. J. Mol. Biol. 334, 639-652 (2003).

67. Su, L. J., Waldsich, C. & Pyle, A. M. An obligate intermediate along the slow folding pathway of a group II intron ribozyme. Nucleic Acids Res 33, 6674-87 (2005).

68. Swisher, J., Duarte, C., Su, L. & Pyle, A. Visualizing the solvent-inaccessible core of a group II intron ribozyme. EMBO J. 20, 2051-2061 (2001).

69. Tinoco, I., Jr., Uhlenbeck, O. C. & Levine, M. D. Estimation of secondary structure in ribonucleic acids. Nature 230, 362-7 (1971).

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

71. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. Kinetic intermediates trapped by native interactions in RNA folding. Science 279, 1943-1946 (1998).

72. Valadkhan, S. The spliceosome: a ribozyme at heart? Biol Chem 388, 693-7 (2007).

73. Valadkhan, S. & Manley, J. L. Splicing-related catalysis by protein-free snRNAs. Nature 413, 701-7 (2001).

74. Waldsich, C., Grossberger, R. & Schroeder, R. RNA chaperone StpA loosens interactions of the tertiary structure in the *td* group I intron in vivo. Genes & Dev. 16, 2300-2312 (2002).

75. Waldsich, C. & Pyle, A. M. A folding control element for tertiary collapse of a group II intron ribozyme. Nat. Struct. Mol. Biol. 14, 37-44 (2007).

76. Waldsich, C. & Pyle, A. M. A kinetic intermediate that regulates proper folding of a group II intron RNA. J. Mol. Biol. 375, 572-580 (2008).

77. Watanabe, K. & Lambowitz, A. M. High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. RNA 10, 1433-43 (2004).

78. Webb, A. E. & Weeks, K. M. A collapsed state functions to self-chaperone RNA folding into a native ribonucleoprotein complex. Nat. Struct. Biol. 8, 135-140 (2001).

79. Weeks, K. M. & Cech, T. R. Assembly of a ribonucleoprotein catalyst by tertiary structure capture. Science 271, 345-348 (1996).

80. Wells, S. E., Hughes, J. M., Igel, A. H. & Ares, M., Jr. Use of dimethyl sulfate to probe RNA structure in vivo. Methods Enzymol 318, 479-93. (2000).

81. Williamson, J. R. Induced fit in RNA-protein recognition. Nat Struct Biol 7, 834-7. (2000).

82. Woodson, S. A. Structure and assembly of group I introns. Curr Opin Struct Biol (2005).

83. Yang, Q. & Jankowsky, E. ATP- and ADP-Dependent Modulation of RNA Unwinding and Strand Annealing Activities by the DEAD-Box Protein DED1. Biochemistry 44, 13591-601 (2005).

84. Zhang, F., Ramsay, E. S. & Woodson, S. A. *In vivo* facilitation of *Tetrahymena* group I intron splicing in *Escherichia coli* pre-ribosomal RNA. RNA 1, 284-292 (1995).

Anhang

Zusammenfassung der vorliegenden Arbeit

Das Gruppe II Intron Ai5γ, welches im mitochondriellen Genom von Hefe zu finden ist, besitzt die Fähigkeit sich selbst aus dem Transkript herauszuschneiden. Um dies bewerkstelligen zu können muss dieses sehr große RNA Molekül eine komplizierte Struktur einnehmen. Die Faltung ist dabei in vitro von einer hohen Konzentration an mono- und divalenten Ionen abhängig. Im zellulären Kontext, wo diese hohen Konzentrationen nicht verfügbar sind, unterstützen Proteine bei dieser Aufgabe. Im Falle von Ai5γ wurde das DEAD-box Protein Mss116p als notwendiger Kofaktor bei der Faltung von Ai5γ identifiziert.

In der vorliegenden Diplomarbeit wurde das Gruppe II Intron Ai5γ in Bezug auf seine Struktur *in vivo* untersucht. Dabei war ich insbesondere an strukturellen Unterschieden von Ai5γ zwischen einem Hefestamm, dessen genetischer Hintergrund dem Wildtyp entspricht, und einer Variante, in welcher Mss116p durch eine Resistenzkassette ausgetauscht wurde, interessiert. Dadurch sollte die Funktion von Mss116p im Zusammenhang mit RNA Faltung beleuchtet werden.

Um strukturelle Information über Ai5γ im zellulären Kontext zu erhalten, benutzte ich Dimethylsulfat (DMS). DMS penetriert die Hefezelle und ihre Kompartimente rasch und methyliert N1 von Adenosin und N3 von Cytosin, wenn diese nicht durch Wasserstoffbrücken oder gebundenen Kofaktoren geschützt werden. Diese Informationen geben Aufschluss sowohl über Interaktionen auf der Ebene der Sekundär- als auch der Tertiärstruktur.

Nach der Optimierung der Methode habe ich zuerst die Struktur von Ai5γ in vivo untersucht. Dabei stellte sich heraus, dass, bis auf eine Ausnahme in Domäne 2, alle Sekundarstrukturelemente, wie zuvor postuliert, ausgebildet sind. Außerdem sind alle Tertiärinteraktionen bis auf eine Ausnahme vorhanden, was darauf schließen lässt, dass auch das aktive Zentrum des Ribozyms in vivo ausgebildet ist. Modifikationen der betreffenden Bereiche zeigten, dass der größte Teil der Ai5γ Moleküle nicht mehr an ihr Substrat gebunden sind. Konsistent damit fanden wir in einem “in vivo splicing assay” heraus, dass der Großteil der Population im geschnittenen Zustand vorliegt. Daraus lässt

sich schließen, dass das Gruppe II intron Ai5γ nach ihrem Herausschneiden ihre Struktur inklusive aktivem Zentrum erhält und für den Vorgang der Reinsertion keine bedeutenden Änderungen der Struktur nötig sind.

In einem Hefestamm, bei dem Mss116p entfernt wurde, zeigte sich, dass die Struktur von Ai5g stark gestört vorliegt, was sich auch im Fehlen von geschnittener mRNA zeigt. Die Sekundärstruktur bildet sich hier nicht aus und bleibt sogar unter dem Level von ungefaltetem Ai5γ in vitro. Sämtliche Tertiärinteraktionen fehlen ebenso. Außerdem wurden keine Anzeichen gefunden, dass sich Ai5γ in vivo ohne Mss116p falsch faltet und in einer sogenannten kinetischen Falle verharrt. Ebenso wurde kein Fußabdruck von Mss116p identifiziert, was darauf hinweist, dass Mss116p nicht mehr am nativ gefalteten Gruppe II Intron gebunden ist.

Meines Erachtens deuten die Ergebnisse dieser Diplomarbeit darauf hin, dass Mss116p in einem frühen Stadium der Faltung von Ai5γ aktiv ist, da bereits die Sekundärstruktur stark betroffen ist. Da bekannt ist, dass mehrere transiente Zwischenprodukte in der Faltung von Ai5γ entstehen, scheint Mss116p eines dieser Intermediate zu stabilisieren. Um die Faltung in den nativen Zustand zu ermöglichen muss zuerst Mss116p das Gruppe II intron freigeben.

Lebenslauf

Name: Andreas Liebeg
Geburtsjahr: 1978
Geburtsort: Aarau (CH)
Staatsbürgerschaft: österreichisch

Schulische Ausbildung

1985-1989: Besuch der Volkschule Deutsch Kaltenbrunn
1989-1997: Besuch des BG/BRG Fürstenfeld

Zivildienst

1997-1998: Auslandsdienst in der Gedenkstätte für Opfer der Euthanasie, Bernburg (D)

Akademische Ausbildung:

1999-2003: Studium der Heilpädagogik an der FH Zittau/Görlitz (D)
Akademischer Titel: Dipl.Heilpäd. (FH)

2003-2009: Studium der Molekularbiologie an der Universität Wien
Diplomarbeit: *DEAD-box protein facilitated folding of the group II intron ai5γ in vivo.*