MASTERARBEIT

„The Activity of Recombinant DEAD-box Protein Mss116p: Towards its Role in the Collapse of the Ai5gamma Group II Intron"

Eva Maria Steiner, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2012

Table of Contents

Abbreviations

APS	ammoniumpersulfat
BSA	bovine serum albumin
BS12	bottom strand 12 nucleotides
CP	cell pellet
COB1	cytochrome b gene-1
COX1	cyclooxygenase-1
CspA	cold-shock protein A
DbpA	DEAD-box protein A
DEAD	amino acid code for D – aspartic acid, E – glutamic acid, A – alanine
DMS	dimethylsulfid
dsRNA	double-stranded ribonucleic acid
DTT	dithiothreitol
EBS	exon binding site
EDTA	ethylenediaminetetraacetic acid
GTS	glycerol tris sodiumchloride buffer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIS	histidine tag
IBS	intron binding site
IGEPAL	synonym for octylphenyl-polyethylene glycol
IPTG	isopropyl-β-D-thiogalactopyranosid
KAN	kanamycin
LB	lysogeny broth medium
ME	MOPS EDTA buffer

MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
Mss116p	mitochondrial splicing system *Saccharomyces cerevisiae*, protein number 116
NAIM	nucleotide analog interference mapping
NAIS	nucleotide analog interference suppression
Ni-NTA	nickel nitrolotriacetic acid
OD600	optical density at 600 nm
ORF	open reading frame
o/n	overnight
PAGE	polyacrylamide gel electrophoresis
PCI	phenol chloroform isoamylalcohol
PNK	polynucleotide kinase
RNA	ribonucleic acid
RNP	ribonucleoprotein
rpm	rounds per minute
RT	reverse transcription
Sc160	saccharomyces cerevisiae primer
SDS	sodium dodecyl sulfate
SF1/2	super family 1/2 helicase
SN	supernatant
snRNA	small nuclear ribonucleic acid
SOC	super optimal broth medium
ssRNA	single stranded ribonucleic acid
StpA	DNA-binding protein stpA in *E. coli*
SUMO	small ubiquitin-like modifiers

TBE	tris borate EDTA buffer
TEMED	tetramethylethylendiamin
TEV	tobacco etch virus
TRA	transcription buffer
TRIS	trisamine
TS12	top strand 12 nucleotides
w/t	wild type

Introduction

An RNA World

RNA is a powerful and highly fascinating molecule which played a striking role during evolution throughout billions of years. Therefore the idea of a primordial RNA molecule has already been hypothesized in the 1960ies by Crick (1968), Orgel (1968) and Woese (1967), and resulted in an 'RNA world' concept two decades thereafter (Gilbert 1986) – with the discovery of catalytic RNA by Cech and Altman (*7, 8*). This concept (Figure 1), where RNA appeared first in time during evolution of life on earth, is supported by a number of arguments and is currently widely accepted. The basis of this
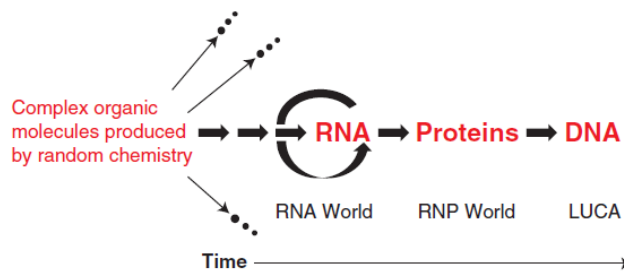
Figure 1: The 'RNA world' model. First complex organic molecules which proceeded a primordial RNA world evolved. Supported by organic molecules self-replicating RNA systems developed. Due to the presence of amino acids and peptides it is hypothesized that proteins followed in time during evolution. LUCA (Last Universal Common Ancestor) is believed to evolve recently < 1 billion years, which already carried a DNA genome (*2*).

model states a set of complex organic molecules which were produced by random chemistry. The majority of the molecules died out but one pathway proceeded and led to self-replicating RNA. The additional arrow around the term "RNA world" indicates potential self-replicating systems which further supported the development of self-replicating RNA. Proteins came about after RNA was present, although amino acis and peptides were present in an RNA world, to catalyze peptide bond formation and amino acid polymerization. DNA is believed to have evolved more recently and this aspect is supported by the fact that LUCA (Last Universal Common Ancestor) already had a DNA genome using biochemical catalysis, enzymes and RNP complexes (*2*).

Many arguments support an RNA world model and one of them is that RNA can be both - genotype and phenotype, and that RNA is able to perform self-replication chemistry (*9*). Therefore the likelihood seems greater that one molecule, RNA, appeared first which is able to perform both functions than two separate molecules, RNA and DNA, were synthesized at the same time and place (*2*). The most direct evidence for RNAs pioneering role would be a crystallographic approach which shows that a ribosome's core activity (peptide bond formation) is driven by an RNA molecule (*10*). In this regard, it has to be noted that the crystal structures of the ribosome revealed that the peptidyl transferase center is devoid of proteins (*11*). Aside from the naturally occurring large and small ribozymes, which catalyze transesterification reactions, dozens of other RNA catalyzing activities were discovered by SELEX (Systematic Evolution of Ligands by Exponential Enrichment), which could have been crucial for RNA in a primordial RNA world (*2*) – e.g. catalyzing peptide bond formation or short RNA molecules that can duplicate others (*12, 13*) . Finally, the abundant enzymes dedicated to RNA biosynthesis and the continuity of the system comprises its first appearance compared to only two additional enzymes (thymidylate synthase and ribonucleotide reductase) which evolved for DNA (*2*). The above mentioned arguments support an RNA world concept, however, an "RNA-first" view also triggers a lot of unsolved problems and opinions vary as to whether an RNA self-replicating system appeared autonomous or is a derivative of an earlier system (*2*). The complexity of modern RNA systems raised circumspect voices which doubt its arising *de novo* (*14, 15*). Today a lot of different approaches, like prebiotic chemistry, synthetic biology or paleogenetics, try to offer experimentally accessible data and a model to explain how life might have emerged e.g. *in vitro* evolution studies (experiments on simple liposomes which grow and divide into 'protocells' upon agitation) (*15-17*).

As RNA is not only a medium to store information but also capable of performing an orchestra of different catalytic functions, it seems obvious why it is worth the effort and energy trying to decipher the basic biochemistry and structural principles of this strikingly diverse molecule.

1. Basic RNA Biochemistry

The structure of an RNA macromolecule is, as in proteins, characterized by its three organization levels. The first level presents the sequence, where nucleotides are connected via their sugar-phosphate backbone. Under ionic conditions RNA folds back on itself via Watson-Crick base pairing between the bases constituting RNA secondary structure. The second organization level compasses double stranded helices interrupted by single stranded regions in internal loops or hairpin loops. Highly dependent on ionic and temperature conditions an RNA molecule is able to perform the transition into a 3-dimensional fold. Within this 3D RNA macromolecule the helices and single stranded regions are very well organized and are instrumental in specific biological functions (11). In the following chapters we take a closer look into RNA organization levels.

2. RNA Primary Structure

The basis of RNAs' first organization level, the primary structure, is alternating occurrence of

nucleobases adenine (A), cytosine (C), guanine (G) and uracil (U). Each so called nucleotide element (5'-NMP) consists of a negatively charged phosphate group, a D-ribose sugar ring and an aromatic nucleobase, pyrimidine or purine, respectively (Figure 2). Several
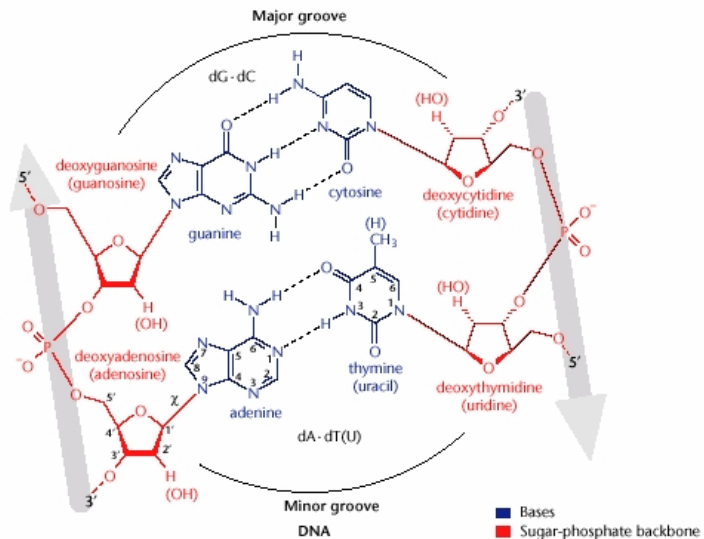
Figure 2: DNA and RNA Watson-Crick base pairing of DNA and RNA duplexes (1).

nucleotide units are connected by its phosphates over 3', 5'- phosphodiester bonds resulting in a

nucleotide chain which shows 5'- to 3'- chain polarity. The chain therefore offers an array of

biochemical features like a negatively charged sugar-phosphate backbone and a set of

hydrophobic nucleobases with amphilic features which naturally drive the occurrence of

secondary and tertiary RNA structures (18).

3. RNA Secondary Structure

In principle RNA bases offer three edges for H-bonding. They are depicted in the left panel of Figure 3 and comprise Watson-Crick edge, Hoogsteen edge (for purine and CH edge for pyrimidine) and sugar edge (includes 2'OH group). The bases interact over its edges by stacking, which account for 90% of interactions within an RNA macromolecule and represents the major force driving RNA folding (*4, 18*). Another way of base interaction is to join edge-to-edge, which is triggered by hydrogen bonds and gives the macromolecule directionality and specificity (*4, 18*).

Figure 3: The three edges for base pair interaction. Left panel: edges for adenosine and cytosine. The Watson-Crick edge, the Hoogsteen edge, the sugar edge and the ribose hydroxyl group O2'. Right panel: showing a cis and trans orientation which is defined by a parallel line drawn between the bases hydrogen bonds (*4*).

In the right panel of Figure 3 glycosidic bond orientations, that can appear as *cis* or *trans* orientation, are illustrated. As a matter of fact, 12 distinct edge-to-edge orientations (Table 1) are possible and over indicating interaction edge and orientation of the glycosidic bond of each base, pairing geometry can be categorized. A canonical Watson-Crick base pair (A-U or G=C) for example would belong to family No. 1. The so called 'wobble' pairs, which terms base pairing between non-complementary bases (G and U) (3) is characterized by a geometrical shift of one base relative to the other. Attention should be paid to the geometry of wobble base pairs as they are not strictly isosteric (in contrast to Watson-Crick base pairs), meaning that a GU wobble is not geometrically identical to a UG wobble (11). The overall stability of canonical base pairing increases as follows: G-C>A-U>G-U (wiki.case.edu).

The sum of Watson-Crick base pairs and intervening unpaired regions can appear as distinct structural elements which is given attention in the following chapter. This nomenclature provides a useful tool to present RNA structure as a two-dimensional system (*19*).

Table 1: displays the 12 distinct families of edge-to-edge orientations between two base pairs. This table assumes that all bases are default *anti* conformation with respect to the sugars, as *syn* conformation is rather rare (*4*).

No.	Glycosidic bond orientation	Interacting edges	Local strand orientation
1	*Cis*	Watson–Crick/Watson–Crick	Antiparallel
2	*Trans*	Watson–Crick/Watson–Crick	Parallel
3	*Cis*	Watson–Crick/Hoogsteen	Parallel
4	*Trans*	Watson–Crick/Hoogsteen	Antiparallel
5	*Cis*	Watson–Crick/Sugar Edge	Antiparallel
6	*Trans*	Watson–Crick/Sugar Edge	Parallel
7	*Cis*	Hoogsteen/Hoogsteen	Antiparallel
8	*Trans*	Hoogsteen/Hoogsteen	Parallel
9	*Cis*	Hoogsteen/Sugar Edge	Parallel
10	*Trans*	Hoogsteen/Sugar Edge	Antiparallel
11	*Cis*	Sugar Edge/Sugar Edge	Antiparallel
12	*Trans*	Sugar Edge/Sugar Edge	Parallel

Secondary Structure Elements

A set of secondary structure elements is presented in Figure 4. The aminoacyl-tRNA offers a descriptive model for secondary structure element interpretation, as it is one of the best understood and most studied molecules due to its central role in gene expression and a high appearance in the cell. In a 2D and 3D model of yeast tRNA^{Ala} is given as an example. In panel (A) a secondary structure model demonstrates the typical clover leaf shape, consisting of three hairpin structures (D-, T-, anti-codon arm) and the acceptor stem, a large multi loop (= four way junction) in the center and a set of additional secondary structure elements (*19*).
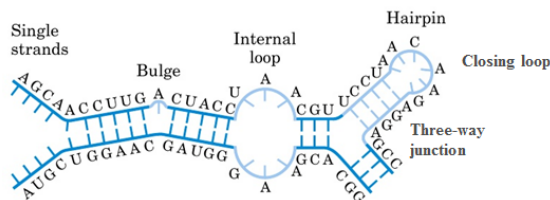
Figure 4: Secondary structural elements. They are defined by a pattern of paired and unpaired bases and represented by bulges, internal loops, and hairpin structures.

2ndry Structure Prediction

A way to determine such a secondary level model is structure prediction with the help of programs (e.g. RNAfold or Mfold). Most of the programs use algorithms that predict the lowest free energy potential for all possible base-pairing combinations and structural elements. Reliable programs take into account of comparative sequence analysis (covariance of nucleotides e.g. exchange of one Watson-Crick pair for another: CG to an AU pair) and experimental data, like chemical probing and/or mutational analysis, among others. Nevertheless, such a secondary structure has to be verified over structural probing techniques that exploit chemicals or enzymes which specifically target RNA structure (e.g. DMS probing). Chemical modification sites are, like in our experiments, revealed by reverse transcription with a radioactive labeled DNA primer (^{32}P end-labeled). Structure prediction data combined with biochemical techniques then yields to important information about the tertiary structure of an RNA macromolecule (*19*).
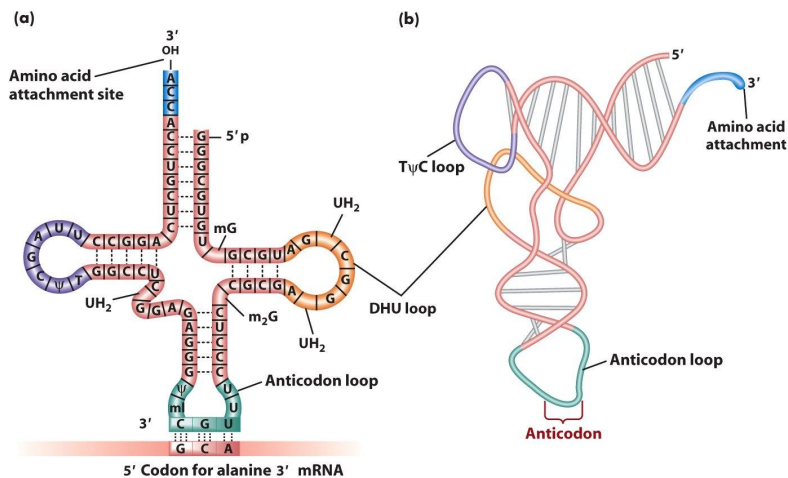
Figure 5: A secondary and tertiary structure representation of aminoacyl-tRNA. (a) Clover leaf representation of yeast alanine tRNA. Watson-Crick base pairs are indicated as blue dots. The World of the Cell by Becker et al, published by Pearson Education

4. RNA Tertiary Structure
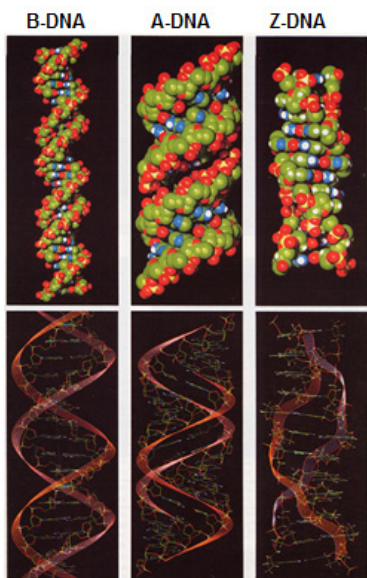
Figure 6: the three forms of helical DNA A, B and Z. The characteristic A-form (middle panel) is characterized by the presence of a 3'-endo sugar pucker, a compact shape (P-P distance 5.9 Å), right-handed orientation and tilted bases to helical axis.

RNA reveals many additional edge-to-edge interactions which give rise to diversified tertiary motifs. The enumeration of these paired regions or helices, displays a description of a secondary structure level. As a consequence of growing literature on RNA biology and the rapid progress of RNA crystallography during the 1990ies, Leontis and Westhof (2001) established a geometric nomenclature and classification of RNA base pairing which unfolded a variety of base-pairing geometries. The nomenclature offers a way to communicate RNA structural information in a two-dimensional representation of complex 3D RNA structures. 3D motifs can be specified by indicating interaction edges and orientation of glycosidic bonds of two bases.

The tertiary organization level resulting from secondary interactions appears in a strikingly diverse manner and depends highly on divalent ions, temperature and cofactors. Panel (A) and (B) of Figure 5 show that the previously predicted secondary clover leaf structure leads to a complete different appearance in the 3D model. The RNA folds by co-axial stacking (tertiary motifs) of the D-stem onto the anticodon stem and the T-stem onto the acceptor stem, resulting in two helical segments. The D- and T-loops which contained unpaired bases only in the 2D structure, form several non-canonical base pairs enable the proper orientation of the two helical stacks to each other. Concerning the overall stability of a 3D RNA molecule RNA helices are not much affected by environmental changes. In contrast, the lack of divalent cations causes a significant change in tertiary interactions and structure; in particular non-canonical base pairs appear very sensitive

('wobble' base pairs; structurally described in previous chapter). As a matter of fact they often play an important function serving as recognition elements for proteins, nucleic acids or ion-binding sites (18). As demonstrated in the tRNA example above RNA adopts quite often helical structures. While DNA duplexes prefer the classical B-form helix, RNA adopts the very characteristic A-form in the third organization level which results of canonical Watson-Crick base pairing showing a remarkable isostericity with a strict tertiary helical structure. The structure is characterized by a narrow, deep major groove and a wide, exposed minor groove (Figure 6). The major groove has a strong negative charge and high potential for binding metal ions or protein side-chains. Therefore the major groove often needs help in widening the narrow groove to allow protein ligand interaction. Compared to the B-form, A-form nucleobases are tilted to the helical axis and stacking is enforced by their $\prod$- electrons and plane aromatic rings, below and above respectively, resulting in a more compact shape. This is achieved by the C3'-endo sugar pucker, predominant for A-form of RNA, which allows a closer phosphate-phosphate distance (5.9 Å) than induced by the C2'-endo sugar pucker in B-form helices (*20*).

Other examples for tertiary structure motifs for example is the 'ribose zipper' which has been found in the P4-P6 structure within the *Tetrahymena* self-splicing intron (*21*). This tertiary motif is characterized by consecutive hydrogen-bonding between 2'-OH groups of the ribose from different regions of an RNA chain or between two RNA chains (*22*). The 'tetraloop-helix interaction' has also been found in the P4-P6 domain of *Tetrahymena* group I intron and in the crystal structure of the hammerhead ribozyme (*21, 23*). In the hammerhead structure of ribozymes a GAAA tetraloop interacts with a helical stem capped with a GNRA tetraloop that ends with Watson Crick base pairs (G-C). The Gs belonging to these Watson Crick pairs the final A of the capping loop; interact with all three A's of the GAAA tetraloop. Tetraloop-helix interactions stabilize over hydrogen bonding interactions between NRA bases of GNRA tetraloops and groups found in the minor groove of helical stems (*24*).

The third organization level is the decisive factor for biological function of an RNA molecule and several molecular biological techniques have to be combined to get an overall image. Tertiary structures are identified by comparative sequence analysis close together with structural probing, chemogenetics, mutational analysis, homology modeling and, importantly, biophysical techniques, crystallography and NMR spectroscopy. But this is a very tricky challenge for RNA molecules as a lot of them are quite huge, highly charged and need high ion concentration to maintain their native, functional structure (without inducing spontaneous metal-induced cleavage) (*25*). A combination of all available approaches is therefore suggested to unveil RNA architecture, folding and function.

RNA Folding and the Problem

The biological activity of an RNA molecule is highly dependent on its proper folding. Over the past decades the *Tetrahymena* group I ribozyme was thermodynamically examined and presents one of the best understood models for RNA folding. All studies and facts could be unified and the output was presented exactly 10 years ago (*26*). A folding pathway to the native structure is interspersed by folding intermediates and most of them are on-path with an increased amount of native contacts. However, an RNA molecule can be trapped in non-native conformations and escaping such kinetic traps is often rate-limiting during a folding pathway (*5*). A multidimensional
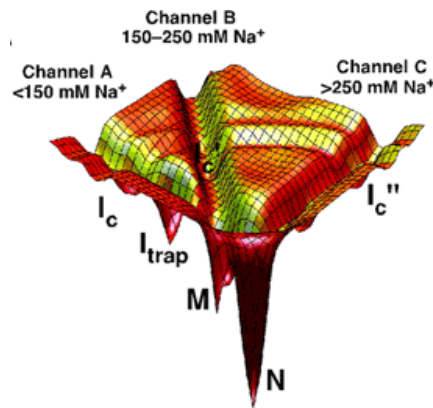
Figure 7: Folding landscape of wt Tetrahymena group I ribozyme. Likely folding pathways are indicated as folding channels (A,B,C). The landscape is contoured by internal free energy. Folding properties depend highly on initial conditions and 'walls' need to be overcome to interconverse different folding pathways (Russel 2001). (N) native ribozyme conformation; (M) misfolded structure; (Itrap) intermediate trap; (Ic, Ic'') initial folding conditions at different ionic concentrations) (*3*).

energetic landscape is able to summarize and describe intermediate states and parallel folding pathways. The overall shape of a landscape provides information about thermodynamics and kinetics of the RNA molecule (3). Figure 7 presents a schematic of the *Tetrahymena* group I ribozyme energetic folding landscape contoured by its internal free energy. Every point in the landscape states information about the free energy, degrees of freedom and surrounding conditions of the ribozyme in a specific structural conformation. Someone can see folding funnels with kinetic traps (I_{trap} or M) along a folding pathway energetically downhill to the final folding state of the molecule (N). The final state can be described as the conformation with least free energy and is a very narrow region within the energy landscape. In general folding landscapes for RNAs are substantially more rugged than protein landscapes but for both a pathway is mainly dependent on the initial conditions (3).

As previously mentioned the native state is a very narrow region in an energetic folding landscape. Therefore it is quite likely that a lot of molecules adopt non-native conformations that are energetically favored. Trapped folding intermediates (I_{trap}) and misfolded structures (M) are very common in folding pathways; the escape from such kinetic traps requires structural reorganization and is often the rate-limiting event in folding. In other words, RNA faces a substantial problem which is termed as 'RNA folding problem'. Assistance to overcome the folding problem is provided by nature: i.) proteins and ii.) metal ions (27).

1. **Proteins Involved in RNA Folding**

In a living cell metal ion concentration is well below that of *in vitro* folding studies but two types of proteins showed profound features for guiding RNA to its catalytically active structure. One type acts as non-specific RNA chaperones without recognizing a specific sequence or structural element within their target RNA(s) (27). The other type though binds very tight to RNA specifically recognizing either RNA sequence or structural elements, thereby stabilizing RNA structure (28).

The *E. coli* proteins StpA or S12 had been shown to act as RNA chaperones, both *in vitro* and *in vivo* (*29-31*). Structural analysis using DMS modification showed that StpA offers its support by resolving tertiary interactions within its target RNA, the *td* group I intron *in vivo,* thereby destabilizing the intron structure (*32*). There are several other proteins which showed RNA chaperone activity *in vitro* like the cold-shock protein CspA , the yeast La protein or the tumor suppressor protein p53 (*33, 34*). Proteins like Cyt18, Cbp2 and LtrA for example were observed to bind specifically to their target RNAs, thereby decreasing the Mg^{2+} concentration required for proper folding *in vitro* (*28, 35-37*). For some proteins like the DEAD-box Mss116p the specific role remains unclear. The yeast mitochondrial splicing system (MSS) component has been shown to be essential for efficient splicing of all yeast group I and group II introns (*6, 38*). *In vitro* folding studies revealed that Mss116p stabilizes, acting as a cofactor and reducing Mg^{2+} concentration for intron folding, as well displaying unwinding and annealing activity *in vitro* (*6, 38, 39*). Unclear remains whether unwinding activity is essential for stimulating intron splicing and two distinct models for Mss116p have evolved – Mss116p functions as splicing cofactor either by acting in a chaperone-like manner or by providing stability to RNA structure through specific binding of intron regions (*6, 38, 40*). Meanwhile a dual role for Mss116p in assisting intron folding has been proposed *in vitro*, namely stabilization of the intron core and resolving exonic structures that interfere with splicing *in vitro (41).*

2. Metal Ions and RNA folding

The earliest event in folding of large multi-domain ribozymes is the transition from the unfolded state to a compact structure (*42*). The transition is significantly preceded by the presence of counterions (non-specific electrostatic interactions with 'condensed' metal cations) and their charge reduction on the negative phosphate backbone with more than 90% (*43*). This condensation around RNA reduces the electrostatic repulsion between the phosphate groups,

reduces substantially the radius of gyration (R_g) and as a result drives the collapse of RNA (counterion-mediated collapse). The counterion-mediated collapse has been proven with small angle X-ray and small angle neutron scattering experiments with various RNAs (*43-45*). Monovalent ions such as Na^+ or K^+ lack preferred coordination geometry that is different to divalent cations such as Mg^{2+} (*46*). In the *H. marismortui* structure of the large ribosomal subunit 88 monovalent cations led to the classification of four different monovalent ion categories according to their binding sites: a) those that bind to the major groove of RNA helices, b) those that bind ribosomal proteins, c) those that bind in the interface of ribosomal protein and RNA and d) those that bind RNA motifs (*46*). The coordination distance for Mg^{2+} hence, is much smaller than e.g. for Na^+ (between 2.8 Å to 3.2 Å) and although it never appears fully hydrated, Mg^{2+} it is the preferred partner for inner-sphere coordination and as a result for tertiary RNA interactions (*46*). For example, the autocatalytic group II intron ai5γ requires 500mM monovalent ions, 100mM Mg^{2+} and 42°C for optimal folding and splicing *in vitro* (*47*). Notably, *in vitro* conditions are typically non-physiological meaning that ion concentration and temperature are considerably increased (*48*). Figure 6 further underlines the utmost importance of metal ions, for the *Tetrahymena* ribozyme where at low Na^+ (<150mM) concentration the molecule can easily be caught in either a misfolded conformation or intermediate trap. Higher initial Na^+ conditions obviously allow bypassing these traps and a folding funnel is chosen, in which the native RNA conformation is reached directly and more rapidly (*3*). Metal ions and RNA folding are examined in more detailed in chapter 'Metal Ions and RNA folding'.

3. Metabolites

Furthermore, there are metabolites which promote RNA folding. *UREA* for example has been long known to resolve misfolded RNA conformations at sub-denaturing concentrations (*49*). It can be used for probing stability and unfolding kinetics of proteins (*50*) and plays a helpful tool for exploring RNA folding parameters as well (*51*). Folding studies in the presence of urea further confirmed that the D135 ai5γ ribozyme, for example, follows a folding pathway devoid of kinetic traps (*52*). However up to now it had been unknown how urea destabilizes RNA. Mason et al. (2004) used extensive atom molecular dynamics (MD) simulations to examine the mechanism in which urea molecules participate in stacking interaction with purine bases of RNA. The findings strongly indicate that destabilization of RNA is caused by a disruption of base-pair interaction over multiple hydrogen bonding and formation of stacking interactions of urea with the bases, thereby increasing the folding rate of misfolded, kinetically trapped RNA molecules and acting like a 'chemical' chaperone (*53*).

Polyamines are directly affecting gene expression and its role in RNA folding has been illustrated for the *Tetrahymena* ribozyme, for which they induce near-native conformations through structural stabilization. The effect increases with charge and decreases with size of the polyamines; thus changes in intracellular polyamine concentrations might alter RNA folding pathways. Furthermore polyamines enhance the synthesis of key regulators of gene expression at translational level (*43*). In light of these findings, an additional study, in which polyamines were found to enhance the synthesis of RpoE and StpA at translational level, further underlines the positive effect of polyamines on RNA folding. Polyamines are thought to enhance synthesis of key regulators in gene expression at translational level over specific interaction with mRNA regions (*54*).

In the following chapter we had a closer look on metal ions and how they contribute to RNA folding.

Metal ions

Large multi-domain RNA in low salt buffer should have an extended conformation because of electrostatic repulsion of its negatively charged sugar-phosphate backbone. Metal ions influence RNA folding into a specific tertiary structure over thermodynamic and structural forces (55). *In vivo* the most relevant metal ions which support RNA folding of group I and group II introns are K^+ and Mg^{2+} and these therefore had been examined extensively (56). In the following chapter a short overview of three ion environments that differ in ion-RNA contacts, the degree of ion hydration and to what extent ion-environments contribute to

Figure 8: Mg^{2+} ions and the RNA folding problem using tRNA structure formation. (A) Monovalent ions: black dots, divalent ions: circles. Two transitions are represented; first starting transition from an unfolded tRNA state (U). Mg^{2+} is involved mainly in promoting tertiary structure. (B) Illustrating three types of Ion-RNA environments. Left: direct ion-RNA contact; middle: water-mediated contact; right: diffuse ions and hydration. Ion: small circle; larger circles and shades: RNA backbone (Draper 2005).

RNA folding is given (55). Figure 8 (B) illustrates the three general kinds of ion-environments, how ions stabilize RNA and why some ions are more effective than others.

1. Ion-RNA Environments

In the right panel of Figure 8 (B), ions are diffuse (*diffuse ions*) and only coulomb interaction forces are present. Diffuse ions show no contact to RNA surface and their ion hydration shell is not perturbed. Ions can be captured by the electrostatic field of RNA; however they are not restricted to an exact location in space. The nonlinear Poisson-Boltzmann (NLPB) equation provides accurate description and reliable results to calculate energy potentials of ions within RNA regions (57). Water-positioned ions (*delocalized ions*) are illustrated in the middle panel. Still

there is a layer of water molecules between RNA and ion but a lot more forces affect and bend the RNA backbone. Coulomb forces between RNA and ion, perturbation of hydration layers and unsatisfied water molecules cause polarization and charge transfer (*55, 58*). In the right panel chelated ions (*site-specific ions*) interact with at least two negatively charged RNA groups. The electrostatic field surrounding ion and RNA favors a (partially) dehydrated state of the ion resulting in a very close ion-RNA environment (*55*). Notably, interactions between ion and RNA structure cannot be simplified to certain classes it is more a concert of all ion 'classes' which contribute to ion-RNA electrostatic field.

2. The Special Role of Mg^{2+}

The special stabilizing effect of Mg^{2+} on tRNA tertiary structure has been discovered in the 1970ies (*59*) and marked a remarkable change in the past decades leading to discovery of new RNA functions such as ribozyme catalysis (*55*). Together with K^+(~140mM), Mg^{2+} (~1mM) has the highest concentration of all ions within a living cell and is thus the most relevant ion for physiological function (*56*). It has a smaller ionic radius (0.65 Å) than its monovalent and divalent companions (K^+ has an ionic radius of 1.3 Å) and with its 10 electrons it is able to bind six ligands in octahedral conformation. Mg^{2+} is preferred to charge neutralization of RNA structures due to its greater charge density and lower entropic cost for localizing ions in a 3D RNA structure (*60*). Nevertheless it is a very difficult task to study Mg^{2+}-RNA interactions with crystallographic approaches. The reason is that in crystal structures Mg^{2+} is difficult to distinguish from Na^+ ion and water molecules, thus a high structure resolution (<1.5A) is necessary to identify them (*56*). In general Mg^{2+} forms both outer and inner sphere contacts with RNA and the following chapters examine energetic interaction modes in more detail.

Outer Sphere Localization

Preferred outer sphere ligands for Mg^{2+} ions are anionic phosphate oxygen or electronegative atoms of the bases (*56*). This kind of interaction is often seen in the major groove of A-form helices by interacting with the Hoogsteen face of mostly non-canonical base pairs. Guanine presents a preferred outer sphere interaction partner (favored atoms are G(N7) and G(O6)) this fact has been shown in studies on HIV-1dimer initiation site (*61*) or in the P4-P6 domain of *Tetrahymena* (at a tandem 5'GG3'/3'UU5') (*62*). Outer sphere ions are very mobile in contrast to partially dehydrated Mg^{2+} ions that directly bond to RNA atoms (*56*). NMR and crystallographic studies often present useful data, in particular if no substitutes had been used to demonstrate magnesium binding (*56*).

Inner Sphere

Partial dehydration and therefore direct bonding to RNA atoms is more likely in non-helical regions (*56*). Neighboring phosphates take a closer position to each other enabling an electrostatic environment for Mg^{2+} to dehydrate its water molecules and chelate to RNA atoms. Mg^{2+} offers charge neutralization to RNA, this can be observed for example in the crystal structures of P4-P6 group I intron and 58-nucleotide rRNA (*63, 64*). A classification of Mg^{2+} had been derived from 116 ions bound to the 50S large ribosomal subunit of *Haloarcula marismortui* by Klein et al. (2004) (*46*). They observed ten binding modes between Mg^{2+} and RNA but complete dehydration of Mg^{2+} has not yet been observed (*56*). The loop E motif of the ribosomal RNA has been solved in high resolution (1.5Å) and represents a good model for inner sphere coordination. In this structure water molecules and Mg^{2+} ions could be distinguished by their bonding distance (2.1 A for Mg^{2+} and 2.7 for water).

Studies on small RNAs or nucleotides, computer modeling (e.g. Poisson-Boltzmann calculations or molecular dynamic simulations) together with crystallography and NMR, or more a concert of

all approaches provide seminal strategies for providing further insight of Mg^{2+}-RNA binding behavior (*65*) .

DEAD-box Proteins

In 1989, primary structure alignments of eight different proteins, ranging from E. coli (SrmB) to a human antigen (p68), to the eukaryotic initiation factor eIF-4A, lead to the discovery of stunning sequence homologies. Regional similarities of these proteins revealed common motifs like the D-E-A-D box or the S-A-T motif. With the help of additional biochemical data they defined a new protein family where proteins share several motifs, as well as functions like ATPase activity, unwinding of and interaction with nucleic acids (*66*).
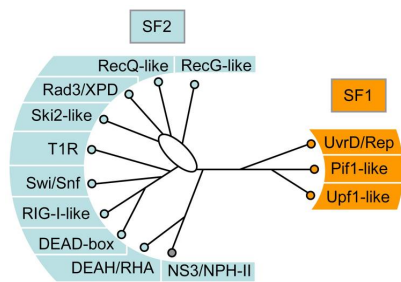
Figure 9: Unrooted cladogram showing different helicase families identified by a phylogenetic alignment of all SF1 and 2 helicases of human, yeast and selected viruses. (Fairman-Williams et al. 2010)

DEAD-box proteins are ubiquitous proteins mediating numerous RNA processes and depend on ATP binding and hydrolysis for exerting their helicase function. They belong to the group of superfamily 2 (SF2) helicases, of which DEAD-box proteins represent the largest family (*67, 68*). Currently, at least 25 yeast DEAD-box proteins, 37 in humans and 5 in *E. coli* are known (*69*). They perform a diverse set of functions such as they are associated with molecular machines, like the ribosome and they play a role in nuclear export of mRNA, folding of self-splicing introns or even in quality control mechanisms (Figure 10) (*70*). Besides their broad range of functions they also possess considerable diversity in their mechanisms. In principle their central activity is to bind and hydrolyze ATP in an ssRNA or dsRNA
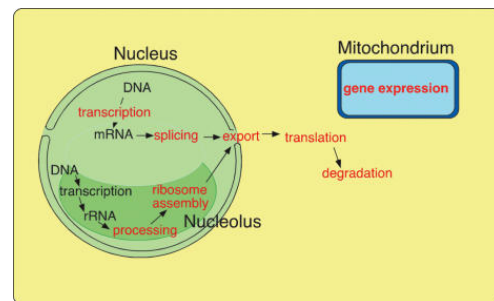
Figure 10: processes in which DEAD-box proteins are required. (Linder et al. 2006)

dependent manner. This biochemical activity is then used by different DEAD-box proteins in different ways. They either separate short dsRNA regions to promote structural rearrangements, remodel RNP complexes or form highly stable complexes on ssRNA (*70*). Due to these effects on RNA structure DEAD-box proteins play a highly interesting role on RNA folding. In the following the main structural features and conserved sequence motifs of DEAD-box proteins are described.

DEAD-box Proteins – Their Structure and Motifs

What we do know about the biochemical activities of DEAD-box proteins is that they are often referred to play a role in unwinding double-stranded RNA. Recent studies suggest that DEAD-box proteins unwind RNA duplexes *in vivo* (*71, 72*) therefore they are often designated as helicases. Such energy related unwinding of double-stranded RNA has been demonstrated but in most cases this activity is not processive, meaning they only unwind very short RNA duplexes (*69*). The most likely explanation for the weak action as a helicase would be that they are only required for local unwinding of a limited length of dsRNA or rapid dissociation of the protein to allow downstream processes (*69*). There has also been shown that the helicase activity of DEAD-box proteins declines with increasing dsRNA length or stability. Some proteins e.g. eIF4A show equivalent activity on blunt-ended duplexes and on helices with an 5' or 3' extension (*73*), implying that their function is to generate local strand separation within a helix. This has been as well shown for the putative orthologs Mss116p and Ded1p (*39, 40, 70*). Many other proteins though show enhanced unwinding for extended helices. As it is the case for the *Neurospora crassa* protein CYT-19 which promotes folding of mitochondrial group I and II introns and shows increased unwinding activity upon extended helices within the highly structured group I introns (*40, 74*). Thus, it has been proposed that extensions provide a site of interaction for the DEAD-box protein to disrupt nearby RNA structure (*70*).

Solving the structure of yeast translation initiation factor eIF4A (Caruthers et al. 2000) and subsequent structural alignments of all annotated sequences revealed at least nine conserved motifs that span across the helicase core of the DEAD-box protein family and represent the main center for protein activity (*66, 75*). The helicase core consists of two RecA like domains arranged in tandem and connected by a flexible linker. The highest degree of conservation across both helicase families had been observed to in motifs that are crucial for nucleotide-triphosphate binding and hydrolysis and for residues located within cleft between two RecA-like domains (*67*). More precisely, Motif II (also called Walker B motif) which carries the amino acid sequence D-E-A-D, together with motif I (Walker A motif), Q-motif and motif VI are required for ATP processing (*75, 76*). Less well characterized are motifs Ia/b, III, IV and V but these have been proposed to be involved in binding of the RNA (*77*). In a recent study, in which they used the Ded1 DEAD-box protein as a model, it has been shown that mutations in motif III (also known as S-A-T motif) severely reduce the binding affinity for ssRNA, the ability to unwind duplexes and to hydrolyze ATP but these mutations have no effect on ATP binding both *in vivo* and *in vitro* (*78*). Thus it has been concluded that motif III and the energy from ATP binding is used to coordinate the two RecA-like domains, in particular motifs I, II and VI, to create a high-affinity ssRNA binding site (*78*).

Not many DEAD-box proteins have been biochemically characterized yet compared to the enormous amount of proteins present in databases (*69*). One very critically examined is Mss116p. It is still unclear whether Mss116p provides stabilization of the intron structure, therefore acting as an RNA cofactor or if it supports RNA folding rather in a chaperone-like manner on group I and II introns (*6, 38, 40*). Therefore we chose the DEAD-box protein Mss116p in our studies and the chapter below provides detailed information about Mss116p.

The DEAD-box Protein Mss116p – a Group I and Group II Intron Splicing Factor

Mss116p is a member of the DEAD-box protein family and belongs to the mitochondrial splicing system of *Saccharomyces cerevisiae*. It is essential for splicing of 13 group I and group II introns *in vivo* but how Mss116p facilitates splicing remains enigmatic (*79-82*). Like all DEAD-box proteins, Mss116p is able to unwind short RNA duplexes by local strand separation, thus inducing conformational changes on RNA and RNP complexes. It contains a conserved helicase core consisting of two RecA-like domains linked by a flexible linker and flanked by an N-terminal extension (NTE) as well as a α-helical C-terminal extension (CTE) and a basic hydrophilic C-tail (Figure 11 and Figure 12) (*83*). The helicase core of Mss116p encompasses 9 conserved sequence motifs and in the closed conformation of the folded protein motifs important for ATP and RNA binding are brought together forming the 'binding cleft' (*83*).
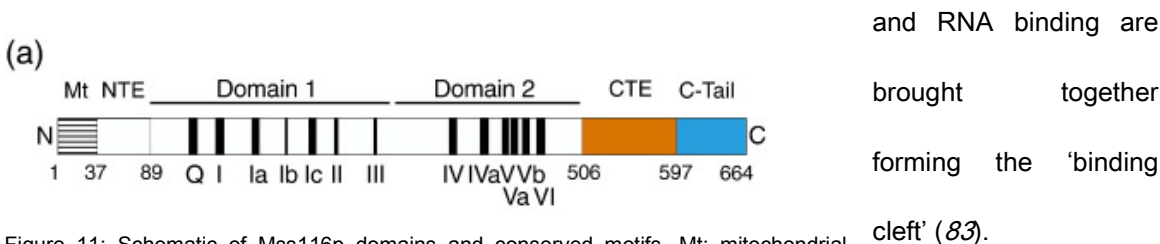
Figure 11: Schematic of Mss116p domains and conserved motifs. Mt: mitochondrial localization sequence, NTE: N-terminal extension, CTE: C-terminal extension. (Mohr et al. 2011)

N-terminal Extension (NTE)

The NTE is located upstream of domain 1 of the helicase core. This sequence region has been of minor interest during the past years and only a few studies on the NTE exist. One presents SAXS solution structures for Mss116p, showing that the NTE emerges from domain 1 pointing away from RNA binding regions (*84*). In addition, high-throughput genetic analysis combined with *in vivo* splicing assays showed that mutations in this region (except for regions close its boundary to the helicase core) had only minimal effect on Mss116p-promoted intron splicing *in vivo*. Mss116p mutants completely lacking the NTE function nearly as efficient as the w/t protein in promoting

ai5γ group II intron splicing *in vitro* (*84*). Taken together, it has been suggested that the NTE is a flexible attachment to the helicase core which has only minor effect on Mss116p's role as splicing factor and does not alter folding behavior of the protein (*83*).

Helicase Core

The helicase core of Mss116p is a highly conserved region and very similar to other DEAD box proteins like CYT-19, Vasa, DDX19 or eIF4AIII (*85*). It

Figure 12: structure of Mss116p with ssRNA and AMP-PNP showing two different views of domains and motifs. (Del Campo 2009)

comprises 9 conserved motifs that span across the two RecA-like domains which are connected via a flexible linker region. Mss116p had been successfully crystallized in complex with a 10mer poly-U RNA oligonucleotide and an ATP analog (AMP-PNP) in 2009 with a resolution of 1.9 Å. This crystal structure shows the entire closed-state helicase core and C-terminal extension (CTE) of the protein. In an open conformation the RecA-like domains are separated and moveable over the linker. The compact closed state is induced upon binding of ATP and RNA. When substrates are bound the conserved motifs are brought together in close proximity and are the main interaction points for ATP and RNA binding (*85*). Each motif participates in a certain manner to RNA and ATP binding and below we summarized their specific functions.

Motif Q has a specific function in ATP binding; residue F126 stacks with the adenine base (*86*). High-throughput genetic identification studies support this, as only two functional variants, which provide adequate hydrophobic residues this position, were found (*67, 87*). *Motif II b* (GG) and
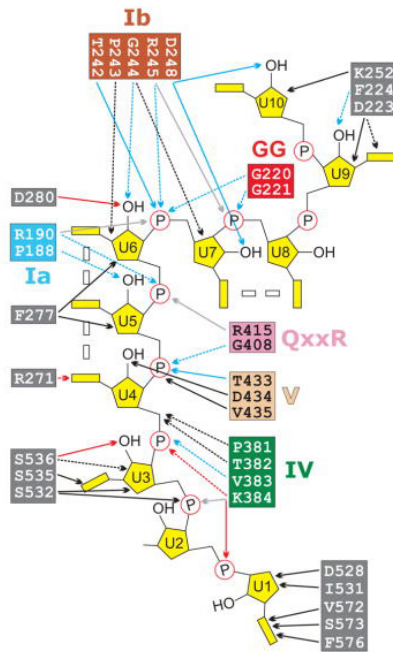
Figure 14: showing contacts between U10 RNA and Mss116p residues. Blue: H-bonds; light-grey: ionic contacts; black: hydrophobic contacts; red: water contacts; solid and clashed lines: side and main-chain interactions; double rectangles: base stacking; blue bases: positive potential, red bases: negative potential. (Del Campo 2009)

motif Ic (TPGRxxD) interact with the bound ssRNA at the bending point in the helicase core (*83*). Motif Ic is thought to further exaggerate RNA bending by H-bonding of the G244 amino group with the 2' oxygen of U6 of the ssRNA oligonucleotide (*83*). *Motif II* (DEAD) and *motif VI* are the main constituents of the ATPase active site and both are highly conserved (*67*). *Motif III* (SAT) is involved in a series of interactions across domain 1 and 2 of the helicase core. Through mutational analysis of the yeast DEAD-box protein Ded1p showed that mutations in motif III inhibited unwinding and weakened ATP-dependent ssRNA binding,

resulting in a decreased ATP hydrolysis (*78*).

The crystal structure of Mss116p shows eight amino acids (R190, T242, R245, D248, D280, K384, R415 and T433) whose side chains directly contact with poly U RNA (Figure 13) (*86*). Surprisingly, only two of the eight showed strong
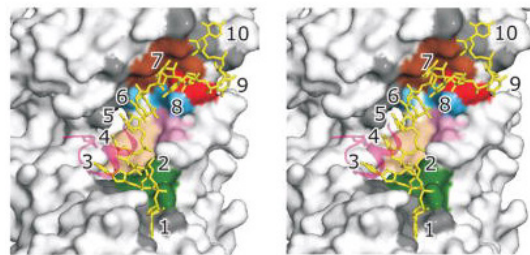
Figure 13: Mss116p residues that contacting U10 RNA. (Del Campo 2009)

conservation in the high-throughput genetic identification (R415 and T433) (*83*). The side chains of the two conserved residues are supposed to be essential for Mss116p function. Two other

residues (R190 and R245) make both side-chain and main-chain interaction with U10 RNA but selected variants (mostly hydrophobic) replacing them indicates that their side-chain contacts are not essential. For R245 a mutant was tested (R245A) showing only main-chain contacts which slowed mutant growth indicating that the side-chain contact enhances Mss116p function. Similar results had been shown for the R190A mutant (*83*). All four residues are very conserved and crystal structures of Vasa and eIF4AIII show the same contacts of these amino acids with ssRNA substrate (*83, 88*).

C-terminal Extension (CTE) and Basic Tail (C-tail)

The crystal structure of Mss116p consisting of the helicase core and a structured CTE (truncated basic C-tail; Mss116p/Δ598-664) showed that the CTE interacts with RNA binding side of domain
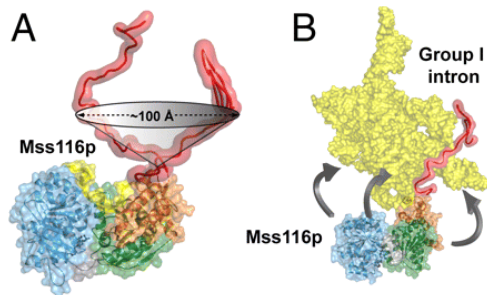
Figure 15: Panel (A): Mss116p tethering range of the C-tail, ~100 Å is the lower limit for space over which the C-tail can bind nucleic acid. Panel (B): Model of Mss116p binding to Tetrahymena group I intron showing that anchored over its C-tail Mss116p is able to act at various over a wide range of the substrate. (Mallam et al. 2011)

2, thereby stabilizing the core and introducing a second bend in the RNA chain acting like a 'crimp' (Figure 15) (*86*). Also, the CTE was found to enhance efficiency of promoting group II intron self-splicing *in vitro* (*6, 38*). Also, comparing the sequence and proteolysis pattern of Mss116p and CYT-19 revealed that Cyt19 has a helicase core, CTE and basic tail similar to Mss116p. For both DEAD-box proteins the C-tail enhances RNA binding

and ability to promote intron splicing (*89, 90*). This interaction stabilizes binding of the helicase core to structured RNAs, giving the protein the ability to sample neighboring dsRNA regions while remaining near substrate. However, the flexibility of the C-tail relative to the helicase core has not been determined for any DEAD-box protein. Mallam et al. used SAXS to obtain structures of Mss116p and CYT-19 in open and closed states. Their results indicate the location of NTE and C-

tail and visualize the conformational changes upon RNA and ATP binding. The SAXS data further revealed a mechanism in which the C-tail is flexibly tethered the core of large RNA substrates (*84*). Notably, the afore mentioned crystal structure of Mss116p lacked a C-tail and the NTE; thus it remained unclear how these substructures contribute to Mss116p function (*84*).

Despite the amount of information present for Mss116p there are unanswered questions on Mss116p functions as group I and group II intron splicing factor. It still has to be addressed which mechanism DEAD-box proteins apply to bind double stranded RNA substrates and to trigger ATP binding and hydrolysis coupled with RNA binding and release. It is as well unknown whether regions on the opposite strand within the helicase core are responsible for strand sequestration, thereby contributing to RNA unwinding and what nature of mechanism causes local unwinding. Furthermore there is need to answer the function of ancillary domains and partner proteins (*83*).

In our study we want to offer additional structural information about Mss116p binding to group II introns and as a model system we chose the autocatalytic ai5γ group II intron.

Group II introns

Group II introns are self-splicing RNAs, meaning that they catalyze their own excision, in most case with the help of proteins *in vivo* (*91*). They can be found in plants, fungi, yeast and bacteria (*92, 93*) where splicing is essential for genes that are involved in metabolism pathways (*94*). Many group II introns do not contain coding regions, though many are mobile elements related to retro-transposons (*95*) and therefore encode a reverse transcriptase designated as maturase. This protein facilitates the insertion of their cognate group II intron into DNA and RNA targets through reverse–splicing (*96*). The open reading frame for such a maturases is located within intron domain 4 of group II introns, like *Lactococcus lactis* LI.LtrB intron (*91, 97*) or the yeast mt introns ai1 and ai2 (*98*). Aside from their reverse transcriptase activity maturase, also called intron-encoded proteins (IEP), show DNA endonuclease activity guaranteeing splicing and intron mobility (*97, 99*). Through their ability to hop into new genomic locations it is believed that group II introns played a major role in the spreading of non-coding RNAs and that they still continue shaping the evolution of host genomes (*100*).

Group II Intron Splicing

The group II intron splicing reaction requires a conserved RNA architecture formed by six domains coming together in space through the formation of long-range tertiary interactions (*101, 102*). Notably, only a few nucleotides are conserved between group II introns and phylogenetic co-variations are rare as well (*103*). This secondary structure organization is shared by the three major group II intron families that are classified
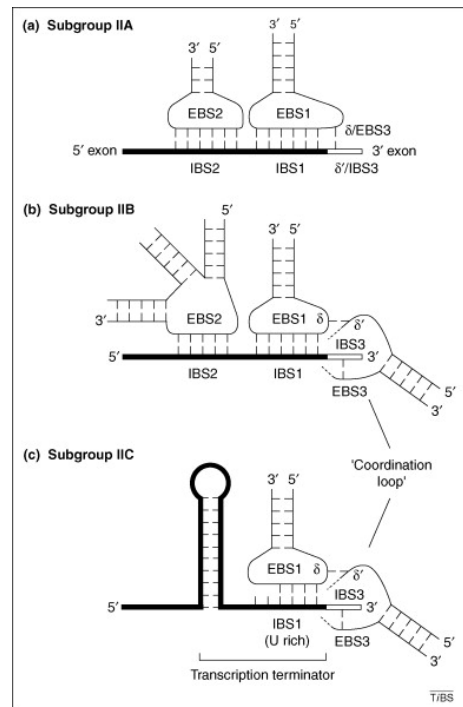
Figure 16: The three classes of group II introns classified according to how exons are bound by the ribozyme. (Michel et al. 2009)

35

according to how exons are bound and positioned (Figure 16) and by their differences in the structural composition of subdomains (*104*).

The six domains with various helical stems contain motifs important for intron assembly and catalysis (Figure 17) (*105*). The catalytic core and architecture of introns are now well-defined and 3D models of IIA and IIB have been created (*106, 107*) and with the crystal structure of a IIC intron recently the active site of a group II intron has been solved (*108*). The mechanism of self-splicing and related reactions by the active sites has been examined extensively and make it an interesting subject for studying RNA folding and catalysis (*94*). Mechanistic
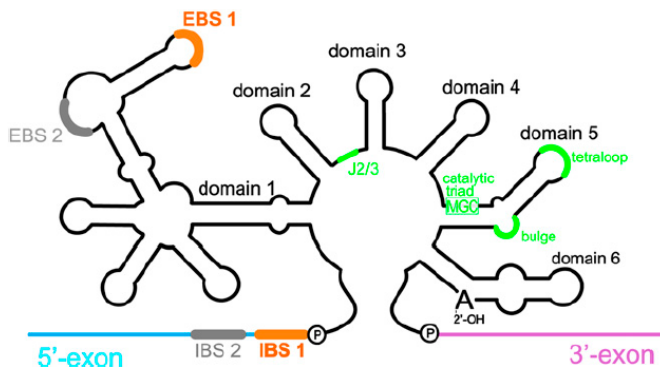
Figure 17: Schematic of group II intron secondary structure with molecular features involved in the active site. EBS: exon-binding sites; IBS: intron-binding sites. Conserved components are highlighted in green. Note that MGC (M= A or C) sequence within the catalytic triad forms a triple helix together with J2/3. (Keating et al. 2010)

parallels with the eukaryotic spliceosome had been observed and led to speculations that group II introns are evolutionary predecessors and simple models of the spliceosomal active site. The ai5γ group II B intron is the fifth intron in the COX1 gene of yeast mitochondria (*109*) and has been widely used as a model system. In the following we describe splicing reactions that have been observed during *in vitro* self-splicing of the ai5γ group II intron (Figure 18).

Branching Pathway of Splicing

The most prominent self-splicing pathway in group II introns is the *branching one* involving a two-step trans-esterification reaction. The first step is carried out by the nucleophilic 2'-OH group of the branch point adenosine located in the intron domain D6 that attacks the 5'splice site. This results in a release of the upstream exon and the formation of an intron-3'-exon intermediate. In the second step of splicing, the 3'-OH group of the 5'exon performs a nucleophilic attack at the 3' splice site thereby releasing the intron in a lariat form and in the ligation of the exons (Panel A). This mechanism is quite similar to the spliceosomal intron splicing and has therefore significantly influenced the development of group II intron reactivity and structure models (*94*).

Figure 18: three types of splicing reactions observed during *in vitro* self-splicing of ai5γ group II intron. Panel A illustrating branch-point splicing; Panel B hydrolytic splicing; Panel C shows spliced-exon reopening. The first step is the rate-limiting step and step 2 happens very fast. (Daniels et al. 1996)
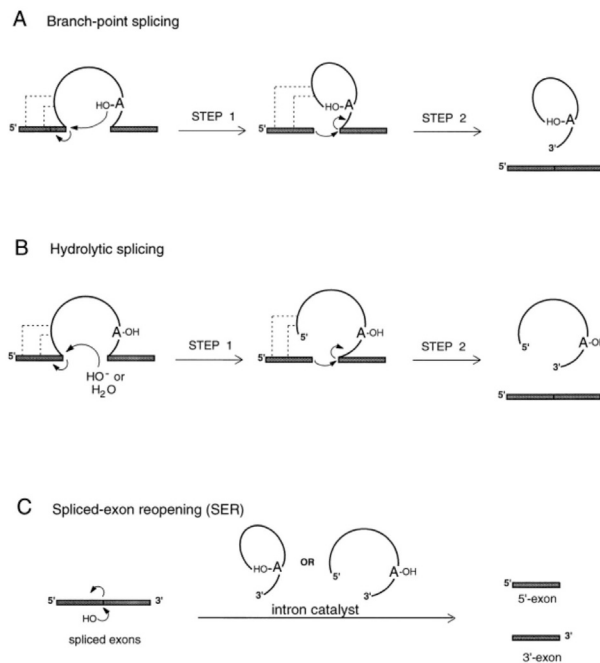
Hydrolytic Pathway of Splicing

In parallel to the branching reaction splicing can occur by water or hydroxide ions acting as a nucleophile, as depicted in Figure 18 B) (*94*). An external 2'-OH group or water molecule herein carries out site-specific phosphodiester hydrolysis reaction at the 5'-splice site, resulting in a released 5'exon, which attacks the 3'splice site in the second step. Eventually this result in ligated exons and the intron is excised in a linear form, which is the only difference to branch-point splicing. A special role for the *hydrolytic pathway* has been suggested *in vivo* inasmuch as it may represent a mechanism for regulating the reactive lariat intron

species, as these are potentially deleterious because of their ability to reverse-splice and reintegrate into other genomic regions (*94, 110, 111*).

Spliced-Exon Reopening (SER)

During *in vitro* splicing of ai5γ free 5' and 3' exons can be observed (Panel C) caused by so called spliced-exon reopening (SER). This reaction had been observed under several different splicing conditions regardless of the branching or hydrolysis pathways. SER can take place immediately after splicing. The two joined exons are hydrolyzed by the intron RNA acting like a ribozyme (*94, 112*). This cleavage reaction was shown to be a mechanistic analog of the first step of reverse splicing (*113*).

Reverse Splicing

Lariat and linear introns are able to perform the reverse of the splicing reaction. Generally it was believed that only lariat introns perform this mechanism but a recent study on the ai5γ group II intron showed that linear introns is able to reverse splice rapidly and with high precision as well (*110, 114*). In addition, it has been observed that reverse splicing by lariat intron is surprisingly unfavorable and that the linear intron was not observed to undergo both steps of reverse splicing suggesting a restricted mobility mechanism (*114*).

Circle Formation

Circle formation during group II intron splicing was first discovered in 1984 by Osiewcaz and Esser. At the beginning it was believed that circles are more likely to be produced than lariat introns (*115*). Interestingly, when RNA circles are observed in a living cell, DNA circles are detectable as well (*116*). In *Podospora anserina* the amount of circular intron DNA had been

found to be increased in senescent cells (*115, 117*). However, it is suggested that circular introns represent a conserved side reaction of group II intron splicing *in vivo* with unknown function.

Structural Group II Intron Domains
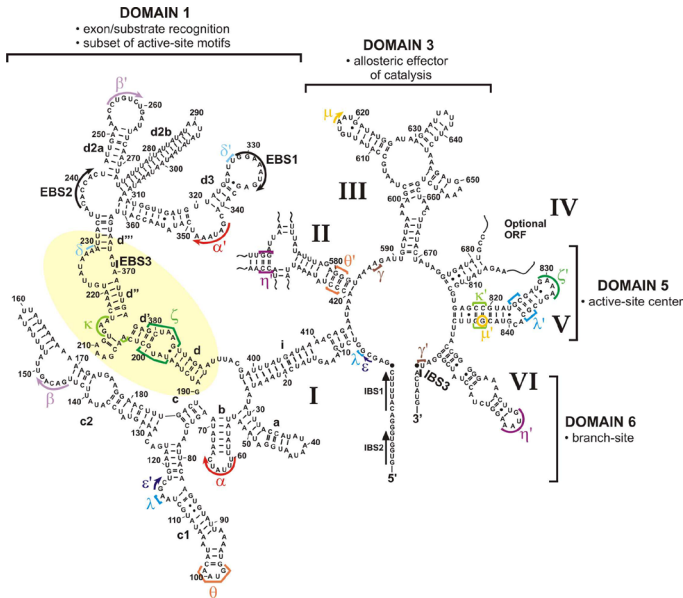
Figure 19: Ai5γ group II intron sequence map. EBS/IBS: exon/intron binding sites; Greek letters indicate long-range tertiary contacts; yellow: folding control element. (Liebeg et al. 2010).

Since their discovery the structure, folding pathway and function of group II introns has been studied biochemically. After ~ 30 years the crystal structure of the *Oceanobacillus iheyensis* (Toor et al 2008), made it possible to obtain profound insights into the group II intron architecture. The group II C intron is ~ 1/3 smaller and has been less extensively studied than the introns belonging to the phylogenetic families IIA and IIB introns but they share the same domain structure (*118*). Despite these differences, numerous of the tertiary contacts identified phylogenetically or biochemically in the IIA or IIB introns were observed in the *Oi.* group IIC intron as well. In brief, due to the presence of most predicted interactions in the crystal structure there is little doubt left that the presented structure is that of a functional molecule (*118*). In the following an overview of the structural organization and the intricate architecture of group II introns is given. While there is little sequence conservation, the six domains that radiate from the central wheel, are structurally highly conserved for group II introns (Figure 19).

Domain 1 (D1)

D1 is by far the largest of all domains (*119*), folds autonomously and serves as a scaffold for docking of the other domains (*120*). Together with D5 it is indispensable for chemical catalysis and represents the minimal structure capable of performing activity (*121*). D1 recognizes the 5'-exon through exonic substrate recognition sites (EBS1/2) and contains elements that recognize the 3'-exon via the δ-δ' interaction (*115, 122*). Furthermore, D1 contains three essential long-range tertiary contacts between D1 and the 'catalytic center' which had been identified by phylogenetic analysis and nucleotide analog interference suppression (NAIS)(*123*). The tertiary interaction ζ-ζ' and κ-κ' had been shown to be important for D5 docking (*103*) and λ-λ' interaction positions D5 in close proximity to the 5'-splice site and is directly involved in catalysis (*123*). Phylogenetic analysis has shown that the interaction ε-ε' is also important for the catalytic function of the intron (*124*). The α-α' contact has been implicated in facilitating binding of 5'-exon and has been shown to play an important role in the folding process as well (*125, 126*). The tetraloop-receptor interaction θ-θ' is suggested to play a role in structural stabilization of the intron core and in recruiting the catalytic effector D3 together with J2/3 to the active site (*127, 128*). In a recent study it was shown that subdomain IC1 within D1 might serve as a first-step-specific RNA receptor for the branchpoint-carrying D6 (*129*).

Domain 2 (D2)

D2 lacks phylogenetic conservation, but has been shown to play an important function in intron activity and tertiary structure stabilization together with D3 (*128*). It forms long-range interactions with D1 and D6 (θ-θ' and η-η', respectively) (*127, 130*). NAIM (nucleotide analog interference mapping) experiments suggest an interaction between D2 stem and D3 that builds on the θ-θ' interaction, suggesting that these two components are key elements of group II intron core and for active-site assembly (*128*).The η-η' interaction is suggested to serve as a switch between the two steps of splicing (*130*). However, up to now not much information about D2 serving as a conformational switch for the second splicing step is available and further experiments are required to provide more information about the functionality of D2 (*115*).

Domain 3 (D3)

D3 is known as the catalytic effector (*131*), however it is not strictly required for catalysis but enhances group II intron ribozyme activity (*128, 132*). D3 is suggested to dock via multiple tertiary interactions and high affinity (*133*). However, not much phylogenetic co-variation between D3 and other domains could be observed, suggesting that the tertiary contacts are not involved in Watson-Crick base pairing or tetraloop-receptor interactions (*115*). The tertiary contact μ-μ' between D3 and D5 has been identified by NAIS experiments and was shown to be important for catalysis (*134*). Nucleotides within the internal bulge are the most conserved residues within D3 and NAIM studies suggest that these are involved in tertiary contacts and might be part of a larger structural motif (*134*).

Domain 4 (D4)

D4 does not contribute to self-splicing but contains the ORF encoding the maturase involved in intron mobility *in vivo* in some group II introns (*121, 135*). It has been suggested that D4 acts as a protein-binding platform due to its location on the surface of the intron (*115*).

Domain 5 (D5)

Domain 5 is the phylogenetically most conserved domain and *trans*-splicing analysis demonstrated that this domain is the most important component for catalysis (*136, 137*). Mutational analysis showed that certain important regions, including ζ-ζ', κ-κ' and λ-λ', are essential for linking Domain 5 to other domains (*131*). As a matter of fact this domain has been most extensively studied and two important themes emerged. First, a set of three nucleotides, so called 'catalytic triad' with the conserved AGC sequence, in the basal stem of D5 together with nucleotides located in the D5 bulge are essential for ribozyme catalysis (*138, 139*). The *Oi.* crystal structure reveals this interaction in more detail, where the J2/3 element is composed of three nucleotides A287, G288 and C289. Nucleotide G and C emerge into the major groove of D5 and forms a stacked triple array to the nucleotides of the catalytic triad located in D5 forming a completely merged triple helix that brings together catalytically crucial residues of the intron (catalytic triplex) (*108*). Second, D5 shares stunning sequence and structure similarity to U6 small nuclear RNA (snRNA) of the spliceosome involved in nuclear pre-mRNA splicing (*140-142*). Therefore D5 represents an important link between group II introns and the spliceosomal machinery (*143*).

Domain 6 (D6)

D6 contains the conserved bulged branch-point adenosine that is essential for the branching pathway of group II intron splicing (*144*). D6 is missing in most crystal structures but comparative

sequence analysis, kinetic analysis, modeling and oligonucleotide anchoring point to subdomain IC1 (ι-ι' interaction) and subdomain II A (η-η' interaction which positions D6 for exon ligation (*130*)) to be the first step RNA receptor to be the successively bound location of D6 and has been shown to be essential for the efficiency of lariat formation (*129*). In the model of Li et al. D6 and subdomain IC1 are optimally tethered over a single nucleotide (*129*) where in another recently proposed arrangement a coordination loop serves as a receptor for branch-point interaction (*145*). The recently published sketch by Pyle et al. displays a model of possible first-step conformation in between the two previous models (*146*) but all three models clearly differ and a major rotation of D6 is necessary after branching step to dock to D2 (second-step receptor) (*129*). It is as well highly interesting to study the atomic environment around the branch point and it has to be kept in mind that tinkering with RNA architecture might alter the ribozymes function and further studies are necessary to find the link between spliceosomal and group II intron splicing.

Metal Ions Involved in the Ai5γ Folding Pathway

Magnesium ions are essential for the large ai5γ ribozyme to fold into the active conformation and for catalysis (*147*). A powerful technique, which provides information about location of metal ion binding sites within folded RNAs, is the metal ion-induced cleavage assay (*147*). In early approaches Pb^{2+} (*148*) or Fe^{2+} ions (*149*) served for probing for divalent ion binding sites, but the main disadvantage is that their geometry differs from Mg^{2+} or these are in need of non-physiological conditions (*147*). As a consequence, lanthanide(III) ions, particularly Tb^{3+} ions have more recently been applied for detecting binding pockets (*147*). The radius of hexa-coordinated Tb^{3+} with 0.92 Å is close to that of Mg^{2+} with 0.72 Å (*150*), and has the same coordination geometry for ligands (*151*).
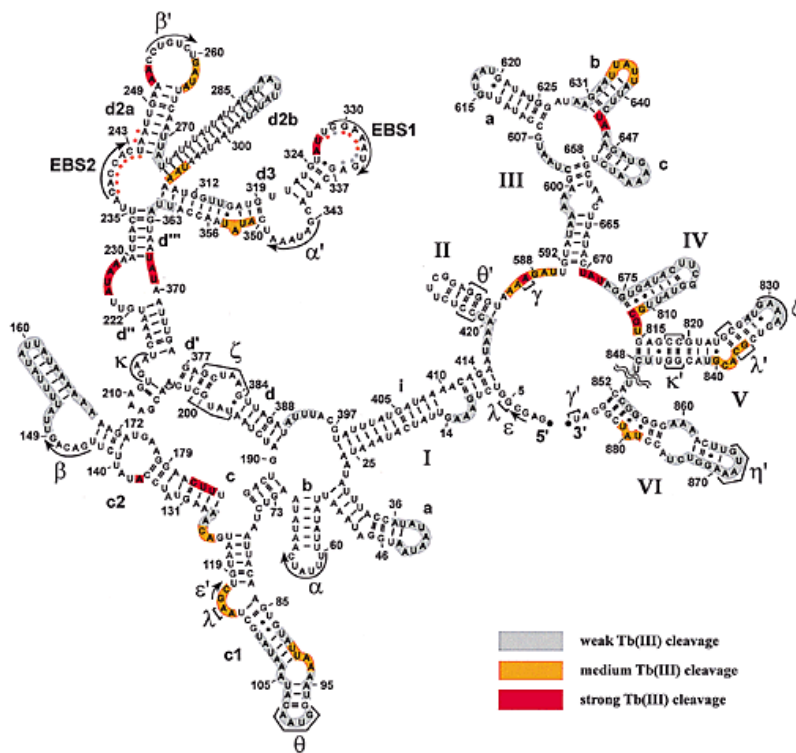
Figure 20: Ai5γ terbium and magnesium cleavage sites superimposed on a secondary structure representation. (Sigel et al. 2000)

Sigel et al. 2000 applied the Tb^{3+}-induced cleavage assay to the ai5γ ribozyme to map the metal ion binding pockets in the intron in its native, active conformation. It has been observed is that most Mg^{2+} binding sites are close to the ribozyme active site in line with the fact that group II introns are metallo-enzymes (Figure 20). In fact, in the crystal structure of the *Oi.* group IIC intron metal ions within the active site have been observed as well (*108*). Two metal ions (M_1 and M_2) are bound by the catalytic triad and the bulge within D5 which results in the formation of a negatively charged pocket (*108*). The crystal structure provides further evidence assigning the two metal ions as catalytic metal ions showing that they reside on the surface of D5, with a distance of 3.9 Å between M1 and M2, which is the ideal distance for metal ion reaction of group I intron and protein phosphate-transferases (*152*) and thus are accessible for both splice sites (5' and 3') (Figure 21) (*108*). Further bindings sites were found close to nucleotides near or within the

EBS1, the Jd''/d''' motif, the λ and ε' interaction site and close to the branch site (*147*). Some prominent cation binding sites were detected close to α-α', β-β' interaction sites, the three-way junction within D1 and close to non-conserved regions in D3 and D4 (*147*). These are likely to play an important structural role, in coordinating the interaction of substructures.

The identified metal ion binding sites correlate well with those regions that have previously been observed to become protected from hydroxyl radical cleavage after folding to the native state (*108*). In other words, close packing of the backbone at these regions due to Mg^{2+} charge neutralization renders these elements inaccessible to the attack by OH radicals. (*153*). Interestingly, metal ion binding pockets near the κ-ζ element could be only detect during ai5γ intron collapse and were not observed in the native state (148). This could be explained by the fact that metal ions during the collapsed state are detectable in the minor groove of ribozyme structure and that Tb^{3+} cleavage assays need a perfect Tb^{3+} location to the 2'-OH ribose which might not be present any more in the native state (*154*). Sigel et al. pointed out a another slight drawback of detecting metal ion binding sites via lanthanides; it is quite likely that additional important binding sites (G-C rich duplexes such as the basal stem of D5 or the tandem GU wobble pairs (*155*)) remain undetected because of RNA hydrolysis caused by metal hydroxides or because they are located in the major groove (*147*).

Figure 21: Schematic of catalytic triplex structure. Hydrogen bonds are shown as dashed yellow lines; Conserved linker element J2/3 (blue) forms base triple interaction with catalytic triad (orange) and stacks beneath the bulge (red) to position the binuclear metal cluster (orange spheres) for splicing reaction. (Toor et al. 2008)

The Ai5γ Group II Intron Folding Pathway *In Vitro*

Despite its size (0.9kb) and multi-domain organization the ai5γ ribozyme appears a smooth and autonomous folding pathway *in vitro*. At high magnesium and monovalent ion concentrations the ai5γ ribozyme folds slowly to the native state, whereby its folding pathway is interspersed by two over on-pathway intermediates (*5, 154, 156*). Kinetic and equilibrium folding studies on the ai5γ D135 and D1356 ribozyme under *in vitro* and near-physiological conditions (Figure 23) provided important insight into the ai5γ folding pathway and showed that transition to the near-native state is direct, autonomously and lacking kinetic traps (*115, 154*).
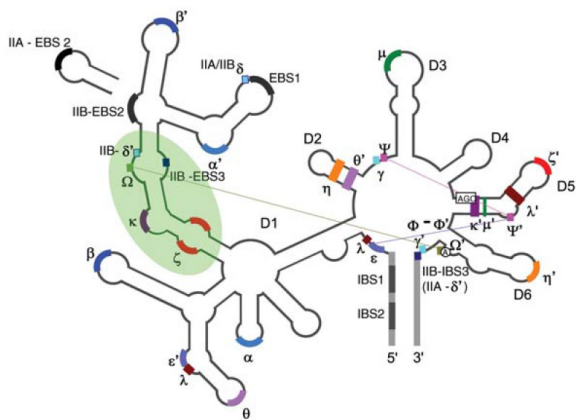
Figure 22: Secondary structure and long-range tertiary contacts within ai5γ intron. Small Greek letters: long-range contacts; Greek capital letters: Proximity contacts detected by UV crosslinking; green: folding control element. (Fedorova and Zingler 2007).

In the first place D1 of the D135 ribozyme collapses slowly to a transient intermediate and forms a highly structured and compact scaffold onto which D5 and D3 rapidly dock into preassembled binding sites (*122*). Very important during the slow collapse is the rate-limiting formation of the κ−ζ element
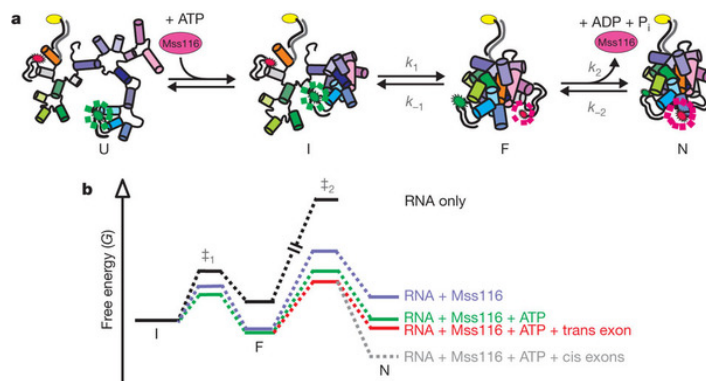
Figure 23: D135 minimal folding pathway. (U) Unfolded; (I) Extended intermediate state; (F) Folded intermediate state; (N) Native conformation. The addition of Mss116p and ATP enhances native intron formation. (Fedorova et. al 2010)

within D1, which offers a docking site for D5 and was found to induce the collapse (*154*). Folding of the κ−ζ element is specifically associated with metal ion binding events that lead to the tertiary

collapse (*5, 126, 154*). The compact substructure is then locked by intra-domain interactions (α–α', β–β'), allowing subsequent docking of D3, D5 and D6, thereby completing intron assembly (*122, 156*). Recent folding single molecule FRET (smFRET) studies on the folding pathway of the D1356 ribozyme under near-physiological conditions revealed a novel intermediate (F) during transition from the extended intermediate (I) to native state (N) (Figure 23)(*157*). It is worth mentioning that studies on the D135 and D1356 constructs lack other domains and the flanking exons, therefore possibly presenting a simplified folding event (*156*). Stabilization of an early folding intermediate by intron-associated proteins is likely to play the most important role in promoting productive group II intron assembly (*105*).

Protein Facilitated Group II Intron Folding

Proteins can facilitate folding of RNAs, like group II introns via several mechanisms. First they facilitate the collapse of large RNA by relieving unfavorable charge-charge interactions when RNA compacts (*105*). Alternatively, proteins interact non-specifically with RNA and are able to unfold misfolded RNAs allowing RNA to refold (*27, 32*). This mechanism is considered as RNA chaperone activity (*27*). Most molecular chaperones , which facilitate folding of proteins, are ATP-dependent (*158*), while RNA chaperones act in an ATP-independent manner facilitating RNA folding by resolving kinetic traps (*48, 159, 160*). Third proteins could stabilize correctly folded RNA in two ways. Either they may bind a substructure of large RNA to form a platform for proper RNA assembly (*161*) or they bind and stabilize the correctly folded form (tertiary-structure capture) (*48, 162*) or over solving misfolded RNA conformations thereby acting like RNA helicases. A very prominent DEAD-box helicase, Mss116p is important for folding of all group I and group II introns *in vivo* (*79*) and facilitates splicing in presence of ADP or ATP (*6*). Mss116p like other helicases (CYT-19 or Ded1) offers a non-conserved C-terminal domain that is positively charged (*6*) suggesting that it could facilitate intron collapse by relieving unfavorable charge-

charge interactions (102). The ai5γ group II intron self-splices under high-salt conditions (~100mM Mg^{2+}, 0.5M monovalent salts and 42°C), but efficient splicing *in vivo* and under near physiological conditions *in vitro* requires Mss116p and ATP (Figure 23) (*82*).

Possible Mechanism of Mss116p on Ai5γ Group II Intron Structure

Mss116p is essential for splicing of all, group I and group II, yeast mitochondrial introns which are nearly all located in the *COX1* and *COB1* genes (one group I intron is found in the large ribosomal RNA) (*38, 79*). Mss116p was observed to reduce the Mg^{2+} concentration necessary for proper intron folding over offering unwinding and annealing activity *in vitro* (*6, 38, 39, 157*). While Mss116p's ATPase activity is essential to promote intron splicing both *in vitro* and in a living cell (*38, 79*), it is not clear whether its unwinding activity is also required for the intron folding process (*40*). Therefore two distinct models for Mss116p facilitating RNA folding have emerged where in the first it might function as a splicing factor by acting in a RNA chaperone-like manner (*6, 40*) and in the second it offers stability to the intron RNA (*38*). Recent DEAD-box protein-facilitated kinetic folding studies showed that Mss116p directly stimulates *Sc.* ai5γ folding by accelerating the collapse to the near-native state (U ↔ I) *in vitro* (*157*). It has been observed that this reaction is highly ATP-dependent, where ATP is required for protein turnover, thereby stabilizing an early folding intermediate (Figure 24, Panel C) (*157*). Interestingly, Mss116p is not required to stabilize the native state of the intron. Unwinding of the long exon sequences of ai5γ *in vitro* is essential for exon unfolding, but not for intron unfolding (*41*) thereby showing unwinding activity as well as acting like a stabilizing cofactor. Mss116p contains conserved helicase motifs which are located on the N-terminal domain and an arginine-rich, positively charged C-terminus which is not shared by all DEAD-box protein (*90*). Cyt19 and Ded1 are able to functionally substitute for Mss116p *in vitro* and *in vivo* (*38, 86*) and carry as well a positively charged C-terminus which might facilitates intron collapse by relieving unfavorable charge-charge interactions (102). DEAD-

box proteins are prominent as helicases and the mechanism appears to be related to RNA chaperones (*154*). However, Mss116p requires ATP as an external energy source to unwind local strand regions (*69, 163-166*). Finding that Mss116p is also able to stabilize RNA structures broadens the known repertoire of function suggesting multiple roles that shape the architecture and folding landscape of RNA complexes.
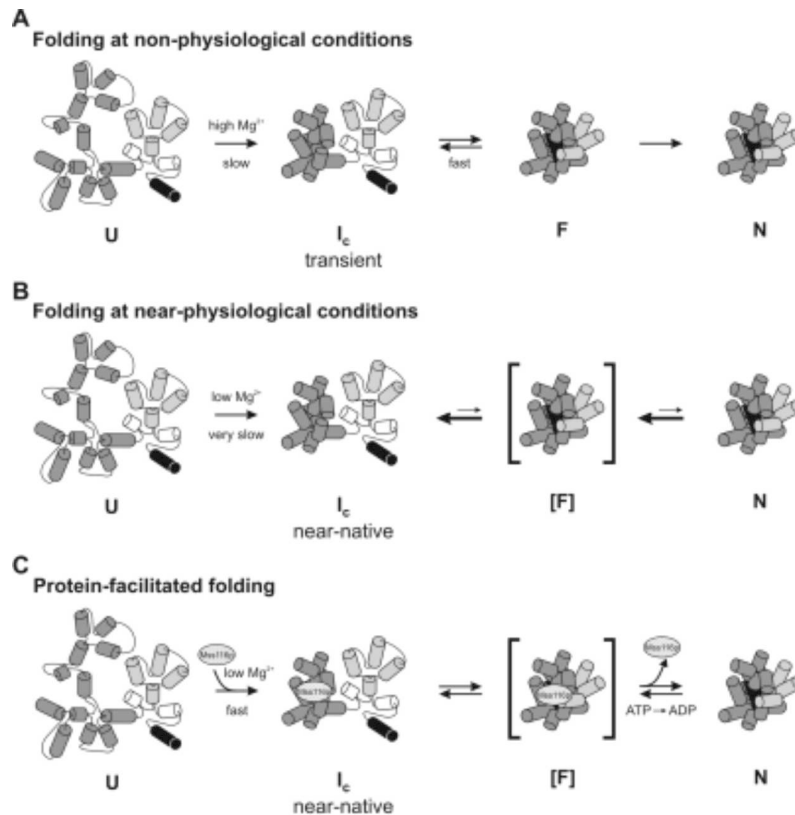
Figure 24: *In vitro* folding of the *Sc.* ai5γ intron. Folding pathways illustrated at (A) non-physiological conditions, (B) near-physiological conditions and (C) in the presence of Mss116p. The cylinders represent intron elements and were sketched: DI (dark gray), DIII (light gray) and DV (black), and DII and DIV (white) and DVI and the exons (omitted). Mss116p is represented as ellipse. The unfolded, compact intermediate, folded and native states are abbreviated as U, Ic, F and N, respectively. The F state has not been studied at near-physiological conditions or in the presence of Mss116p. This conformation is hypothetical model and was therefore put in brackets.

The Scientific Aim of the Project

The aim of the project is to detect metal ion binding sites within the ai5γ group II intron *in vitro* in the presence and absence of the DEAD-box protein Mss116p. As Mss116p significantly lowers magnesium requirements for folding of ai5γ *in vitro* (*6, 38*) and since it does not alter the metal ion binding sites *in vivo* (Wildauer and Waldsich, unpublished), I wanted to monitor whether Mss116p influences the formation of metal ion binding pockets in the in the collapsed and native states of the ai5γ intron *in vitro* at near-physiological conditions. In other words, does Mss116p take over the role of some magnesium ions during compaction of the ai5γ intron? To accomplish this task, recombinant Mss116p had to be purified from *E. coli* and subsequently its functionality has to be confirmed by means of an ATPase assay and an *in vitro* splicing assay. The last milestone is to establish and adapt the powerful technique of a Tb^{3+}-induced cleavage assay to detect magnesium binding sites within the ai5γ intron *in vitro*.

Materials

1. Strains

Competent Cell Kit

The Rosetta 2(DE3)pLysS Singles Competent Cells (Novagen; Cat. No.71401-3) kit uses an *E. coli* strain which is a BL21 derivative and supplies tRNAs for rare codons. This makes the cell kit suitable for eukaryotic protein expression. The cells carry a plasmid with chloramphenicol-resistance and carrying the tRNA genes, which can be induced over T7 lysozyme and regulation of *lac* operon control elements (novagen.com). We used IPTG, a lactose derivate and inducer of the *lac* operon over binding to the *lac* repressor. It cannot be metabolized by *E. coli* and therefore concentration within the cell remains the same. As a result genes underlying the lac operon are expressed and level of expression is regulated over IPTG concentration. Genotype: F^- ompT hsdS$_B$(R$_B^-$ m$_B^-$) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cam^R) (opennetware.org).

DH5α Competent Cells

DH5α cells were made competent over CaCl$_2$ /PIPES method (by Nora Sachsenmaier). In brief cells were grown to an OD600nm of 0.5 – 0.7 with an overnight culture. Afterwards cells were put on ice for 15 min, harvested and resuspended in ½ x Vol. cold 100mM CaCl$_2$ solution and put on ice for 5 min. Cells were again harvested and resuspended in 1/10 x Vol. cold 100mM CaCl$_2$ and put on ice for 90 min. After a spin down cells were resuspended in 1/10 x Vol. ice cold 100mM PIPES and put on ice for 2 min. After another centrifugation cells were resuspended in 1/12.5 x Vol. 100mM CaCl$_2$ / 15% (v/v) glycerol, split in aliquots and immediately stored at -80°C.

This *E. coli* host strain lacks endonuclease A (*endA*) which increases the yield of plasmid DNA produced and lowers the production of nicked plasmid DNA (*167*). Genotype: F^- endA1 glnV44

thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r$_K^-$ m$_K^+$), λ– (opennetware.org).

100mM CaCl$_2$ –PIPES-glycerol pH 7.0

   7.3 g   CaCl$_2$

   1.73 g   PIPES

    75 ml   glycerol

PIPES won´t dissolve until pH is adjusted to 7.0. Adjust pH with NaOH and fill up to 500ml with ddH$_2$0. Autoclave buffer and store at 4°C.

100mM CaCl$_2$

   7.3 g   CaCl$_2$

Fill up to 500ml with ddH$_2$0, autoclave it and store at room temperature.

100mM PIPES pH 7.0

   15.11 g   PIPES

PIPES won´t dissolve until pH is adjusted to 7.0. Adjust pH with NaOH and fill up to 500ml with ddH$_2$0. Autoclave buffer and store at 4°C.

A detailed list of genotype nomenclature and abbreviations can be found at http://openwetware.org/wiki/E._coli_genotypes.

2. Plasmids

pAS02

In this plasmid the wild type (w/t) Mss116p sequence lacking the mitochondrial localization signal
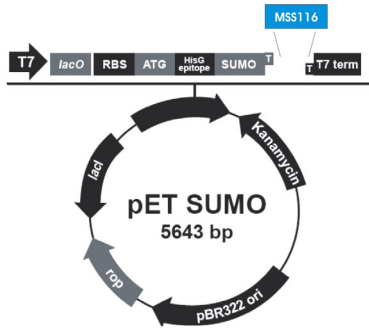
Figure 25: pAS02 plasmid map
The figure shows a plasmid map of the parental vector pET SUMO which carries the full length Mss116p (ΔMLS) sequence (invitrogen.com).

was subcloned into the parental vector pET SUMO TA cloning® site (Invitrogen). pAS02 encodes Mss116p without MLS (Mss116p/ΔMLS, 85.2 kDa). The parental vector carries a T7 promoter which allows IPTG inducible expression of recombinant Mss116p in *E. coli.* After purification via the 6x HIS tag the SUMO protease is able to cleave off the N-terminal tag (SUMO and 6 x HIS tag ~ 13 kDa, see Figure 25), producing untagged, native Mss116p/ΔMLS protein. The plasmid also carries a kanamycin resistance gene. The

plasmid was a generous gift from Anna Marie Pyle (Yale University).

pEL85

The full length group II intron *Sc*. ai5γ D123456 (841 nt) without exons was subcloned into the parental vector pBS⁻ΔT7. This vector has a deleted T7 RNA polymerase promoter region. The plasmid carries an ampicillin resistance gene. We also obtained the plasmid from Anna Marie Pyle (Yale University).
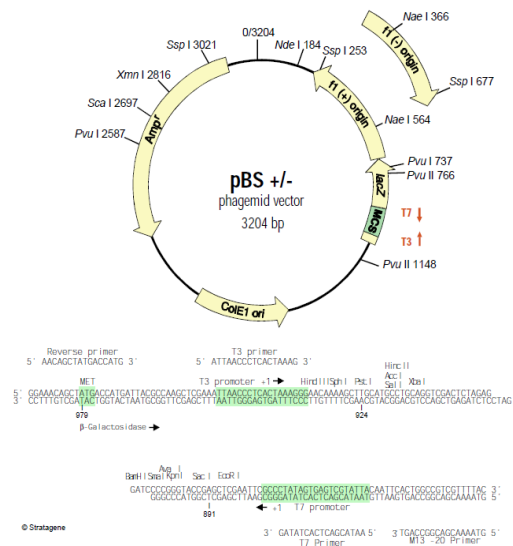
Figure 26: pBluescript II SK- phagemid vector map. The full length ai5γ group II intron sequence was subcloned into pBS⁻ΔT7.

pJD20

In this plasmid the full length group II intron Sc. ai5γ D123456 with part of its flanking exons (1139 nt) was subcloned into the parental vector (Stratagene). The insert is flanked by *SacI* and *HindIII* restriction sites and carries an ampicillin resistance gene (*136*). The plasmid was a kind gift from Philip S. Perlman.

3. Oligonucleotide Specification and Quantification

Overview

Assay	Name	Sequence (5' to 3')	Length (mere)	Tm (°C)
Unwinding assay	TS12	GCCUCGCUGCCG	12	44
	BS12	CGGCAGCGAGGC	12	44
	TS12-cold	GCCTCGCTGCCGAAAAAA	18	56
ATPase Assay	BS12	CGGCAGCGAGGC	12	44
RT primers	Sc 82	TGTTACCATTTTAATACACTTG	22	56
	Sc 160	GTTCCTTCATCTTTTTTTTATAA	23	58
	Sc 234	GTTTTCAATTAGTGGTGTAAG	21	56
	Sc 312	TTCCAATACATAACATCAACC	21	56
	Sc 375	GTAAATATCTAACTTAGCTCTC	22	58
	Sc 440	TTGGTATTATTATTATTTTTTATTATTA	28	60
	Sc 520	TATTTCTTAATCCAATAATTATTTATA	27	60
	Sc 588	GCATTAGCTTTTTATACAATC	21	58
	Sc 656	CACCTATAGTATAAGTTAGCAG	22	58
	Sc 682	ATAATAATAATATATATATAAATAAA GATAGGC	33	53
	Sc 754	ATTTATCAGTTATATATAAACCTCC	25	62
	Sc 805	CGGCTCACGCAATATC	16	58
	Sc 866	CCCGATAGGTAGACCTTTAC	20	60
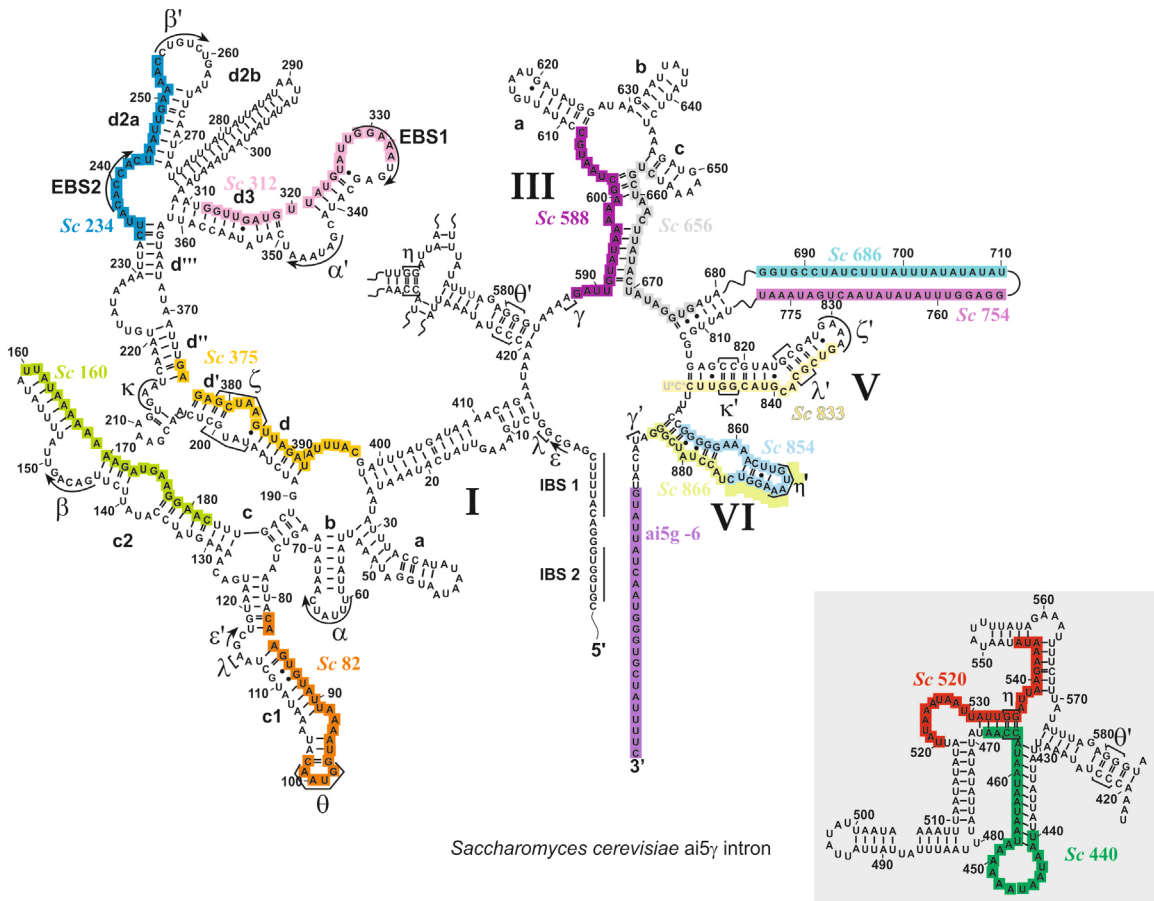
Figure 27: *Sc* primer binding sites in the ai5γ group II intron. The figure shows a 2D structure map of ai5γ intron with its *Sc* primer binding sites highlighted in different colors. Roman numerals: intron domains. Lower case letters: subdomains; greek letters: tertiary intron interactions; wavy lines indicate truncated Domain II which is separately presented in the grey box. EBS: exon binding sites; IBS: intron binding sites.

4. Extinction Coefficients

For *short sequences* we calculated the wavelength dependent extinction coefficient ε [L•mol^{-1}•cm^{-1}] over the following equation (nucleotide extinction coefficients and equation taken from www.operon.com and www.owczarzy.net). The calculation over the below stated equation is one of the most accurate methods to determine ε.

$$\varepsilon_{260} = \sum_{i=1}^{n-1} \varepsilon\, N\, i, i+1 - \sum_{i=2}^{n-1} \varepsilon\, Ii$$

εN … nearest neighbor extinction coefficient of neighboring nucleotides (i,i+1)

εI … is the individual extinction coefficient for each nucleotide (i)

Equation 1: Molar Extinction Coefficient. This formula was used to calculate the molar extinction coefficient at 260 nm of oligonucleotides.

	i = C	i = G	i = A	i = U	i = dC	i = dG	i = dA	i = dT
i + 1 = C	14200	17800	21000	16200	14400	17900	21050	15200
i + 1 = G	17400	21600	25200	21200	17500	21600	25200	20000
i + 1 = A	21000	25000	27400	24000	21100	25000	27400	22800
i + 1 = U	17200	20000	24600	19600	17200	20000	24600	18200
i + 1 = dC	14400	17900	21050	16200	14600	18000	21100	15200
i + 1 = dG	175000	21600	25200	21200	17600	21600	25200	20000
i + 1 = dA	21100	25000	27400	24000	21200	25000	27400	22800
i + 1 = dT	16200	19000	23400	18200	16200	19000	23400	16800

i = C	i = G	i = A	i = U	i = dC	i = dG	i = dA	i = dT
7200	11500	15400	9900	7400	11500	15400	8700

Figure 28: Summary of extinction coefficients at 260nm. The values in the upper table state nearest-neighbor extinction coefficients (i, i+1); the lower table indicates extinction coefficients for each individual nucleotide (i); note that purine nucleotides absorb stronger than pyrimidine nucleotides, table was taken from www.operon.com.

For *longer sequences* we used the following equation which is a mathematical approximation of Equation 1.

$$\varepsilon_{260} = \sum(15{,}346.6 * nA;\ 11{,}751.8 * nG;\ 7{,}625.2 * nC\ ;\ 9{,}929.7 * nU)$$

The table below summarizes extinction coefficients of RNA and DNA oligos used in our experiments.

Name	Molar extinction coefficient [L•mol^{-1}•cm^{-1}]
TS12	69,600
BS12	98,600
TS12-cold	146,400
Sc 82	209,200 [1]
Sc 160	208,200
Sc 234	209,900
Sc 312	206,400
Sc 375	212,900
Sc 440	274,000
Sc 520	269,700
Sc 588	200,800
Sc 656	225,700
Sc 682	369,500
Sc 754	249,100
Sc 833	179,000
Sc 805	148,600
Sc 866	191,900
D123456 RNA	9,569,100[2] / 1,0855,157.5[3]
D123456ex RNA	12,756,600 / XXXX fehlt noch

---

[1] *Sc* oligos were calculated over the online calculator http://biophysics.idtdna.com/cgi-bin/uvCalculator.cgi. The calculator predicts extinction coefficients for single stranded DNAs and is based on Equation 1.

[2] Extinction coefficients for RNAs on the left were calculated with Equation 1 and the help of MatLab 7.12.0.

[3] Extinction coefficients for RNAs on the right were calculated with Equation 2

5. Kits, Devices and Enzyme Overview

Identification	Name	Company	Catalog #
Kits	Rosetta 2(DE3)pLysS Singles Competent Cell Kit	Novagen	71401-3
	PureYield™ Plasmid Midiprep System	Promega	A2492
	MEGAshortscript™ T7 Kit	Ambion	AM1354
	RNA Purification MEGAClear™ Kit	Ambion	AM1908
	Malachite Green Phosphate Detection Kit	R&D Systems	DY996
Devices	Microsep™ Centrifugal Devices (30K)	Pall Life Sciences	OD030C41
	PerfectPro Ni-NTA (Nickel-nitrilotriacetic acid) Agarose Nickel charged resin; max. pressure 2.8 psi (25 ml)	5prime	2400010
	Slide-A-Lyzer Dialysis Casette 10K (orange)	Piercenet	66380
	Storage Phosphor Screen & Cassette Large Screen 35 x 43 cm, for Storm™, mounted GP[4]	G&E Healthcare	63-0034-82
	Image Eraser	Amersham Biosciences	-
	Molecular Dynamics Storm Phosphorimager 820	Molecular Dynamics	-
	U-2000 Spectrophotometer	Hitachi	
	NanoDrop (ND-1000)	Thermo Scientific	-
	Gel Dryer Model GD-5040	SCIE-PLAS	-
	µCLEAR plate, chimney, 96 well – black, clear bottom	Greiner Bio-one	655087
	Infinite F500 microplate reader	Tecan	-
	Illustra™ Microspin™ G-25 Columns	G&E Healthcare	27-5325-01
Enzymes	AMV Reverse Transcriptase	Promega	M5108
	SUMO protease	Invitrogen	12588-018
	T4 Polynucleotide Kinase	NEB	M0201S

---

[4] GP: General purpose screen: for ^{125}I, ^{32}P, ^{33}P, ^{35}S, and ^{14}C
(http://www.gelifesciences.com/aptrix/upp01077.nsf/content/Products?OpenDocument&parentid=957136&moduleid=163735)

Identification	Name	Company	Catalog #
Software	Image Quant™ Version 7.0 and ImageQuant TL SecurITy 8.0 Software Package	G&E Healthcare	28-9380-94
	Molecular Dynamics Storm Scanner Control Version 5.01	Molecular Dynamics	-
	MATLAB 7.12.0; License # 165049	Mathworks®	-
	PyMOL 1.1r1	PyMOL	freeware

Methods

1. Producing Recombinant Mss116p

1.1. Transformation of pAS02 into Competent *E. coli*

The transformation protocol is a modified version of the Novagen Manual User Protocol TB009 Rev. F0104. We transformed 1 µl of pAS02 (c = 44.5 ng•µl^{-1}) into 40 µl Rosetta 2(DE3)pLysS Singles Competent Cells (Novagen). For a negative control we used 10 µl of competent cells.

- Cells were put on ice for 5 min, followed by a heat shock for 30 sec at 42 °C.

- Then 250 µl of SOC medium (delivered with competent cell kit) was added and cells were incubated for 60 min at 37 °C with vigorous shaking.

- The cells were centrifuged for 2 min, 6,000 rpm and supernatant was removed, except for ~200 µl where the cells were resuspended carefully.

- We plated 130 µl of cells on LB KAN plates, put plates at 37°C for overnight incubation and finally stored them at 4°C.

For transformation the following reagents were used:

LB KAN plates

10 g	Bacto tryptone
5 g	Yeast extract
5 g	NaCl
15 g	Agar

Fill up to 1 L with ddH$_2$O and autoclave the medium. After stirring and cooling media down to about 65°C, plates could be poured under a Lamina Flow Hood. Plates were stored at 4°C. Kanamycin (50 µl kanamycin [50 mg·ml^{-1}] plus 50 µl ddH$_2$O) was added directly on the plate before usage.

1.2. Expression of Mss116p in *E. coli*

Overnight (o/n) Culture

As the transformation plates were stored at 4°C a preparative culture is recommended before starting high level protein expression of Mss116p in *E. coli Rosetta.* Therefore 5 ml LB medium, 5 µl kanamycin [50 mg·ml^{-1}] and a single colony of the transformation plate was prepared. The inoculated culture was then incubated on a rotary shaker (200 rpm) at 37°C for o/n (about 12 – 16 hours), allowing the cells to reach the stationary phase.

Growth Cultures

For high level Mss116p expression in *E. coli Rosetta* we set up three large growth cultures each consisting of 600 ml LB medium, 600 µl kanamycin [50 mg·ml^{-1}] and 300 µl of o/n culture. The cells were incubated at 37°C with vigorous shaking to an optical density at 600 nm (OD$_{600}$) of 0.4 – 0.6. The OD$_{600}$ was measured using a Hitachi U-2000 Spectrophotometer.

IPTG-induced Protein Expression

After growing bacterial cultures to the appropriate optical density (OD$_{600}$ 0.4 – 0.6) a 1 ml sample was withdrawn and centrifuged for 2 min at 8.500 rpm. The pellet was resuspended in 100 µl SDS PAGE sample buffer and stored at -20°C for subsequent polyacrylamide gel analysis. At this point protein expression was chemically induced with fresh 0.5mM IPTG. We incubated the bacterial culture overnight at 18°C with vigorous shaking. The next day a 1 ml sample was taken and again centrifuged for 2 min at 8.500 rpm. The pellet was resuspended in 100 µl SDS PAGE sample buffer and stored at -20°C for subsequent polyacrylamide gel analysis.

For expression of Mss116p in *E. coli* the following reagents were used:

LB Medium

           10 g   Bacto tryptone

            5 g   Yeast extract

            5 g   NaCl

Fill up to 1 L with ddH_2O and autoclave it.

1M IPTG

          238 mg IPTG

Fill up to 1 ml with ddH_2O and store at -20°C.

1.3. Recombinant Protein Purification

Preparation of *E. coli* Lysate Under Native Conditions

We harvested the cells by centrifugation at 6.500 rpm for 20 min at 4°C and the pellets were frozen for 30 min at -80°C. After thawing the pellets for 15 min on ice, the cells were resuspended in a total of 20 ml Lysis Buffer. 1 ml lysozyme (10 mg ml^{-1}) was added and the sample was incubated on ice for 30 min. We further sonicated cells on ice with 6 x 10sec bursts and cooling samples on ice for 10 sec inbetween. The cells were harvested at 10.000 g for 30 min at 4 °C. We transferred the supernatant (SN) into a fresh 50ml falcon tube. The cell pellet (CP) was slightly resuspended in 5 ml lysis buffer. 20 µl samples from SN as well as from the CP were withdrawn and 80 µl SDS PAGE sample buffer was added. The samples were then analyzed on a 10% SDS PAGE protein gel.

Batch Purification of 6 x HIS Tagged Mss116p from *E. coli* Under Native Conditions

The PerfectPro Ni-NTA Agarose (5prime) was prepared by washing the resin two times with about 15 ml lysis buffer. We added 4 ml Ni-NTA agarose to cleared supernatant (1 ml to 4ml lysate) and mixed gently for 1 h on a rotary shaker at 4 °C. The lysate-agarose mixture was then loaded onto a disposable polypropylene column (BioRad) with 20 ml capacity. The following purification steps were used to purify recombinant Mss116p:

- Prior to purification 4ml of the flow through were collected for analysis

- Afterwards, the resin was washed 4 x with 4 ml Wash Buffer I

- Followed by 4 x washing with 4ml Wash Buffer II

- Recombinant protein was then eluted 4 x with 4ml Elution Buffer

From each purification fraction a 10 µl sample was taken and 10 µl SDS PAGE sample buffer added. The samples were analyzed on a 10% SDS PAGE protein gel.

For recombinant protein purification the following reagents were used:

0.5M Tris·HCl pH 6.8

     15.1 g   Tris

Adjust pH with HCl. Afterwards fill up to 250 ml with ddH$_2$O and autoclave it.

3M KOH

     8.4 g   KOH

Resuspend in 50 ml ddH$_2$O.

1M Imidazole

     6.8 g   Imidazole

Resuspend in 100 ml ddH$_2$O and store at 4°C.

3M KCl

      111.4 g   Potassium chloride

Resuspend in 500 ml ddH$_2$O and autoclave it.

Lysis Buffer

      10 ml   0.5M Tris·HCl pH 6.8   [f. c. 50mM]

      10 ml   3M KCl   [f. c. 300mM]

       1 ml   1M Imidazole   [f. c. 10mM]

Adjust pH to 8.0 with KOH, fill up to 100 ml with ddH$_2$O and store at 4°C.

Wash Buffer I

      10 ml   0.5M Tris·HCl pH 6.8   [f. c. 50mM]

      10 ml   3M KCl   [f. c. 300mM]

       5 ml   1M Imidazole   [f. c. 50mM]

Adjust pH to 8.0 with KOH, fill up to 100 ml with ddH$_2$O and store at 4°C.

Wash buffer II

      10 ml   0.5M Tris·HCl pH 6.8   [f. c. 50mM]

      10 ml   3M KCl   [f. c. 300mM]

      10 ml   1M Imidazole   [f. c. 100mM]

Adjust pH to 8.0 with KOH, fill up to 100 ml with ddH$_2$O and store at 4°C.

250mM Elution buffer

10	ml	0.5M Tris·HCl pH 6.8	[f. c. 50mM]
10	ml	3M KCl	[f. c. 300mM]
25	ml	1M Imidazole	[f. c. 250mM]

Adjust pH to 8.0 with KOH, fill up to 100 ml with ddH$_2$O and store at 4°C.

1.4. Tag Cleavage, Concentration and Dialysis Against Storage Buffer

Removal of the Tag Using the SUMO Protease

The peak fractions from our batch purification were unified and protein concentration was measured by means of a Bradford Assay. To increase protein solubility after cleavage (*85*) we added[5]

4	µl	1M DTT	[f. c. 1mM]
8	µl	0.5M EDTA pH 8.0 (Ambion)	[f. c. 1mM]
400	µl	0.5M L-Arginine	[f. c. 50mM]
400	µl	0.5M L-Glutamine	[f. c. 50mM]
400	µl	glycerol	[f. c. 10%]

to the protein pool. For cleavage of the tag we prepared the following reaction:

4 ml	Mss116p pool [~0.5 mg·ml^{-1}]
100 µl	+ salt buffer 10 x (Invitrogen)
100 µl	- salt buffer 10 x (Invitrogen)
20 µl	SUMO protease (Invitrogen) [20 units]

---

[5] Addition of glycerol as well as the amino acids arginine and glutamine can increase solubility of cleaved protein from < 1mg ml^{-1} up to 10 mg ml^{-1}. The SUMO protease is a cysteinyl protease (Ulp1) and therefore EDTA presence has no inhibiting effect to the enzyme.

The cleavage reaction was incubated at 4°C overnight. With a spin-down for 10 min at 11.000 rpm at 4°C we got rid of any precipitate. A 20 µl sample was withdrawn and mixed with 20 µl SDS PAGE sample buffer for analysis on a 10% SDS PAGE protein gel. Protein concentration was measured via a Bradford Assay.

Concentration of Mss116p and Dialysis Against Storage Buffer

Concentration

To reach a higher protein concentration and to get rid of the cleaved SUMO tag (~13kDa) we used a Microsep Centrifugal Device 30 K (Pall Life Sciences). The device offers a cut off for molecules smaller than 30 kDa. We centrifuged for 2 hours and 7.000 rcf at 4°C to reach a concentration of ~ 10 µM. Protein concentration was determined by a Bradford Assay. A 20 µl sample was withdrawn and mixed with 20 µl SDS PAGE sample buffer for analysis on a 10% SDS PAGE protein gel.

Dialysis

We used Slide-A-Lyzer Dialysis Casettes (10K) for dialyzing protein against its storage buffer. 1 liter protein storage buffer contains:

50	ml	1M	Tris·HCl pH 7.5	[f. c. 50mM]
1	ml	1M	DTT	[f. c. 1mM]
100	ml	3M	KCl	[f. c. 300mM]
100	ml		glycerol	[f. c. 10%]

The concentrated protein samples were loaded into the cassette with a syringe as described in the manual of the manufacturer (see www.piercenet.com) and dialyzed at 4°C overnight with one time buffer change.

On the next day, the protein sample was collected from the cassette with a syringe. We prepared 20 µl aliquots and after shock freezing them in liquid nitrogen the samples were stored at -80°C.

1M Tris·HCl pH 7.5

     30.3 g   Tris

Adjust pH with HCl. Afterwards fill up to 250 ml with ddH$_2$O and autoclave it.

1M DTT

     154 mg   DTT

Add 1ml ddH$_2$O, resuspend and store at -20°C.

0.5M L-Arginine

     4.3   g   L-Arginine (Sigma)

Fill up to 50 ml with ddH$_2$O and store at room temperature.

0.5M L-Glutamine

     3.6   g   L-Arginine (Sigma)

Fill up to 50 ml with ddH$_2$O and store at room temperature.

3M KCl

> 111.4   g   Potassium chloride

Resuspend in 500 ml ddH$_2$O and autoclave it.

1.5.   Bradford Assay

The Bradford Assay was used to determine Mss116p protein concentration. We prepared a 6 point standard curve from a 100 µg·ml^{-1} BSA stock. A dilution series of the protein was prepared with 1, 2, 3 and 4 µl of protein in 20 µl ddH$_2$O. The dilutions were added into the µCLEAR 96-well plate to 200 µl of Bradford Reagent (AppliChem). After incubating samples for ~20 minutes at room temperature the absorbance at 595nm (A$_{595}$) was measured with the Infinite F500 microplate reader (Tecan).

µCLEAR Plate Assay - BSA Standard Dilution			
BSA standard	BSA concentration [ng·µl^{-1}]	ddH$_2$O [µl]	BSA of stock [µl]
1	0	20	-
2	0.9	18	2
3	1.81	16	4
4	2.72	14	6
5	4.54	10	10
6	6.18	5	15
7	9.09	-	20

The microplate reader out-put was processed in Excel and protein concentration was calculated by comparing OD$_{595nm}$ of the protein sample to the BSA straight calibration line.

1.6.   SDS-PAGE 10% Protein Gels

Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) is a technique to separate proteins of different molecular weight. SDS is a detergent and reduces disulfide bonds which break protein tertiary structures. Heating up samples to 97°C further denatures proteins and allows coating by SDS which results in an egg-shaped conformation and an increased negative charge proportional to protein size. Smaller proteins therefore migrate much faster through the acrylamide network than larger ones. Proteins of different sizes can then be identified over Coomassie staining. Gels for protein electrophoresis were prepared as described in the following chapters.

1.6.1.   Preparation of Gels

The amount of gel prepared was enough for two gels (0.75mm). First the separating gel was prepared and poured into a Dual Gel Caster (SE245, Hoefer). The gel was covered on top with a thin layer of Ethanol (to smoothen the surface) and allowed to polymerize (by adding 100 µl 10% APS and 10 µl TEMED). Afterwards the stacking gel was poured and a comb inserted. When stacking gel was polymerized the comb was removed and gels were installed on a Mini-vertical gel electrophoresis unit (SE250, Hoefer) in 1 x GTS buffer. Protein samples were boiled at 97°C for 5 minutes before loading into slots. We run gels with 140 Volts for about 1 hour and 20 minutes. For all samples the below mentioned SDS-PAGE sample buffer was used.

Separating gel

Reagent	10% SDS-PAGE
Water	4.87 ml
1.5M T Tris·HCl pH 8.8	2.5 ml
10% SDS	100 µl
40% Acrylamide/bisacrylamide	2.48 ml

Stacking Gel

Reagent	10% SDS-PAGE
Water	3.83 ml
1.5M T Tris·HCl pH 6.8	625 µl
10% SDS	50 µl
40% Acrylamide/bisacrylamide	487.5 µl

The reagents listed below were used for preparing the gels.

1.5M Tris·HCl pH 8.8

    45.4 g   Tris

Adjust pH with HCl. Afterwards fill up to 250 ml with ddH$_2$O and autoclave it.

0.5M Tris·HCl pH 6.8

    15.1 g   Tris

Adjust pH with HCl. Afterwards fill up to 250 ml with ddH$_2$O and autoclave it.

10% SDS

      5 g   SDS

Dissolve carefully in 50 ml ddH$_2$O.

10% APS

     1.5 g   APS

Suspend in 15 ml ddH$_2$O and store at 4°C protected from light.

10 x GTS

     60 g   TRIS base   [f. c. 250mM]

    288 g   glycine   [f. c. 1.92M]

     20 g   SDS   [f. c. 1%]

Add ddH$_2$O up to 2 L and filter but do not autoclave.

SDS-PAGE sample buffer

     10 ml   0.5M Tris·HCl pH 6.8   [f. c. 0.1M]

     10 ml   glycerol   [f. c. 20%]

     10 ml   10% SDS   [f. c. 2%]

     10 mg Bromphenol blue   [f. c. 0.02%]

Fill up to 50ml with ddH$_2$O and add ß-Mercaptoethanol always fresh before use [f. c. 5%].

1.6.2. Coomassie Staining of Protein Gels

After gel run, the Mini-vertical gel electrophoresis unit was disassembled, the stacking gel was cut off and separating gel was stained with about 10 ml staining solution with gentle shaking on a shaker platform for 10 minutes. Protein bands were visualized by subsequent de-staining.

Staining Solution

125	ml	Methanol	[f. c. 50%]
25	ml	Acetic acid	[f. c. 10%]
625	mg	Coomassie Brilliant Blue R250	[f. c. 0.25%]

Fill up to 250ml with ddH$_2$O.

De-staining Solution

50	ml	Methanol	[f. c. 5%]
75	ml	Acetic acid	[f. c. 7.5%]

Fill up to 1 L with ddH$_2$O.

2. Preparing *In Vitro* Transcribed RNA

2.1. Plasmid Purification and Preparation

In order to produce plasmid DNA the following transformation was set up:

Transformation of pEL85 and pJD20

For plasmid preparation we transformed ~ 50 ng·µl^{-1} of plasmid DNA into DH5α competent cells.

A typical transformation protocol was performed as follows:

- We used 50 µl DH5α competent cells and added 1 µl of plasmid DNA (c ~ 50 ng·µl^{-1}).

- For the negative control we used 50 µl DH5α competent cells.

- The tubes were carefully flipped and incubated on ice for 30 min.

- We performed a 30 sec heat shock at 42°C.

- After placing cells immediately on ice for 5 min we added 950 µl LB medium and incubated the cells at 37°C for 1 hour with vigorous shaking.

- After a short spin down for 2 min at 5.000 rpm, the cells were resuspended in 100 µl medium and plated on selective LB plates.

- The plates were incubated at 37°C overnight and afterwards stored at 4°C until usage.

Preparative Cultures

We prepared 5 ml LB and added one colony of the transformation plate. 5 µl of a 1000x stock of antibiotics (1:1000) were added and culture was incubated at 37° C with vigorous shaking for about 5 hours. This culture was then added to 250 ml LB containing 250 µl 1000x antibiotic and afterwards incubated for 17-21 hours at 37°C shaking vigorously.

Plasmid Midiprep

We used a PureYield™ Plasmid Midiprep System (Promega) to isolate plasmid DNA of preparative culture. DNA purification was performed as described in the PureYield™ Plasmid Midiprep System Technical Manual (Quick Protocol). DNA concentration was measured using the NanoDrop device.

Linearizing the Plasmid

To linearize the plasmid we performed a restriction digest with a single cutting enzyme. We cut the pEL85 construct with *BamHI* (NEB; 20,000 units·ml^{-1}) and pJD20 with *HindIII* (NEB; 20,000 units•ml^{-1}). A typical restriction digest was prepared as follows:

100	µg	plasmid DNA
50	µl	10x NEBuffer 2
5	µl	100x BSA (NEB)
3	µl	restriction enzyme (20 U/µl)

Fill up to 500 µl with ddH$_2$O and incubate the digest on 37°C overnight.

Purification of the Linearized Vector

After the overnight digest the linearized plasmid was purified via Phenol Chloroform Isoamylalcohol (PCI) extraction. For this we added

- 500 µl PCI (250 µl Phenol/250 µl CI) vortexed well and centrifuged for 5 min at 13,000 rpm.

- We transferred the upper phase in a new 1.5ml Eppendorf tube and added 500 µl CI followed by vortexing and centrifuging for 3 min at 13,000 rpm.

- The upper phase was again carefully transferred into a new 1.5 ml Eppendorf tube followed by precipitation with 2.5 x EtOH/0.3M NaOAc at -20°C for 1 hour.

- After precipitation we centrifuged at full speed at 4°C for 30 min and washed the pellet with 70% Ethanol.

- After letting the pellet dry we resuspended it in 50 µl ddH$_2$O and measured DNA concentration using a NanoDrop.

LB Medium and selective plates

10	g	Bacto tryptone
5	g	Yeast extract
5	g	NaCl
15	g	Agar (for LB plates)

For LB medium we filled up to 1 L with ddH$_2$O and autoclaved.

For LB plates we stirred medium and let it cool down to a temperature of about 65°C. Then the antibiotic was added and plates were poured under a Lamina Flow Hood. Plates were stored at 4°C.

2.2. *In Vitro* T7 Transcription

For pEL85 and pJD20 a typical 1 ml transcription reaction was prepared as described below:

5	µl	linearized plasmid DNA [c ~ 25 µg]
100	µl	10 x TRA buffer
10	µl	1 M DTT
20	µl	ATP (100mM)
20	µl	UTP (100mM)
20	µl	GTP (100mM)
20	µl	CTP (100mM)
5	µl	rRNasin (Promega; 40 u·µl^{-1})
40	µl	T7 RNA polymerase (lab made)
760	µl	ddH$_2$O

The reaction was assembled on ice and incubated for 3-4 hours at 37°C. Finally the RNA was precipitated by adding 50 µl 0.5M EDTA pH 8.0 (Ambion) and 3 ml 2x vol. EtOH/0.3M NaOAc and freezing it at -20°C for 1 hour. Afterwards we centrifuged at 4°C for 30 min at maximum speed, discarded the supernatant, air-dried the pellet and resuspended the RNA in 30 µl ddH$_2$O and 100 µl 7 M urea loading dye. The RNA was then purified using a 5% denaturing polyacrylamide gel.

1M Spermidine

    254.6 mg Spermidine (AppliChem)

Fill up to 1ml with ddH$_2$O and store at -20°C.

10 x Transcription buffer (TRA)

400	µl	1M Tris·HCl pH 7.5 (AppliChem)	[f. c. 400mM]
260	µl	1M MgCl$_2$ (Ambion)	[f. c. 260mM]
30	µl	1M Spermidine	[f. c. 30mM]

Fill up to 1ml with ddH$_2$O and store at -20°C

100mM Nucleotide (AppliChem) Preparation

We prepared nucleotides the following way: we made 1 ml of an approximately 200mM solution. For this we took dry NTP, in the below listed amount, added 20 µl 1M Tris•HCl pH 7.5 (AppliChem) and approximately 50-90 µl freshly prepared 5M NaOH to adjust the pH to 7. We took the below stated amounts of NTP, NaOH and water for the stock solutions.

	GTP	CTP	UTP	ATP	
NTP	128	134	140	130	mg
5 M NaOH	48	84	48	88	µl
ddH$_2$O	932	896	932	892	µl

The pH was checked using pH-indicator strips (Merck) and the concentration measured using a NanoDrop device. The NTP stock solutions were stored at -80°C. For 100mM NTP solutions we diluted the stock to a final concentration of 100mM and stored them at -20°C.

5M NaOH

    10 g   Sodium Hydroxide Pellets

Fill up with ddH$_2$O to 50 ml.

1M Dithiothreitol (DTT)

    154 mg   Dithiothreitol

Add 1 ml ddH$_2$O and store at -20°C.

10 x TBE

    540 g   Tris base   [f. c. 890mM]

    275 g   Boric acid   [f. c. 890mM]

    29.2 g   EDTA   [f. c. 20mM]

Fill up to 5 L with ddH$_2$O, stir and autoclave it.

7M Urea Loading Dye

    6.3 g   UREA   [f. c. 7M]

    3.75 ml   D(+)-Sucrose   [f. c. 25%]

    37 mg   Bromphenol Blue [f. c. 0.25%]

    37 mg   Xylene Cyanol   [f. c. 0.25%]

    1.5 ml 10x TBE

Fill up to 15 ml with H$_2$O and store at room temperature.

2.3. RNA Purification via a 5% Denaturing Polyacrylamide Gel

Size-exclusion chromatography was used to purify the reverse transcribed RNA. We prepared a 5 % Polyacrylamide (PAA) gel. For this we took 60 ml of 5% AA, added 500 µl APS and 50 µl TEMED. The gel was mixed briefly, poured and allowed to polymerize. We pre-heated the gel by running it for 20 min with 25 W. With a syringe and 1 x TBE buffer we flushed away diffused urea from the wells. The RNA samples were loaded and the gel was run by 25 W for ~ 2 hours, until the Xylene Cyanol dye reached the bottom of the gel. After removing gel from both of the glass plates and covering it with saran wrap we cut bands under UV light and transferred the gel pieces into 2 ml tubes. We added 1 ml elution buffer and incubated the samples at -20°C overnight. The next day we let the samples sit for 10 min at room temperature and incubated them afterwards on a rotating wheel at 4°C for 3-4 hours. Samples were centrifuged shortly and the supernatant was transferred to a 5 ml syringe (Omnifix) connected to a sterile syringe filter (0.2 µm). The filter was rinsed by adding another 200 µl of elution buffer to the syringe. We split the filtrate in two 2 ml tubes and added 4 µl glycogen [10 mg·ml^{-1}] and 600 µl 100% Ethanol. We precipitated at -20°C overnight. After centrifugation for 30 – 60 minutes at 4°C full speed we air-dried the pellet and resuspended the RNA in 20 µl ME buffer. The RNA concentration was measured over NanoDrop device, RNA quantity was calculated as described and the quality of RNA was checked on a 1% agarose gel as described below.

5% Acrylamide (AA)

62.5	ml	40 % Acrylamide-Solution Mix 19:1
210	g	Urea
50	ml	10 x TBE

Fill up to 500 ml with ddH$_2$O, stir, filter and store at 4°C.

10% APS

     1 g   APS   [f. c. 10%]

Add 10 ml ddH$_2$O and store at 4°C protected from light.

1M MOPS buffer

     52.3 g   MOPS

Add ~100 ml ddH$_2$O and adjust pH with NaOH. Then fill up to 250 ml with ddH$_2$O, filter over

Nalgene filter units and store buffer at 4°C protected from light.

Elution buffer

     500 µl   1M MOPS pH 7.0   [f. c. 10mM]

      4.2 ml   3M NaOAc pH 5.0 (Ambion)   [f. c. 250mM]

     200 µl   0.5M EDTA pH 8.0 (Ambion)   [f. c. 2mM]

Complete with ddH$_2$O to 50 ml.

ME buffer

     500 µl   1M MOPS pH 6.0   [f. c. 10mM]

     100 µl   0.5M EDTA pH 8.0 (Ambion)   [f. c. 1mM]

Complete with ddH$_2$O to 50 ml.

Glycogen [10 µg·µl^{-1}]

     10 mg glycogen

Add 1ml ddH$_2$O, filter through a syringe connected to a filter and store at -20°C.

10 x TBE

540	g	Tris base	[f. c. 890mM]
275	g	Boric acid	[f. c. 890mM]
29.2	g	EDTA	[f. c. 20mM]

Fill up to 5 L with ddH$_2$O, stir and autoclave it.

2.4. Determination of RNA Quantity and Quality

RNA Quantity

The final RNA concentration [μM] was calculated using a modified Beer-Lambert equation:

$$[μM] = \frac{(A_{260} * dilution * 10^6)}{\varepsilon}$$

ε ... *wavelength dependent extinction coefficient [L·mol⁻¹·cm⁻¹]*

A_{260} ... *absorbance at 260 nm*

The wavelength dependent extinction coefficients are listed in section Oligonucleotide Specification and Quantification.

Visualizing Presence of RNA

To visualize presence of RNA, a small sample (0.2 μg) was loaded on a 1% agarose gel. RNA concentration was measured using a NanoDrop device and 0.2 μg of RNA was prepared in 4 μl ddH$_2$O together with 6 μl 7 M UREA loading dye. The samples were loaded on the gel was run for ~ 40 min at 120 V in 1 X TBE. Presence of RNA was checked with EtBr staining. If required RNA

quality can simply be determined over a denaturing polyacrylamide gel. RNA of good quality shows a sharp band without smear.

1% agarose gel

For 250 ml 1 % agarose we used 2.5 g agarose and 250 ml 1 x TBE. The agarose was boiled 2 - 3 times in the microwave until all agarose particles were melted. We stored it at 65°C until its usage. For one small agarose gel we used 50 ml agarose containing 1 µl ethidium bromide 1 % solution [10 mg·µl^{-1}].

7M Urea Loading Dye

6.3	g	UREA	[f. c. 7M]
3.75	ml	D(+)-Sucrose	[f. c. 25%]
37	mg	Bromphenol Blue	[f. c. 0.25%]
37	mg	Xylene Cyanol	[f. c. 0.25%]
1.5	ml	10 x TBE	

Fill up to 15 ml with ddH$_2$O and store at room temperature.

3. Oligonucleotide and RNA Labeling

3.1. Primer 5' End Labeling

Using the T4 Polynucleotide Kinase (PNK) and γ-^{32}P ATP, radiolabeled γ-phosphate group is transferred to the 5'OH group of oligos or RNA. For a reaction we used equimolar amounts of oligo and γ-P^{32}-ATP [6000 Ci·mmol^{-1}]:

2	µl	10x PNK buffer
1	µl	10µM oligo
6	µl	γ-^{32}P ATP [10 µCi·µl^{-1}]
		added ddH$_2$O up to 19 µl
1	µl	T4-PNK [10 u·µl^{-1}] (NEB)

- We incubated the samples at 37°C for 40 minutes.

- 1 µl 0.5 M EDTA pH 8.0 was added and samples were incubated at 95°C for 1 min and then put immediately on ice for 2 min.

- The primers were precipitated with 1 µl glycogen [10 µg·µl^{-1}] and 2.5x vol. Ethanol/0.3M NaOAc pH 5.0 and frozen at -20°C for 1 hour.

- After centrifugation at 4°C full speed for 30 min we discarded the supernatant, washed the pellet with 70 % Ethanol and dried it carefully for 5 min at room temperature.

- Finally the pellet was resuspended in 40 µl ddH$_2$O and stored at -20°C.

3.2. 5' End Labeling of TS12

For 5' labeling of the top strand TS12 we used:

20	pmol	TS12 RNA
5	µl	T4 PNK buffer (NEB)
12	µl	γ-^{32}P ATP [10 µCi·µl^{-1}]
2	µl	T4-Polynucleotide Kinase [10 u·µl^{-1}] (NEB)
		We increased the volume to 20 µl with ddH$_2$O.

- The reaction mixture was incubated for 40 min at 37°C.

- Afterwards we added 1 µl 0.5 M EDTA pH 8.0 (Ambion), placed the reaction on 95°C for 1 min and precipitated with 1 µl glycogen [10 µg·µl^{-1}] and 150 µl EtOH/0.3M NaOAc (Ambion) pH 5.0 for 1 hour at -20°C.

- After a 30 min centrifugation step (4°C, maximum speed) the pellet was resuspend in 10 µl ddH$_2$O. We added 30 pmol of cold TS12 RNA as there is a 5x molar excess of top strand over the bottom strand BS12 required for efficient duplex formation (*168*).

3.3. Body Labeling of Ai5γ Pre-RNA for the *Cis*-splicing Assay

We used α-P^{32} UTP and T7 RNA polymerase were used to randomly incorporate the labeled-nucleotides (as α-P^{32}-UMP) into the transcript, thereby producing body-labeled ai5γ pre-RNA.

~ 2.5	µg	linearized pJD20 plasmid DNA
5	µl	10 x TRA buffer
0.5	µl	1M DTT
3	µl	ATP, GTP, CTP [100mM]
1.5	µl	UTP [100mM]
5	µl	α-P^{32} UTP [10 µCi·µl^{-1}]
1	µl	rRNasin [40 U·µl^{-1}] (Promega)
4	µl	T7 RNA Polymerase (lab made)

We filled up to 50 µl with ddH$_2$O and incubated the reaction at 37 °C for 1 ½ hours. Afterwards DNA was digested with 2 µl DNAse I [1 U·µl^{-1}] (Fermentas) for 15 min at 37 °C. Ai5γ pre-RNA was then purified over the MEGAclear™ Kit.

Purification of the Ai5γ pre-RNA Using the MEGAclear™ Kit

ai5γ pre-RNA was purified as described in the user manual (www.ambion.com/techlib/prot/bp_1908.pdf). 100% Ethanol was required as it was not provided with the kit. The RNA was resuspended finally in 100 µl ddH$_2$O and stored at -20°C.

4. Reverse Transcription

To examine protein dependent structural changes within RNA we used metal-induced cleavage assays in combination with reverse transcription. On the basis of the RNA fragments of different length produced in the metal-induced cleavage assay, a cDNA pool was created during reverse trancription. The resulting cDNA pattern reflected the changes in the RNA molecule structure. In our experiments we used 0.5 pmol of RNA, ^{32}P-labeled primers and AMV reverse transcriptase.

4.1.   Primer Annealing Reaction

The annealing reactions contained:

2.5	µl	purified RNA (0.5 pmol modified RNA and RNA for A/C lane, 0.75 pmol for samples that containg Mss116p during the metal-induced cleavage assay[6])
1	µl	^{32}P-labeled primer [0.25µM]
1	µl	4.5 x Hybridization buffer

The samples were incubated at 95°C for 1 minute. Afterwards we adjusted the heating block to 55°C and let it cool slowly.

4.2.   Extension of Primer

For primer extension we used

2	µl	5 x AMV Extension buffer
0.7	µl	10mM dNTP mix
1.4	µl	ddH$_2$O
0.3	µl	AMV Reverse Transcriptase [10 u·µl^{-1}]

---

[6] Purified Mss116p protein is not completely without impurities (RNAses). To balance the RNA loss a larger amount of RNA (0.75 instead of 0.5 pmol) was used for reverse transcription.

For A and C lane we added 1 µl of 1mM ddNTP in addition to the extension mix (Amersham Biosciences).

- The samples were incubated at 55°C for 20 min.

- Then 1.5 µl 1 M NaOH was added in order to degrade the RNA and samples were incubated for further 20 min at 55°C.

- We added 1.5 µl 1 M HCl to neutralize the pH and

- precipitated with

1 µl	glycogen [10 mg·ml^{-1}]
1 µl	0.5M EDTA pH 8.0 (Ambion)
45 µl	Ethanol/0.3M NaOAc pH5.0

- The samples were frozen at -20°C 1 hour to overnight (alternatively at -80°C)

- Centrifuge the samples at 4°C at full speed for 30 – 60 min.

- The supernatant was discarded; the pellet washed with 70% Ethanol and the dried pellet was resuspended in 10 µl 7 M Urea Loading buffer.

- We loaded samples on an 8% denaturing PAGE for analysis.

1M NaOH

    6 g   NaOH

Dissolve pellets in 15 ml   ddH$_2$O. Dilute 10M NaOH solution 1:10 with ddH$_2$O.

1M HCl

    4.6 ml   33% HCl

    55.4 ml   ddH$_2$O

dNTP mix

10	µl	100mM dATP	[f. c. 10mM]
10	µl	100mM dGTP	[f. c. 10mM]
10	µl	100mM dCTP	[f. c. 10mM]
10	µl	100mM dTTP	[f. c. 10mM]
60	µl	ddH$_2$O	

We stored dNTPs at -20°C. We used a 100mM dNTP set (as Li-salt) from Roche.

10 x Extension buffer

650	µl	2M TRIS·HCl pH 8.0	[f. c. 1.3M]
100	µl	1M MgCl$_2$ (Ambion)	[f. c. 100mM]
100	µl	1M DTT	[f. c. 100mM]

Fill up with ddH$_2$O to 1ml, aliquot and store at -20°C.

4.5 x Hybridization buffer

225	µl	1M K-Hepes pH 7.0 (AppliChem)	[f. c. 225mM]
225	µl	2M KCl (Ambion)	[f. c. 450mM]

fill up with ddH$_2$O to 1ml, aliquot and store at -20°C.

7M Urea Loading Dye

    6.3 g   UREA   [f. c. 7M]

   3.75 ml   D(+)-Sucrose   [f. c. 25%]

    37 mg Bromphenol Blue [f. c. 0.25%]

    37 mg Xylene Cyanol   [f. c. 0.25%]

    1.5 ml 10 x TBE   [f. c. 1x]

Fill up to 15 ml with ddH$_2$O and store at -20°C.

4.3. 8% Denaturing Polyacrylamide Gel-Electrophoresis (PAGE)

We used an 8 % Polyacrylamide (PAA) gel to separate the cDNA pool resulting from the reverse transcription. For this we used 60 ml of 8% AA, 500 µl 10% APS and 50 µl TEMED were mixed together and the gel was poured and allowed to polymerize (for at least 1 hour). We pre-heated the gel by running it for 30 min with 40 W. With a syringe and 1 x TBE buffer we flushed away diffused urea from the wells. The RNA samples were loaded and the gel was run by 40 W for ~ 1.5 hours, until the Bromphenol Blue dye reached the bottom of the gel. After removing the upper glass plate the gel was transferred on Whatman 3M® paper. The second glass plate was carefully withdrawn and the other side of the gel was covered with saran wrap. After drying the gel ~ 1.5 hours in a Gel Dryer (80°C, vaccum) it was exposed to a Storage Phosphor Screen (GE Healthcare) and stored overnight in a cassette. The next day the screen was scanned with a Phosphor Imager (GE Healthcare) and the result was analyzed with Image Quant TL software (GE Healthcare).

8% Acrylamide (AA)

       100 ml   40 % Acrylamide-Solution Mix 19:1

       210 g   Urea

        50 ml   10 x TBE

Fill up to 500 ml with ddH$_2$O, stir, filter and store at 4°C.

10% APS

         1 g   APS   [f. c. 10% (w/v)]

Add 10 ml ddH$_2$O and store at 4°C protected from light.

10 x TBE

       540 g   Tris base   [f. c. 890mM]

       275 g   Boric acid   [f. c. 890mM]

     29.2 g   EDTA   [f. c. 20mM]

Fill up to 5 L with ddH$_2$O, stir and autoclave.

5. Protein Activity Assays

Before we started structural probing of ai5γ in the presence and absence of Mss116p we wanted

to verify that our purified recombinant Mss116p is able to perform its enzymatic functions:

promoting self-splicing of the ai5γ intron, unwinding of RNA duplex and ATP hydrolysis.

5.1. *In vitro* Splicing Assays

By performing *cis*-splicing assays with the ai5γ pre-RNA (full length intron flanked by part of the exons) in the absence and presence of Mss116p, we tested the activity of the recombinant Mss116p/ΔMLS protein. Notably, it has previously been reported that Mss116p can promote group II intron splicing at near-physiological ionic and temperature conditions. Below we describe the experimental set up. Notably, a positive control for which ai5γ pre-RNA was folded and spliced under high-salt conditions and elevated temperature (100mM $MgCl_2$, 500mM NH_4Cl, 42°C). As negative control the ai5γ pre-RNA was folded at near-physiological conditions in the absence of Mss116p (8mM $MgCl_2$, 100mM NH_4Cl, 30°C), which does not promote self-splicing. In brief, we incubated the full-length body labeled ai5γ precursor RNA in 40mM MOPS, pH 7.5 at 95°C for 1 min and subsequently at 30°C for 3 min. For reactions in the presence of Mss116p, the reaction mixtures contained 600nM Mss116p, 8mM $MgCl_2$, 1mM ATP (in a mix with 1mM $MgCl_2$). The reaction was incubated at 30°C, aliquots were withdrawn at selected time points and the reaction was stopped with Proteinase K and 15 µl stop buffer. Half of the sample was analyzed on a 5% denaturing polyacrylamide gel (see section 2.3). The autoradiography was performed as described in section 0.

	Mss116p and ai5γ pre-RNA (exD123456) *in vitro* splicing assay			
Action	**Reagent**	**+ control**	**- control**	**+ Mss116p**
	ai5γ pre-RNA ^{32}P labeled	1 µl	1 µl	1 µl
ADD	1M MOPS pH 7.0 [f. c. 40mM]	1 µl	1 µl	1 µl
ADD	ddH$_2$O	12.25 µl	19.97 µl	16.05 µl
	denature RNA for 1 min at 95 °C 3 min at 30 °C spin down			
ADD	1M MgCl$_2$ (Ambion) *+ control [f. c. 100mM]* 0.1M MgCl$_2$ (Ambion) *- control and Mss116p [f. c. 8mM]*	2.5 µl	2 µl	2 µl
ADD	600 nM Mss116p [6.66µM]	-	-	2.25 µl
ADD	2M NH$_4$Cl [f.c. 500mM]	6.25 µl	-	-
ADD	3M KCl (Ambion) *- control [f. c. 100mM]*	-	0.83 µl	-
ADD	10mM ATP/Mg^{2+} mix [f. c. 1mM] (AppliChem) *includes 10 mM MgCl$_2$ (Ambion)*	-	-	2.5 µl
ADD	rRNasin [40 U/µl] (Promega)	2 µl	2 µl	2 µl
total		25 µl	25 µl	25 µl
	mix gently (+) control: incubate at 42 °C, (-) control and Mss116p sample: incubate at 30°C withdraw aliquots (3 µl) after 0, 10, 20, 30, 60 and 120 min			
	aliquot treatment To each 3 µl aliquot add 1 µl of Proteinkinase K (AppliChem) in case of Mss116p-containing samples and Incubate for 1 min at room temperature and then add 15 µl Stop buffer (to all aliquots) store at -20 °C analyze half of sample on a 5 % PAA gel (note: the second dye (xylene cyanol) should reach the bottom of the gel!)			

2M NH4Cl

  5.3 g Ammoniumchlorid

Add 50 ml ddH$_2$O and filter through syringe with an acrodisc filter (0.2µm) attached.

10 x TBE

  540 g Tris base   [f. c. 890mM]

  275 g Boric acid   [f. c. 890mM]

  29.2 g EDTA   [f. c. 20mM]

Fill up to 5 L with ddH$_2$O, stir and autoclave it.

Stop buffer

The stop buffer used in the above described *in vitro* splicing assay is a combination of the following three solutions.

<u>Solution 1</u>

  500 µl sucrose (AppliChem)   [f. c. 50%]

  100 µl 10xTBE   [f. c. 1xTBE]

  500 µl ddH$_2$O

<u>Solution 2</u>

  800 µl formamide   [f. c. 80%]

  200 µl solution1   [f. c. 20%]

<u>Solution 3</u>

  100 µl 0.5M EDTA pH 8.0 (Ambion)   [f. c. 50mM]

  900 µl solution 2   [f. c. 100%]

1M MOPS pH 7.0

        52.3 g   MOPS

Add ~200 ml ddH$_2$O and adjust pH to 7.0 with 10M KOH. Then fill up to 250 ml with ddH$_2$O, filter over Nalgene filter units and store buffer at 4°C protected from light. Take the dilution effect into account!

5.2.   Unwinding Assays

Duplex Formation

For BS/TS12 duplex formation we used a 5x molar excess of labeled TS12 strand over the complementary BS12 strand (*168*). We prepared a reaction mixture containing 50pmol ^{32}P labeled TS12 strand (is a mixture of 30pmol cold TS12 and 20pmol labeled TS12) and 10pmol of BS12 in 10mM MOPS pH 6.5 and 1mM EDTA pH 8.0 (Ambion). The reaction was heated up to 95°C for 2 min and cooled down to room temperature slowly (takes ~ 90 minutes). The duplexes were separated over Illustra™ Microspin™ G-25 Columns (for working protocol see protocol of manufacturer) and duplex was stored at -20°C.

1M MOPS pH 6.5

        52.3 g   MOPS

Add ~200 ml ddH$_2$O and adjust pH to 6.5 with 10M KOH. Then fill up to 250 ml with ddH$_2$O, filter over Nalgene filter units and store buffer at 4°C protected from light.

Unwinding Assay of 12bp Blunt End Duplex

For the unwinding assays we used 5nM labeled TS/BS12 RNA duplex and 400nM Mss116p in 1 x Helicase Reaction Buffer (HRB). The mixture was incubated for 10min at 30°C and reaction was initiated with the addition of 20mM ATP/MgCl$_2$-mix [f.c.]. As a negative control we prepared one reaction without MgCl$_2$ and one without Mss116p. To create a single stranded size marker an additional aliquot was withdrawn, heated up to 95°C for 2 min and placed on ice. Samples were taken at time point 0, 30 and 60 minutes and the reaction was stopped with Helicase Reaction Stop Buffer (HRSB). We loaded the samples on a 15% non-denaturing polyacrylamide gel (*68*).

1M Tris·HCl pH 8.0

     30.2   g   Tris

Dissolve salt in 200 ml ddH$_2$O. Adjust pH to 8.0 with HCl. Afterwards fill up to 250 ml with ddH$_2$O and autoclave it.

10% SDS

     5   g   SDS

Dissolve carefully in 50 ml ddH$_2$O.

0.5% IGEPAL

     75   µl   IGEPAL® CA-630 (Sigma)

Fill up to 15 ml with ddH$_2$O.

1M DTT

     154   mg   Dithiothreitol

Add 1 ml ddH$_2$O and store at -20°C.

2 x Helicase Reaction SDS Stop Buffer (HRSB)

1	ml	0.5M EDTA pH 8.0	[f. c. 50mM]
1	ml	10% SDS	[f. c. 1%]
10	mg	Bromphenol blue	[f. c. 0.1%]
10	mg	Xylene Cyanol	[f. c. 0.1%]
2	ml	glycerol	[f. c. 20%]

Fill up to 10 ml with ddH$_2$O and store at room temperature.

10 x Helicase Reaction Buffer (HRB)

4	ml	1M Tris·HCl pH 8.0	[f. c. 400mM]
50	µl	1M MgCl$_2$ (Ambion)	[f. c. 5mM]
2	ml	0.5% IGEPAL	[f. c. 0.1%]
200	µl	1M DTT	[f. c. 20mM]

Fill up to 10 ml with ddH$_2$O, aliquot in 1 ml tubes and store at -20°C.

15% Non-Denaturing PAGE

The 15 % non-denaturing PAGE was used to visualize unwinding of the 12bp blunt end duplex TS/BS12 in the presence of Mss116p. For this we used 60 ml of 15% AA, 500 µl 10% APS and 50 µl TEMED were mixed together and the gel was poured and allowed to polymerize. We pre-heated the gel by running it for 20 min with 40 W. Using a syringe filled with 1 x TBE buffer we flushed away diffused urea from the wells. The samples were loaded and the gel was run by 30 W for ~ 1 hour in 1 x TBE buffer. After removing the upper glass plate the gel was covered with Saran wrap first. This is important because otherwise the gel can be easily destroyed due to the dryness of a 15% PAGE gel. The second glass plate was carefully withdrawn and the other side

of the gel was covered with Whatman 3M® paper. After drying the gel ~ 1 hour in a Gel Dryer (80°C, vaccum) it was exposed to a Storage Phosphor Screen (GE Healthcare) and stored overnight in a cassette. The next day the screen was scanned with a Phosphor Imager (GE Healthcare) and the result was analyzed with Image Quant TL software (GE Healthcare).

10 x TBE

540	g	Tris base	[f. c. 890mM]
275	g	Boric acid	[f. c. 890mM]
29.2	g	EDTA	[f. c. 20mM]

Fill up to 5 L with ddH$_2$O, stir and autoclave it.

15% Acrylamide (AA)

93	ml	40% Acrylamide/bisacrylamide	[f. c. 15%]
25	ml	10 x TBE	[f. c. 1x TBE]

Fill up to 250 ml with ddH$_2$O, filter and store at 4°C.

5.3. Malachite Green ATPase Assays

The Malachite Green Phosphate Detection Kit (R&D Systems) is a method which determines the concentration of free inorganic phosphate (P$_i$). It is based on the complexation of malachite green, ammonium molybdate and free inorganic phosphate under acidic conditions. First the free orthophosphate forms a complex with ammonium molybdate in a sulfuric acid solution. The complexation of the phosphomolybdate with malachite green can be measured at 620nm and is directly related to ATP hydrolyzed by the Mss116p protein (bitesizebio.com).

$$\text{Enzyme + Substrate} \longrightarrow \text{[ES]} \longrightarrow \text{Enzyme} + P_i$$

$$P_i + (NH_4)_2MoO_4 \xrightarrow{H^+} H_3PMo_{12}O_{40}$$

$$\underset{\text{(yellow)}}{H_3PMo_{12}O_{40}} + \underset{\substack{\text{(malachite green)} \\ \text{(yellow, } \lambda_{max} \text{ 446 nm)}}}{HMG^{2+}} \xrightarrow{H^+} \underset{\text{(green, } \lambda_{max} \text{ 640 nm)}}{(MG^+)(H_2PMo_{12}O_{40})} + 2H^+$$
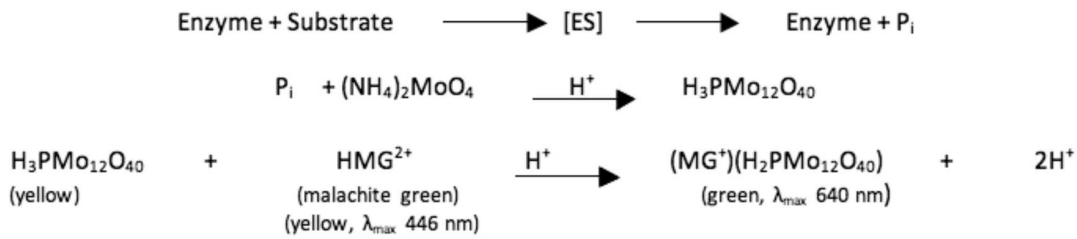
Figure 29: The Malachite Green Assay Mechanism. It is based on the complexation of malachite green, ammonium molybdate and free inorganic phosphate (orthophosphate, P_i) in 3M sulfuric acid (bitesizebio.com; thanks to the author Megan Hogan).

For our reactions we pre-incubated 250nM Mss116p, rRNasin Inhibitor [1 U·µl^{-1}] (Promega) with 5µM BS12 ssRNA or 0.5µM ai5γ pre-RNA at 30°C for 10min. The reaction buffer mixture consisting of 55mM KCl (Ambion), 40mM HEPES pH 7.5 (AppliChem) and 0.85mM DTT and the samples were incubated at 30°C for another 10min. We started the reaction by adding an equimolar concentration of 1mM MgCl$_2$ (Ambion) and ATP (AppliChem). At specific time points 5 µl samples were withdrawn and diluted in a µCLEAR plate with 45 µl ddH$_2$O. The reaction was stopped with 10 µl Malachite Green Reagent A to the 50 µl sample (10 min, room temperature). Then 10 µl Malachite Green Reagent B was added and the samples were incubated at room temperature finally the absorbance was measured with the Infinite F500 microplate reader at 620nm. Results were evaluated with Excel. For controls we digested Mss116p with 0.5 µl Proteinase K (AppliChem) and RNA with 0.25 µl RNAse A at the first incubation step.

1M DTT

     154 mg Dithiothreitol

Add 1ml ddH$_2$O, resuspend, make 1:100 dilution and store at -20°C.

ATP

For preparation of ATP see chapter 2.2.

6. *In Vitro* Metal Ion Cleavage Reactions

6.1. Lead Cleavage Reactions

We used 2pM of full-length ai5γ RNA in 50mM MOPS pH 7.0 and 500mM KCl (Ambion), denatured the RNA for 1min at 95°C and let it cool at room temperature for 2min. 8mM MgCl$_2$ (Ambion) was added together with 250nM Mss116p and 1mM ATP (including 1mM MgCl$_2$) and incubated at 30°C for 60min. Afterwards we added 5mM Pb(OAc)$_2$ (Aldrich) to the reaction mixture and incubated it for 5min at 30°C. The protein was degraded with 0.5 µl Proteinase K (AppliChem) for 1min at room temperature. As a control we performed Pb^{2+}-induced cleavage of D123456 RNA which was folded at optimal conditions: 0.5M KCl, 100mM MgCl$_2$ and 42°C (high salt and temperature conditions promoting self-splicing). We precipitated the RNA with 1 µl glycogen [10 mg·ml^{-1}], 20 µl 0.5M EDTA pH 8.0 (Ambion) and 2.5x vol. EtOH/0.3M NaOAc for 1 hour at -20°C. We centrifuged samples at 4°C, at maximum speed for 1 hour and after drying the pellets were resuspended in 25 µl ddH$_2$O. The sites of cleavage were mapped by reverse transcription (see chapter 4).

6.2. Terbium Cleavage Reactions

For *in vitro* Tb^{3+} cleavage reactions 50nM full-length ai5γ intron RNA (D123456) was denatured at 95°C for 1 min in 40mM MOPS pH 7.5. After cooling samples at 30°C for 3min and a short spin we added 8mM MgCl$_2$ (Ambion), 250nM Mss116p, KCl (Ambion) to a total concentration of 100mM (depends on [KCl] in Mss116p storage buffer), 200 units rRNasin (Promega) and 1mM ATP (including 1mM MgCl$_2$). The RNA-protein complex was formed by incubating the reaction for 1 hour at 30°C. Subsequently different Tb^{3+} concentrations ranging from 1mM to 5mM Terbium (III) chloride (Sigma) were added to the sample. The tubes were placed on ice for 1 hour. After digesting the protein with 1 µl Proteinase K (AppliChem) at room temperature for 1 min, we

stopped the reaction with 25mM EDTA pH 8.0 (Ambion) and precipitated the RNA with 2 µl

glycogen [10 mg·ml^{-1}] and 2.5 x EtOH/0.3M NaOAc for 1 hour at -20°C. The pellets were

resuspended in a final volume of 50 µl ddH$_2$O and analyzed by reverse transcription (see chapter

4).

1M MOPS pH 7.5

    20.9 g   MOPS

Add ~ 70 ml ddH$_2$O, adjust pH to 7.5 with KOH and bring volume to 100 ml with ddH$_2$O, sterile

filter and store at 4°C protected from light.

0.1M Cacodylate buffer pH 5.0

    1.07 g   Sodium cacodylate trihydrate (Sigma)

Dissolve salt in 20 ml ddH$_2$O and adjust pH to 5.0 with HCl. Afterwards fill up to 50 ml with ddH$_2$O

and store at 4°C.

ATP

For preparation of ATP see chapter 2.2

5M NaOH

    10 g   Sodium Hydroxide Pellets

Fill up with ddH$_2$O to 50 ml.

Results

In order to map ai5γ group II intron structure in collapsed and native conformation at near physiological conditions *in vitro*, recombinant Mss116p had to be produced under native conditions, meaning that protein is soluble in supernatant after cell lysis and protein activity is preserved. Consequently the functionality of recombinant Mss116p was examined by means of an *in vitro* splicing assay, an unwinding assay and an ATPase assay. Once Mss116p functionality had been proven, we performed Pb^{2+} and Tb^{3+}-induced cleavage reactions on the ai5γ intron under near-physiological conditions (8mM Mg^{2+} and 30°C) in the absence and presence of Mss116p. Metal ion cleavage reactions were extensively used before (*147, 169*) and on the basis of the described protocols we established *in vitro* Pb^{2+} and Tb^{3+} cleavage reactions for mapping metal ion binding sites within the ai5γ intron.

Producing Recombinant Mss116p

Mss116p had been successfully crystallized in 2009 by Del Campo and Lambowitz (*85*). As our plasmid was a generous gift from Anna M. Pyle we established a protein purification protocol on the basis of information given by several publications coming from the Lambowitz and Pyle laboratories. Despite the information given in papers purification of Mss116p was a challenge, as detailed information about purification strategies and distinct steps were not indicated. Mss116p without mitochondrial localization sequence (Mss116p/ΔMLS), deletion of mitochondrial localization sequence (*79*)) was produced as a fusion protein to a 6xHIS-tag as it has been sub-cloned into a pET SUMO® expression system (pAS02 from (*170*)). Mss116p was overexpressed in *Escherichia coli* and expression was induced with IPTG. First optimal expression conditions had to be tested by studying protein expression and solubility at different growth parameters like temperature, IPTG concentration and incubation time.

1. **Optimizing Protein Expression**

Induction of Mss116p/∆MLS Expression

Mss116p/∆MLS expression has been performed by other groups at 16°C overnight (HIS-Mss116p ∆MLS fusion and cleavage with SUMO protease) (*38*) or 15°C overnight (MalE-Mss116p fusion and cleavage with TEV protease) (*6*). To confirm these parameters different temperature together with various IPTG concentrations and induction periods were analyzed. The combinations are illustrated in Table 2 and expression was found to be optimal with 0.5mM IPTG at 18°C and overnight growth in our hands as well. The amount of Mss116p/∆MLS was compared before and after induction with IPTG (at the specified incubation times and temperature) for the total extract as well as, in case of the induced aliquot, for the cell pellet and the supernatant (Figure 33). It is worth mentioning that the expression of Mss116p/∆MLS varies slightly between individual colonies, as shown in Figure 30, which displays the protein expression pattern of five different colonies induced under same conditions. In this case, colony 3 shows a slightly better expression pattern for Mss116p/∆MLS.

Table 2: Parameters tested for protein expression, including IPTG concentration, temperature and time of growth.

Temperature (°C)	IPTG [mM]	Growth (h)
16	0.7	overnight
18	5,1,0.7, 0.5, 0.25	3;4;overnight
20	1, 0.5	overnight
26	1	4

IPTG [0.25mM]
3h growth at 18°C

◀ Mss116p ~85 kDa

Colony 1   Colony 2   Colony 3   Colony 4   Colony 5
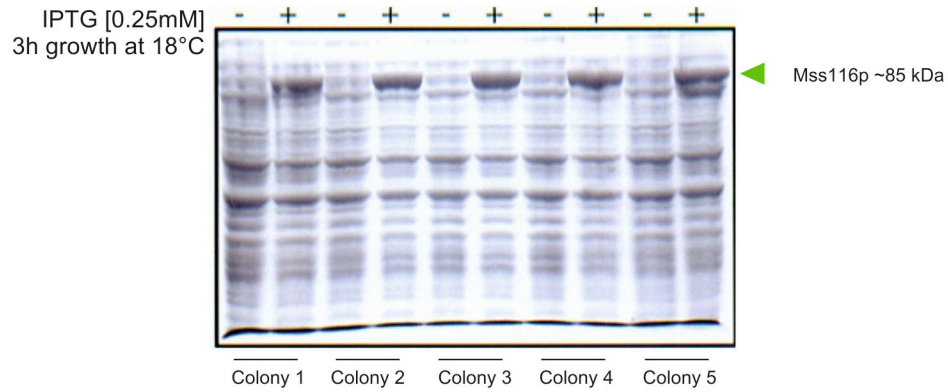
Figure 30: Testing five different colonies for induction parameters. Cells were induced with 0.25mM IPTG and grown at 18°C for 3 hours. Colony 3 shows best expression indicating that there is as well a slight difference in expression pattern with different colonies selected.

Figure 31 below illustrates three bacterial cultures where Mss116p/ΔMLS was expressed under optimal conditions (0.5mM IPTG and overnight growth at 18°C).

Figure 31: Mss116p/ΔMLS expression with 0.5mM IPTG and growth at 18°C overnight. (-) cell lysate before induction; (+) cell lysate after IPTG induction; green arrow indicates Mss116p/ΔMLS expression.

2. Purification and Washing

During our first trial the Ni-NTA column was sequentially washed with five cycles of Wash buffer I (including 50mM imidazole) after loading of the crude protein extract to remove nonspecifically bound proteins. While the control SDS-PAGE did not show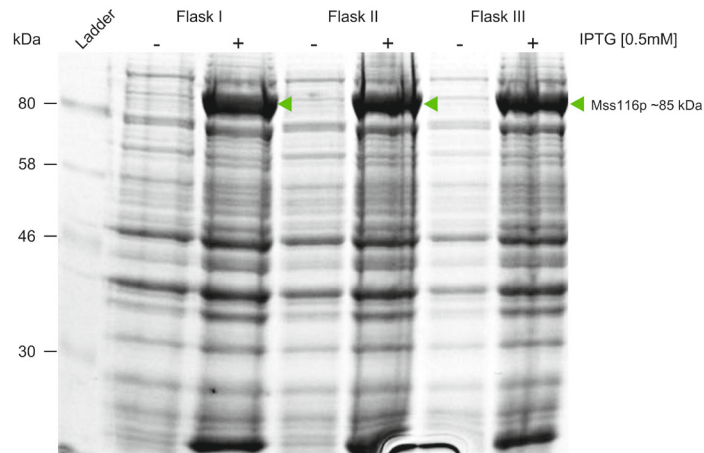 any impurities within the Mss116p/ΔMLS protein sample, substantial contamination with RNases was detected upon performing an *in vitro* splicing assay (Figure 32). Impurities completely degraded ai5γ precursor RNA within 1 hour (compared to the control reactions). The RNase activity could not be balanced by increasing the amount of RNase inhibitor. On the bright side, we also observed that the recombinant Mss116p/ΔMLS was active, as indicated by the spliced intron in linear and lariat forms. Besides, the drawback of using higher RNase inhibitor concentrations is that storage buffer of the enzyme contains large amounts of glycerol, up to 50% (v/v). Together with the Mss116p/ΔMLS storage buffer, which includes 10% glycerol as well, the final concentration of glycerol was greater than 10% in the enzymatic reaction, potentially inhibiting Mss116p/ΔMLS in the metal ion-induced cleavage assay.
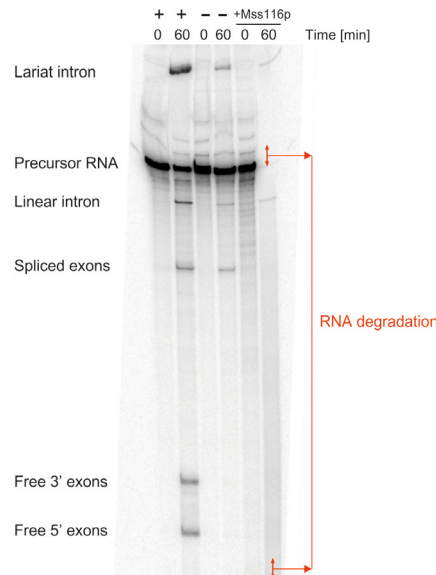

Figure 32: Testing protein sample quality over an *in vitro* splicing assay. Recombinant protein added to ai5γ pre-RNA (D123456) at near physiological conditions (8mM Mg^{2+} and 30°C) fully degrades RNA after 1 hour. Positive control indicates pre-RNA self-splicing under *in vitro* conditions (100mM Mg^{2+} and 42°C).

The majority of impurities were removed by additional washing cycles with increased imidazole concentration. Ni-NTA column was therefore sequentially washed with a) four cycles of Wash buffer I, including 50mM imidazole and subsequently by b) four additional washing steps with Wash buffer II which includes 100mM imidazole. Washing nickel-affinity column with Wash buffer II (Figure 33) removed small amounts of, presumably, Mss116p/ΔMLS as well, but also further removed impurities which were not washed away with Wash buffer I. Applying these four additional washing steps significantly improved the purity of the Mss116p/ΔMLS protein, such that only minimal RNA degradation was visible during the *in vitro* splicing assay (see Figure 39).
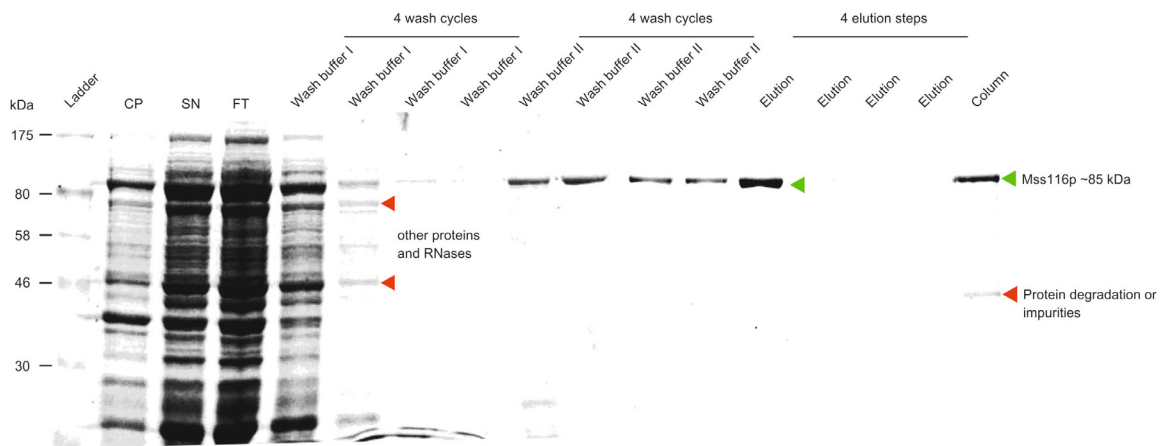
Figure 33: Illustrating protein purification steps and protein levels within cell pellet (CP), Supernatant after clarifying lysate over centrifugation (SN) and the flow through after incubating Ni-NTA resin with cleared lysate (FT). Green arrows: indicate Mss116p/ΔMLS levels; Red arrows: illustrate impurities.

RNases Expressed by BL21

A 'Uniprot' database query of RNAses expressed in the *E. coli* BL21 strain revealed several RNA degrading proteins which have similar size as Mss116p/ΔMLS and also harbor numerous histidines within their protein sequence. Exoribonuclease 2 for example which is highlighted in grey in Table 3 carries 12 histidine residues, with a size of 72.4 kDa and exactly same amino acid sequence length as Mss16p/ΔMLS which carries as well 12 histidine within protein sequence. Over using Centrifugal Devices like in our case the 30K cut-off device was able to exclude impurities with a molecular size smaller than 30 kDa e.g. the Ribonuclease HII which has a size

of 21.5 kDa. To remove remaining RNases a gel filtration experiment should therefore be performed to obtain higher protein quality.

Table 3: Abridgment of Uniprot query of RNAses expressed by *E. coli* strain B/ BL21-DE3. The table indicates accession number of the protein, number of histidines in protein sequence (HIS), protein size in Dalton and amino acid sequence length (AA).

Accession	Protein names	Gene names	HIS	Size DA	Length AA
C6ECH0	Exoribonuclease R, RNase R (EC 3.1.-.-) (VacB and RNase II family 3'-5' exoribonuclease) (EC 3.1.13.1)	rnr ECBD_3855 ECD_04046	18	92.109	813
C6EFU6	Exoribonuclease 2 (EC 3.1.13.1) (Exoribonuclease II)	rnb ECBD_2331 ECD_01263	17	72.476	644
C6EHC9	Fused ribonucleaseE: endoribonuclease/RNA-binding protein/RNA degradosome binding protein (EC 3.1.4.-) (Ribonuclease, Rne/Rng family)	rne ECBD_2516 ECD_01080	24	118.287	1061
C6EAA2	Ribonuclease HII (RNase HII) (EC 3.1.26.4)	rnhB ECBD_3436 ECD_00181	8	21.543	198
C6EBY5	Exodeoxyribonuclease X (EC 3.1.11.-) (Exonuclease RNase T and DNA polymerase III)	exoX ECBD_1796 ECD_01815	8	25.115	220
C6EK20	Ribonuclease 3 (EC 3.1.26.3) (Ribonuclease III)	rncS rnc ECBD_1114 ECD_02461	7	25.550	226
C6EF15	Ribonuclease BN	rbn ECBD_4141 ECD_03771	2	32.839	290
C6EK83	Enterobacter ribonuclease (EC 3.1.27.6) (Ribonuclease I) (EC 3.1.27.6)	rna ECBD_3041 ECD_00579	4	29.618	268
C6E9T1	Ribonuclease Z (RNase Z) (EC 3.1.26.11) (tRNA 3 endonuclease) (tRNase Z)	elaC rnz ECBD_1390 ECD_02195	12	32.930	305
C6EGL4	Ribonuclease G (EC 3.1.4.-) (Ribonuclease, Rne/Rng family)	rng ECBD_0500 ECD_03106	13	55.364	489

Removing Remaining Impurities By Gel Filtration

Solem et. al 2006 and Del Campo and Lambowitz 2009 (*38, 85*) performed a gel filtration as a final step before dialysis to remove remaining impurities. In collaboration with Kristina Djinovic-Carugo and co-workers a gel filtration using a Hiload Superdex-200 column was performed. Gel filtration though was not successful as protein came off in very late elution fractions which could not be collected. This might have happened due to high glycerol and/or salt conditions and as a result the protein molecules formed aggregates resulting in precipitation. Thus, filtration conditions would have to be adjusted in a very time-consuming process. Meanwhile, we had done some activity assays with the Mss116p/ΔMLS prepared with stringent washing steps. As the purity of the protein preparation seemed to be sufficient for successfully performing the activity assays, in which we only observed a minimum of RNA degradation (Figure 39), we refrained from

processing with optimizing the gel filtration set up. However, as gel filtration is known to improve the purity of recombinant protein preparations, this leaves of course room for further improvement of the quality of our Mss116p/ΔMLS preparations.

3. Preparing untagged Mss116p/ΔMLS

For producing recombinant protein six consecutive histidines and a protease cleavage site were introduced to the N-terminus of Mss116p. His tags bind selectively to nickel-affinity columns and tag can be is removed by a protease, requiring an additional purification step to eliminate the cleaved tag (~13kDa) afterwards. To facilitate protease cleavage of Mss116p/ΔMLS Solem et al. added an N-terminal serine to Mss116p/ΔMLS (*38*). Unexpectedly, removal of the 6x-His tag decreased the solubility of Mss116p/ΔMLS. Thus, the final purification conditions have to be adjusted to avoid aggregation and in turn precipitation of untagged Mss116p/ΔMLS protein. Del Campo and Lambowitz added 10% glycerol and L-arginine and L-glutamate to increase protein solubility from 1 mg·ml^{-1} up to 10 mg·ml^{-1} (*85*).

Cleaving off the 6x-His Tag

As mentioned before, the construct contains a cleavage site for the SUMO protease between the 6x His tag and the Mss116p/ΔMLS sequence. The SUMO protease is a cysteinyl protease (Ubl-specific protease 1) from *S. cerevisiae* and cleaves highly specific by binding to tertiary structure of ubiquitin-like protein, SUMO rather than to specific amino acid sequence (Invitrogen). Performance of the SUMO protease is optimal at 30°C; therefore the cleavage was tested in a time-dependent manner at this temperature. At 30°C cleavage was completed after 3 hours of incubation (Figure 34). Performing the cleavage overnight at 30°C has no additional positive effect on removing the tag.
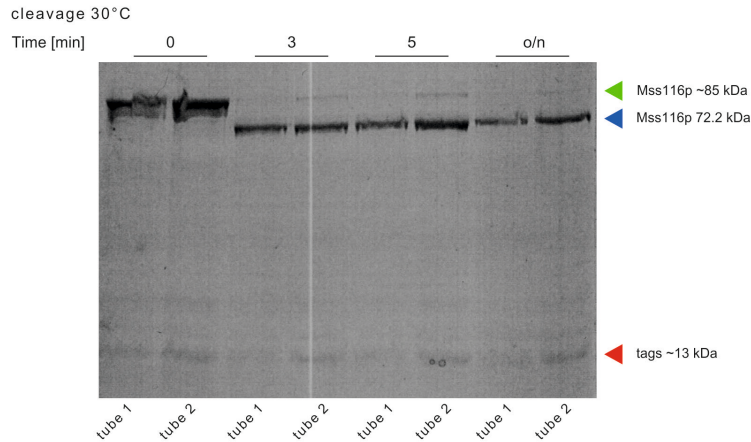
Figure 34: Tag cleavage at 30°C for 3, 5 hours and overnight. Reaction mixture was split in two parts (tube 1 and 2) to reach a maximum of 2ml reaction volume. Full cleavage can be already observed after 3 hours at 30°C. Red arrow indicates cleaved tags with an approximate size of 13 kDa.

Additionally, SUMO-protease cleavage was performed at 4°C overnight, as suggested by Solem et al. 2006. The cleavage reaction at 4°C was slightly less efficient than at 30°C (compare Figure 35). Since washing and elution steps during purification of the recombinant Mss116p/ΔMLS

protein were performed at 4°C, we decided to apply the same temperature for the SUMO-cleavage step as well, offering constant low temperature conditions, which might be favorable for protein stability. To this end, the tag was cleaved off by the SUMO protease at 4°C overnight.

Figure 35: Tag cleavage at 4°C overnight and subsequent purification steps. Precipitate was removed over centrifugation. Protein was concentrated with 30K Centrifugal Devices and dialyzed against storage buffer. (SN) visualizes cleared untagged protein without precipitate.

Increasing Protein Solubility Over 10% Glycerol, L-Arg and L-Glu

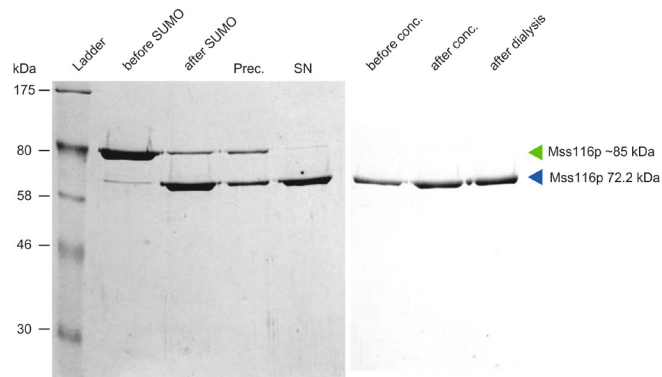After cleavage of fusion protein, precipitation of the untagged Mss116p/ΔMLS was observed at both reaction temperatures, 4°C and 30°C, respectively. The precipitate was analyzed on SDS PAGE and showed that ~60% of recombinant protein Mss116p/ΔMLS was precipitated after removal of the tag (Figure 36). Therefore, following the Lambowitz and coworkers (*85*), we tried to increase the solubility of untagged Mss116p/ΔMLS by adding 10% glycerol, 50mM L-arginine and 50mM L-glutamate to SUMO protease cleavage reaction. Figure 36 illustrates precipitation of protein before addition of additives where we calculated protein levels over ImageQuant. Only 40% of protein could be collected in the supernatant. After adding supportive agents protein solubility could be increased to ~ 65% which can be seen in lane SN of Figure 35.
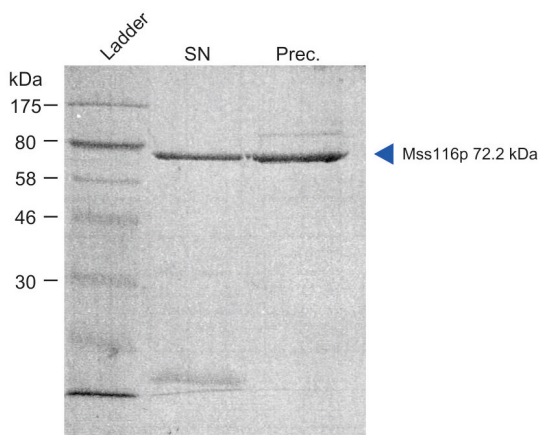
Figure 36: Protein precipitation after tag cleavage. ~60 % of untagged protein precipitates after SUMO cleavage. (SN) supernatant after removing precipitation by centrifugation; (Prec.) Precipitate resuspended in adequate volume.

Removal the Cleaved Tag and the SUMO Protease

At the beginning, we reasoned that by loading the SUMO cleavage reaction on the Ni-NTA column, the cleaved tag as well as remnants of tagged Mss116p/ΔMLS would remain on the matrix, while the untagged Mss116p/ΔMLS would be in the flow through. However, the untagged Mss116p/ΔMLS did not come off the column due to its 11 histidine residues dispersed over its primary sequence. Consequently, we tried to elute the untagged protein with increasing imidazole

or KCl concentrations, hoping that the untagged protein would come off the column at a different concentration compared to the tagged one and the cleaved tag. However, only very low levels of protein could be collected. Checking an aliquot of column material on SDS PAGE revealed that most of protein was still bound to the resin. Elution with buffers containing between 50mM and 500mM imidazole or between 0.5 and 1M KCl was not effective at all (Figure 37). Protein is hardly removable from resin neither by an imidazole nor a high salt gradient. In a different approach we tried to cleave off protein directly from the column resin after the washing steps. The result showed similar results as in the previous experiment where Mss116p/ΔMLS was tightly bound over its exposed histidines to the nickel-charged resin and not removable over high imidazole and salt gradients.
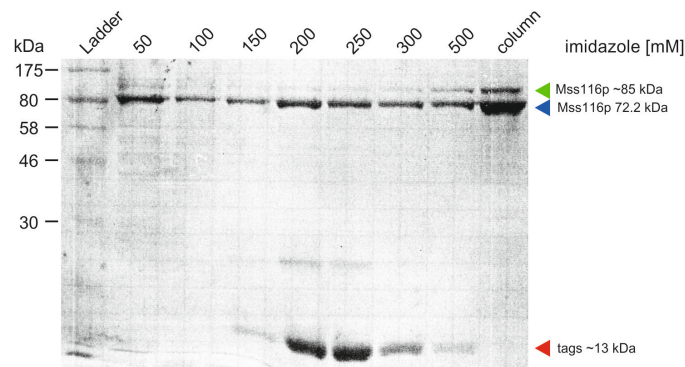
Figure 37: elution with increasing imidazole concentration ranging from 50mM to 500mM. Elution fractions 200mM and 250mM indicate major removal of fusion tags. Lane 'column' indicates that majority of protein is still bound to nickel-charged resin

In fact it is tempting to suggest that there might be a structural change after tag cleavage which increases protein affinity for the nickel-charged resin. However, another explanation would be that Mss116p/ΔMLS is not traversing a structural change but gets tightly bound by several other exposed histidines on its protein surface (w/t Mss116p carries 12 histidines in protein sequence and Mss116p/ΔMLS carries 11 histidine residues). And in fact, checking location of histidines in crystal structure (Mss116p/Δ598-664) (*86*) revealed that most of the 11 histidines are exposed on the surface (Figure 38).
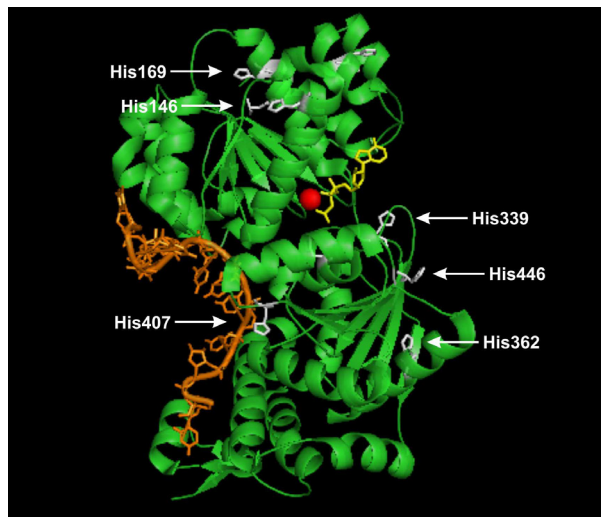
Figure 38: Mss116p/Δ598-664 crystal structure (green) complexed with ssRNA (orange) and AMP-PNP (yellow). Histidine residues were highlighted in white using PyMOL showing that most of them protruding from the surface. Mg^{2+} is indicated as a red sphere.

As a result 30K centrifugal devices were used in order to remove the tag (consisting of 6x His and cleavage site for SUMO protease) and the SUMO protease, which have an approximate size of 13 kDa and 26 kDa, respectively (lifesensors.com). Protein degradation fragments and RNases smaller than 30 kDa (see Table 2) are removed by centrifugation as well. Additionally, a twofold increase in protein concentration could be achieved (Figure 35) resulting in a protein yield ~1mg per liter of culture.

Protein Activity Assays

Mss116p-Promoted Splicing Activity of Ai5γ pre-RNA

Mss116p promotes splicing of ai5γ group II precursor RNA, resulting in the appearance of ligated exons and intron lariat with a rate constant k_{obs} of $30 \times 10^{-3} - 35 \times 10^{-3}$ min^{-1} (*6*). To show that the recombinant Mss116p/ΔMLS, which we prepared, is able to promote splicing with similar activity, we established a protocol on basis of previously published splicing assays (*6*). A minimum of 6mM Mg^{2+} is required for the observing Mss116p-stimulated ai5γ splicing and this activity is observable up to 35mM Mg^{2+}, while concentrations above 35mM inhibit Mss116p in its ability to promote intron splicing (*38*). Indeed, at near-physiological conditions (30°C; 8mM MgCl$_2$, 100mM KCl, 40mM MOPS [pH.7.5]) Mss116p-mediated self-splicing was observed to generate the same reaction products, as self-splicing promoting at high-salt

Figure 39: Protein-mediated splicing of Ai5γ. Self-splicing reactions under high-salt and temperature conditions (in the absence of Mss116p) (+) and Mss116p-assisted intron splicing at near physiological conditions. (-) The negative control shows lack of splicing at near physiological conditions in the absence of Mss116p. I-Lin may contain a mixture of linear intron and broken lariat intron RNA.

conditions (42°C, 100mM MgCl$_2$, 0.5M NH$_4$Cl) (*38*). As illustrated in Figure 39 recombinant Mss116p/ΔMLS efficiently promotes splicing of *Sc.* ai5γ group II intron plus 1mM ATP. Comparable to the optimal splicing conditions, after 120 min almost all of the pre-RNA has been converted into spliced introns, in both linear and lariat forms, and ligated exons in the presence of Mss116p/ΔMLS as well. On the contrary, little splicing is observed at near-physiological

conditions and in the absence of Mss116p/ΔMLS. The gel was quantified with ImageQuant and data was fit to a single turnover equation using the program KaleidaGraph. Figure 40 illustrates the precursor RNA levels plotted with respect to time and quantification revealed a comparable rate constant of k_{obs} of $30 \pm 0.009 \times 10^{-3}$ min^{-1}. Despite these positive results, it has to be mentioned that there is still significant amount of degradation of precursor RNA, spliced introns, ligated exons and free 3'- and 5' exon taking place in reactions containing Mss116p/ΔMLS. This is mainly caused by RNases that are still present in the recombinant protein Mss116p/ΔMLS preparation.

$y = m1-m3*(1-exp(-m2*m0))$		
	Value	Error
m1	1.0117	0.091586
m2	0.030424	0.0091382
m3	1.0339	0.1317
Chisq	0.017421	NA
R	0.9861	NA
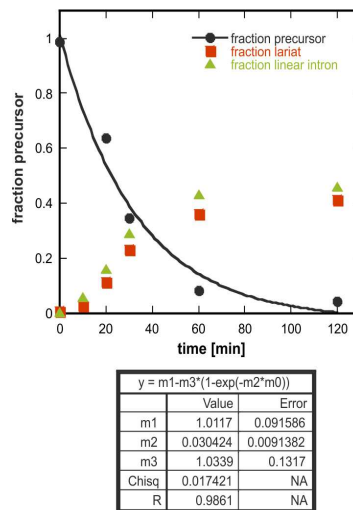
Figure 40: *in vitro* preRNA splicing by Mss116p/ΔMLS at near-physiological conditions. To evaluate the rate constant of splicing reaction the gel was quantified with ImageQuant and data was fit to a first-order rate equation with KaleidaGraph (has been described by (*5*)). A rate constant k_{obs} of $30 \pm 0.009 \times 10^{-3}$ min^{-1} could be calculated which is comparable to earlier kinetic studies on Mss116p (*6*). m1: value at time point 0; m2: k_{obs}, rate constant; m3: amplitude; R: correlation coefficient.

ATPase Coupled Unwinding of Short Duplex RNAs

Mss116p was shown to perform unwinding activity of short duplex RNA oligonucleotides (6-16 bps) in the presence of ATP with a rate constant k_{ATP} of 4.9 ± 1.1 ($x10^9$ M^{-1} min^{-1}) (6). Reactions were carried out at saturating Mss116p conditions and rate constants for blunt end duplexes had been shown to be similar to that of duplexes with a 3' or 5' overhang. It is estimated that a minimum of 1.4 ATP molecules are necessary for strand displacement (171). For showing Mss116p duplex-unwinding, reactions were performed on the basis of previously published protocols (6, 38). Mss116p/ΔMLS showed no unwinding activity with reactions including 0.1nM duplex, 400nM Mss116p at 100mM KCl, 40mM MOPS pH 7.5, 2mM DTT, 0.01%IGEPAL conditions started with 4mM $MgCl_2$/ATP mix.
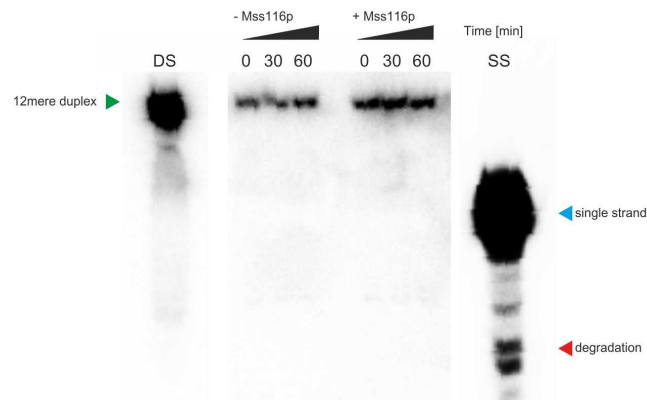
Figure 41: 12mere blunt end duplex in the absence and presence of Mss116p. (DS) 12mere duplex serving as a negative control; (SS) heat denatured control indicating position of labeled single strand. Unwinding was performed at 30°C and reaction started with 1mM $MgCl_2$/ATP.

As reaction should not include magnesium concentrations greater than 1mM (6), because Mg^{2+} ions stabilize RNA duplexes and decrease functional binding of Mss116p, a 1mM $MgCl_2$/ATP mix was used. Again no unwinding activity was detectable. Subsequently, we modified the protocol according to the recently published unwinding protocol by Jankowsky and Putnam 2010 (68). To prevent reformation of duplex after unwinding we tried to capture the labeled complementary strand after unwinding by adding a 2000-fold molar excess of unlabeled top strand directly into

unwinding reaction mixture. Additionally reaction was stopped with an increased amount of EDTA (from 50mM up to 200mM) and withdrawn aliquots were immediately put on ice or, alternatively, loaded immediately on a running native gel. However, unwinding could not be detected with either protocol (Figure 41). As a result, it is uncertain whether our Mss116p/ΔMLS preparation is capable of unwinding short duplex RNAs or if the problem resides in how we perform the unwinding assay. After personal communication with Dr. Fedorova (Yale University), who is familiar with this type of assay and Mss116p, performing an unwinding assay with Mss116p is everything but trivial. In other words, it is likely that our preparation will be active in unwinding and the assay needs further troubleshooting and experience to be successfully performed.

RNA-Dependent Activation of Mss116p ATPase Activity

RNA enhances the slow intrinsic ATPase activity of Mss116p ~7 fold by accelerating rate limiting ATP hydrolysis (*171*). The ATP hydrolysis of Mss116p progressively increases with increasing amounts of RNA bound to protein and at saturating ATP concentrations Mss116p shows different ATPase activity for different RNAs (*6*). For *Sc*. ai5γ ATPase activity was measured with a k_{obs} of 112 ± 6 (min^{-1}) and no activity was detectable in the absence of RNA (*6*). For determining the RNA-dependent ATPase activity of Mss116p/ΔMLS *in vitro*, the concentration of the free inorganic phosphate was measured by complexation with malachite green and ammonium molybdate under acidic conditions. The centration of this complex can be measured at 620nm and is directly related to levels of ATP hydrolyzed by Mss116p/ΔMLS. Reaction conditions were chosen on basis of previous published protocols (*6, 38*). Nevertheless, the initial protocol had to be optimized due to the occurrence of precipitate in reaction mixture. Consequently, we diluted the reaction volume 1:10, thereby improving the stability of the assay and decreasing the amount of precipitate. Furthermore, the Mss116p/ΔMLS protein had to be hydrolyzed or inactivated before measurement (suggested by manufacturer). To hydrolyze the Mss116p/ΔMLS protein it

was suggested to heat the reaction mixture at 100°C for 30min. However, incubating ATP at temperatures above 56°C in the presence of Mg^{2+} induces its hydrolysis as well. Therefore, we modified the manufacturer's instructions and tried to degrade protein with denaturing agents. First proteinase K was added but caused substantial precipitation within reaction. Proteinase K storage buffer includes 40% glycerol and DTT, which both reduce the sensitivity of reaction. In a different approach EDTA (10mM) was used to stop ATP hydrolysis by Mss116p/ΔMLS, but adding EDTA caused precipitation as well. At next we used β-mercaptoethanol to terminate ATP hydrolyzing reaction by Mss116p/ΔMLS. Again, this substance had a negative effect causing precipitation in the reaction mixture. Screening the literature we found in an earlier publication that proteins must be removed or stabilized prior to Pi determination with 0.25N sulfuric acid or 3% (w/w) perchloric

A

B

$y = 0.80566 - 0.00010129x \quad R = 0.10234$

$y = m1+m3*(1-exp(-m2*m0))$		
	Value	Error
m1	0.68386	0.10623
m2	0.13222	0.037693
m3	1.2263	0.1237
Chisq	0.022592	NA
R	0.99001	NA

$y = m1+m3*(1-exp(-m2*m0))$		
	Value	Error
m1	0.65656	0.069405
m2	0.039726	0.0085617
m3	1.1521	0.091657
Chisq	0.010969	NA
R	0.99407	NA

$y = m1+m3*(1-exp(-m2*m0))$		
	Value	Error
m1	0.62456	0.08227
m2	0.016226	0.0065019
m3	1.4037	0.23841
Chisq	0.018212	NA
R	0.99054	NA

Figure 42: ATP hydrolyzed by Mss116p. (A) black: ATP hydrolysis without Mss116p; green: ATP hydrolysis in the presence of Mss116p; red: ATP hydrolysis in the presence of Mss116p and 12mere RNA; blue: hydrolysis by Mss116p in the presence of ai5γ full length intron. (B) Illustrating reaction parameters. m1: value at time point 0; m2: k_{obs}, rate constant; m3: amplitude; R: correlation coefficient

acid (*172*). As Malachite Green Reagent A contains 3M sulfuric acid, we assumed that no specific step to stop the reaction was necessary. Indeed, we successfully tested this hypothesis and proceeded with the assay.

The ATPase activity of Mss116p/ΔMLS was determined in the absence of RNA as well as in the presence of a short single-stranded RNA (12mer) and the large, full-length ai5γ intron RNA at 30°C, pH 7.5 and the presence of 500mM KCl and 10mM $MgCl_2$. Plotting the fraction of Pi (and in turn of ADP) as a function of time, we derived observed rate constants k_{obs} of 0.132; 0.039 and 0.016 (min⁻¹) in the presence of full-length ai5γ intron RNA, of 12mere RNA

oligonucleotide and without RNA, respectively. In the presence of full length ai5γ Mss116p/ΔMLS shows ~8 fold higher ATPase activity than in the absence of RNA. The presence of a 12mer RNA oligonucleotide enhances the ATPase activity of Mss116p/ΔMLS by ~2.4 fold (Figure 42). This is in excellent agreement with previously published data (*6, 171*).

Towards Assessing the Influence of Mss116p on Metal Ion Binding Sites in the Collapsed Ai5γ Intron

Established a metal-induced cleavage assay – Pb^{2+} and Tb^{3+} Cleavage Reactions

At first, to examine optimal concentration range for performing Lead-induced cleavage reactions a concentration series from 1mM to 5mM was performed on the full-length ai5γ intron RNA folded at high-salt conditions (known to promote self-splicing). Cleavage reactions with 5mM $Pb(OAc)_2$ showed best results and therefore this concentration was chosen for subsequent metal-ion induced cleavage reactions. Importantly, since the intron RNA is rather large (886 nts) and we aimed to compare our results with those from studying metal ion binding sites *in vivo* using Pb^{2+}-induced cleavage (*173*), the cleavage sites were mapped using primer extension analysis. Thus, a control in which RNA that was not cleaved by lead was reverse transcribed, was performed to distinguish between natural stops of the reverse transcriptase and lead-induced stops. However, numerous RT-stops were observed in this negative control in the absence of lead (0mM Pb^{2+}). Therefore entire lead-induced cleavage assay including the primer extension protocol had to be optimized to reduce background. First, cleavage reactions were quenched with an increased amount of EDTA (from 100mM to 200mM) but had no positive effect on RNA stability. Checking pH of reactions in the presence and absence of Pb^{2+} revealed a pH of 7.0 in all mixtures excluding the possibility that an increased pH could have caused adverse effects in the negative controls. An increased amount of RNase inhibitor within reaction mixture did not improve the data

117

as well. With increased experience, I was able to detect metal ion binding sites within D1 of the full length intron (Figure 43). Weak Pb^{2+} cleavage could be observed close to the θ interaction site, around nucleotides A125-A128 and at the A139 in the internal loop of the D1c2 branch. Stronger Pb^{2+} cleavage could be detected at the λ and ε' regions in the D1c1 stem and within the regions flanking the β element. Despite the difficulties with the assay, the observed sites correlate nicely with previously published data (*126, 147*).
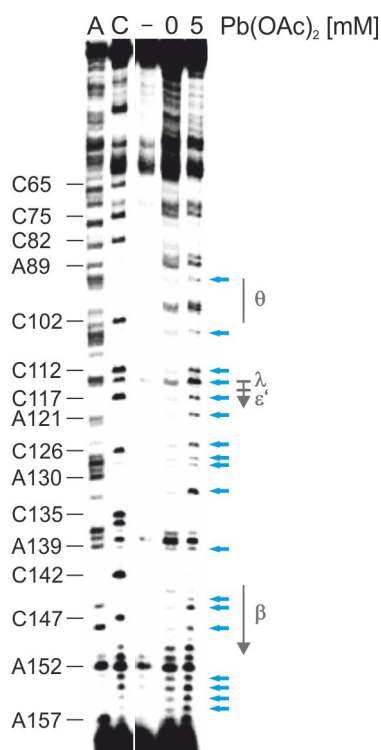
Figure 43: Pb^{2+} cleavage reactions within D1. Schematic views lead cleavage sites within full length ai5γ intron under high-salt conditions (100mM $MgCl_2$, 500mM KCl and 42°C) which were visualized over reverse transcription with end labeled *Sc160* primer and resolved on an 8% polyacrylamide gel. Strong cleavage occurs around ε' and λ regions in the D1c1 stem

As Tb^{3+} has a smaller ionic radius (0.92 Å) than Pb^{2+} (1.20 Å), it has been of advantage to use terbium as a probe due to its radius being closer to that of Mg^{2+} (0.72 Å) (*147*). Therefore, we decided to switch to a Tb^{3+}-induced cleavage assay to improve the signal-to-noise ratio when mapping the cleavage sites by reverse transcription. To examine optimal Tb^{3+} concentration for cleavage reactions different concentrations 1, 2.5, and 5mM were tested. It could be observed

that cleavage with 5mM Tb^{3+} caused increased degradation of RNA and is therefore not suitable for *in vitro* cleavage reactions (data not shown). At last we decided to simply try and compare metal ion binding sites at near-physiological conditions in the absence and presence of Mss116p/ΔMLS (Figure 44). It is evident that only very few cleavage sites were observed and not significant differences are visible between the reactions with or without Mss116p/ΔMLS. By performing the metal-induced cleavage reactions at lower temperature (4°C for 1 hour) and using snap cooling during reverse transcription only showed only a minor improvement in mapping the cleavage sites by primer extension.

As a result only a few cleavage sites could be detected over Tb^{3+} cleavage reactions (Figure 44). Cleavage was observed at nucleotides A155-A157 near β interaction site. Weak cleavage was observed as well at position A339 and A351, but cleavage sites are ambiguous, as they are only clearly visible in reactions containing 2mM Tb^{3+}. Within EBS1 RNA degradation can be observed within all reactions and are caused by the addition of Mss116p/ΔMLS. In other words the problems are multi-fold. On the one hand, Mss116p/ΔMLS preparation still contains some RNases and, second, working with such a large complex RNA like the ai5γ intron requires a lot of experience to avoid a strong background in the negative control of the primer extension (0mM Tb^{3+} lanes), making it difficult to map metal-induced cleavage sites. In sum metal-induced cleavage reactions are in need of further improvement to map metal ion binding sites within the ai5γ intron by reverse transcription.
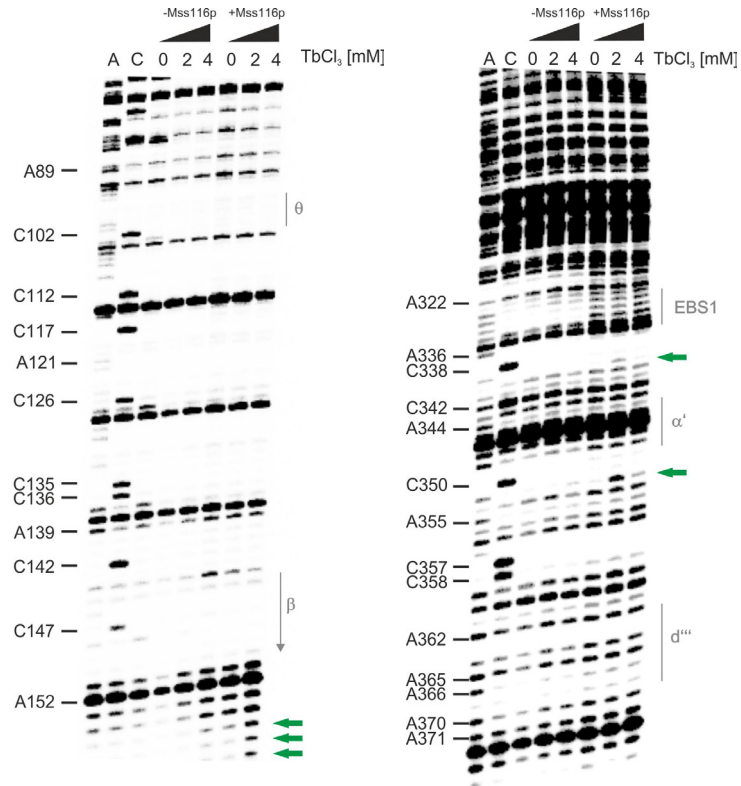
Figure 44: Tb³⁺ cleavage site throughout D1 of full length ai5γ ribozyme. Each schematic views cleavage pattern in collapsed (-Mss116p) and native intron conformation (+Mss116p).

Discussion

We expressed the yeast DEAD-box protein Mss116p in *E. coli* and characterized the biochemical activity of the purified protein. The recombinant Mss116p/ΔMLS protein was active and promoted intron splicing and hydrolyzed ATP comparable to previous reports (*6, 171*). Afterwards we attempted to map metal ion binding sites within the ai5γ group II intron was mapped over by performing a metal ion cleavage reaction in the absence and presence of Mss116p/ΔMLS under near physiological conditions, yielding the collapsed and native intron structure *in vitro*, respectively. It was our goal to examine whether Mss116p is able to take over the role of Mg^{2+} ions in certain structural elements of the intron. However, due to time constraints as well as the inherent difficulty of successfully applying this approach to large RNAs, additional experiments are required to determine whether Mss116p indeed influence metal ion binding sites *in vitro*. These data are then compared with those metal ion binding sites observed previously in the native ai5γ intron conformation *in vitro* (*147*) as well as with those found *in vivo* in the spliced intron which maintains is functional conformation (*80, 173*). This comparison allows us to determine whether Mss116p facilitates formation of the folding control element, which is captured by Mg^{2+} ions *in vitro* at high salt conditions and in the absence of Mss116p, and in turn accelerates the rate constant for the collapse (*41, 126, 173*).

Producing Recombinant Mss116p/ΔMLS

Purification of untagged Mss116p/ΔMLS posed numerous challenges, like eliminating unwanted impurities as well as stabilizing the protein after removing the tag and preserving protein activity. The majority of impurities could be removed by applying more stringent washing steps including increased imidazole concentrations to an extent that subsequent assays requiring high levels of RNA and its stability could be performed and. The second challenge was to stabilize protein and preserving its native structure. After the tag has been cleaved off, the protein precipitated; this

adverse effect could be largely diminished by adding 10% glycerol, L-arginine and L-glutamate during the SUMO cleavage reaction. Together with an N-terminal serine, which was inserted into sequence protein solubility can be increased (Invitrogen). After removing the affinity tag the untagged Mss116p/ΔMLS protein exposed a significant affinity for the nickel-charged resin and was hardly removable by imidazole and high salt gradients. A tempting point of view is now that such affinity tags can have an effect on structure and function of the protein (*174*). Comparative studies from tagged and untagged protein structures revealed however only minor structural differences and these differ from protein to protein (*175*). Also, it has been found that C-terminal His-tags have more frequently a negative effect on protein solubility than N-terminal His-tags (*176*), suggesting that tag has no profound effect on Mss116p native structure. A reason for the high affinity to nickel-charged resin therefore might be that protein gets tightly bound to the matrix through its 11 histidines dispersed within its primary sequence where most histidines residues are exposed at the protein surface. For future studies we suggest optimization of gel filtration to further improve protein quality.

Is the recombinant Mss116p/ΔMLS Protein Active?

We showed that untagged Mss116p/ΔMLS promotes group II intron splicing reactions at near physiological conditions in an ATP-dependent manner comparative to activity determined in earlier studies (*6*). However, slight RNA degradation is still observed during assays due to RNases present in the protein preparation. To further improve the splicing assays we suggest slightly increasing the amount of RNase inhibitor (<10% of glycerol v/v) or improving quality of the Mss116p/ΔMLS protein sample by applying a gel filtration prior to dialysis. In spite of the successful stimulation of intron splicing, unwinding of short RNA duplex by our untagged Mss116p/ΔMLS preparation in the presence of ATP could not be detected. In a previous study a rate constant k_{ATP} of 4.9 ± 1.1 (x 10^9 M^{-1} min^{-1}) was measured for a 16mer blunt-end substrate

(*6*). Like many other DEAD-box proteins, Mss116p does not display a preference for single stranded RNA regions or particular orientations therefore a 12mer blunt-end substrate was chosen in our unwinding studies (*38*). The rate constant should be even higher as activity increases with shorter duplexes (*6*). Notably, Mss116p and other DEAD-box proteins are known to be poor, non-processive RNA helicases (*68*). However, it has been proposed that because Mss116p unwinding is coupled with ATP and RNA binding, this raises the possibility that ATP binding can promote local unwinding of the bound duplex RNA, suggesting a coupled mechanism between ATP and RNA binding (*171*). Consequently, the lack of unwinding activity could simply imply the need for further optimization of the activity assays. Secondly, it remains possible that Mss116p could also be inactivated by impurities, which are present in the protein preparation and which block RNA or ATP binding activity. Interestingly, Mss116p/ΔMLS displayed an ATPase activity, which is ~8 fold increased in the presence of full length ai5γ intron RNA. Cao et al. revealed that RNA accelerates the Mss116p steady-state ATPase activity ~7 fold by promoting rate-limiting ATP hydrolysis (*171*). Current models suggest that Mss116p bound to ATP and ADP-Pi strongly binds and reorganizes RNA structure; Pi release promotes weak RNA binding, thereby dissociating the complex and allowing intron self-splicing and regenerating free Mss116p for additional rounds of catalysis (*81, 157, 171*). For Mss116p and DbpA, the ATPase cycling is completely characterized at kinetic and thermodynamic level and the reaction pathway seems to be a conserved one (*171*). Therefore additional mechanistic details of kinetic and equilibrium constants of wild-type Mss116p are could be helpful in understanding Mss116p-promoted intron splicing.

Mapping Metal Ion Binding Sites within the ai5γ Intron

D1 of ai5γ group II intron harbors several important long-tertiary interactions as well as secondary structure elements like bulges. The β-β' interaction between positions G145-G149 in the c2 stem and C254-C259 at d2a is known to be important during intron compaction (*126*). We observed Pb^{2+}-induced cleavage at residues close to the β element within the natively folded ai5γ intron (D123456) supporting the probable presence of a metal ion pocket. This correlates well with the weak Tb^{3+} cleavage observed at residues flanking the β region and medium to strong Tb^{3+} cleavage at nucleotides flanking β' region within the D135 ribozyme folded to the native state *in vitro (147)*. Thus, the presence of D2 and D4 in the D123456 construct (compared to D135) has no influence on this particular metal ion binding site. In contrast, recent *in vivo* lead cleavage assays (Wildauer & Waldsich, unpublished) did not detect cleavage within the β-β' intra-domain interaction in the presence of Mss116p. This potentially indicates an influence of the Mss116p protein on this metal ion binding site. Similarly, weak lead cleavage was detected close to the θ interaction site in both D135 (*147*) and D123456 ribozymes, while cleavage sites were not visible *in vivo (80)*. The θ-θ' tetraloop-receptor interaction is suggested to play a role in stabilization of the intron core and in recruiting the catalytic effector D3 together with J2/3 to the active site (*128*). Additional Pb^{2+} cleavages were detected around ε' and λ regions in the D1c1 stem of the native ai5γ intron (D123456), which had been also observed in the native D135 ribozyme (*147*). The subdomain IC1 of D1 has been proposed to serve as a first-step-specific RNA receptor for the branchpoint-carrying D6 substructure and conserved elements like the ε-ε' and λ-λ' interaction (*177*) and therefore represents a highly interesting domain element (*123*). While no lead-induced cleavage was observed at the ε' and λ sites *in vivo*, cleavage within Jc1/c2 was detected *in vivo* as well (*173*).

Since the Pb^{2+}-induced cleavage had a low signal-to-noise-ratio, we tried to improve the data by employing Tb^{3+} cleavage to map the metal ion binding sites within the ai5γ intron in the absence

and presence of Mss116p, thereby shedding further light on its mechanism of action. However, this endeavor was not very successful, as *in vitro* the cleavages might be hidden due to high RNA degradation (even in the absence of Mss116p). Due to the time constraints this project remains to be finished (see outlook below). At last, it is noteworthy to mention that some metal-ion binding sites might remain undetected mainly within RNA helices (major groove of G-C rich duplexes or GU wobble pairs). This is inherent to the assay, as Pb^{2+} and Tb^{3+} mediate RNA hydrolysis through the formation of hydroxides, which cleave RNA dependent on geometry (2'OH in minor groove) (*147, 148*). Additional biochemical characterization of Mss116p-K158E protein, which carries a mutation in the ATPase active site, thereby showing no detectable ATPase activity, would be a further step as well as mapping the structural effect of the ATP-deficient mutant to the ai5γ group II intron.

Conclusion and Outlook

With my work I offered a purification protocol for recombinant Mss116p, biochemical assays to validate Mss116p activity and set the first milestone for metal-ion cleavage reactions *in vitro*. Visualizing cleavage sites by reverse transcription is very sensible and requires as little as possible of background RNA degradation. As a matter of fact many cleavages may remain undetected and at this point there is significant room for improvement. Also, it might be worth considering to map the metal-induced cleavage products directly by end labeling of the full length intron, as it has been done in previous studies with the truncated ai5γ group II ribozyme (D135) (*126, 147*). Notably, mapping the full length ribozyme using end labeled RNA is a challenge due to the size of the full-length intron, which is likely to result in the need of an internal site-specific label of the intron to be able to map the entire RNA. In any event, to understand the role of Mss116p during the intron collapse, it is necessary to study the metal ion binding sites in the collapsed ai5γ intron in the absence and presence of Mss116p in order to determine whether Mss116p compensates for the high Mg^{2+} requirements during compaction, e.g. by facilitating formation of the folding control elements, which is very slow even at high salt and then captured by Mg^{2+} ions, and in turn accelerating the collapse. Since *in vivo* the metal ion binding sites in the presence of Mss116p are very similar to those observed *in vitro* for the native D135 ribozyme, this implies that Mss116p does not influence metal ion binding in the native ai5γ intron. Thus, in the future it would be highly interesting to either produce recombinant Mss116p carrying the K158R mutation to study metal ion binding sites within the collapsed ai5γ intron that remains bound to the ATPase-deficient Mss116p mutant *in vitro* or to apply the *in vivo* Pb^{2+}-cleavage assay to a yeast strain encoding the K158R Mss116p variant.

References

1. P. Acharya, P. Cheruku, S. Chatterjee, S. Acharya, J. Chattopadhyaya, Measurement of Nucleobase pKa Values in Model Mononucleotides Shows RNA-RNA Duplexes To Be More Stable than DNA-DNA Duplexes. *JACS Articles* **126**, 2862 (2004).

2. J. F. Atkins, R. F. Gesteland, T. R. Cech, Eds., *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*, vol. IV (2011), vol. IV.

3. R. Russell *et al.*, Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci. USA* **99**, 155 (2002).

4. N. B. Leontis, E. Westhof, Geometric nomenclature and classification of RNA base pairs. *Rna* **7**, 499 (Apr, 2001).

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

6. C. Halls *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. *Journal of Molecular Biology* **365**, 835 (2007).

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

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

9. A. J. Zaug, M. D. Been, T. R. Cech, The *Tetrahymena* ribozyme acts like an RNA restriction enzyme. *Nature* **334**, 429 (1986).

10. P. B. Moore, T. A. Steitz, The Structural Basis of Large Ribosomal Subunit Function. *Annual Review of Biochemistry* **72**, 813 (2003).

11. J. H. Cate, M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, H. F. Noller, X-ray crystal structure of 70S ribosome functional complexes. *Science* **285**, 2095 (1999).

12. B. Zhang, T. R. Cech, Peptide bond formation by in vitro selected ribozymes. *Nature* **390**, 96 (1997).

13. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **292**, 1319 (2001).

14. M. P. Robertson, G. F. Joyce. (CSH Laboratory Press, New York, 2010).

15. S. A. Brenner, H.-J. Kim, Z. Yang, in *Cold Spring Harbor Laboratory Press*. (CSH Laboratory Press, New York, 2010).

16. J. P. Schrum, T. F. Zhu, J. W. Szostak, Eds., *The origins of cellular life. In RNA Worlds*, (CSH Laboratory Press, New York, 2010).

17. D. M. Shechner *et al.*, "Crystal Structure of the Catalytic Core of an RNA-polymerase Ribozyme" (2009).

18. E. C. Long, in *Bioorganic Chemistry: Nucleic Acids,* S. M. Hecht, Ed. (Oxford University Press, USA 1996), pp. 512.

19. R. T. Batey, R. P. Rambo, J. A. Dounda, Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed.* **38**, 2326 (1999).

20. N. C. Horton, B. C. Finzel, The Structure of an RNA/DNA Hybrid: A Substrate of the Ribonuclease Activity of HIV-1 Reverse Transcriptase. *Journal of Molecular Biology* **264**, 521 (1996).

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

22. M. Tamura, S. R. Holbrook, Sequence and structural conservation in RNA ribose zippers. *J Mol Biol* **320**, 455 (Jul 12, 2002).

23. H. W. Pley, D. S. Lindes, C. DeLuca-Flaherty, D. B. McKay, Crystals of a hammerhead ribozyme [published erratum appears in J Biol Chem 1994 Feb 11;269(6):4692]. *J Biol Chem* **268**, 19656 (1993).

24. P. B. Moore, Structural Motifs in RNA. *Annual Review of Biochemistry* **68**, 287 (1999).

25. E. Westhof, P. Auffinger, in *Encyclopedia of Analytical Chemistry,* R. A. Meyers, Ed. (John Wiley & Sons Ltd, Chichester, 2000).

26. R. Russell, D. Herschlag, Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway. *J Mol Biol* **308**, 839 (2001).

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

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

29. T. Coetzee, D. Herschlag, M. Belfort, *Escherichia coli* proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev.* **8**, 1575 (1994).

30. A. Zhang, V. Derbyshire, J. L. Salvo, M. Belfort, Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *Rna* **1**, 783 (1995).

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

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

33. W. Jiang, Y. Hou, M. Inouye, CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J Biol Chem* **272**, 196 (1997).

34.	B. K. Pannone, D. Xue, S. L. Wolin, A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *Embo J* **17**, 7442 (1998).

35.	M. Matsuura *et al.*, A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* **11**, 2910 (1997).

36.	M. G. Caprara, G. Mohr, A. M. Lambowitz, A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J. Mol. Biol.* **257**, 512 (1996).

37.	A. Gampel, T. R. Cech, Binding of the CBP2 protein to a yeast mitochondrial group I intron requires the catalytic core of the RNA. *Genes Dev.* **5**, 1870 (1991).

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

39.	Q. Yang, M. Del Campo, A. M. Lambowitz, E. Jankowsky, DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell* **28**, 253 (Oct 26, 2007).

40.	M. Del Campo *et al.*, Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol. Cell* **28**, 159 (Oct 12, 2007).

41.	N. Zingler, A. Solem, A. M. Pyle, Dual roles for the Mss116 cofactor during splicing of the ai5γ group II intron. *Nucleic Acids Research* **38**, 6602 (2010).

42.	D. Thirumalai, N. Lee, S. A. Woodson, D. Klimov, Early events in rna folding. *Annu Rev Phys Chem* **52**, 751 (2001).

43.	E. Koculi, N. K. Lee, D. Thirumalai, S. A. Woodson, Folding of the Tetrahymena ribozyme by polyamines: importance of counterion valence and size. *J Mol Biol* **341**, 27 (Jul 30, 2004).

44.	R. Russell *et al.*, Rapid compaction during RNA folding. *Proc. Natl. Acad. Sci. USA* **99**, 4266 (2002).

45.	R. Russell, I. S. Millett, S. Doniach, D. Herschlag, Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* **7**, 367 (2000).

46.	D. J. Klein, P. B. Moore, T. A. Steitz, The contribution of metal ions to the structural stability of the large ribosomal subunit. *Rna* **10**, 1366 (Sep, 2004).

47.	O. Fedorova, L. Julie Su, A. M. Pyle, Group II introns: highly specific endonucleases with modular structures and diverse catalytic functions. *Methods* **28**, 323 (Nov, 2002).

48.	G. Zemora, C. Waldsich, RNA folding in living cells. *RNA Biol.* **7(6):6**, 34 (2010).

49.	J. Pan, D. Thirumalai, S. A. Woodson, Folding of RNA involves parallel pathways. *J. Mol. Biol.* **273**, 7 (1997).

50.	J. Myers, C. Pace, J. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci 1995* **4**, 2138 (1995).

51.   V. M. Shelton, T. R. Sosnick, T. Pan, Applicability of urea in the thermodynamic analysis of secondary and tertiary RNA folding. *Biochemistry* **38**, 16831 (Dec 21, 1999).

52.   L. J. Su, M. Brenowitz, A. M. Pyle, An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *J. Mol. Biol.* **334**, 639 (Dec 5, 2003).

53.   D. U. Priyakumar, C. Hyeon, D. Thirumalai, A. D. Mac Kerell Jr., Urea destabilizes RNA by forming stacking interaction and multiple hydrogen bonds with nucleic acid bases. *Journal of the American Chemical Society* **131**, 17759 (2010).

54.   Y. Terui, K. Higashi, Y. Tabei, H. Tomitori, K. Yamamoto, Enhancement of the Synthesis of RpoE and StpA by Polyamines at the Level of Translation in Escherichia coli under Heat Shock Conditions. *Journal of Bacteriology* **191**, 5348 (2009).

55.   D. E. Draper, D. Grilley, A. M. Soto, Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **34**, 221 (2005).

56.   P. Auffinger, N. Grover, E. Westhof, Metal Ion Binding to RNA. *Met. Ions Life Sci.* **9**, 1 (2011).

57.   V. K. Misra, D. E. Draper, Mg2+ Binding to rRNA Revisited: The Nonlinear Poisson-Boltzmann Model. *Journal of Molecular Biology* **299**, 813 (2000).

58.   A. S. Petrov, G. R. Pack, G. Lamm, Calculations of Magnesium-Nucleic Acid Site Binding in Solution. *Journal of Physical Chemistry* **108**, 6072 (2004).

59.   A. Stein, D. M. Crothers, Conformational changes of transfer RNA. The role of magnesium(II). *Biochemistry* **15**, 160 (1976).

60.   V. K. Misra, R. Shiman, D. E. Draper, A Thermodynamic Framework for the Magnesium-Dependent Folding of RNA. *Biopolymers* **69**, 118 (2002).

61.   E. Ennifar, P. Walter, P. Duams, Cation-dependent cleavage of the duplex form of the subtype-B HIV-1 RNA dimerization initiation site. *Nucleic Acids Research* **38**, 5807 (2010).

62.   J. H. Cate, R. L. Hanna, J. A. Doudna, A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.* **4**, 553 (1997).

63.   D. E. Draper, RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys J* **95**, 5489 (Dec 15, 2008).

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

65.   I. Carter-O'Connell, D. Booth, B. Eason, N. Grover, Thermodynamic examination of trinulceotide bulged RNA in the context of HIV-1 TAR RNA. *RNA* **14**, 2550 (2008).

66.   P. Linder *et al.*, Birth of the D-E-A-D box. *Nature* **337**, 121 (1989).

67.   M. E. Fairman-Williams, U.-P. Guenther, E. Jankowsky, SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* **20**, 313 (2010).

68. E. P. Jankowsky, A., Duplex unwinding with DEAD-box proteins. *Methods Mol Biol.* , 245 (2010).

69. O. Cordin, J. Banroques, N. K. Tanner, P. Linder, The DEAD-box protein family of RNA helicases. *Gene* **367**, 17 (Feb 15, 2006).

70. I. Jarmoskaite, R. Russell, DEAD-box proteins as RNA helicases and chaperones. *John Wiley & Sons, Ltd.* **2**, 135 (2010).

71. X.-h. Liang, M. J. Fournier, The Helicase Has1p Is Required for snoRNA Release from Pre-rRNA. *Molecular and Cellular Biology* **26**, 7437 (2006).

72. M. Kos, D. Tollervey, The Putative RNA Helicase Is Required for Release of the U14 snoRNA from Preribosomes in Saccharomyces cerevisiae. *Molecular Cell* **20**, 53 (2005).

73. G. W. Rogers, J. W. F. Lima, W. C. Merrick, Further Characterization of the Helicase Activity of eIF4A. *Journal of Biological Chemistry* **276**, 12598 (2001).

74. P. Tijerina, H. Bhaskaran, R. Russell, Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *PNAS* **103**, 16698 (2006).

75. O. Cordin, N. K. Tanner, M. Doere, P. Linder, J. Banroques, The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *Embo J* **23**, 2478 (Jul 7, 2004).

76. A. Pause, N. Sonenberg, Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *Embo J* **11**, 2643 (Jul, 1992).

77. Y. V. Svitkin *et al.*, The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7**, 382 (2001).

78. J. Banroques, M. Doère, M. Dreyfus, P. Linder, K. N. Tanner, Motif III in Superfamily 2 'Helicases' Helps Convert the Binding Energy of ATP into a High-Affinity RNA Binding Site in the Yeast DEAD-Box Protein Ded1. *Journal of Molecular Biology* **396**, 949 (2009).

79. H. R. Huang *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 (Jan 4, 2005).

80. A. Liebeg, O. Mayer, C. Waldsich, DEAD-box protein facilitated RNA folding *in vivo*. *RNA Biol.* **7**, 803 (2010).

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

82. N. Zingler, A. Solem, A. M. Pyle, Dual roles for the Mss116 cofactor during splicing of the ai5{gamma} group II intron. *Nucleic Acids*, 1 (2010).

83. G. Mohr *et al.*, High-Throughput Genetic Identification of Functionally Important Regions of the Yeast DEAD-Box Protein Mss116p. *Journal of Molecular Biology* **413**, 952 (2011).

84. A. L. Mallam *et al.*, Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *PNAS* **108**, 12254 (2011).

85. M. Del Campo, A. M. Lambowitz, Crystallization and preliminary X-ray diffraction of the DEAD-box protein Mss116p complexed with an RNA oligonucleotide and AMP-PNP. *Acta Crystallogr Sect F Struct Biol Cryst Commun.*, 832 (2009, 2009).

86. M. Del Campo, A. M. Lambowitz, Structure of the Yeast DEAD-box protein Mss116p reveals two wedges that crimp RNA. *Molecular Cell* **35**, 598 (2009).

87. P. Schütz *et al.*, Comparative Structural Analysis of Human DEAD-Box RNA Helicases. *PLoS* **5**, (2010).

88. C. B. F. Andersen *et al.*, Structure of the Exon Junction Core Complex with a Trapped DEAD-Box ATPase Bound to RNA. *Science* **29**, 1986 (2006).

89. J. K. Grohman *et al.*, Probing the Mechanisms of DEAD-box Proteins as General RNA Chaperones: The C-terminal Domain of CYT-19 Mediates General Recognition of RNA. *Biochemistry* **46**, 3013 (2007).

90. G. Mohr *et al.*, Function of the C-terminal Domain of the DEAD-Box Protein Mss116p Analyzed In Vivo and In Vitro. *Journal of Molecular Biology* **375**, 1344 (2008).

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

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

93. J. L. Ferat, F. Michel, Group II self-splicing introns in bacteria. *Nature* **364**, 358 (1993).

94. D. L. Daniels, W. J. Michels, Jr., A. M. Pyle, 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 (1996).

95. T. H. Eickbush, Retrohoming by complete reverse splicing. *Current Biology* **9**, 11 (1999).

96. T. Huang *et al.*, The group II intron ribonucleoprotein precursor is a large, loosley packed structure. *Nucleic Acids Research* **39**, 2845 (2010).

97. M. Matsuura, J. W. Noah, A. M. Lambowitz, Mechanism of maturase-promoted group II intron splicing. *Embo J* **20**, 7259 (Dec 17, 2001).

98. J. Kennell, J. Moran, P. Perlman, R. Butow, A. Lambowitz, Reverse Transcriptase Activity Associated with Maturase-encoding Group II Introns in Yeast Mitochondria. *Cell* **73**, 133 (1993).

99. H. Wank, J. SanFilippo, R. N. Singh, M. Matsuura, A. M. Lambowitz, A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell* **4**, 239 (1999).

100. E. V. Koonin, The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biology Direct* **1**, (2006).

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

102. R. W. Davies, R. B. Waring, J. A. Ray, T. A. Brown, C. Scazzochio, Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300**, 719 (1982).

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

104. N. Toor, G. Hausner, S. Zimmerly, Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**, 1142 (2001).

105. A. Solem, N. Zingler, A. M. Pyle, L.-P.-T. J., in *Non-Protein Coding RNAs,* N. G. Walter, S. A. Woodson, R. T. Batey, Eds. (Springer-Verlag, Berlin Heidelberg, 2009).

106. A. de Lencastre, A. M. Pyle, Three essential and conserved regions of the group II intron are proximal to the 5'-splice site. *Rna* **14**, 11 (Jan, 2008).

107. J. W. Noah, A. M. Lambowitz, Effects of maturase binding and Mg2+ concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry* **42**, 12466 (Nov 4, 2003).

108. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, Crystal structure of a self-spliced group II intron. *Science* **320**, 77 (Apr 4, 2008).

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

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

111. M. W. Müller, P. Stocker, M. Hetzer, R. J. Schweyen, Fate of the junction phosphate in alternating forward and reverse self-splicing reaction of Group II intron RNA. *JMB* **222**, 145 (1991).

112. M. Podar, P. S. Perlman, R. A. Padgett, Stereochemical selectivity of group II intron splicing, reverse-splicing and hydrolysis reactions. *Mol. Cell Biol.* **15**, 4466 (1995).

113. K. A. Jarrell, C. L. Peebles, R. C. Dietrich, S. L. Romiti, P. S. Perlman, Group II intron self-splicing: Alternative reaction conditions yield novel products. *J. Biol. Chem.* **263**, 3432 (1988).

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

115. O. Fedorova, N. Zingler, Group II introns: structure, folding and splicing mechanism. *Biol Chem* **388**, 665 (Jul, 2007).

116. M. D. Molina-Sánchez, F. Martinez-Abarca, N. Toro, Excision of the *Sinorhizobium meliloti* Group II Intron RmInt1 as Circles *in Vivo*. *Journal of Biological Chemistry* **281**, 28737 (2006).

117. H. D. Osiewacz, K. Esser, The mitochondrial plasmid of *Podospora anserina:* a mobile intron of a mitochondrial gene. *Curr. Genet.* **8**, 299 (1984).

118. F. Michel, M. Costa, E. Westhof, The ribozyme core of group II introns: a structure in want of partners. *Trends in Biochemical Sciences* **34**, 189 (2009).

119. W. J. Michels, Jr., A. M. Pyle, Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry* **34**, 2965 (1995).

120. P. Z. Qin, A. M. Pyle, 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 (1997).

121. K. T. Dayie, R. A. Padgett, A glimpse into the active site of a group II intron and maybe the spliceosome, too. *RNA* **14**, 1697 (2008).

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

123. M. Boudvillain, A. Delencastre, A. M. Pyle, 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 (2000).

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

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

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

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

128. O. Fedorova, T. Mitros, A. M. Pyle, Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330**, 197 (Jul 4, 2003).

129. C.-F. Li, M. Costa, F. Michel, Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *The EMBO Journal* **30**, 3040 (2011).

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

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

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

133.   M. Podar, S. Dib-Hajj, P. S. Perlman, A UV-induced Mg^{2+}-dependent cross-link traps an active form of Domain 3 of a self-splicing group II intron. *RNA* **1**, 828 (1995).

134.   O. Fedorova, A. M. Pyle, Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3. *Embo J* **24**, 3906 (Nov 16, 2005).

135.   A. M. Lambowitz, S. Zimmerly, Mobile group II introns. *Annu Rev Genet* **38**, 1 (2004).

136.   K. Jarrell, R. Dietrich, P. Perlman, Group II Intron Domain 5 facilitates a trans-splicing reaction. *Mol. Cell Biol.* **8**, 2361 (1988).

137.   C. L. Peebles, M. Zhang, P. S. Perlman, J. F. Franzen, Identification of a catalytically critical trinucleotide in Domain 5 of a group II intron. *PNAS* **92**, 4422 (1995).

138.   G. Chanfreau, A. Jacquier, Catalytic site components common to both splicing steps of a group II intron. *Science* **266**, 1383 (1994).

139.   P. M. Gordon, J. A. Piccirilli, Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat Struct Biol* **8**, 893 (Oct, 2001).

140.   S. L. Yean, G. Wuenschell, J. Termini, R. J. Lin, Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881 (Dec 14, 2000).

141.   H. D. Madhani, C. Guthrie, A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803 (1992).

142.   G. C. Shukla, R. A. Padgett, A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell* **9**, 1145 (May, 2002).

143.   K. S. Keating, N. Toor, P. S. Perlman, A. M. Pyle, A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**, 1 (2010).

144.   C. Waldsich, R. Schroeder, in *Handbook of RNA Biochemistry,* R. K. Hartmann, A. Bindereif, A. Schön, E. Westhof, Eds. (Wiley-VCH, Weinheim, 2005), vol. 1, pp. 229-237.

145.   J.-L. Wang, Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Biological Crystallography* **66**, 988 (2010).

146.   A. M. Pyle, The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.* **45**, 215 (2010).

147.   R. Sigel, A. Vaidya, A. Pyle, Metal ion binding sites in a group II intron core. *Nat Struct Biol* **7**, 1111 (2000).

148.   R. S. Brown, B. E. Hingerty, J. C. Dewan, A. Klug, Pb(II)-catalysed cleavage of the sugar-phosphate backbone of yeast tRNA[Phe] - implications for lead toxicity and self-splicing RNA. *Nature* **303**, 543 (1983).

149. C. Berens, B. Streicher, R. Schroeder, W. Hillen, Visualizing metal-ion-binding sites in group I introns by iron(II)- mediated Fenton reactions. *Chem. Biol.* **5**, 163 (1998).

150. R. D. Shannon, Revised effective ionic radii and systematic stdies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A* **32**, 751 (1976).

151. E. Nieboer, Lanthanide ions as structural probes in biological model systems. *Structure and Bonding* **22**, 1 (1975).

152. M. R. Stahley, S. A. Strobel, Structural Evidence for a Two-Metal-Ion Mechanism of Group I Intron Splicing. *Science* **309**, 1587 (2005).

153. S. A. Strobel, J. A. Doudna, RNA seeing double: close-packing of helices in RNA tertiary structure. *Trends Biochem. Sci.* **22**, 262 (1997).

154. C. Waldsich, A. M. Pyle, A kinetic intermediate that regulates proper folding of a group II intron RNA. *J. Mol. Biol.* **375**, 572 (Jan 11, 2008).

155. R. Hanna, J. A. Doudna, Metal ions in ribozyme folding and catalysis. *Curr Opin Chem Biol* **4**, 166 (2000).

156. O. Fedorova, C. Waldsich, A. M. Pyle, Group II intron folding under near-physiological conditions: collapsing to the near-native state. *J. Mol. Biol.* **366**, 1099 (Mar 2, 2007).

157. O. Fedorova, A. Solem, A. M. Pyle, Protein-facilitated folding of group II intron ribozymes. *J Mol Biol.* **3**, 799 (2010).

158. F. U. Hartl, M. Hayer-Hartl, Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852 (2002).

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

160. L. Rajkowitsch, R. Schroeder, Dissecting RNA chaperone activity. *Rna* **13**, 2053 (Dec, 2007).

161. X. Chen, R. R. Gutell, A. M. Lambowitz, Function of tyrosyl-tRNA synthetase in splicing group I introns: an induced-fit model for binding to the P4-P6 domain based on analysis of mutations at the junction of the P4-P6 stacked helices. *J Mol Biol* **301**, 265 (2000).

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

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

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

165. L. Rajkowitsch *et al.*, RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4**, 118 (Nov, 2007).

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

167. G. T. Robin, Walker D. C, M. R. R., E.coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research* **21**, 1677 (February 18, 1993, 1993).

168. E. Jankowsky, C. H. Gross, S. Shuman, A. M. Pyle, The DExH protein NPH-II is a processive and directional motor for unwinding RNA. *Nature* **403**, 447 (Jan 27, 2000).

169. M. Lindell, P. Romby, G. H. Wagner, Lead (II) as a probe for investigating RNA structure in vivo. *RNA* **8**, 534 (2002).

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

171. W. Cao *et al.*, Mechanism of Mss116 ATPase Reveals Functional Diversity of DEAD-Box Proteins. *Journal of Molecular Biology* **409**, 399 (2011).

172. T. P. Geladopoulos, T. G. Sotiroudis, A. E. Evangelopoulos, A malachite green colorimetric assay for protein phosphatase activity. *Analytical Biochemistry* **192**, 112 (1991).

173. W. Michael, C. Waldsich. (2010).

174. A. Chant, C. M. Kraemer-Pecore, R. Watkin, G. G. Kneale, Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-bining site. *Protein Expression and Purification* **39**, 152 (2005).

175. M. Carson, D. H. Johnson, H. McDonald, C. Brouillette, L. J. DeLucas, His-tag impact on structure. *Biological Crystallography* **63**, 295 (2007).

176. D. Busso, R. Kim, S.-H. KIm, Expression of soluble recombinant proteins in a cell-free system using a 96-well format. *Journal of Biochemical and Biophysical Methods* **55**, 233 (2003).

177. M. Costa, F. Michel, Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site. *EMBO J.* **18**, 1025 (1999).

Abstract

The self-splicing *Sc.* ai5γ group II intron has long been a powerful and representative model to study RNA folding *in vitro*. Like for other RNAs, metal ions play an essential role for ai5γ folding *in vitro* and in yeast mitochondria. In contrast to the intracellular ionic concentrations, the ai5γ intron depends on high salt conditions for splicing *in vitro.* Specifically, compaction is limited by the formation of an unstable folding intermediate that is captured by the binding of Mg^{2+}. Mg^{2+} ions are known to play a crucial role in promoting the slow collapse of the ai5γ ribozyme *in vitro*. It has however been observed that the nuclear-encoded DEAD-box helicase, which serves as a splicing factor for all yeast mitochondrial introns, is able to lower the Mg^{2+} requirements *in vitro* to near physiological concentrations. Since Mss116p has recently been described to accelerate the compaction of the ai5γ intron *in vitro,* we are interested in understanding whether Mss116p is able to stimulate intron folding by stabilizing the unstable, rate-limiting folding intermediate. The first step to accomplish this goal was to produce recombinant Mss116p protein. After establishing an appropriate expression and purification protocol for Mss116p we analyzed whether the recombinant protein is active. In particular, we demonstrated that the recombinant Mss116p is capable of hydrolyzing ATP stimulated by RNA and it promoted group II intron self-splicing at near-physiological ionic and temperature conditions, leaving little doubt to be a fully functional protein. At last, we established a method to probe the ai5γ intron structure in the absence and presence of Mss116p *in vitro* employing a metal ion-induced cleavage assay. Our preliminary data suggest further optimization of the Tb^{3+} induced cleavage assay is necessary in order to map metal-induced cleavage sites by means of reverse transcription and eventually to understand the role of Mss116p in the collapse of the ai5γ intron.

Zusammenfassung

Das selbst-spleißende *Sc.* Gruppe II Intron ai5γ stellt seit Langem ein überaus wichtiges und mächtiges Modelsystem für RNA Faltung *in vitro* dar. Wie für alle anderen RNAs, spielen Metall Ionen eine entscheidende Rolle im ai5γ *in vitro* Faltungsprozess und in den Hefemitochondrien. Im Gegensatz zu den vorherrschenden intrazellulären Ionenkonzentrationen, ist ai5γ *in vitro* auf hohe Salzkonzentrationen angewiesen um den selbst-spleiß Prozess durchzuführen. Besonders die Komprimierung des Introns ist limitiert von der Formation eines instabilen Faltungsintermediates, welches durch die Bindung von Mg^{2+} Ionen stabilisiert wird. Mg^{2+} Ionen sind bekannt dafür eine überaus wichtige Rolle in der Förderung des langsamen ai5γ Faltungskollapses *in vitro* zu spielen. Es wurde beobachtet, dass eine kernkodierte DEAD-box Helikase, welche als Spleißfaktor für alle mitochondrialen Introns in Hefe dient, die Mg^{2+} Anforderung *in vitro* auf zelluläre Konzentration senken kann. Kürzlich wurde Mss116p die Funktion zugewiesen, die Komprimierung von ai5γ *in vitro* zu beschleunigen. Deshalb sind wir überaus interessiert daran, ob Mss116p den Intron-Faltungsprozess über die Stabilisation des instabilen geschwindigkeitsbeschränkenden Faltungsintermediates stimuliert. Um dieses Ziel zu erreichen muss zuerst rekombinantes Mss116p Protein hergestellt werden. Nach Entwicklung eines adäquaten Expressions- und Aufreinigungsprotokolls wird Mss116p auf die Aktivität und Funktionalität überprüft. Im Detail zeigen wir, dass Mss116p im Stande ist ATP, in Abhängigkeit von RNA, zu hydrolysieren und dass es den ai5γ Gruppe II Intron Selbst-Spleiß-Prozess unter beinah-physiologischen Ionenkonzentrationen und Temperaturbedingungen fördert. Zuletzt wurde eine Methode entwickelt um die Struktur von ai5γ in Ab- und Anwesenheit von Mss116p *in vitro* zu ermitteln wobei wir Metall-Ionen-induzierte Schnitte als mächtiges Instrument zur Hilfe nehmen. Unsere bislang erlangten Daten deuten auf weitere wichtige Optimierungsschritte der Tb^{3+} Schnittbedingungen hin um die Metall-induzierten Schnittstellen über eine Reverse Transkription visualisieren zu können.

Curriculum Vitae

Personal information

First name(s) / Surname(s)	**Eva Maria Steiner**
Address(es)	Holzhäuser 13, 4632 Pichl/Wels, Austria
Mobile:	+43 (0)680 21 59 006
E-mail	eva.m.steiner@gmail.com
Nationality	Austrian
Date of birth	04.12.1984

Work experience

Dates	October 2010 – Present
Occupation or position held	Master Student at Max F. Perutz Laboratories (MFPL)
Main activities and responsibilities	'The Activity of recombinant DEAD-box Protein Mss116p: Towards its Role in the Collapse of the Ai5γ Group II Intron Protein purification and functionality assays, structural probing of RNA
Name and address of employer	Christina Waldsich Dr. Bohrgasse 9/5, 1030 Vienna, Austria

Dates	August 2010 – September 2010
Occupation or position held	Internship at MFPL
Main activities and responsibilities	'Do DEAD Box Proteins Assist *td* Group I Intron Splicing *in vivo*?' cloning, *in vivo* splicing assays
Name and address of employer	Christina Waldsich Dr. Bohrgasse 9/5, 1030 Vienna, Austria

Dates	September and December 2009
Occupation or position held	Internship at MFPL
Main activities and responsibilities	Mss116p Protein Purification
Name and address of employer	Christina Waldsich Dr. Bohrgasse 9/5, 1030 Vienna, Austria

Dates	October 2008 – November 2008
Occupation or position held	Bachelor Student at MFPL
Main activities and responsibilities	'Rad61 Promotes the Association of Meiotic Cohesin with Chromatin' FISH, yeast genomics
Name and address of employer	Franz Klein Dr.-Bohr-Gasse 1, 1030 Vienna, Austria

Education and training

Dates	April 2009 – Present
occupational skills covered	Master Student Biochemistry, Bioinformatics, Molecular Biology
Name and type of organisation	University of Vienna, Austria Received Performance-based Scholarship in 2010
Dates	2007 – 2009
Title of qualification awarded	BSc
occupational skills covered	Biology, Molecular Biology
Name and type of organisation	University of Vienna, Austria
Dates	2005 – 2007
occupational skills covered	Molecular Biology
Name and type of organisation	Paris Lodron University of Salzburg and Johannes Kepler University of Linz
Dates	2005 SS
occupational skills covered	Economy and Law
Name and type of organisation	Paris Lodron University of Salzburg
Dates	Oct. 2004 – Feb. 2005
Name and type of organisation	A1 Mobilkom Austria at Linz Taubenmarkt, Upper Austria Sales and Advisory Service
Dates	1999 – 2004
Name and type of organisation	Commercial Academy for Information Technology Kirchdorf an der Krems, Upper Austria
Dates	1995 – 1999
Name and type of organisation	Bundesrealgymnasium Kirchdorf an der Krems, Upper Austria
Dates	1990 – 1995
Name and type of organisation	Elementary School Kirchdorf an der Krems, Upper Austria

Personal skills and competences

Mother tongue(s)	**German**
Other language(s)	English (Proficient User), Italian (Independent User), Swedish (Basic User)
Social skills and competences	Reliable Communicative
Organisational skills and competences	Creative Well-organised - During my master thesis I supervised an RNA practical course at the FH Campus Vienna and students in the laboratory of Christina Waldsich
Computer skills and competences	- Good command of Microsoft Office™ tools (Word™, Excel™, Access™ and PowerPoint™); - Basic knowledge of graphic design applications (Adobe Products like Adobe Photoshop™); - Basic knowledge in programming (Perl and C++) - Familiar with Linux and Unix operating systems
Artistic skills and competences	Tennis Trainer License, Classical Voice Education - I regularly organized and held kids tennis lessons
Driving licence	B

References

Dr. Christina Waldsich
Max F. Perutz Laboratories
Department of Biochemistry & Cell Biology
Dr. Bohr-Gasse 9, 1030 Vienna | Room: 5.512/5.513
Mail: christina.waldsich@univie.ac.at
Phone: +43-1-4277-52818
Web: http://www.mfpl.ac.at/mfpl-group/group/waldsich.html

Prof. Dr. Franz Klein
Max F. Perutz Laboratories
Department of Chromosome Biology
Dr. Bohr-Gasse 1(VBC2), 1030 Vienna | Room: 1. OG 7
Mail: franz.klein@univie.ac.at
Phone: +43-1-4277-56220
Web: http://www.mfpl.ac.at/mfpl-group/group/klein.html