

**Function of the RNA helicase Sub2 from  
*Saccharomyces cerevisiae* and its  
regulation by Tho1**

**Dissertation**

Zur Erlangung des Doktorgrades der Naturwissenschaften

-Dr. rer. nat.-

am Fachbereich 08 für Biologie und Chemie

der Justus-Liebig-Universität Gießen

vorgelegt von

**Matthias Bastian Miosga**

Gießen, 2022

Die vorliegende Arbeit wurde am Institut für Biochemie (Fachbereich 08) der Justus-Liebig-Universität Gießen, unter Leitung von Apl. Prof. Dr. Peter Friedhoff, erstellt.

Dissertation eingereicht am: 07.03.2022

Erstgutachter: Apl. Prof. Dr. Peter Friedhoff  
Fachbereich 08: Biologie und Chemie  
Institut für Biochemie  
Justus-Liebig-Universität Giessen

Zweitgutachter: Prof. Dr. Roland K. Hartmann  
Fachbereich 16: Pharmazie  
Institut für Pharmazeutische Chemie  
Philipps Universität Marburg

# Table of contents

<b>Zusammenfassung.....</b>	<b>1</b>
<b>Abstract .....</b>	<b>3</b>
<b>Introduction .....</b>	<b>5</b>
Transcription and mRNA Export .....	5
DEAD box RNA helicases .....	6
Sub2/UAP56.....	7
Tho1.....	8
Aim of this study.....	10
<b>Material .....</b>	<b>11</b>
Chemicals and consumables .....	11
Equipment and devices.....	13
Buffers and media .....	15
Media.....	15
Purification .....	15
SDS-PAGE .....	17
Cloning.....	17
Yeast transformation .....	18
Dot blot assay .....	18
Enzymatic assays .....	18
<b>Work material .....</b>	<b>19</b>
Kits .....	19
Enzymes.....	19
Antibodies .....	19
Markers .....	20
Software .....	20
<b>Organisms .....</b>	<b>20</b>
<i>E. coli</i> strains .....	20
Yeast strains .....	21
<b>Oligonucleotides and plasmids .....</b>	<b>21</b>
Plasmids .....	21
Primers .....	23
Substrates.....	26
<b>Methods .....</b>	<b>28</b>
Production of Z-competent cells.....	28
Transformation.....	28

Plasmid preparation .....	28
Yeast viability assay (Dot spots) .....	29
Dot blot assay to determine level of R-Loops.....	30
Cloning and Mutagenesis .....	31
Purification of Sub2 .....	32
Purification of Tho1 .....	34
Gel electrophoresis (SDS-PAGE and agarose gel-electrophoresis).....	35
Bioconjugation.....	36
Annealing of Substrates.....	37
Electrophoretic Mobility Shift Assay (EMSA) .....	37
ATPase assay.....	38
Helicase assay.....	39
Annealing Assay .....	39
Structure and amino acid sequence analysis.....	40
<b>Results .....</b>	<b>41</b>
<b>Chapter 1: Characterization of the DECD-Box RNA helicase Sub2 from <i>S. cerevisiae</i> .....</b>	<b>41</b>
Development of suitable assays to investigate DEAD-Box helicases .....	41
Magnesium and ADP are inhibitory for all activities of Sub2 .....	44
Sub2 unwinds RNA/DNA hybrids .....	48
<b>Chapter 2: Sub2 has several conserved cysteines with unknown function.....</b>	<b>52</b>
Structural and sequential analysis of Sub2.....	52
Reducing agent is necessary to maintain Sub2's activity .....	53
Introducing additional cysteines to Sub2 .....	54
Deletion of all cysteines do not show an altered phenotype <i>in vivo</i> .....	57
<b>Chapter 3: Modulation of the activity of Sub2 by Tho1. ....</b>	<b>59</b>
Sub2 and Tho1 form a ternary complex on RNA .....	59
Tho1 strongly inhibits unwinding of partial duplex RNA by Sub2 .....	60
<b>Chapter 4: Tho1 has a novel annealing activity.....</b>	<b>62</b>
Discovery of a novel annealing activity of Tho1.....	62
Tho1 is a DNA annealing protein .....	63
Sub2 enhances the annealing activity of Tho1 .....	65
Annealing activity is conserved among Tho1's orthologues.....	67
<b>Chapter 5: Tho1 domain analysis .....</b>	<b>70</b>
Domain analysis of Tho1 to discover their functionalities.....	70
The RBD and the C-terminal end are mainly responsible for binding of nucleic acids .....	71
The C-terminal end plays an important role in annealing activity.....	72
Annealing activity correlates with inhibition of Sub2's unwinding activity .....	74
Tho1 interacts via its RNA-binding domain with Sub2.....	74

<b>Chapter 6: Sub2 and Tho1 might resolve R-Loops <i>in vitro</i> and <i>in vivo</i></b>	<b>77</b>
Tho1 is able to displace RNA-strands <i>in vitro</i>	77
Tho1 prevents R-Loop formation <i>in vivo</i>	79
<b>Discussion</b>	<b>81</b>
<b>Chapter 1: Characterization of the DECD-Box RNA helicase Sub2 from <i>S. cerevisiae</i></b>	<b>81</b>
Investigation of the activity of Sub2	81
Sub2 is a sensing protein	83
Substrate specificity of Sub2	83
<b>Chapter 2: Sub2 has several conserved cysteines with unknown function</b>	<b>85</b>
Introduction of additional cysteines	85
Function of cysteines	86
<b>Chapter 3: Modulation of the activity of Sub2 by Tho1</b>	<b>87</b>
Sub2 and Tho1 interact with each other	87
<b>Chapter 4: Tho1 has a novel annealing activity</b>	<b>88</b>
Detailed investigation of the accelerated annealing of nucleic acids by Tho1	88
Sub2 enhances the annealing activity of Tho1	90
<b>Chapter 5: Tho1 domain analysis</b>	<b>92</b>
The C-terminal end is needed for binding activity, the role of the RNA-binding domain remains unclear	92
The C-terminal end is crucial for the annealing activity	92
Tho1 interacts via its RNA-binding domain with Sub2	93
Inhibition of Sub2 correlates with annealing activity	95
The RNA-binding domain of Tho1 is necessary for stimulation by Sub2	95
Summary of the domain analysis	96
<b>Chapter 6: Sub2 and Tho1 might resolve R-Loops <i>in vitro</i> and <i>in vivo</i></b>	<b>97</b>
A model for R-Loop resolution by Sub2 and Tho1	98
<b>Conclusions</b>	<b>100</b>
<b>References</b>	<b>101</b>
<b>List of figures</b>	<b>108</b>
<b>List of tables</b>	<b>110</b>
<b>Abbreviations</b>	<b>111</b>
<b>Publications</b>	<b>114</b>
<b>Eidesstattliche Erklärung</b>	<b>115</b>
<b>Appendix</b>	<b>116</b>
<b>Quickguide</b>	<b>117</b>

# Zusammenfassung

Die Transkription von mRNA findet in Eukaryonten im Nukleus jeder Zelle statt. Bereits während der Transkription binden verschiedene Proteine die mRNA, um diese auf den Export aus dem Zellkern vorzubereiten, damit die Information des Genoms in ein Protein translatiert werden kann. Ein wesentlicher Bestandteil dieser Maschinerie ist der TREX Komplex, der zusammen mit der mRNA ein mRNP bildet und diese auf den Export vorbereitet. Dieser ist in Tieren, Pilzen und Pflanzen hochkonserviert und besteht in *Saccharomyces cerevisiae* aus dem heteropentamerischen THO Komplex und der RNA Helikase Sub2 (UAP56/DDX39 im Menschen). Zusätzlich werden weitere RNA-Bindungsproteine (RBPs) wie zum Beispiel das RNA-Annealing Protein Yra1 (AlyRef im Menschen) rekrutiert. Im Menschen interagiert zusätzlich noch CIP29/SARNP (in *Arabidopsis thaliana* MOS11) mit dem Komplex. Diese RBPs spielen eine wichtige Rolle für den mRNA Export, ihre genauen Funktionen sind jedoch weitestgehend nicht bekannt.

Sub2 ist eine konservierte DEAD-Box RNA Helikase aus *S. cerevisiae* und ist an Spleißen und mRNA Export beteiligt. Während die Deletion von Sub2 zum Zelltod führt, behebt die Überexpression den mRNA-Export Defekt eines nicht funktionierenden THO-Komplexes.

Obwohl Sub2 eine essenzielle Rolle beim mRNA-Export und Spleißen innehat, sind dessen biochemische Aktivitäten in diesen Prozessen weitgehend unklar. In dieser Arbeit wurden deshalb die biochemischen Eigenschaften von Sub2 charakterisiert und die Regulation durch Interaktionspartner detailliert *in vitro* untersucht. Für dieses Vorhaben wurden geeignete biochemische Echtzeit-Assays entwickelt. Um wichtige Konformationsänderung im Zuge der Regulation von Sub2 in Zukunft verfolgen zu können, wurden die Voraussetzungen für Förster-Resonanz-Energie-Transfer (FRET) Experimente geschaffen. Für die notwendige ortsspezifische Markierung von Sub2 wurden die vorhandenen, konservierten Cysteinreste ersetzt und zusätzliche Cysteinreste an strategischen Positionen auf der Oberfläche eingeführt. Die Funktion der konservierten Cysteine ist derzeit noch unklar, allerdings führt die cystein-freie Sub2 Variante *in vivo* zu keinem offensichtlichen Phänotyp. Erste Markierungsexperimente mit Fluoreszenzfarbstoffen zeigen, dass die Voraussetzungen für FRET Experimente geschaffen wurden.

Ausgangspunkt für die Untersuchungen zur Regulation von Sub2 durch Tho1 war die Tatsache, dass die humanen Orthologe UAP56 und CIP29/SARNP einen ATP-abhängigen Komplex bilden und zur Stimulierung der ATPase und Helikase-Aktivität führt. In dieser Arbeit konnte die ATP-abhängige ternäre Komplexbildung von Sub2 mit Tho1 und RNA nachgewiesen werden.

Allerdings inhibiert Tho1 die Helikase-Aktivität von Sub2, was durch eine bis dato unbekannte Annealingaktivität von Tho1 bewirkt wird. Diese neue Aktivität ist evolutionär konserviert und konnte auch für das humane CIP29/SARNP, als auch für das SARNP/Tho1 Ortholog aus dem thermophilen Pilz *Chaetomium thermophilum* nachgewiesen werden.

Durch eine eingehende Mutationsanalyse von Tho1 konnte zudem die konservierte RNA-Bindungs-Domäne als Interaktionsdomäne mit Sub2 experimentell identifiziert werden. Dieser Befund deckt sich mit der kürzlich in Science (Humphreys *et al.*, 2021) veröffentlichten Vorhersage der Strukturen von relevanten Protein-Komplexen in *S. cerevisiae*.

Mit Hilfe weitere biochemischer und zellbiologischer Assays konnten zudem erste Hinweise auf eine neue Funktion von Tho1 im mRNA Export gefunden werden. Während die Deletion von Tho1 keinen Phänotyp zeigt, behebt die Überexpression von Tho1 (ähnlich wie von Sub2) verschiedene Phänotypen (Hyperrekombination, Transkriptions- und mRNA Export Defekt) eines nicht funktionierenden THO-Komplexes. Der Phänotyp kann sich durch eine Anhäufung von R-Loops manifestieren und in der Tat verringert die Überexpression von Tho1 die Anzahl der R-Loops *in vivo*. *In vitro* Experimente zeigen, dass sowohl Sub2 als auch überraschenderweise Tho1 alleine bereits in der Lage sind, R-Loop ähnliche Strukturen aufzulösen. Während die Beseitigung von R-Loops durch Sub2 zu erwarten war, da das humane Ortholog UAP56/DDX39B bereits als wichtige kotranskriptionelle RNA/DNA-Helikase identifiziert wurde, ist die Rolle von Tho1 überraschend und neu. Inwiefern die Interaktion von Tho1 und Sub2 für diese Funktion relevant ist und ob diese evolutiv konserviert ist, bleibt jedoch abzuwarten.

# Abstract

In eukaryotes, the transcription of mRNA takes place in the nucleus of each cell. Already during transcription, various proteins bind the mRNA to prepare it for export from the nucleus so that the information of the genome can be translated into a protein. An essential part of this machinery is the TREX complex, which forms an mRNP together with the mRNA and prepares it for export. This is highly conserved in animals, fungi and plants and consists in *Saccharomyces cerevisiae* of the heteropentameric THO complex and the RNA helicase Sub2 (UAP56/DDX39 in humans). In addition, other RNA binding proteins (RBPs) such as the RNA annealing protein Yra1 (AlyRef in humans) are recruited. In humans, CIP29/SARNP (in *Arabidopsis thaliana* MOS11) also interacts with the complex. These RBPs play an important role in mRNA export, but their exact functions are not yet known.

Sub2 is a conserved DEAD-box RNA helicase from *S. cerevisiae* and is involved in splicing and mRNA export. While deletion of Sub2 leads to cell death, overexpression rescues the mRNA export defect of a non-functioning THO complex.

Although Sub2 has an essential role in mRNA export and splicing, its biochemical activities in these processes are largely unclear. In this work, therefore, the biochemical properties of Sub2 were characterised and its regulation by interaction partners was studied in detail *in vitro*. Suitable biochemical real-time assays were developed for this purpose. In order to be able to follow important conformational changes in the course of the regulation of Sub2 in the future, the prerequisites for Förster resonance energy transfer (FRET) experiments were created. For the necessary site-specific labelling of Sub2, the existing conserved cysteine residues were replaced and additional cysteine residues were introduced at strategic positions on the surface. The function of the conserved cysteines is currently unclear, but the cysteine-free Sub2 variant does not lead to an obvious phenotype *in vivo*. First labelling experiments with fluorescent dyes show that the prerequisites for FRET experiments have been created.

The starting point for the investigations on the regulation of Sub2 by Tho1 was the fact that the human orthologues UAP56 and CIP29/SARNP form an ATP-dependent complex which leads to the stimulation of ATPase and helicase activity. In this work, ATP-dependent ternary complex formation of Sub2 with Tho1 and RNA was demonstrated.

However, Tho1 inhibits the helicase activity of Sub2, which is caused by a previously unknown annealing activity of Tho1. This new activity is evolutionarily conserved and could also be



demonstrated for the human CIP29/SARNP, as well as the SARNP/Tho1 orthologue from the thermophilic fungus *Chaetomium thermophilum*.

Furthermore, a detailed mutational analysis of Tho1 experimentally identified the conserved RNA-binding domain as an interaction domain for Sub2. This finding is consistent with the recent prediction of the structures of relevant protein complexes in *S. cerevisiae* published in Science (Humphreys *et al.*, 2021).

With the help of further biochemical and cell biological assays, first indications for a new function of Tho1 in mRNA export could also be found. While deletion of Tho1 does not show any phenotype, overexpression of Tho1 (similar to Sub2) rescues several phenotypes (hyperrecombination, transcription and mRNA export defect) of a non-functioning THO complex. The phenotype can be manifested by an accumulation of R-loops and indeed overexpression of Tho1 decreases the number of R-loops *in vivo*. *In vitro* experiments show that both Sub2 and, surprisingly, Tho1 alone are already capable of resolving R-loop-like structures. While the resolution of R-loops by Sub2 was expected, as the human orthologue UAP56/DDX39B has already been identified as an important cotranscriptional RNA/DNA helicase, the role of Tho1 is surprising and novel. However, to what extent the interaction of Tho1 and Sub2 is relevant for this function and whether this is evolutionarily conserved is yet to be seen.

# Introduction

## Transcription and mRNA Export

The cell nucleus is the place where genetic information in the form of chromosomal DNA is located, genes are transcribed and RNA is processed. Since the RNA is only processed in the cell nucleus, but the protein is translated in the cytoplasm, the mRNA must be exported across the nuclear membrane into the cytoplasm (Katahira, 2012). The "gateway" between the cell nucleus and the cytoplasm is the nuclear pore complexes (NPC) (Bjork and Wieslander, 2014). The NPC also serves as a filter to prevent unnecessary exchange of macromolecules such as RNA or proteins (Hetzer and Wente, 2009). Some RNA molecules, such as tRNA or microRNAs, are recognised directly by the adaptor protein exportin via certain structures, such as T-loops or double-stranded regions, and are exported from the cell nucleus (Katahira, 2012). Compared to tRNA or microRNAs, mRNAs have a wide range of length, sequence and structure. For this reason, other strategies must be applied in order to selectively transport mRNA out of the cell nucleus (Wickramasinghe and Laskey, 2015).

An important step in gene expression is the formation of a messenger ribonucleoprotein particle (mRNP). The messenger RNA (mRNA) is synthesized by RNA polymerase II (RNAPII) transcribing a protein-coding gene. The mRNA is processed, i.e., capped, spliced and polyadenylated, and packaged into an mRNP by binding of RNA-binding proteins (RBPs) (Wende *et al.*, 2019). The evolutionarily conserved TRanscript-EXport (TREX) complex has key functions during many steps from mRNP formation to export from the nucleus to the cytoplasm (Chavez *et al.*, 2000; Katahira *et al.*, 2002; Strasser and Hurt, 2001). In *Saccharomyces cerevisiae*, the TREX complex is formed by the heteropentameric THO complex (Hpr1, Mft1, Tex1, Tho2, and Thp2) together with the DEAD-box RNA helicase Sub2, the yeast RNA annealing protein Yra1 and the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 (Hurt *et al.*, 2004; Strasser *et al.*, 2002).

After the TREX complex is formed, Mex67-Mtr2, a heterodimer, can be recruited by ubiquitinated Hpr1 (Gwizdek *et al.*, 2005; Strasser *et al.*, 2000). Activity of Sub2 and Yra1 allows Mex67-Mtr2 to be loaded onto the mRNA (Gwizdek *et al.*, 2006). Mex67-Mtr2 can subsequently be recognised by other proteins of the nuclear pore complex, so-called nucleoporins, and initiates the export of mRNA from the nucleus into the cytoplasm (Katahira, 2012; Strasser *et al.*, 2000). Finally, the mRNA is exported from the nucleus and can be translated in the cytoplasm (Katahira, 2012).

In humans and *Arabidopsis thaliana* additional proteins are recruited to the TREX complex via interaction with the Sub2 orthologues UAP56/DDX39B, e.g. the SAP domain containing ribonucleoprotein (SARNP) CIP29 in humans or its orthologue Mos11 in *Arabidopsis thaliana*, which lacks the SAP domain (Dufu *et al.*, 2010; Sørensen *et al.*, 2017). In *S. cerevisiae*, the orthologue of SARNP/CIP29 Tho1 has been shown in high-throughput two-hybrid analysis to interact with Sub2 (Ito *et al.*, 2001) and may also be a component of the yeast TREX complex (Jimeno *et al.*, 2006). Additional TREX interactors are the mammalian protein ZC3H11A and ALYREF-like proteins UIF, LUZP4, POLDIP3, and CHTOP (Chang *et al.*, 2013; Folco *et al.*, 2012; Heath *et al.*, 2016; Younis *et al.*, 2018).

Defects in the THO complex cause several phenotypes, e.g. transcriptional defects, mRNA export defects and hyper-recombination (Chavez *et al.*, 2001; Chavez and Aguilera, 1997; Jimeno *et al.*, 2002). Interestingly, these phenotypes can be suppressed by overexpression of either the DEAD-box RNA helicase Sub2 or Tho1 (Jimeno *et al.*, 2006). These phenotypes are caused by the cotranscriptional formation of R-loops (RNA:DNA hybrids) between the nascent mRNA and the template DNA (Luna *et al.*, 2019).

Between transcription and export of the mRNA lies the step of splicing, in which the introns are cut out of the mRNA. However, it is still unclear whether the recruitment of the TREX complex occurs independently of splicing, or whether there is a dependency (Bjork and Wieslander, 2017; Katahira, 2012). Both possibilities were described, transcription-dependent recruitment of the TREX complex in yeast and splice-dependent recruitment in metazoans (Rother and Straesser, 2009).

## DEAD box RNA helicases

DEAD-box RNA helicases represent the largest family of RNA helicases. They are characterised by their conserved sequence DEAD. A subfamily, to which Sub2 also belongs, have the sequence DECD (Fan *et al.*, 2001). DEAD-box RNA helicases locally destabilise double-stranded RNA in an ATP-dependent manner and play a central role in RNA metabolism (Rudolph and Klostermeier, 2015). They are involved in transcription, translation, mRNA splicing, RNA transport and RNA degradation (Rudolph and Klostermeier, 2015). Malfunctions of DEAD-box helicases are associated with cancer and other diseases (Steimer and Klostermeier, 2012). DEAD-box helicases consist of two RecA-like domains connected by a non-conserved linker sequence and a DEAD-box helicase core. The DEAD-box helicase core is made of an approximately 400-long amino acid sequence that all members of the family share. The core consists of conserved motifs which are

responsible for nucleotide binding and hydrolysis, as well as binding and unwinding of RNA (Saguez *et al.*, 2013). Motif II is the name-giving motif with the amino acid sequence DEAD, and is also known as the Walker B motif (Tanner *et al.*, 2003).

In the absence of a nucleotide, the two RecA-like domains are separated from each other and flexible. Numerous variations in the positions of the RecA-like domains have been observed in crystal structures (Rudolph and Klostermeier, 2015). The binding of ATP and RNA is followed by a conformational change, whereby the conformation changes from an open to a closed one (Sengoku *et al.*, 2006). In the closed state, DEAD-box RNA helicases show an extensive interface. The nucleotide is bound between the two domains and interacts mainly with the N-terminal domain (Sengoku *et al.*, 2006). However, interaction with the C-terminal domain is required for ATP hydrolysis. RNA binding occurs at both RecA-like domains (Sengoku *et al.*, 2006). The conformational change is directly linked to the unwinding of RNA. Single-stranded RNA bound at the phosphate backbone is bent. This bending of the RNA is not compatible with the geometry of a double-stranded RNA. Accordingly, binding of the double strand leads to local destabilisation and unwinding of the double strand (Sengoku *et al.*, 2006). Binding of nucleotide and RNA leads to a closed conformation (Rudolph and Klostermeier, 2015). Subsequently, a conformational change leads to an active and unwinding state (Talavera and La Cruz, 2005). The RNA phosphate backbone is bent in the process, leading to a disruption of some base pairings, which destabilises the entire double strand. In this step, one of the strands dissociates from the complex (Sengoku *et al.*, 2006). The other strand remains in an RNA-protein complex. The temporal separation of the release of both strands prevents reannealing. After ATP hydrolysis the DEAD-box helicase can change to an open state, releasing the remaining RNA strand. (Aregger and Klostermeier, 2009). The protein is now ready to bind an RNA double strand again.

## Sub2/UAP56

UAP56 was identified as a human 56-kD U2AF associated protein (UAP56) (Fleckner *et al.*, 1997). Its orthologue Sub2 from yeast *S. cerevisiae* was initially identified as an essential pre-mRNA splicing factor critical in spliceosome assembly (Kistler and Guthrie, 2001; Libri *et al.*, 2001; Zhang and Green, 2001), but is also required for nuclear mRNA export through its interaction with the yeast RNA annealing protein 1 (Yra1) (Strasser and Hurt, 2001). This was followed by the discovery that both proteins (Sub2 and Yra1) together with the heteropentameric THO complex (Tho2, Hpr1, Tex1, Mft1 and Thp2) form the TREX complex thereby coupling transcription with messenger RNA export (Strasser *et al.*, 2002).

Both depletion and overexpression of UAP56 in mammals, *Drosophila*, *Caenorhabditis* and yeast significantly impair mRNA export and cell growth and survival (Gatfield *et al.*, 2001; Kapadia *et al.*, 2006; Kistler and Guthrie, 2001; MacMorris *et al.*, 2003; Strasser and Hurt, 2001). In detail, depletion of UAP56/DDX39B leads to a genomic instability phenotype and R-loop accumulation (Dominguez-Sanchez *et al.*, 2011; Perez-Calero *et al.*, 2020). In *S. cerevisiae*, Sub2 mutants are also hyper-recombinant (Fan *et al.*, 2001). Recently, it has been shown that the human UAP56/DDX39B helicase prevents R-loops *in vivo* and that *in vitro* UAP56 unwinds RNA/DNA hybrids and resolve RNA/DNA flap structures mimicking R loops (Perez-Calero *et al.*, 2020). RNA/DNA unwinding has also been shown for *S. cerevisiae* Sub2 (Saguez *et al.*, 2013). Hence, whereas the mechanism of suppression of the THO mutant phenotypes by Sub2 (UAP56/DDX39B) seems to be solved (i.e., via its helicase activity), the mechanism by which Tho1 acts is still unclear.

Sub2's (UAP56/DDX39B) activities are regulated by its interaction partners, the THO complex and/or Yra1 (human ALYREF) and Tho1 (human SARNP/CIP29) (Chen *et al.*, 2021; Ren *et al.*, 2017; Schuller *et al.*, 2020; Xie *et al.*, 2021). Structures of Sub2 in complex with the THO complex (Chen *et al.*, 2021; Ren *et al.*, 2017; Schuller *et al.*, 2020; Xie *et al.*, 2021) or with short RNA and the C-terminal fragment of Yra1 (Ren *et al.*, 2017) have been published.

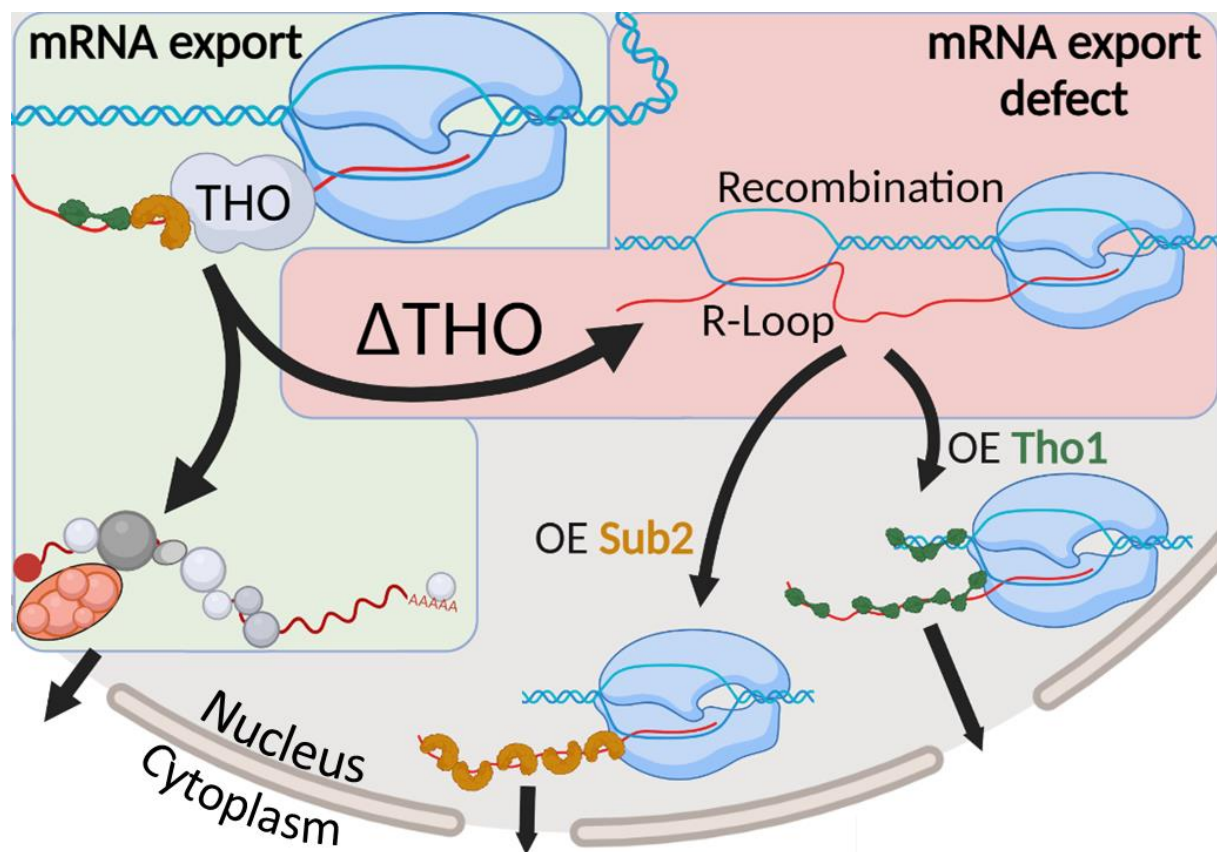
The RNA-dependent ATPase activity of Sub2 is stimulated by the THO complex (up to 4-fold (Chen *et al.*, 2021) and Yra1 (2-fold) (Ren *et al.*, 2017). However, how these interactions affect the helicase activity of yeast Sub2 is still unclear. Similar enhancement (2-3 fold) of the ATPase activity were observed for human UAP56 and ALYREF or SARNP/CIP29 (Chang *et al.*, 2013; Taniguchi and Ohno, 2008). Moreover, the helicase activity of human UAP56 was stimulated by ALYREF and SARNP/CIP29 (Chang *et al.*, 2013; Sugiura *et al.*, 2007).

## Tho1

Recruitment of Tho1 from *S. cerevisiae* to transcribed chromatin is dependent on THO and RNA (Jimeno *et al.*, 2006). *In vitro*, Tho1 and its orthologues bind to single-stranded RNA and double-stranded DNA (Jacobsen *et al.*, 2016; Jimeno *et al.*, 2006). Tho1 contains an N-terminal SAP domain, found in many but not all orthologues (e.g. Mos11) and a putative RNA-binding domain (Jacobsen *et al.*, 2016). NMR-structures of the N-terminal SAP-domain of human (pdb: 2DO1) and *S. cerevisiae* (pdb: 1H1J, 4UZW; 2WQG;) and the RNA-binding domain of *S. cerevisiae* (pdb: 4UZX) have been solved (Dodson *et al.*, 2010; Jacobsen *et al.*, 2016). The SAP domain is sufficient for

DNA binding *in vitro* (Jacobsen *et al.*, 2016; Kipp *et al.*, 2000). On the other hand, deletion of the SAP domain is not affecting RNA binding *in vitro*.

Tho1 was identified as a suppressor of the transcriptional defect of Hpr1 by overexpression and therefore, is like Sub2 able to rescue the phenotype of missing THO (Jimeno *et al.*, 2006; Piruat and Aguilera, 1998). But a role in mRNA export has not been demonstrated. Furthermore, Tho1 is a not essential protein and so far no phenotype was reported for a deletion of Tho1 (Piruat and Aguilera, 1998).



**Figure 1: Overexpression of Sub2 or Tho1 rescue the  $\Delta$ THO phenotype.** The THO-complex plays an important role in mRNA export. Deletion of HPR1 is associated with a missing THO complex and leads to an mRNA export defect and increased R-loop formation. Both, the overexpression of Sub2 or Tho1 are able to rescue the phenotype, but the mechanism remains unclear. Created with BioRender.com.

## Aim of this study

Sub2 plays a crucial role in *S. cerevisiae*. It is the link between mRNA transcription and mRNA export. Although it is already known that Sub2 is a DEAD-box helicase and its deletion in yeast is lethal, its biochemical functions in mRNA export remain unknown.

The first aim of this work was to characterise Sub2 *in vitro* as much detailed as possible in order to draw conclusions about its *in vivo* mechanisms. Even though DEAD-box helicases belong to one family, their biochemical properties may differ. All three activities, the binding and unwinding of nucleic acids and the hydrolysis of ATP were investigated under different conditions. First, suitable assays were introduced and the enzymatic activity of the purified protein was confirmed. Subsequently, substrate and nucleotide specificity were investigated, but also buffer conditions were analysed.

To gain further insight into the biochemical activity, protein-protein interaction and complex formation of Sub2, an attempt was made to track conformational change and interaction via Förster resonance energy transfer (FRET). For this purpose, it is necessary to label a protein specifically with a fluorophore. In this work thiol-maleimide chemistry was used for this reason. However, Sub2 has six cysteines, four of which are highly conserved. Therefore, an attempt was made to add cysteines on the surface and at the same time to remove the cysteines from the wild type in order to finally produce a single-cysteine variant that can be specifically labelled.

Thirdly, the regulation of the activities of Sub2 by and the interaction with Tho1 was investigated. In the process, a yet undescribed activity of Tho1 was revealed. It follows that another aim of this work was to characterise this activity. In addition, Tho1 was dissected and analysed to gain more details about the protein and its activity.

Finally, *in vitro* and *in vivo* experiments were carried out, to determine whether Sub2 and/or Tho1 are involved in the resolution of R-loops.

# Material

## Chemicals and consumables

**Table 1:** Chemicals and consumables

Chemicals and consumables	Supplier
2-Mercaptoethanol	Merck
5-Fluoroorotic acid	Apollo Scientific
Acetic acid	Carl Roth
Acrylamide 40% (29:1)	AppliChem
ADP	Jena Bioscience
ADPNH <sub>2</sub>	Jena Bioscience
ADPNP	Jena Bioscience
Agar bacteriology grade	AppliChem
Agarose	AppliChem
Alexa594	Life Technologies
Ammonium persulfate	AppliChem
ATP	Jena Bioscience
ATP <sub>γ</sub> S	Jena Bioscience
Benzamidine hydrochloride	Sigma Aldrich
BisTris	Sigma Aldrich
Boric acid	Merck
Bromophenol blue	Serva
CheLuminate-HRP PicoDetect	AppliChem
Chloroform	Carl Roth
Coomassie Brilliant Blue R-250 Dye	Thermo Fisher Scientific
DEPC H <sub>2</sub> O	Carl Roth
Dimethyle sulfoxide	Life Technologies
Disodium phosphate	Merck
DTT	Sigma Aldrich
Ethanol	Carl Roth



Ethylenediaminetetraacetic acid	Carl Roth
Gel loading dye, purple (6x)	New England Biolabs
Glucose	AppliChem
Glycerol	Carl Roth
Glycine	Sigma Aldrich
HEPES	AppliChem
IGEPAL	Alfa Aesar
Imidazole	Carl Roth
Isoamylalcohol	Carl Roth
Isopropanol	Carl Roth
Kanamycin	AppliChem
Leupeptin	Merck
Liquid nitrogen	Linde
Magnesium chloride	New England Biolabs
Magnesium sulfate	Merck
MES	Carl Roth
MOPS	Carl Roth
NP-40	Sigma Aldrich
Pepstatin A	Merck
Peptone	Becton, Dickinson and Company
Phenol	Carl Roth
Phenylmethylsulfonyl fluoride	AppliChem
Phosphoenolpyruvate	Sigma Aldrich
Potassium chloride	Carl Roth
Potassium hydrogen phosphate	Merck
Potassium hydroxide	AppliChem
Powdered milk, fat free, blotting grade	Carl Roth
Sodium chloride	Merck
Sodium dodecyl sulfate	Carl Roth
Spermidine	Sigma Aldrich
Sybr Gold	Thermo Fisher Scientific

Tetramethylethylenediamine	AppliChem
Tris	AppliChem
Tris(2-carboxyethylphosphine)	Sigma Aldrich
Tween® 20 detergent	Merck
Yeast extract	AppliChem
β-Nicotinamide adenine dinucleotide	Sigma Aldrich

## Equipment and devices

**Table 2:** Equipment and devices

Device	Supplier
10 mL Falcon	Sarstedt
50 mL Falcon	Sarstedt
Bio-Dot® Microfiltration Apparatus	Biorad
Bio-Link BLX	Vilber Lourmat
Biological Safety Cabinet Herasafe™	Thermo Scientific
Biometra WT12	Biometra
Centrifuge 5424	Eppendorf
Centrifuge Avanti JXN-26	Beckman Coulter
Chemocam Imager	Intas
Chicane flask (50mL, 1L, 5L)	Schott
CVC 3000	Vacuubrand
Dialysis membrane Membra-Cel™	Roth
Dri-Block DB-3	Techne
ECL Chemocam Imager	Intas
FluoroMax-4 Spectrofluorometer	Horiba
Freezer (-20°C)	Bosch
Fridge (4°C)	Bosch
Heraeus Megafuge 40R Centrifuge	Thermo Scientific
Herafreeze HFU T Series (-80°C)	Thermo Scientific
Incubator	Memmert

Infinite F200 pro	TECAN
iX20 Imager	Intas
JLA-8.1 rotor	Beckman Coulter
Milli-Q ® integral water purification system	Merck
Mini-PROTEAN Tetra Handcast System	BIO-RAD
Mobicol	Mo Bi Tec Molecular Biotechnology
NanoDrop® ND-1000	Thermo Scientific
neoLab-Rotator	neoLab
Nylon membrane Amersham Hybond™-N+	GE Healthcare
Parafilm® "M" Laboratory Film	BEMIS®
Pipettetips grey, yellow, blue	Sarstedt
Power Supply - EPS 301	General Electric
Protino® Ni-IDA 2000 Column	Machery-Nagel
Reaction vessels	Eppendorf
Research® multichannel pipette 200	Eppendorf
Research® plus pipettes 2.5, 10, 20, 100, 200, 1000	Eppendorf
Scale 440-53N	Kern
Scale AM100	Mettler
Scale PM2000	Mettler
Shaker KS 4000 ic control	IKA
Shaker Multitron Pro	INFORS HT
Sonifier 250	Branson Ultrasonics™
Sunrise plate reader	TECAN
Superdex™ 75 10/300 GL	GE Healthcare Life Sciences
Table centrifuge Heraeus Pico 17	Thermo Scientific
Typhoon FLA 9500	GE Healthcare Life Sciences
Ultracentrifuge Optima XPN-80	Beckman Coulter

## Buffers and media

### Media

#### SOC media

2% tryptone (w/v)  
0,5% yeast extract (w/v)  
10 mM NaCl  
2,5 mM KCl  
10 mM MgCl<sub>2</sub>  
10 mM MgSO<sub>4</sub>  
20 mM glucose  
pH 6.5

#### LB-media

10% peptone (w/v)  
5% yeast extract (w/v)  
5% NaCl (w/v)  
pH 7.5

#### YPD media

1% yeast extract (w/v)  
2% peptone (w/v)  
2% glucose (w/v)  
pH 5.5

#### SOB media

2% peptone (w/v)  
0,5% yeast extract (w/v)  
10 mM NaCl  
2,5 mM KCl  
10 mM MgCl<sub>2</sub>  
10 mM MgSO<sub>4</sub>  
pH 6.5

#### SDC media for 1 L

6.75 g yeast nitrogen base (w/o aa)  
0.6 g complete synthetic amino acid mix  
20 g glucose  
10 mL of each 100x aa stock except the required drop out  
pH 5.5

### Purification

#### Lysis/wash buffer

150 mM KCl  
50 mM HEPES  
10% glycerol (v/v)  
15 mM imidazole  
0,15% Np-40 (v/v)  
2 mM TCEP  
pH 6.5

#### Wash buffer 1

300 mM KCl  
50 mM HEPES pH 6.5  
10% glycerol (v/v)  
15 mM imidazole  
0,15% Np-40 (v/v)  
2 mM TCEP  
pH 6.5

Wash buffer 2

500 mM KCl  
50 mM HEPES  
10% glycerol (v/v)  
15 mM imidazole  
0,15% Np-40 (v/v)  
2 mM TCEP  
pH 6.5

Wash buffer 4

1000 mM KCl  
50 mM HEPES  
10% glycerol (v/v)  
15 mM imidazole  
0.15% Np-40 (v/v)  
2 mM TCEP  
pH 6.5

Elution buffer 2

50 mM KCl  
50 mM HEPES  
10% glycerol (v/v)  
30 mM imidazole  
2 mM TCEP  
pH 6.5

PBS buffer

137 mM NaCl  
2.7 mM KCl  
10 mM Na<sub>2</sub>HPO<sub>4</sub>  
2 mM KH<sub>2</sub>PO<sub>4</sub>  
pH 7.4

100x Protease inhibitor solution in 50mL Ethanol

6.85 mg pepstatin A  
1.42 mg leupeptin hemisulfat  
0.85g PMSF  
1.65g benzamide HCl

Wash buffer 3

700 mM KCl  
50 mM HEPES  
10% glycerol (v/v)  
15 mM imidazole  
0,15% Np-40 (v/v)  
2 mM TCEP  
pH 6.5

Elution buffer

150 mM KCl  
50 mM HEPES  
10% glycerol (v/v)  
250 mM imidazole  
2 mM TCEP  
pH 6.5

Dialysis buffer

50 mM KCl  
50 mM HEPES  
2 mM TCEP  
pH 6.5

1x STE buffer

100 mM NaCl  
10 mM Tris/HCl  
0.1 mM EDTA  
pH 8.0

## SDS-PAGE

### 6x Sample buffer for SDS-PAGE

375 mM Tris-HCl  
9% SDS (w/v)  
9% 2-mercaptoethanol (v/v)  
50% glycerol (v/v)  
0,03% bromphenol blue (w/v)  
0.5 M DTT  
pH 6.8

### 4x separating gel buffer

1.5 M Tris/HCl  
8 mM EDTA  
0.1% SDS (w/v)  
pH 8.8

### Coomassie staining solution

0.5% CBB R250 (w/v)  
25% isopropanol (v/v)  
10% acetic acid (v/v)

### 10x SDS running buffer

250 mM Tris  
1.9 M glycine  
1% SDS (w/v)  
pH 8.3

### 4x stacking gel buffer

0.5 M Tris/HCl  
8 mM EDTA  
0.1% SDS (w/v)  
pH 6.8

### 10% acetic acid

10% acetic acid (v/v)  
90% Aqua dest. (v/v)

## Cloning

### ISO Buffer

25% PEG-8000  
500 mM Tris  
50 mM MgCl<sub>2</sub>  
50 mM DTT  
1 mM of each dNTP  
5 mM NAD<sup>+</sup>

### 50x TAE buffer

2 M TRIS  
1 M NaOAc  
50 mM EDTA  
pH 8.0

### 6x Agarose loading dye

0.03% bromphenol blue  
0.03% xylen cyanol  
60% glycerin  
60 mM EDTA  
10 mM TRIS

## Yeast transformation

### Solution I

1x TE  
100 mM LiOAc

### 10x TE

100 mM TRIS  
10 mM EDTA  
pH 7.5

## Dot blot assay

### PBST

137 mM NaCl  
2.7 mM KCl  
10 mM Na<sub>2</sub>HPO<sub>4</sub>  
2 mM KH<sub>2</sub>PO<sub>4</sub>  
0.1% (w/v) Tween®  
pH 7.4

### Solution II

1x TE  
100 mM LiOAc  
40% PEG 3800

### 10x TE

100 mM TRIS  
10 mM EDTA  
pH 7.5

## Enzymatic assays

### 10x Annealingbuffer

100 mM MOPS  
0.5 M KCl  
pH 6.5

### 10x TBE buffer

1 M boric acid  
1 M Tris  
25 mM EDTA  
pH 8.3

### Binding buffer

10 mM MES  
5% glycerol  
2 mM TCEP  
pH 6.5

### Helicase reaction buffer

20 mM MES  
4 mM TPCEP  
pH 6.5

## Work material

### Kits

Table 3: Kits

Kit	Supplier	Usage
Nucleospin Gel and PCR Clean-Up	Macherey & Nagel	PCR product purification
NucleoSpin® Plasmid (NoLid)	Macherey & Nagel	plasmid purification

### Enzymes

Table 4: Enzymes

Enzyme	Supplier	Usage
DpnI	New England Biolabs	PCR
Lactatedehydrogenase	Sigma life sciences	ATPase assay
Phusion Polymerase	New England Biolabs	PCR
Pyruvatekinase	Sigma life sciences	ATPase assay
RNase H	Thermo Scientific	Dot blot assay
RNase III	Ambion™	Dot blot assay
RNase T1	Thermo Scientific	Dot blot assay
T5 exonuclease	New England Biolabs	Gibson assembly
Taq DNA ligase	New England Biolabs	Gibson assembly
Taq Polymerase	Own production	PCR
TEV Protease	Own production	His6 cleavage

### Antibodies

Table 5: Antibodies

Antibody	Source	Dilution	Supplier
anti-dsDNA	mouse	1:1000	abcam
anti-mouse-HRP	goat	1:3000	Biorad
S9.6 anti-RNA/DNA	mouse	1:1000	Merck



## Markers

**Table 6:** Markers

Marker	Supplier
1 kb DNA Ladder	New England Biolabs
100 bp DNA Ladder	New England Biolabs
Unstained Protein Standard, Broad Range (10-200 kDa)	New England Biolabs

## Software

**Table 7:** Software

Software	Usage
BioEdit	Analysis conservation
FIJI	Quantitative analysis dot blot
ImageQuant <sup>TL</sup>	Quantitative analysis gel
Microsoft Excel	Quantitative analysis
Origin	Quantitative analysis
PyMol	Analysis protein structure
SnapGene	Cloning

## Organisms

### *E. coli* strains

**Table 8:** *E. coli* strains

<i>E. coli</i> strain	Genotype
<i>E. coli</i> DH5 $\alpha$	F $^-$ $\phi$ 80lacZ $\Delta$ M15 $\Delta$ (lacZYA-argF)U169 recA1 endA1 hsdR17(rK $^-$ , mK $^+$ ) phoA supE44 $\lambda$ -thi-1 gyrA96 relA1
<i>E. coli</i> Rosetta 2	F $^-$ omp <sup>T</sup> hsdSB(rB $^-$ mB $^-$ ) gal dcm (DE3) pRARE2 (CamR)

## Yeast strains

**Table 9:** Yeast strains

Yeast strain	Genotype
RS453	MAT <sup>a</sup> , ade2-1, his3-11, 15, ura3-52. leu2-3, 112, trp1-1, can1-100, GAL <sup>+</sup>
RS453 Sub2-shuffle	MAT <sup>a</sup> , ade2-1, his3-11, 15, ura3-52. leu2-3, 112, trp1-1, can1-100, GAL <sup>+</sup> ; pRS315-Sub2
W303	MAT <sup>a</sup> , ura3-1, trp1-1, his3-11,15, leu2-3,112, ade2-1, can1-100, GAL <sup>+</sup>
W303 $\Delta$ hpr1	MAT <sup>a</sup> , ura3-1, trp1-1, his3-11,15, leu2-3,112, ade2-1, can1-100, GAL <sup>+</sup> , HPR1::HIS3
W303 tho1-shuffle	MAT <sup>a</sup> , ura3-1, trp1-1, his3-11,15, leu2-3,112, ade2-1, can1-100, GAL <sup>+</sup> , THO1::KanMX, pRS316-THO1
Y34	Mat <sup>a</sup> ; ura3-52; HIS3; leu2-1; trp1-63; YDL084w::kanMX4
Y34 Sub2-shuffle	Mat <sup>a</sup> ; ura3-52; HIS3; leu2-1; trp1-63; YDL084w::kanMX4; pRS315-Sub2

## Oligonucleotides and plasmids

### Plasmids

**Table 10:** Plasmids

Plasmid	Description
pCMV SARNP	Ordered plasmid at Origene carrying the gene for SARNP
pET 151 His6 CIP29	Synthesized plasmid by Thermofisher carrying the gene for CIP29
pET 151 His6 UAP56	Synthesized plasmid by Thermofisher carrying the gene for UAP56
pRS315 Sub2	Transformation of Sub2 in yeast Sub2-shuffle strains
pRS315 Sub2 $\Delta$ 6C	Transformation of Sub2 $\Delta$ 6C in yeast Sub2-shuffle strains
pRS315 Sub2 C80A, C360L	Transformation of Sub2 C80A, C360L in yeast Sub2-shuffle strains
pRS315 Sub2 GFP A206K	Transformation of Sub2-GFP fusion protein variant A206K in GFP in yeast Sub2-shuffle strains
pRS315 Sub2 WB	Transformation of Sub2 D215A in yeast Sub2-shuffle strains
pRS424	high copy number plasmid, 2 $\mu$ , with TRP maker
pRS424-SUB2	ORF + 500 bp of 5' and 300 bp of 3' UTR of SUB2 was cloned into pRS425
pRS424-THO1	ORF + 500 bp of 5' and 300 bp of 3' UTR of THO1 was cloned into pRS425

pT7_His6TEV SARNP	Recombinant expression of human SARNP in <i>E. coli</i>
pT7_His6TEV Sub2	Recombinant expression of wildtype Sub2 in <i>E. coli</i>
pT7_His6TEV Sub2 Δ6C	Recombinant expression of Sub2 variant Δ6C in <i>E. coli</i>
pT7_His6TEV Sub2 C80A, C360L	Recombinant expression of Sub2 variant C80A, C360L in <i>E. coli</i>
pT7_His6TEV Sub2 D240C	Recombinant expression of Sub2 variant D240C in <i>E. coli</i>
pT7_His6TEV Sub2 GFP A206K	Recombinant expression of Sub2-GFP fusionprotein variant A206K in GFP in <i>E. coli</i>
pT7_His6TEV Sub2 K177C	Recombinant expression of Sub2 variant K177C in <i>E. coli</i>
pT7_His6TEV Sub2 K341C	Recombinant expression of Sub2 variant K341C in <i>E. coli</i>
pT7_His6TEV Sub2 N173C	Recombinant expression of Sub2 variant N173C in <i>E. coli</i>
pT7_His6TEV Sub2 N320C	Recombinant expression of Sub2 variant N320C in <i>E. coli</i>
pT7_His6TEV Sub2 T62A	Recombinant expression of Sub2 variant T62A in <i>E. coli</i>
pT7_His6TEV Sub2 T62D	Recombinant expression of Sub2 variant T62D in <i>E. coli</i>
pT7_His6TEV Sub2 WA	Recombinant expression of Sub2 variant K112A in <i>E. coli</i>
pT7_His6TEV Sub2 WB	Recombinant expression of Sub2 variant D215A in <i>E. coli</i>
pT7_His6TEV Tho1	Recombinant expression of yeast wildtype Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 Δ1st half CTE	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 Δ2nd half CTE	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 ΔCTE	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 ΔLinker	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 ΔRBD	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 ΔSAP	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 F143D	Recombinant expression of Tho1 variant F143D in <i>E. coli</i>
pT7_His6TEV Tho1 K135D	Recombinant expression of Tho1 variant K135D in <i>E. coli</i>
pT7_His6TEV Tho1 linker+RBD	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 RBD	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 RBD+CTE	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>

pT7_His6TEV Tho1 SAP	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1 SAP+linker	Recombinant expression of a truncated version of Tho1 in <i>E. coli</i>
pT7_His6TEV Tho1-yoTag-RFP	Recombinant expression of Tho1-yoTag-RFP fusionprotein in <i>E. coli</i>
pT7_His6TEV UAP56	Recombinant expression of wildtype UAP56 in <i>E. coli</i>

## Primers

Table 11: Primers

Number	Name	Sequence (5'-3')
MM1	Sub2-K341C fwd	CATGGTCATATGTGTCAAGAAGAACGTATTGCTCGTTAC
MM2	Sub2-K341C rev	ACGTTCTTCTTGACACATATGACCATGAACGGTG
MM3	Sub2-K173C fwd	ACTCCAATTCTTGTGATGCTGAACCTTTAAAGAATAAAG
MM4	Sub2-K173C rev	AAGTTCAGCATCACAAGAAATTGGAGTACCACCG
MM5	Sub2-D240C fwd	GCTACTCCAAGATGCAAACAAGTCATGATGTTTTC
MM6	Sub2-D240C rev	CATGACTTGTTCGTCATCTTGGAGTAGCTCTG
MM7	Sub2-L177C fwd	AAGGATGCTGAATGTTAAAGAATAAAGATACTG
MM8	Sub2-L177C rev	TTTATTCTTTAAACATTCAGCATCCTTAGAAATTG
MM9	Sub2-N320C fwd	ACCACAAGAGCTTGTGAGTTGACCAAACCTATTAAAC
MM10	Sub2-N320C rev	GTTTGGTCAACTCACAAGCTCTTGTGGTAGATTAAAC
MM11	pT7-Sub2-Vector rev	TTTACTGCATGCATTATTCAAATAAGTGGACGGATCAATG
MM12	pT7-Sub2-Vector fwd	CAGCCAAGCTAAGGATCCGGCTGCTAACAAAG
MM13	pNOP-Sub2-GFP fwd	TATTTGAATAATGCATGCAGTAAAGGAGAAG
MM14	pNOP-Sub2-GFP rev	GCAGCCGGATCCTTAGCTTGGCTGCAGTTTGTATAGTTCA
MM15	Sub2-GFP-A206K fwd	GTCCACACAATCTAAGCTTTCCAAAGATC
MM16	Sub2-GFP-A206K rev	GATCTTTGGAAAGCTTAGATTGTGTGGAC
MM17	pRS315 fwd	GCCAAGCTAATAAAGGAAAAAAGATACGTTTTT ATATAGATTATATAAAGATTTTG
MM18	pRS315 rev	CGTGTGACATCGTTGCTGTTTT

MM19	Sub2 GFP fwd	AACAGCAACGATGTCACACGAAGGTGAAGAAG
MM20	Sub2 GFP rev	TTTTTCCTTTATTAGCTTGGCTGCAGTTTGTATAGT
MM21	pT7 FOR	GGTTACAGAAGATAAGGATCCGGCTGCTAACA
MM22	pT7 REV	ATAATCTGCCATGCCCTGAAAATACAGGTTTTCG
MM23	Tho1 FOR	GCCGGATCCTTATCTTCTGTAACCAGAGCGGTTCCTCT
MM24	Tho1 REV	TATTTTCAGGGCATGGCAGATTATTCTTCCCTTACTG
MM25	pT7 Tho1 FOR	TTGTATAAGTAAGGATCCGGCTGCTAACAAAGCCC
MM26	pT7 Tho1 REV	TAATTAACCAGCTCTTCTGTAACCAGAGCGGTTCG
MM27	yoTagRFP FOR	GGTTACAGAAGAGCTGGTTTAATTAACATGGTATCTA AAGGT
MM28	yoTagRFP REV	GCAGCCGGATCCTTACTTATACAATTCATCCATACCATTCAG
MM29	Vektor deltaNTE FOR	TATTTTCAGGGCGGCGACAAGAAAGGTTCCCTATGTTG
MM30	KanR Vektor REV	CTCACCCTGCGATCCCCGGGAAA
MM31	KanR Fragment FOR	TTTCCCGGGGATCGCAGTGGTGAG
MM32	Fragment deltaNTE REV	TTTCTTGTCGCCGCCCTGAAAATACAGGTTTTCGGT
MM33	Vektor Thrombin FOR	GGTGATAATAATAACCTCGTTCACGTGGCTCCAAGA AAGGTTCCCTATGTTGG
MM34	Fragment Thrombin REV	GAACCTTTCTTGGAGCCACGTGGAACGAGGTTATTATTATC ACCTTCAGTTGCA
MM35	Vec SAP end fwd	GTCTCTCCACAGTAAGGATCCGGCTGCTAACAAAGC
MM36	Vec SAP beg rev	AGAATAATCTGCCATGCCCTGAAA
MM37	In SAP beg fwd	TTTCAGGGCATGGCAGATTATTCT
MM38	In SAP end rev	GCCGGATCCTTACTGTGGAGAACTTCACTTTCACC
MM39	Vec RBD end fwd	AGGAAAAATGAATAAGGATCCGGCTGCTAACA
MM40	Vec RBD beg rev	CAAAGCGGGAGCCATGCCCTGAAAATACAG
MM41	In RBD beg fwd	TTTCAGGGCATGGCTCCCGCTTGTTCACC
MM42	In RBD end rev	GCCGGATCCTTATTCAATTTTCCTACTGACAAGGCC

MM43	Vec CTE end fwd	GGTTACAGAAGATAAGGATCCGGC
MM44	In CTE end rev	GCCGGATCCTTATCTTCTGTAACC
MM45	Vec Linker beg rev	CTGGT*TTTGTTCATGCCCTGAAAATACAGGT
MM46	In Linker beg fwd	T*TCAGGGCATGGAACAAAACCAGGAACAAGGG
MM47	Vec Linker end fwd	GCT*TCCTACTGCT*TAAGGATCCGGCTGCTAACAAAG
MM48	In Linker end rev	GCCGGATCCTTAAGCAGTGGAAGCAGCAG
MM49	pT7 UAP56 V for	GAACAGACCCGCTAAGGATCCGGCTGCTAACAA
MM50	pT7 UAP56 V rev	ATT*TCGGCCATGCCCTGAAAATACAGGT*TT
MM51	pT7 UAP56 I for	TATT*TCAGGGCATGGCCGAAAATGATGTGGATAAT
MM52	pT7 UAP56 I rev	GCCGGATCCTTAGCGGGTCTGT*TCATATAGCT
MM53	pT7 SARNP V for	TTTGGGATTGCCTAAGGATCCGGCTGCTAACAA
MM54	pT7 SARNP V rev	CTCGGTCGCCATGCCCTGAAAATACAGGT*TT
MM55	pT7 SARNP I for	TATT*TCAGGGCATGGCGACCGAGAC
MM56	pT7 SARNP I rev	GCCGGATCCTTAGGCAATCCCAAAGCG
MM57	pT7 Tho1 V7 Vek REV	CAAAGCGGGAGCCTGTGGAGAAACTTCACTT
MM58	pT7 Tho1 V7 Ins FOR	GT*TTCTCCACAGGCTCCCGCT*TTGTCAC
MM59	pT7 Tho1 V8 Vek REV	GCCACT*TTCAGGAGCAGTGGAAGCAGCA
MM60	pT7 Tho1 V8 Ins FOR	GCTTCCTACTGCTCCTGAAAGTGGCAATAACGG
MM61	pT7 Tho1 V9 Vek REV	GCCGGATCCTTAGTTGT*TTGCAT*TTTGT*TTCTA T*TTTAAATT*ACC
MM62	pT7 Tho1 V9 Ins FOR	AATGCAAACAATAAGGATCCGGCTGCTAACAA
MM63	pT7 Tho1 V10 Vek REV	T*TA*CTCTGGAGCGTT*CAT*TTT*CTACTGACAAGGC
MM64	pT7 Tho1 V10 Ins FOR	TAGGAAAAATGAACGCTCCAGAGTAAGTAAAAACAG
MM65	Sub2 Vek FOR	GAATTCCCAGAAGAAGGCATTGAT
MM66	Sub2 Vek REV	GGAGTACTCCAATAAATCTTCTTCACC
MM67	Sub2 Ins FOR	GAAGAAGATT*ATTTGGAGTACTCCGA

MM68	Sub2 Ins REV	ATCAATGCCTTCTTCTGGGAATTC
PK75	T7 Promotor	AATACGACTCACTATAGG

## Substrates

All substrates listed in **Table 12** were synthesized by Eurogentec. Putnam substrates were taken from (Putnam and Jankowsky, 2013), Ma substrates are derived from (Ma *et al.*, 2013).

**Table 12:** Substrates

Name	Sequence (5'-3')	Modification
DNA_Ma_Top long	GTT-TTG-TTT-TGT-TTT-GTT-TTA- AGC-ACC-GTA-AAG-ACG-C	
DNA_MA_16_21	GCG-TCT-TTA-CGG-TGC-TTA-AAA- CAA-AAC-AAA-ACA-AAA-C	
DNA_Ma_16_21_Cy5	GCG-TCT-TTA-CGG-TGC-TTA-AAA- CAA-AAC-AAA-ACA-AAA-C	Cy5 5'
DNA_Ma_16_BHQ2	AGC-ACC-GTA-AAG-ACG-C	BHQ2 3'
DNA_Putnam_13	AGC-ACC-GTA-AAG-A	
DNA_Putnam_13_25	TCT-TTA-CGG-TGC-TTA-AAA-CAA- AAC-AAA-ACA-AAA-CAA-AA	
DNA_Putnam_13_25 Cy5	TCT-TTA-CGG-TGC-TTA-AAA-CAA- AAC-AAA-ACA-AAA-CAA-AA	Cy5 5'
DNA-Putnam_13_BHQ2	AGC-ACC-GTA-AAG-A	BHQ2 3'
RNA Putnam13	AGC-ACC-GUA-AAG-A	
RNA_Ma_16_21_Cy5	GCG-UCU-UUA-CGG-UGC-UUA- AAA-CAA-AAC-AAA-ACA-AAA-C	Cy5 5'
RNA_Ma_16_BHQ2	AGC-ACC-GUA-AAG-ACG-C	BHQ2 3'
RNA_Putnam_10_BHQ2	AGC-ACC-GUA-A	BHQ2 3'
RNA_Putnam_13_25_Cy5	UCU-UUA-CGG-UGC-UUA-AAA- CAA-AAC-AAA-ACA-AAA-CAA-AA	Cy5 5'
RNA_Putnam_13_bottom	UCU-UUA-CGG-UGC-U	

RNA_Putnam_13_bottom_Cy5	UCU-UUA-CGG-UGC-U	Cy5 5'
RNA_Putnam_13_Cy3	AGC-ACC-GUA-AAG-A	Cy3 3'
RNA_Putnam_13_U25	UCU-UUA-CGG-UGC-UUU-UUU- UUU-UUU-UUU-UUU-UUU-UUU-UU	
RNA_Putnam_38 BHQ2	UUU-UGU-UUU-GUU-UUG-UUU- UGU-UUU-AAG-CAC-CGU-AAA-GA	BHQ2 3'
RNA_Putnam13_25	UCU-UUA-CGG-UGC-UUA-AAA- CAA-AAC-AAA-ACA-AAA-CAA-AA	
RNA_Putnam13_25 rev	AAA-ACA-AAA-CAA-AAC-AAA-ACA- AAA-UUC-GUG-GCA-UUU-CU	
RNA_Putnam13_BHQ2	AGC-ACC-GUA-AAG-A	BHQ2 3'
U13	UUU-UUU-UUU-UUU-U	
UvrD_18top_Cy5	GCC-TCG-CTG-CCG-TCG-CCA	Cy5 3'



# Methods

## Production of Z-competent cells

Two kinds of *E. coli* cells were used in this work: Rosetta 2 for overexpression and a subsequent purification of a target protein and DH5 $\alpha$  for expression of a plasmid to perform, for example, a mutagenesis. Competent cells, such as those already mentioned, are able to receive DNA in form of plasmids. This method was modified according to the original protocol of Zymo Research.

First, a 3 mL pre-culture was prepared and was inoculated with the desired cells from an agar plate. This was followed by overnight incubation at 37°C and 200 rpm. A 50mL main culture of SOB medium is inoculated with 0.5 mL of the pre-culture in a chicane flask. The main culture was incubated at 28°C and 200 rpm to an OD<sub>600</sub> of 0.4-0.6. The main culture was then transferred to 50 mL Falcons and incubated on ice for 10 minutes. Centrifugation for 10 minutes at 4°C and 2500 x g follows. The supernatant was removed and discarded. The pellet was resuspended in 5 mL ice-cold washing buffer. Centrifugation for 10 minutes at 4°C and 2500 x g follows. The supernatant was discarded again and the cell pellet resuspended in 5 mL ice-cold competent buffer. The cells were aliquoted to 50  $\mu$ L each and stored at -80°C.

## Transformation

For transformation, plasmids are introduced into cells, for example to express a specific protein and then to purify it. The cells were stored at -80°C and were thawed on ice for transformation. Then 1  $\mu$ L of the plasmid is added to the cells and incubated on ice for 15 minutes. Then 400  $\mu$ L of SOC medium is added and incubated for 60 minutes at 37°C and 200 rpm on the shaker to recover the cells. After incubation, a short centrifugation step follows for 15 seconds. The supernatant is discarded and the pellet is dissolved in the remaining liquid (approx. 100  $\mu$ L). The cells were then spread out on an agar plate with kanamycin (c=25  $\mu$ g/mL) and incubated overnight at 37°C in an incubator.

## Plasmid preparation

Plasmid preparation is used to obtain plasmids from bacteria in order to transform them again, to sequence or to use them as a template for various PCRs, for example. First, 3 mL LB medium was placed in incubation tubes and then the antibiotic kanamycin was added to a final concentration of 25  $\mu$ g/mL. Then a colony was streaked from an already incubated agar plate with a sterile pipette

tip and added to the LB medium. This was followed by an overnight incubation at 37°C and 200 rpm.

The plasmid was obtained the next day with the NucleoSpin® Plasmid (NoLid) kit from Macherey & Nagel according to the protocol. The concentration of the obtained sample was measured at the NanoDrop and stored at -20°C.

### **Yeast viability assay (Dot spots)**

To investigate the viability of certain variants of Sub2, a dot-spot assay was performed. In this method, the wild type of Sub2 is replaced in yeast cells with a variant of the protein to determine positive or negative effects on yeast growth. This method is divided into transformation of yeast cells with a plasmid carrying the gene for the target protein, FOA selection to displace the wild-type gene from the cells, and finally culturing the cells at different temperatures.

First, a 50 mL YPD culture was inoculated with the desired Sub2 shuffle strain and grown overnight shaking at 30°C. The next day, the culture was diluted to an OD<sub>600</sub> of 0.2 and grown again to an OD<sub>600</sub> of 0.6-0.8. The culture was transferred to a 50 mL tube and centrifuged at 3600 rpm for 5 minutes (r max<sub>Rotor</sub>: 98 mm). The supernatant was discarded and 10 mL of H<sub>2</sub>O was added to resuspend and wash the pellet. After centrifugation at 3600 rpm for 5 min (r max<sub>Rotor</sub>: 98 mm), the pellet was resuspended in 500 µL of solution I and transferred to a 2 mL reaction tube. This was followed by centrifugation at 3600 rpm for 3 min (r max<sub>Rotor</sub>: 84 mm). The supernatant was then discarded and 250 µL of solution I was added. One transformation mixture contained 100-200 ng of plasmid DNA, 10 µg of carrier DNA, 300 µL of solution II and 50 µL of yeast cells resuspended in solution I. After pipetting the transformation mixture, it was incubated at room temperature on a turning wheel for 30 min. This was followed by a heat shock for 10 min at 42°C. The mixture was then placed on ice for 3 min. Then 1 mL of H<sub>2</sub>O was added and the mixture was centrifuged for 3 min at 3600 rpm (r max<sub>rotor</sub>: 84 mm). The supernatant was discarded and 200 µL H<sub>2</sub>O was added. The entire volume was then applied to an SDC-Leu plate to select for positive clones.

After 3 days at 30°C in the incubator, the plates were checked. Untransformed cells served as negative control and cells transformed with wild type Sub2 as positive control. If the controls looked as expected, three colonies were picked and spread on FOA plates to shuffle the wild-type plasmid out of the cells. The gene URA3 codes for the protein Orotidine 5'-phosphate decarboxylase (ODCase). This protein normally catalyses a reaction in the synthesis pathway of pyrimidine ribonucleotides. However, it also catalyses the conversion of 5-fluoroorotic acid (5-FOA) into the toxic component 5-fluorouracil. The plasmid with the wild type Sub2 which was

already present in the cell before transformation carries the gene for ODCase and URA3 is silenced in the strains used. In order to survive, the transformed cells need to get rid of the wildtype plasmid. Already in this step, a phenotype might become recognisable. To be sure that all wild-type plasmids have been shuffled out of the cells, the cells were again spread on FOA plates to force a new selection. After three more days at 30°C in the incubator, the plates were removed again. Even if lethal phenotypes were already recognisable, thermosensitive variants were examined with a dot-spot assay.

Before applying the dots, the cells were again restreaked on YPD plates and incubated overnight at 30°C. The next day, a sample of these plates was taken and resuspended in 2 mL H<sub>2</sub>O. The OD<sub>600</sub> of this solution was measured and then diluted to an OD<sub>600</sub> of 0.09. A dilution series of 1:10, 1:100 and 1:1000 was then prepared. These dilutions and the undiluted sample were then spotted onto four different YPD plates. The different plates were then incubated at 16°C, 25°C, 30°C and 37°C. The 16°C plate was incubated for 7 days, the 25°C and 30°C plates for 2 days and finally the 37°C plate for 3 days. After incubation, the plates were scanned and digitised.

### **Dot blot assay to determine level of R-Loops**

Yeast cells were grown in appropriate medium and 10 mL culture grown in mid-log phase were harvested. Pellet was washed once with 10 mL water. Standard gDNA extraction method with phenol:chloroform:isoamylalcohol was performed. After ethanol precipitation gDNA was resuspended in 200 µL TE. 10 µg gDNA was digested with 5 units RNase III and 500 units RNase T1 for 2.5 h at 37°C. Half of the reaction mix was additionally digested with 10 units RNase H. Digested samples were spotted on a positively charged nylon membrane in serial dilution in SSC (1:1) using a 96-well Bio-Dot microfiltration unit. The membrane was cross-linked with 254 nm UV for 40 s and blocked for 1 h with 5 % milk in PBST at room temperature. Afterwards, membranes were incubated with anti-DNA-RNA hybrid antibody and anti-dsDNA antibody. Membranes were washed three times with PBST before secondary antibody was added for 1 h at room temperature. Membrane was washed again three times with PBST and chemiluminescence signals were imaged using CheLuminate-HRP ECL-solution and ChemoCam Imager. Signals were quantified using FIJI software.

## Cloning and Mutagenesis

In order to obtain overexpressed protein from *E. coli* and to study it *in vitro*, it must be cloned into a suitable vector to perform the required type of purification. In addition, it is also possible to alter the amino acid sequence of the target protein via mutagenesis. It is a powerful tool to investigate the protein of interest via inactivation of activity or deletion of domains to study the role of that domain. The Gibson assembly were used to clone and mutate, modified after Gibson *et al.*, 2009.

First two, for mutagenesis three PCRs have to be carried out. Usually this involves a vector, the backbone, and the gene of interest, the insert. For Mutagenesis two PCRs for the insert is carried out. The PCR mix contains 1x Phusionbuffer from NEB, 250  $\mu$ M dNTPs, a primer pair, consisting of 500 nM forward and revers primer respectively, 10 ng of the template in form of a plasmid and 2 U of Phusion polymerase. Annealing temperature was calculated via NEB Tm Calculator (version 1.9.10 – version 1.13.0). The program of the thermocycler for a PCR with phusion polymerase is shown in Table 13.

**Table 13:** Thermocycler program for PCR with Phusion polymerase

Step	°C	Time
Initial denaturation	98	30 s
Denaturation	98	10 s
Annealing	45-72	30 s
Elongation	72	45 s/kb
Final extension	72	10 min

} 30x

The products of the PCR were checked on a 1% agarose gel. If the products had the expected length, they were treated with 2 U of DpnI for 1 h at 37°C, followed by a heat inactivation at 80°C for 20 min to digest the starting material. After that, the samples were purified, according to the protocol, with the Nucleospin Gel and PCR Clean-Up Kit by Machery & Nagel.

Subsequently 100 ng of the backbone and the double molar amount of the insert or inserts, were incubated with a Gibson assembly mix (0.08 U T5 exonuclease, 0.5 U Phusion Polymerase, 80 U Taq Ligase in 1x ISO buffer) at 50°C for 30 min. Afterwards DH5 $\alpha$  *E. coli* cells were transformed with 5  $\mu$ L of the reaction.

The next day a colony PCR were carried out to check for positive clones. Clones were picked, restreaked on agar plates and dissolved in 20  $\mu$ L of H<sub>2</sub>O. The PCR mix contained 1x Taq buffer,

250  $\mu$ M dNTPs, a primer pair, consisting of 500 nM forward and reverse primer respectively, 5  $\mu$ L of the solution with one colony and 0.5  $\mu$ L of self-purified Taq polymerase. Annealing temperature was calculated via NEB Tm Calculator (version 1.9.10 – version 1.13.0). The program of the thermocycler for a PCR with Taq polymerase is shown in Table 14.

**Table 14:** Thermocycler program for PCR with Taq polymerase

Step	°C	Time
Initial denaturation	98	30 s
Denaturation	98	30 s
Annealing	45-72	60 s
Elongation	72	1 min/kb
Final extension	72	10 min

} 30x

Colonies that showed an expected length of the PCR product were inoculated in 3 mL LB media. The next day a plasmid preparation with the NucleoSpin® Plasmid (NoLid) kit from Macherey & Nagel was carried out. The plasmid was sequenced (Microsynth Seqlab) to confirm the successful cloning or mutagenesis.

## Purification of Sub2

To study the activity and function of Sub2, it needs to be purified. The gene of Sub2 were cloned into a plasmid for overexpression in *E. coli*. Sub2, its variants and orthologues (UAP56, Sub2 *C. thermophilum*) were expressed with a TEV-cleavable His-Tag. The His-Tag will bind to nickel beads, and after washing and elution, the protein is available in a purified form and can be used for different assays. This method was modified according to the manual of the His-Tag purification kit (Ni-NTA-beads) by Machery Nagel.

First, a pre-culture was prepared. For this purpose, a colony was scraped off an already incubated agar plate (see transformation) and 50 mL LB medium with 25  $\mu$ g/mL kanamycin was inoculated. This was followed by an overnight incubation at 37°C and 200 rpm.

On the following day, a main culture was prepared. For this, 20mL of the pre-culture was added to 500mL LB medium with 25 $\mu$ g/mL kanamycin in a chicane flask. The culture was incubated at 37°C and 180rpm until the culture reaches an OD<sub>600</sub> of about 0.7. When the appropriate optical

density was reached, expression of the protein was induced by 1 mM IPTG. The cultures incubate overnight at 20°C and 150rpm.

The cells were then harvested. The cultures were transferred to centrifuge buckets and centrifuged at 4200rpm and 4°C for 20 min ( $r_{avRotor}$ : 284 mm). The supernatant was discarded and the pellet was dissolved in 20 mL lysis buffer. The mixture of lysis buffer and cells was transferred to a beaker and the cells were lysed via ultrasound on ice (5 min, oc: 5, dc: 30%). The mixture was then transferred to ultracentrifuge tubes and centrifuged under vacuum at 4°C and 40000 rpm for 45 min (rotor diameter: 91.9 mm). After centrifugation, soluble proteins are found in the supernatant and other cell residues in the pellet. The supernatant was collected and further used.

While the centrifugation was running, Ni-NTA-beads were prepared. 600  $\mu$ L of beads (300  $\mu$ L Beads, 300  $\mu$ L ethanol) were transferred to a 15 mL reaction vessel and then centrifuged at 900 g, for 3 minutes at 4°C. The supernatant was removed with a vacuum pump. Afterwards the beads were washed with 10 mL of lysis buffer, centrifuged at 900 g, for 3 minutes at 4°C. The supernatant was removed again and the supernatant from the ultracentrifugation was added to the beads. The lysate and beads were incubated for 1h on a turning wheel in the cold room. After incubation the sample was centrifuged at 900 g, 3 minutes at 4°C. The supernatant was discarded, because the target protein is now bound to the beads. The beads were transferred to a mobicol (a column with a filter that can be closed at the top and bottom) and next washed with 8 mL of each wash buffer 1-4 and re-equilibrated with lysis buffer. After adding 700  $\mu$ L of elution buffer (250 mM imidazole), the mobicol were closed and incubated for 15 minutes in the cold room. The sample was centrifuged in the cold room at 2000 rpm ( $r_{maxRotor}$ : 84 mm). The flow-through incubated for 1-2 hours with TEV-protease (50  $\mu$ L,  $c \approx 1.5$  mg/mL). A dialysis in dialysis buffer followed. The dialysis tubes with a cut-off of 14000 Da, were incubated overnight stirring in the cold-room.

On the next day Ni-NTA-beads were prepared, similar to before. 200  $\mu$ L slurry (100  $\mu$ L beads, 100  $\mu$ L ethanol) were equilibrated with 700  $\mu$ L of dialysis buffer in a mobicol. The samples were removed from the dialysis tubes and put to beads in the mobicol. The samples incubated for 15 minutes and then centrifuged for 1 minute at 2000 rpm ( $r_{maxRotor}$ : 84 mm) in the cold room. The beads were then treated with 500  $\mu$ L of elution buffer 2 (30 mM imidazole) for 15 minutes on a turning wheel in the cold room. The sample were centrifuged for 1 minute at 2000 rpm ( $r_{maxRotor}$ : 84 mm) in the cold room. The flow-through were flash frozen and stored at -80°C.

For determination of the concentration the absorbance at 280 nm were measured. The calculation was made with following formula:

$$\frac{Abs_{280}}{\epsilon} \times 10^6 = c [\mu M]$$

Whereas  $\epsilon$  is the extinction coefficient in  $M^{-1}cm^{-1}$ . The extinction coefficient was calculated with ProtParam (Gasteiger *et al.*, 2005).

## Purification of Tho1

To test the activity and other functions of Tho1, it needs to be purified. Similar to the purification of Sub2, *E. coli* cells were transformed with a certain plasmid to overexpress Tho1, its variants and orthologues (SARNP, Tho1 *C. thermophilum*). The proteins were tagged with a His-tag and can be purified via nickel affinity chromatography. This method was modified according to the manual of the His-tag purification kit (Ni-IDA 2000 columns) by Machery-Nagel.

First, a pre-culture was prepared. For this purpose, a colony was scraped off an already incubated agar plate (see transformation) and 50 mL LB medium with 25  $\mu g/mL$  kanamycin was inoculated. This was followed by an overnight incubation at 37°C and 200 rpm.

On the following day, a main culture was prepared. For this, 20mL of the pre-culture was added to 500mL LB medium with 25 $\mu g/mL$  kanamycin in a chicane flask. The culture was incubated at 37°C and 180rpm until the culture reaches an OD<sub>600</sub> of about 0.7. When the appropriate optical density was reached, expression of Tho1 was induced by 1 mM IPTG. The cultures incubate overnight at 20°C and 150rpm.

The cells were then harvested. The cultures were transferred to centrifuge buckets and centrifuged at 4200rpm and 4°C for 20 min (r aV<sub>Rotor</sub>: 284 mm). The supernatant was discarded and the pellet was first dissolved in 20 mL lysis buffer. The mixture of lysis buffer and cells was transferred to a beaker and the cells were lysed via ultrasound on ice (5 min, oc: 5, dc: 30%). The mixture was then transferred to ultracentrifuge tubes and centrifuged under vacuum at 4°C and 40000 rpm for 45 min (rotor diameter: 91.9 mm). After centrifugation, soluble proteins are found in the supernatant and other cell residues in the pellet. The supernatant was collected and further used.

The supernatant was transferred on a Ni-IDA 2000 column (Machery-Nagel) previously equilibrated with lysis buffer. The His-tagged protein interacts with Ni<sup>2+</sup> ions and binds to the resin of the column. The binding was followed by 4 wash steps with wash buffers 1-4 and then re-equilibration to the previous ionic strength. The elution was performed 4 times with 500  $\mu L$  elution buffer (250 mM imidazole) and the eluate was collected in fractions. The concentrations of the fractions were determined via Bradford assay. For this, 10  $\mu L$  of the eluates were mixed with 90  $\mu L$

Bradford reagent each, incubated for 5 minutes and the absorbance measured at 595 nm with a plate reader (Tecan Sunrise). The fractions with the highest protein concentration were combined, flash-frozen in liquid nitrogen and then stored at -80°C.

A Bradford assay was used to determine the concentration. 20 µL of the diluted sample (1:10 or 1:100 (= dilution factor)) was mixed with 180 µL Bradford reagent, incubated for 5 minutes and the absorbance measured at 595 nm with a plate reader (Tecan Sunrise). The standard curve with the function  $y=0.0622x + 0.0204$  was used to calculate the amount of protein.

$$\frac{(Abs_{sample} - Abs_{Background}) - 0.0204}{0.0622} = x$$

$$\left( \frac{x}{V_{sample}} : Da \right) \times 10^6 \times Dilution\ factor = c [\mu M]$$

Whereas x is the mass of the protein in 20 µL and Da is the size of the protein in Dalton.

### **Gel electrophoresis (SDS-PAGE and agarose gel-electrophoresis)**

The SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) is used to separate proteins according to their size on a polyacrylamide gel. The gel consists of a stacking gel and a separation gel. In this system, which was used in this work, especially the glycinate ions and chloride ions have an important role. Due to the pH value in the stacking gel, the chloride ions run ahead of the proteins and glycinate serves as an initial trailing ion. At the boundary between the collection and separation gels, there is a stagnation effect, because due to the changing pH value from the collection to the separation gel, the glycinate ions partially lose their braking positive charge and become the leading ion as well. SDS has another important role, as it denatures proteins and gives them a uniform charge. Thus, proteins were separated according to their size, and not according to other properties, such as charge.

The stacking gel consists of 4% PAA and the separation gel of 10% PAA. For small proteins 15% PAA was used for the SDS-PAGE. Before the samples were loaded, they were first mixed with sample buffer, boiled at 95°C for 5 minutes and then centrifuged at 13000 rpm for 5 minutes. After application of the samples, the gel was placed under a current of 75 mA per gel and a voltage of 300 V for about 40 minutes. For SDS-PAGEs of fluorescently labelled proteins (see Bioconjugation), the gel was placed under a current of 75 mA per gel used and a voltage of 300 V for about 30-40 minutes.



This was followed by staining of the proteins with Coomassie staining solution. Basic side chains were bound non-specifically and sequence-independently by the substance Coomassie Brilliant Blue. After the running front (consisting of bromophenol blue and other small molecules) reaches the bottom of the gel, the polyacrylamide gel was removed from the glass plates and briefly washed in distilled water. It was then covered with Coomassie staining solution, briefly warmed and then incubated, shaking, for 15 minutes. The Coomassie staining solution was removed, the gel was washed in distilled water and covered with 10% acetic acid solution and briefly heated. A shaking incubation for 15 minutes follows. The destaining solution was removed, the gel washed and again covered with fresh 10% acetic acid solution and warmed up. An overnight incubation follows to allow the gel to destain completely. The gel was digitised with the Intas gel documentation system.

Agarose gel electrophoresis was used to separate DNA by length. For that a 1% (w/v) agarose gel in TAE buffer was made. The amount of INTAS HDGreen™ recommended by the manufacturer was added to the liquid gel. The gel was poured into a plastic mould and hardened. The DNA from PCR or plasmid preparation were mixed with 1x loading buffer and 10 µL per sample was loaded on the agarose gel. The gel was then placed under a voltage of 100 V. The gel was then documented with the Intas gel documentation system under UV light.

## Bioconjugation

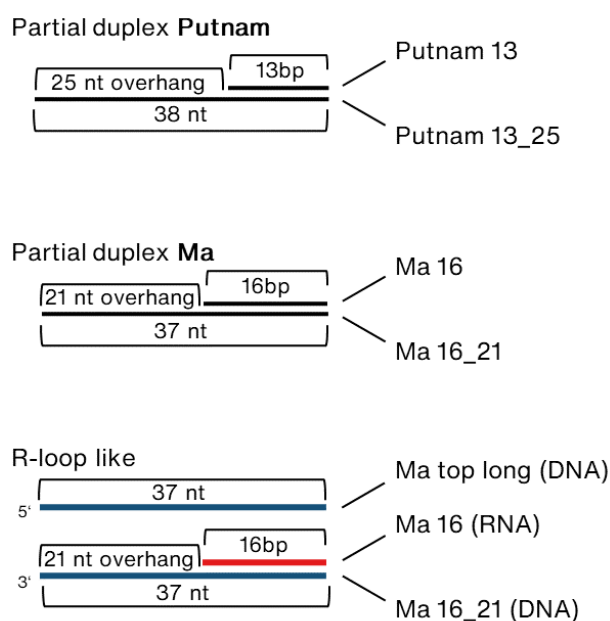
In order to be able to label proteins with a fluorophore, to possibly draw conclusions about the conformation or as a means for future experiments (e.g., FRET, EMSA, etc.), a maleimide coupled to a fluorophore was covalently bound to a cysteine of the target protein.

The dye was dissolved in DMSO and the concentration measured at the NanoDrop. For labelling reactions 2 µM of the target protein was used in one reaction. The reaction was carried out in the dark at room temperature. The dye was added in the indicated concentration. After adding the dye and mixing the reaction, 1 µL 1 M DTT was immediately added to stop the reaction. After the reaction was stopped, SDS sample buffer was added. The samples were then loaded onto a 10% SDS gel (see gel-electrophoresis). The result was documented with the Typhoon FLA 9500. Proteins labelled with Alexa594 were excited at 532 nm and the fluorescence was filtered with the LPG filter ( $\geq 575$  nm). After visualizing the fluorescence, the gel was Coomassie stained.

## Annealing of Substrates

To investigate the binding or unwinding of double stranded substrates, two separated substrates had to be annealed. This method is modified according to Putnam and Jankowsky, 2012.

The two strands were mixed together in equimolar amounts in 1x annealing buffer, and heated to 95°C in a thermocycler, and finally cooled down stepwise to 4°C. The substrates were aliquoted and stored light protected at -80°C (RNA) or -20°C (DNA). In Figure 2 are shown the annealed substrates used in the experimental procedure in this thesis. Putnam substrates were taken from Putnam and Jankowsky, 2013, Ma substrates were derived from Ma *et al.*, 2013.



**Figure 2: Annealed substrates used in this thesis.** The Putnam substrates were the standard substrate to investigate the activity of various helicases. The Ma substrates were used in annealing assays. For the displacement assay a pre-annealed Ma RNA/DNA hybrid were used, and Ma top long, a fully complementary strand to Ma 16\_21 was added to the reaction.

## Electrophoretic Mobility Shift Assay (EMSA)

The substrate binding activity were studied with an Electrophoretic Mobility Shift Assay (EMSA). For that a Cy3 or Cy5 labelled strand were used, or when investigating double strand binding the fluorophore labelled strand were annealed beforehand with an unlabelled complementary strand. The samples were loaded onto a native TBE-Gel. The binding of a substrate changes its running speed in the gel, and is consequently shifted up in the gel. This method is modified according to Saguez *et al.*, 2013.

The binding buffer contains 10 mM MES pH 6.5, 5% Glycerol and 2 mM TCEP. When a helicase was used, 0.5 mM MgCl<sub>2</sub> and 1 mM ADPNP were added and the helicase (1 µM) were preincubated

for 15 min. When only Tho1 (10  $\mu$ M) or an orthologue were used, the reaction mixture did not contain ADPNP or  $MgCl_2$ . After adding the proteins, and in the case of an ATP-dependent enzyme, and a pre-incubation of 15 min, the labelled substrate (100 nM) was added and the reaction incubated light protected for 30 min. After that, the samples were loaded onto a pre-cooled 10% TBE-Gel in a cold room. The gel was placed in 1x TBE buffer under a current of 100 mA at a voltage of 100 V in a cold room for 1.5 hours. The gel was then briefly washed with distilled water and evaluated with the Typhoon FLA 9500. Quantitative analysis with ImageQuant<sup>TL</sup> software followed.

### ATPase assay

The ATPase assay is used to investigate the hydrolysis of ATP by an enzyme, in this case Sub2. In this work a NADH enzyme-coupled absorbance assay was used. The ATP is regenerated to avoid accumulation of ADP, which might have a negative effect on the activity. This method is modified according to Ren *et al.*, 2017.

In this assay, the absorption of NADH is measured at 350 nm. Via a multi-step reaction pathway, one NADH per ADP is oxidised to  $NAD^+$ . NADH absorbs at 340 nm (reduced form), whereas  $NAD^+$  absorbs only at 240 nm. Therefore, one expects a decreasing signal at 350 nm with increasing reaction length. If ADP is formed by the hydrolysis of ATP by Sub2, ADP is converted to ATP by pyruvate kinase (PK), which transfers a phosphate from phosphoenolpyruvate (PEP) to ADP. The resulting pyruvate is converted to lactate by the lactate dehydrogenase (LDH), consuming the cofactor NADH. The signal at 340 nm decreases.

All ATPase experiments were carried out at room temperature using the plate reader Tecan Sunrise. The standard reaction took place in helicase reaction buffer (20 mM MES pH 6.5, 4 mM TCEP). A 100  $\mu$ L reaction mixture, containing 4  $\mu$ M Sub2, 400  $\mu$ M NADH, 4 mM PEP, 2 mM  $MgCl_2$ , 2 mM ATP, 2% v/v enzyme mix (pyruvate kinase and lactate dehydrogenase; Sigma Life Sciences) was prepared in a flat-bottomed clear 96-well plate. Another mix of 100  $\mu$ L containing 1  $\mu$ M substrate (RNA\_Putnam13\_25 rev) was then added to the reaction mixture to start the reaction. After mixing, the measurement of absorbance at 350 nm started immediately. The acquired curve was analysed in Microsoft Excel. With a calibration curve, the amount of hydrolysed ATP in  $\mu$ M per second can be derived from the slope of the curve.

## Helicase assay

The helicase assay is used to study the unwinding of RNA, DNA or hybrid substrates by various helicases. In this thesis, a pair of fluorophore and quencher was used to trace the unwinding reaction. One strand was labelled at the 5' end with a Cy5 fluorophore, the other on at the 3' end with a BHQ2 quencher. In close proximity, BHQ2 quenches the signal from Cy5. After unwinding the substrate, Cy5 can fluoresce and the measured signal increases. The advantage of this setup compared to a gel-based approach is that unwinding can be monitored in real time. This method is modified according to Ordabayev *et al.*, 2018.

All fluorescence helicase experiments were carried out at room temperature using the plate reader Tecan Infinite F200 pro. The assay took place in helicase buffer (20 mM MES pH 6.5, 4 mM TCEP). A 100  $\mu$ L reaction mixture of helicase buffer, with 4  $\mu$ M of Sub2/UAP56, 4 mM ATP, 2 mM  $MgCl_2$  was prepared in a flat-bottomed black 96-well plates. Another mix of 100  $\mu$ L of helicase buffer containing 20 nM of substrate was prepared. To start the reaction the mix containing the RNA was added to the reaction mixture. Measurement of fluorescence was started immediately. An excitation wavelength of 620 nm was used and the fluorescence signal was measured at an emission wavelength of 670 nm. The acquired measurement points were fitted with a single exponential function using Origin (see below).

$$y = y_0 + A_1 e^{-x/t_1}$$

## Annealing Assay

The annealing assay is very analogous to the helicase assay, but in comparison it is used to study the annealing of two nucleic acid strands rather than the unwinding. To do this, the two separate strands are mixed in the presence of an annealing protein. The strands are labelled as in the helicase assay. When both strands are annealing, the fluorophore Cy5 and the quencher BHQ2 are in close proximity and the fluorescence signal decreases.

All fluorescence annealing experiments were carried out at room temperature using the plate reader Tecan Infinite F200 pro. The assay took place in helicase buffer (20 mM MES pH 6.5, 4 mM TCEP). A 100  $\mu$ L reaction mixture of helicase buffer, containing 10  $\mu$ M annealing protein and 20 nM 3' BHQ2-labeled strand (Ma\_16) was prepared in flat-bottomed black 96-well plates. To investigate the effect of helicases on the annealing reaction, 4 mM ATP, 2 mM  $MgCl_2$  and 4  $\mu$ M of the helicase were also added. Another mix of 100  $\mu$ L of helicase buffer containing 20 nM of 5' Cy5-labelled Substrate (Ma\_16\_21) was prepared. To start the reaction the mix containing Cy5-

labelled RNA was added to the reaction mixture. Measurement of fluorescence began immediately. An excitation wavelength of 620 nm was used and the fluorescence signal was measured at an emission wavelength of 670 nm. The acquired measurement points were fitted with a single exponential function using Origin (see helicase assay). The calculated amplitude (A) were divided by the time constant (t) to get an indicator for the activity of the investigated annealing protein.

### **Structure and amino acid sequence analysis**

PyMol is used for the structural analysis. The structures investigated come from the Research Collaboratory for Structural Bioinformatics Protein Databank (RCSB PDB). For analysing the solvent accessible surface (SA) a dot density of 2 and a solvent radius of 1.4 Å were used. The criteria for the chosen cysteine variants were a SA great than 40 Å<sup>2</sup>.

For the sequence analysis, the available data of Sub2 and Tho1 from *Saccharomyces cerevisiae* were first retrieved from the protein database of the National Centre for Biotechnology Information. Then, a BLAST was performed to determine the closest related proteins. The sequence of the nearest 100 relatives were used to create a sequence mask using the software Bioedit on Sub2 and Tho1 from *Saccharomyces cerevisiae*. Finally, a sequence logo was created using WebLogo3. The criteria for the chosen cysteine variants were a conservation < 21%, a conservation grade less than 5 and that the amino acid is not in an already described motif.

# Results

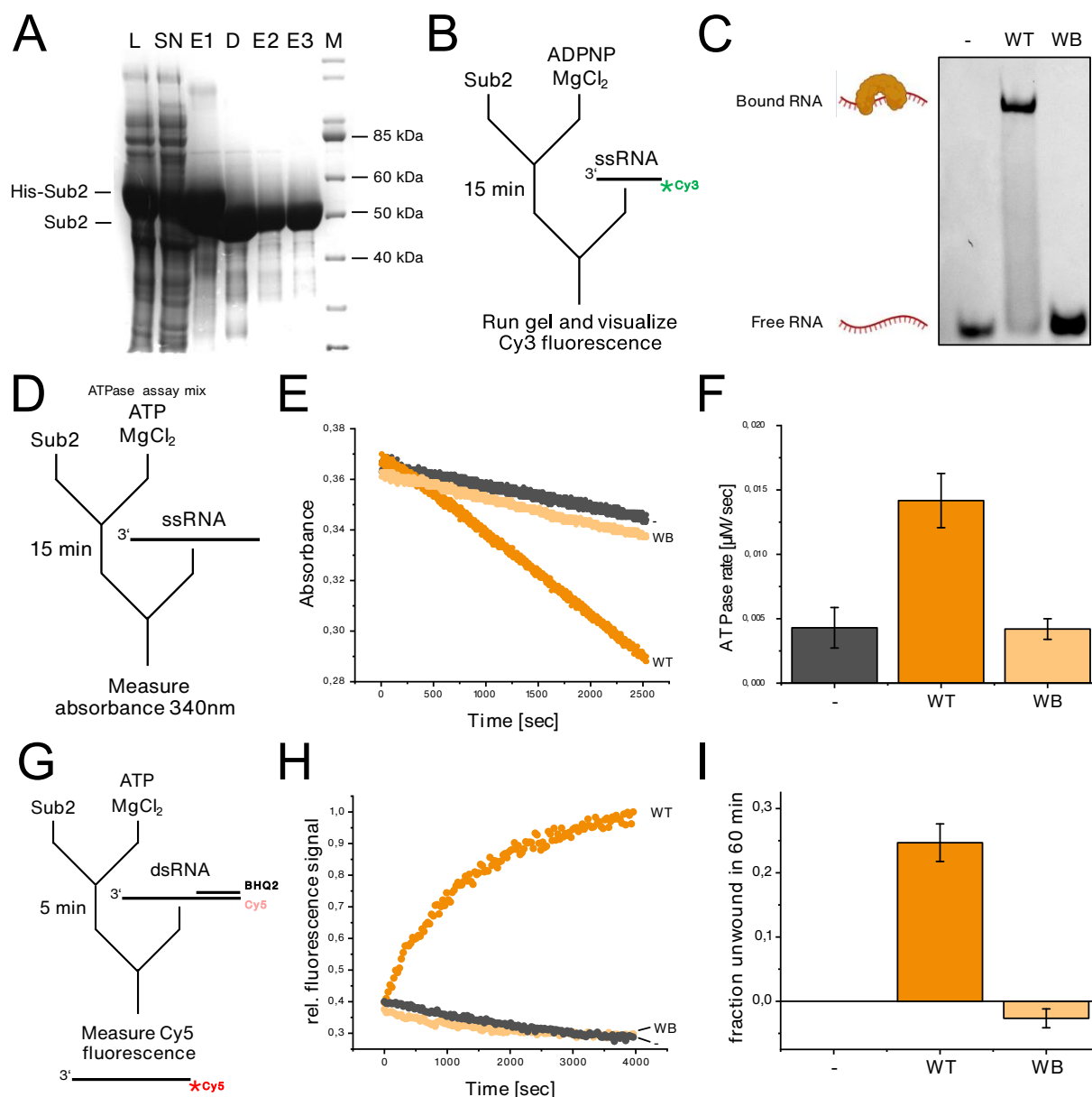
## Chapter 1: Characterization of the DECD-Box RNA helicase Sub2 from *S. cerevisiae*

### Development of suitable assays to investigate DEAD-Box helicases

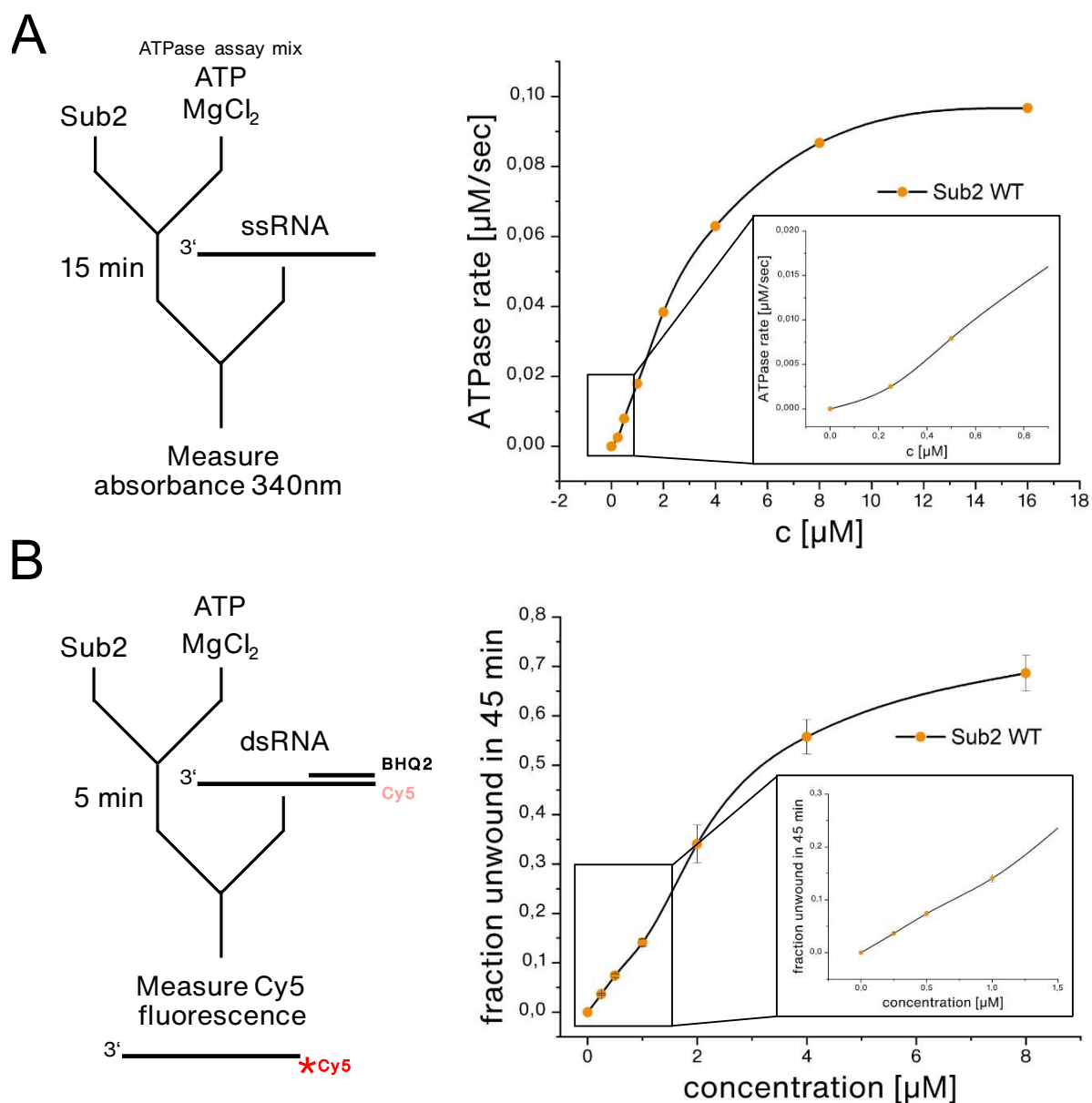
The initial goal was to investigate Sub2 in detail *in vitro* and to study the activity of Sub2 under different conditions. For that, assays had to be developed and be adjusted for the three activities of Sub2: Binding of a substrate, ATP-hydrolysis and unwinding of double stranded RNA. The catalytic dead Sub2 variant D215A (Sub2 Walker B (Sub2 WB)) is used as negative control.

In Figure 3 the activities, studied with different methods, can be seen under standard conditions. To investigate the binding of a substrate an EMSA was used. A fluorescence labelled substrate is incubated in the presence or absence of Sub2 and then loaded on a native TBE gel. The free nucleic acid, in this case RNA, runs faster in the gel, but bound by Sub2 it gets “shifted” upwards in the gel, because it does not run as fast as the free RNA. The inactive variant (Sub2 WB) is not able to bind nucleotides and therefore also no RNA. Furthermore, a NADH-coupled ATPase assay was used. In this method, it is possible to indirectly detect the hydrolysis of ATP to ADP. The absorption of NADH is measured at 350 nm. In a multi-step reaction pathway, one NADH per ADP is oxidised to NAD<sup>+</sup>. In the presence of Sub2 WT, the signal decreases, which means that ATP is hydrolysed. In the presence of Sub2 WB no difference to the sample without any protein is measurable. The slow decrease without protein is merely background.

The last activity to be investigated is the helicase activity. A partially double-stranded substrate is used for this assay. One strand is labelled with a fluorophore (Cy5) and the other strand with a suitable quencher (BHQ2). If the substrate is unwound, an increasing signal can be measured. This activity requires the binding and unwinding of the substrate and the hydrolysis of ATP. In the presence of Sub2 and ATP, an increasing signal can be measured. In the presence of the WB mutant and ATP, no activity can be measured. To be sure that the activity is enzymatic, a concentration series of Sub2 was made in the ATPase assay and the helicase assay as well (Figure 4). Typical for enzymatic activity, an exponential curve results, which saturates at a certain concentration. It should be mentioned, however, that at low concentrations the curve is flatter and linear, and it changes then to an exponential curve.



**Figure 3: Purification and biochemical analysis (RNA-binding, ATPase and helicase) of Sub2 and variants thereof used in this thesis.** (A) SDS-PAGE analysis of progress of His-tag affinity purification of His<sub>6</sub>-Sub2. L = E. coli cell lysate, SN = supernatant, E1 = elution from Ni-beads, D = TEV-cleaved dialysate, E2 = unbound fraction after second Ni-beads, E3 = bound fraction after second Ni-beads, eluted with 30 mM imidazole, for details see Purification of Sub2 in methods; M = molecular weight marker (Unstained Protein Standard, Broad Range (10-200 kDa)). (B) Diagram showing order-of-addition in electrophoretic mobility shift assays (EMSA) of Sub2 with RNA. (C) EMSA of 2  $\mu$ M of wild-type Sub2 (WT) and Sub2 Walker B (D215A) mutant (WB), respectively, with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP 0.5 mM MgCl<sub>2</sub> (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (D) Diagram showing order-of-addition in NADH-coupled RNA-dependent ATPase of Sub2 (E) Time course of decrease in NADH absorbance measured at 340 nm in the absence and presence of wild-type Sub2 (WT) or Sub2 Walker B (D215A) mutant (WB) (see ATPase assay in methods for details). (F) initial velocities ( $\mu$ M/s) of ATPase in absence or presence of the indicated Sub2. Error bars are SE of  $n = 3$  independent experiments. (G) Diagram showing order-of-addition for real-time helicase assay of Sub2 and partial duplex RNA. (H) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 4000 sec) from partial duplex RNA (Putnam 10 nM, for details Annealing of Substrates in methods)) in the absence and presence of 2  $\mu$ M wild-type Sub2 or Sub2 WB variant, respectively (see section Helicase assay in methods for details). (I) Fraction of RNA unwound in 60 min in the absence and presence of wild-type Sub2 or Sub2 WB variant, respectively. Error bars are SE of  $n = 3$  independent experiments.

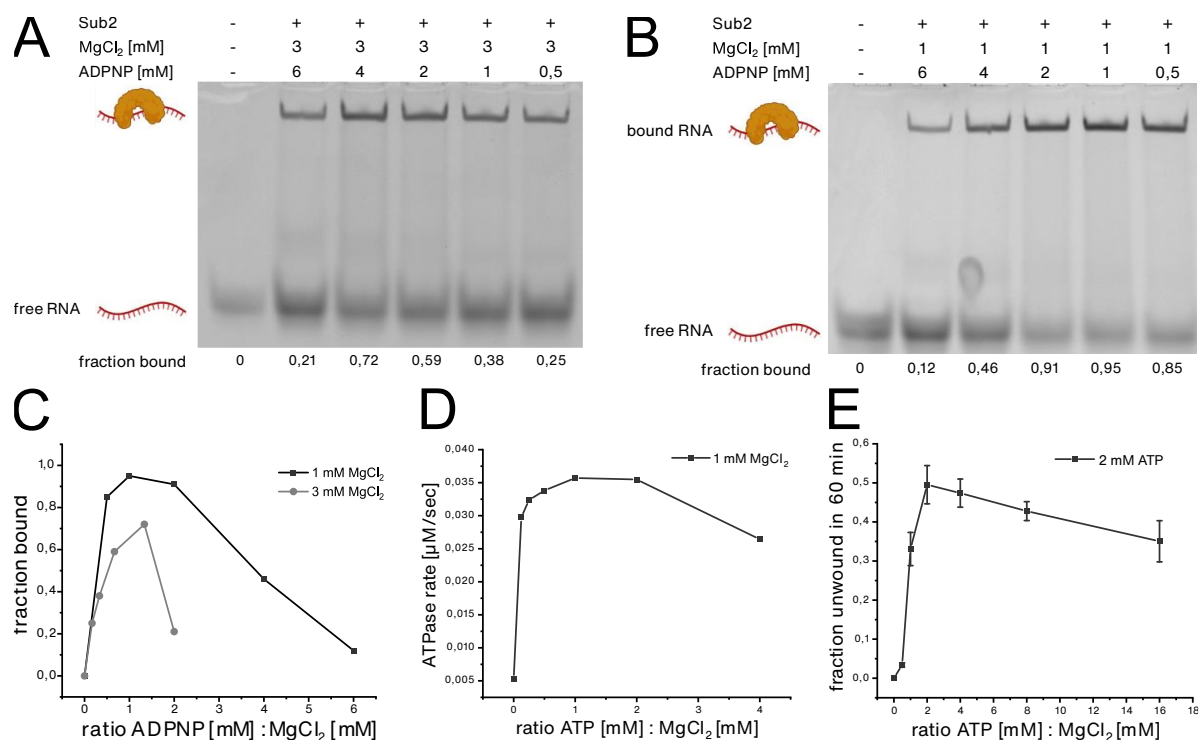


**Figure 4: Concentration dependency of Sub2 wild-type in ATPase and helicase assay.** (A) Diagram showing order-of-addition in NADH-coupled RNA-dependent ATPase of Sub2. Initial velocities ( $\mu\text{M}/\text{s}$ ) of ATPase with the indicated concentrations of Sub2-wildtype (see section ATPase assay in methods for details). Lines between data points are spline connections. (B) Diagram showing order-of-addition for real-time helicase assay of Sub2 and partial duplex RNA. Fraction of 10 nM partial duplex Putnam RNA (see Annealing of Substrates in methods for details) unwound in 45 min with the indicated concentrations of Sub2 wild-type (see section Helicase assay in methods for details). Error bars are SE of  $n = 3$  independent experiments. Lines between data points are spline connections.



## Magnesium and ADP are inhibitory for all activities of Sub2

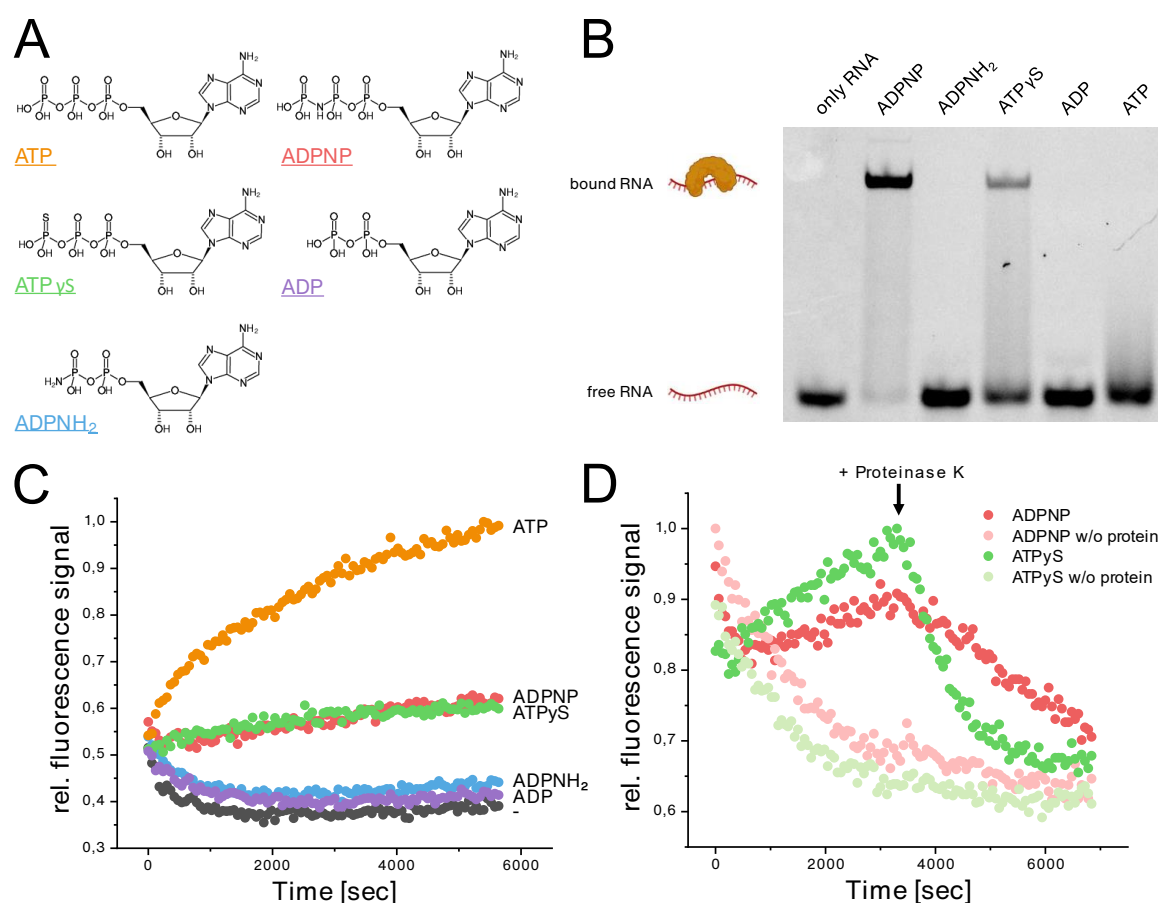
For buffer optimisation, different ATP and  $\text{MgCl}_2$  concentrations were tested. It has already been shown that different DEAD-box helicases require different buffer conditions for maximum activity. This might be connected to a sensing function of DEAD-box proteins (Putnam and Jankowsky, 2013). Figure 5 shows the assays for all three activities with different  $\text{MgCl}_2$  or nucleotide concentrations. Panel C, D and E show the respective ratios of the concentration of nucleotide to  $\text{MgCl}_2$ . In all experiments, an optimum of activity was found, with this being approximately 2 molecules of nucleotide to 1 molecule of  $\text{MgCl}_2$ . From this it can be concluded that an excess of magnesium leads to an inhibition of Sub2, but in the presence of too little magnesium there is also not enough biologically active nucleotide.



**Figure 5: The ratio of magnesium ions to nucleotide is crucial for the activity of Sub2.** (A+B) EMSA of 1  $\mu\text{M}$  of wild-type Sub2 (WT) respectively, with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of indicated concentrations of ADPNP and 3 mM (A) or 1 mM (B)  $\text{MgCl}_2$  (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (C) Quantitative analysis of EMSAs seen in A and B. Shown is the fraction bound at different ratios of ADPNP to  $\text{MgCl}_2$ . (D) Initial velocities ( $\mu\text{M}/\text{s}$ ) of ATPase in the presence of 2  $\mu\text{M}$  Sub2 wild-type, 1 mM  $\text{MgCl}_2$  and indicated concentrations of ATP, shown as ratio of ATP to  $\text{MgCl}_2$  (see section ATPase assay in methods for details). (E) Fraction of 10 nM partial duplex Putnam RNA (see Annealing of Substrates in methods for details) unwound in 60 min with 2  $\mu\text{M}$  of Sub2 wild-type, in the presence of 2 mM ATP and indicated concentrations of  $\text{MgCl}_2$ , shown as ratio of ATP to  $\text{MgCl}_2$  (see section Helicase assay in methods for details). Error bars are SE of  $n = 3$  independent experiments.

Besides the concentration of the nucleotide, the type of nucleotide might also play a role. DEAD-box helicases can not only be sensitive to different buffer conditions, but also to the type of

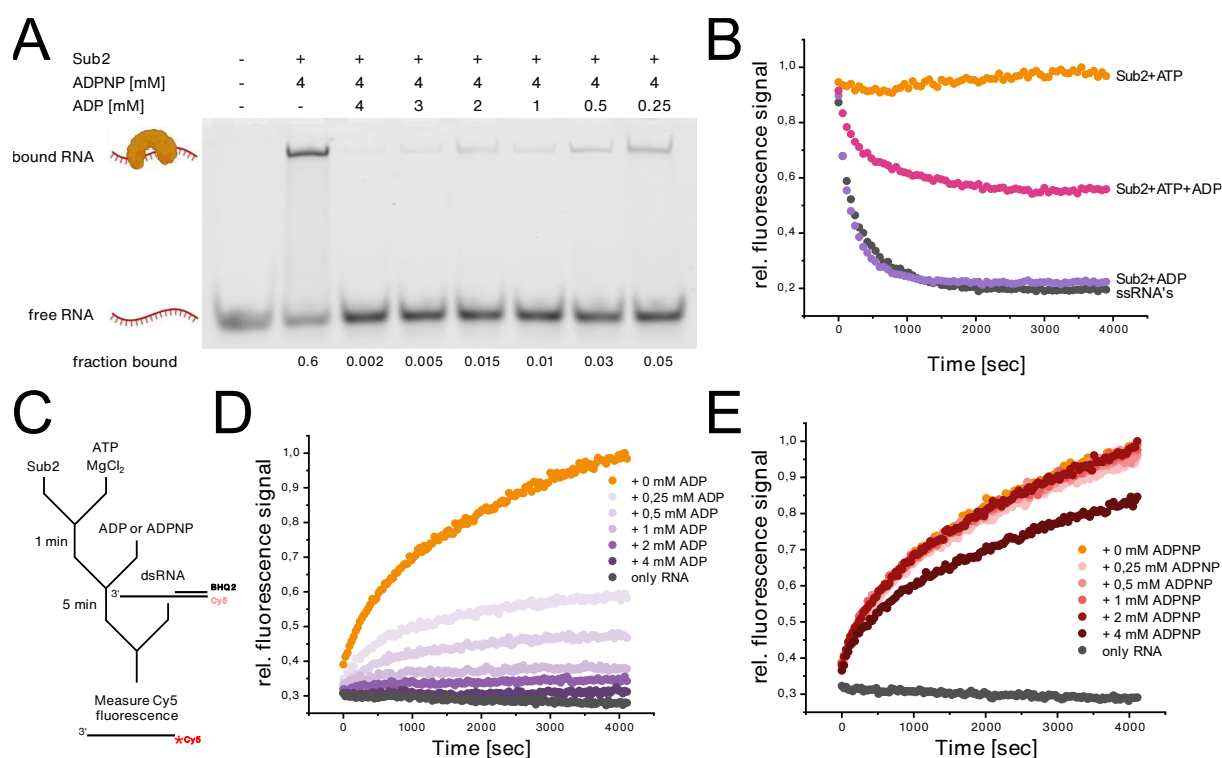
nucleotide (Putnam and Jankowsky, 2013). Therefore, different nucleotides and their analogues were tested (Figure 6). First, the influence of different nucleotides on the binding activity of Sub2 was investigated (panel B). The standard nucleotide used in binding experiments for Sub2 is ADPNP (Saguez *et al.*, 2013), which is bound by the protein, but cannot be hydrolysed and thus leads to an arrested and stable complex. ATP $\gamma$ S is also a non-hydrolysable analogue of ATP, but no published data for Sub2 with ATP $\gamma$ S. In addition to the analogues of ATP, a closer look was also taken at the products of hydrolysis. The influence of ADP on the activity of Sub2 was investigated, but also of ADPNH<sub>2</sub>, which is the hydrolysed form of ADPNP. ADPNP leads to a stable complex of Sub2 with RNA. A shift is also seen with ATP $\gamma$ S, but it is not as effective as with ADPNP. ADP and ADPNH<sub>2</sub> show no complex formation. With ATP, a smear is seen, which is indicative of a complex, but it is not stable enough to remain intact in a gel-shift experiment.



**Figure 6: Impact of different nucleotides on the activity of Sub2.** (A) Structure of the nucleotides used in the following assays (B) EMSA of 2  $\mu$ M of wild-type Sub2, with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM of the indicated nucleotide and 0.5 mM MgCl<sub>2</sub> (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (C) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 5600 sec) from partial duplex RNA (Putnam 10 nM, for details Annealing of Substrates in methods)) in the presence of 2  $\mu$ M wild-type Sub2, 1 mM MgCl<sub>2</sub> and 2 mM of the indicated nucleotide (see section Helicase assay in methods for details). (D) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 10 nM partial duplex Putnam RNA in the presence or absence of 2  $\mu$ M wild-type Sub2, 1 mM MgCl<sub>2</sub> and 2 mM of the indicated nucleotide. Proteinase K was added to all reactions after 3600 sec.

It was already shown that DEAD-box helicases, by binding a double strand, destabilise it locally and separate the double strand. One could therefore say that unwinding occurs by binding. Therefore, the different nucleotides were also tested in the unwinding assay (Figure 6 C). And indeed, the non-hydrolysable ATP analogues show an increasing signal. It is not as strong as with ATP, but clearly higher than ADP and ADPNH<sub>2</sub>, which both do not show any unwinding at all. To confirm that unwinding in the presence of ATP $\gamma$ S and ADPNP are an enzymatic activity of Sub2, the experiment was repeated, but with the difference that a protease is added after a certain time. After the addition of proteinase K, the signal drops significantly, which indicates that Sub2 is digested and thus disassociates from the RNA again and the single strands reanneal. Surprisingly, ADPNP and ATP $\gamma$ S show a similar curve in the helicase assay, although in the EMSA experiment the binding efficiency with ATP $\gamma$ S is far less.

DDX19 is a DEAD-Box helicase involved in mRNA export in *H. sapiens* with an ADP triggered alpha helical switch (Collins *et al.*, 2009). Not only DDX19 is a DEAD-Box protein which is affected by ADP. ADP also activates the annealing function of Ded1 from *S. cerevisiae* (Yang and Jankowsky, 2005). And last but not least, surprisingly the crystal structure of UAP56 could have been achieved with ADP, but not with ATP or an ATP analogue (Shi *et al.*, 2004). Subsequently, the influence of ADP on the activity of Sub2 was tested as well (Figure 7).



**Figure 7: The ATPase Sub2 is inhibited by ADP.** (A) EMSA of 2  $\mu$ M of wild-type Sub2 with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 4 mM ADPNP, 3 mM MgCl<sub>2</sub> and additionally the indicated concentration of ADP (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (B) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from ssRNAs (Putnam13 10 nM and Putnam13\_25 10 nM, for details see Substrates in materials) in the presence of 4  $\mu$ M wild-type Sub2, 1 mM MgCl<sub>2</sub> and 2 mM ATP, respectively 2 mM ADP or both together (see section Helicase assay in methods for details). (C) Diagram showing order-of-addition for real-time helicase assay of Sub2 and partial duplex RNA shown in D and E. (D+E) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 4200 sec) from partial duplex RNA (Putnam 10 nM, for details see section Annealing of Substrates in methods) in the presence of 2  $\mu$ M wild-type Sub2, 2 mM MgCl<sub>2</sub>, 2 mM ATP and in the absence or additionally presence of the indicated concentrations of ADP (D) or ADPNP (E) (see section Helicase assay in methods for details).

First, the influence of ADP on the binding activity of Sub2 was investigated. A fixed concentration of ADPNP was used and additionally ADP was added to the reaction. It is shown that even small concentrations of ADP inhibit the binding activity of Sub2. Since even small concentrations of ADP lead to strong inhibition, ADP seems to have a much higher affinity for the binding site of Sub2 than ADPNP.

After the already mentioned annealing activity of Ded1, triggered by ADP, it was investigated whether an annealing activity can also be triggered for Sub2 after the addition of ADP. The single stranded substrates are added individually to the reaction and the fluorescence is measured. In the presence of ATP, Sub2 is able to keep the two strands separate. When an equal concentration of ADP is added, the signal initially decreases quickly and reaches an equilibrium between the maximum and minimum measured fluorescence signal. In the presence of ADP alone, the RNA anneals, but no acceleration is seen compared to the sample without protein.

Next, the influence of ADP on the helicase activity of Sub2 was investigated. In addition to a fixed concentration of ATP, the indicated concentrations of ADP were added (Figure 7 D). Here, as with the EMSA experiment, it shows that even small concentrations of ADP lead to inhibition of the helicase activity of Sub2. To test whether this is a specific effect by ADP, the same experiment was carried out with ADPNP (Figure 7 E). No inhibition by ADPNP was detected. Even an equimolar concentration does not lead to reduced helicase activity. Only when the concentration of ADPNP (4 mM) exceeds that of ATP (2 mM), the fluorescence signal does not reach the same intensity.

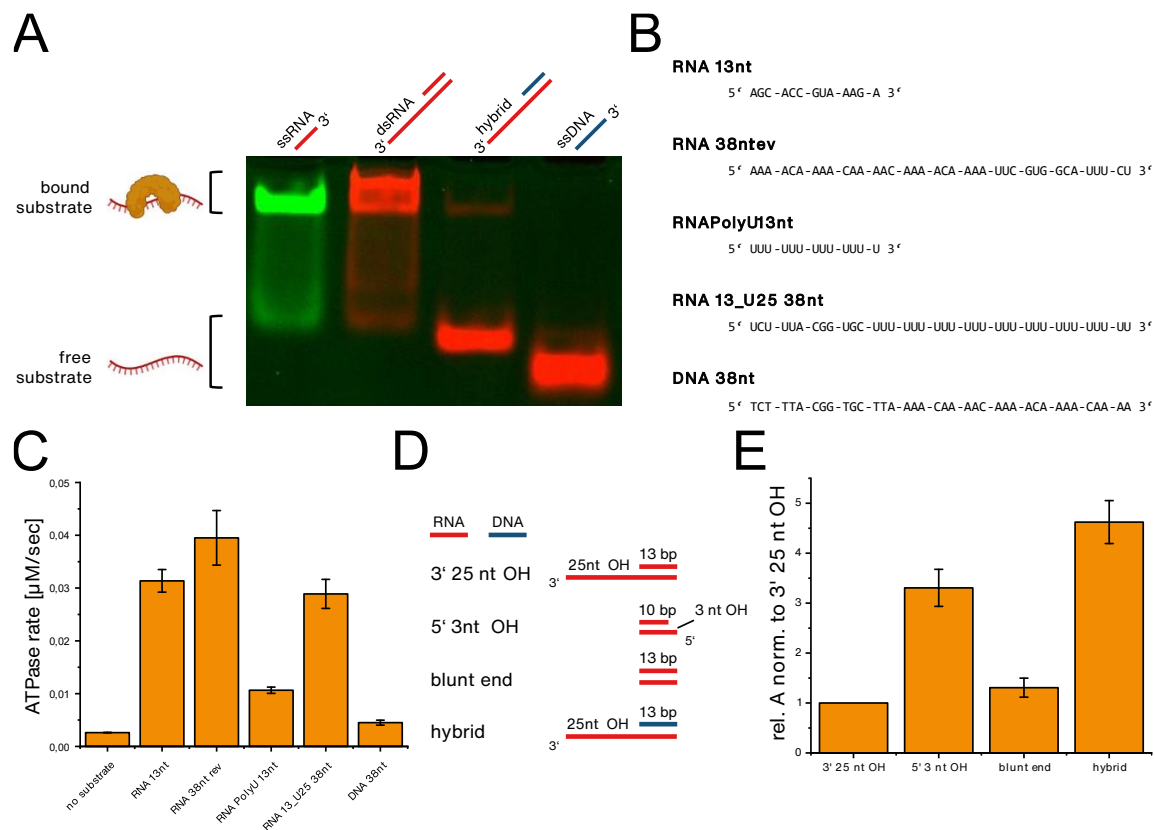
### **Sub2 unwinds RNA/DNA hybrids**

Once the nucleotide specificity has been investigated, the next step was to take a closer look at the substrates to be unwound. Various substrates were tested in the different assays (Figure 8).

In the EMSA experiment, the 13 nt ssRNA, which has always been used up to now, was used. Furthermore, the standard substrate for the helicase assay was investigated (partial duplex Putnam) and the same substrate, but with one strand as DNA (hybrid). Single-stranded DNA was also tested as a control. The single-stranded RNA is bound as expected. The partial duplex is also bound, but shows two bands. One at the level of the ssRNA and one shifted further up. This is probably the unwound single strand and the partial duplex, supported by the data already shown (Figure 6 C), where it has been proved that binding with ADPNP might also lead to partial unwinding. The hybrid is only very weakly bound and the single-stranded DNA is not bound at all.

Furthermore, different substrates were used in the ATPase assay, to investigate if the composition of the nucleotides of the substrate makes a difference in the stimulation of Sub2's ATPase, or if the properties does not matter, as long as it is RNA (Figure 8 B+C). On the one hand, different lengths are compared with each other and on the other hand, whether the content of uracil plays a role. As expected, DNA does not stimulate ATPase activity. Comparing the ATPase activity on a 13 nt and a 38 nt long substrate, one sees that the ATPase activity is stimulated more on the long substrate, but the stimulus is not proportional to the length. A greater influence is observed when 13U is compared with 13\_U25 RNA. However, it must be pointed out here that the long substrate does not consist exclusively of uracil. If one compares the 13 nt long Putnam substrate with the 13 nt substrate consisting exclusively of uracil, one can see that the ATPase activity is stimulated less by uracil. DNA does not stimulate ATPase activity, but surprisingly the activity is reduced when using a substrate consisting only of uracil, an exclusive feature of RNA.

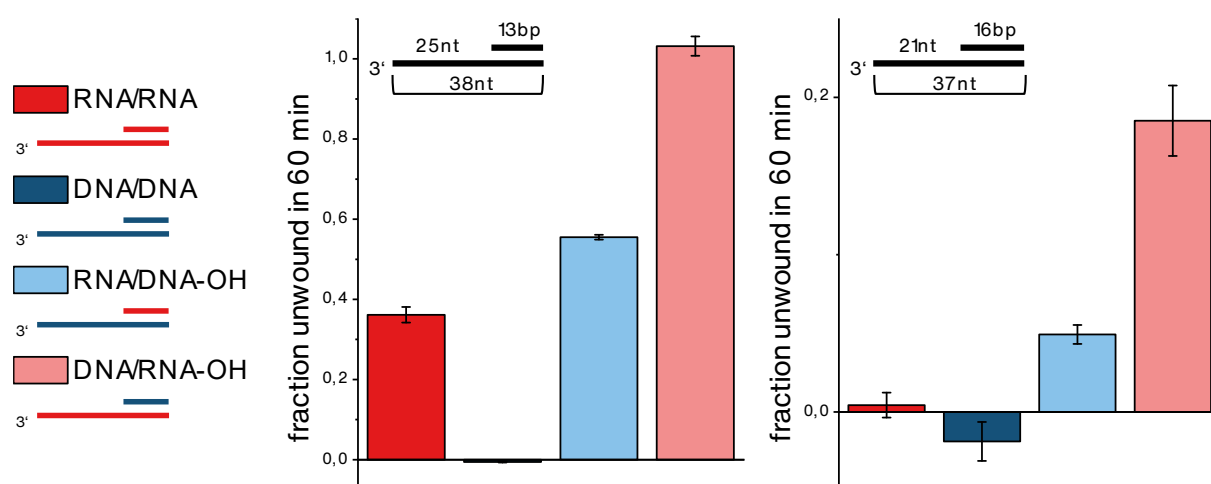
Finally, different substrates were tested in the unwinding assay (Figure 8 D). The question was whether the overhang is needed for unwinding and whether the orientation of it is important. Furthermore, an RNA/DNA hybrid was examined to determine whether Sub2 binds the top strand without overhang or on the bottom strand with overhang to unwind these substrates. All substrates are compared with the standard partial duplex Putnam substrate. Three times more of the substrate with 10 bp duplex and a short 3 nt 5' overhang is unwound than the standard substrate. It is unclear whether the shorter duplex region or the 5' overhang is responsible for this effect. However, the 13 bp blunt end substrate can be seen to be unwound similarly to the standard substrate. This suggests that the overhang plays a secondary role, and mainly the shorter duplex region has a positive effect on unwinding. But the best substrate to unwind is the RNA/DNA hybrid. Four times more of this substrate can be unwound than the standard substrate. Based on this interesting finding, several substrate combinations were tested to follow up on this result



**Figure 8: Investigation of substrate specificity of Sub2.** (A) EMSA of 1 μM of wild-type Sub2 and 100 nM of the indicated substrates (from left to right: RNA Putnam13 (ssRNA), partial duplex RNA Putnam (dsRNA), partial duplex hybrid Putnam (hybrid), UvrD\_top18\_Cy5 (ssDNA), see section Substrates in materials for details) labelled with Cy3 (3', green) or Cy5 (5', red) in the presence of 1 mM ADPNP 0.5 mM MgCl<sub>2</sub> (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details) (B) Substrates used in C to stimulate ATPase activity. (C) Initial velocities (μM/s) of ATPase in the presence of 2 μM Sub2 wild-type, 1 mM MgCl<sub>2</sub> and 1 mM ATP with different substrates (see section ATPase assay in methods for details). Error bars are SE of n = 3 independent experiments. (D) Schematic overview of the double stranded substrates used for the helicase assay shown in E. Sequences are derived from standard Putnam substrates. OH = overhang (E) Calculated amplitude of the time course of the unwinding reaction of several substrates, normalized to standard substrate pdRNA Putnam (see section Annealing of Substrates in methods for details) in the presence of 2 μM wild-type Sub2, 1 mM MgCl<sub>2</sub> and 2 mM ATP (see section Helicase assay in methods for details). Error bars are SE of n = 3 independent experiments.

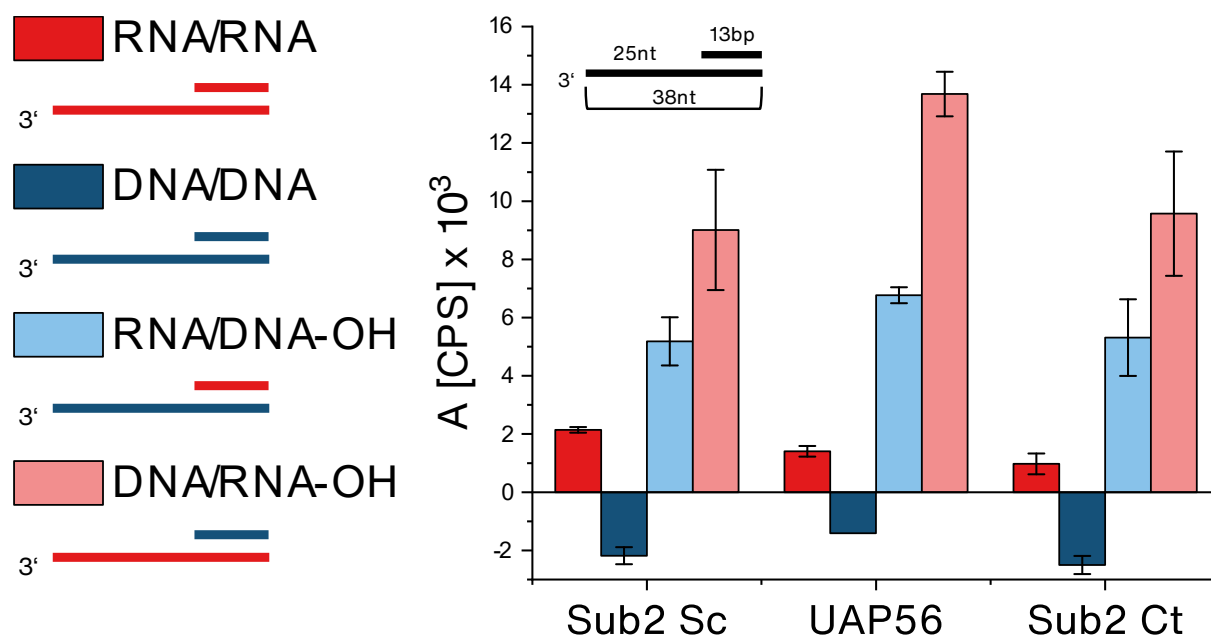
RNA/RNA, DNA/DNA and hybrid substrates, each with overhangs, were tested (Figure 9). The first substrates were the Putnam substrates (13 nt duplex with 25 nt overhang). Sub2 is able to unwind all substrates except dsDNA. As already shown in Figure 8 the RNA helicase Sub2 unwinds the RNA/DNA hybrid with RNA overhang best. The hybrid substrate with DNA overhang is similarly well unwound as the RNA/RNA substrate.

After observing that Sub2 almost completely unwinds the hybrid substrate with an RNA overhang, a substrate was tested that has a partial duplex of 16 nt. Sub2 unwinds significantly less of the substrate, but is still able to unwind it. It is striking that the pattern does not change in comparison to the 13 nt duplex substrate. DNA/DNA is not unwound at all, RNA/RNA and the hybrid with DNA overhang are unwound similarly well. The hybrid with RNA overhang is unwound best.



**Figure 9: Sub2 unwinds RNA/DNA hybrids.** Fraction of 10 nM partial duplex Putnam or Ma RNA/RNA, DNA/DNA and hybrids (see Annealing of Substrates in methods for details) unwound in 60 min in the presence of 2  $\mu$ M Sub2 wild-type, 1 mM  $MgCl_2$  and 2 mM ATP (see section Helicase assay in methods for details). Error bars are SE of  $n = 3$  independent experiments

To check whether this specificity is conserved at least within the Sub2 orthologues, the orthologues from *H. sapiens* (UAP56) and *C. thermophilum* (Sub2 Ct) were purified and tested for their unwinding activity (Figure 10). And indeed, all proteins show the same pattern. The hybrid with RNA overhang is best unwound. Sub2 from *S. cerevisiae* and *C. thermophilum* show very similar activity. In the case of UAP56, it can be assumed that the activity in comparison to its orthologues is lower than shown, since 4  $\mu$ M instead of 2  $\mu$ M was used.



**Figure 10: Unwinding of RNA/DNA hybrids is conserved among *Homo sapiens* and *Chaetomium thermophilum*.** Calculated amplitude of the time course of the unwinding reaction of 10 nM partial duplex Putnam or Ma RNA/RNA, DNA/DNA and hybrids (see Annealing of Substrates in methods for details) in the presence of 2  $\mu$ M Sub2 from *Saccharomyces cerevisiae* or *Chaetomium thermophilum* or 4  $\mu$ M UAP56 from *Homo sapiens* (all purified recombinant from *E. coli*) with 1 mM MgCl<sub>2</sub> and 2 mM ATP (see section Helicase assay in methods for details). Error bars are SE of  $n = 3$  independent experiments.

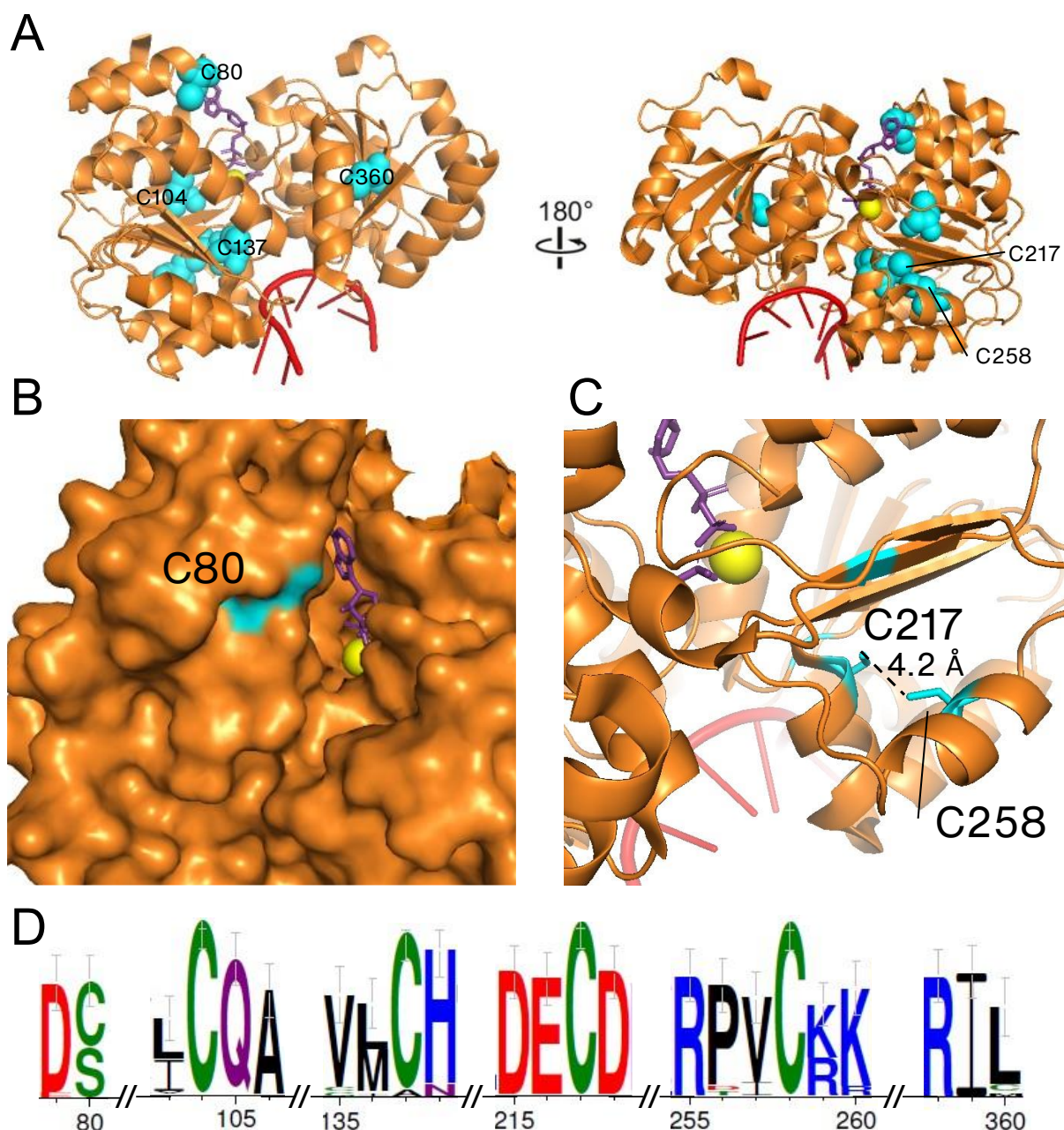


## Chapter 2: Sub2 has several conserved cysteines with unknown function.

### Structural and sequential analysis of Sub2

Structural and sequential analysis of Sub2 revealed that this protein not only contains six cysteines, but that four of them are highly conserved. Cysteines take on a special role as an amino acid. It is the only amino acid with a thiol group and, together with methionine, the only amino acid that contains sulphur. Because of the thiol group, cysteines have properties that make them unique and often highly conserved (Marino and Gladyshev, 2010). Cysteines are known for their ability to form disulphide bridges, a covalent bond between two thiol groups, such as between two cysteines (Poole, 2015). They are also often found in catalytic centres, where their unique properties allow them to carry out various reactions (Marino and Gladyshev, 2008). Furthermore, they serve as antioxidants (Marino and Gladyshev, 2009) or can be tagged in the cell as a signal for subsequent signal pathways (Paulsen and Carroll, 2013). In addition to the biologically relevant functions, the presence of a cysteine can be utilised for experiments. It is possible to let the thiol react with a maleimide so that they form a covalent bond (Hermanson, 2013). The maleimide can in turn be coupled with a fluorophore or other molecule. In this way, a protein can be specifically labelled by introducing a cysteine on the surface. For the most precise results, the existing cysteines must be removed and a new one inserted on the surface at an unfunctional position.

Analysis of the sequence shows that the cysteines at positions 104, 137, 217 and 258 are highly conserved (Figure 11 D). C217 is the position that gives its name to the subgroup of DECD-box helicases. The other three highly conserved cysteines are all located in the N-terminal Rec-A like domain (Figure 11 A). In addition, C217 and C258 can potentially form a disulphide bridge. Specific to Sub2 are C80 in the N-terminal domain and C360. The cysteine at position 80 is located on the surface (Figure 11 B) and is in close proximity to the nucleotide binding pocket. It is semi-conserved and another amino acid that frequently occurs at this position in related proteins is serine. Serine and cysteine are structurally very similar and differ only in one atom. Instead of a sulphur, serine has an oxygen atom. The question is whether this cysteine plays a functional role on the surface of Sub2. The other highly specific cysteine in Sub2 is the one at position 360. It is hardly conserved, is located within the protein, near the surface but without exposure, and is the only cysteine in the c-terminal domain. Its function is also completely unknown. These two cysteines at position 80 and 360 are therefore of great interest in the studies on Sub2.

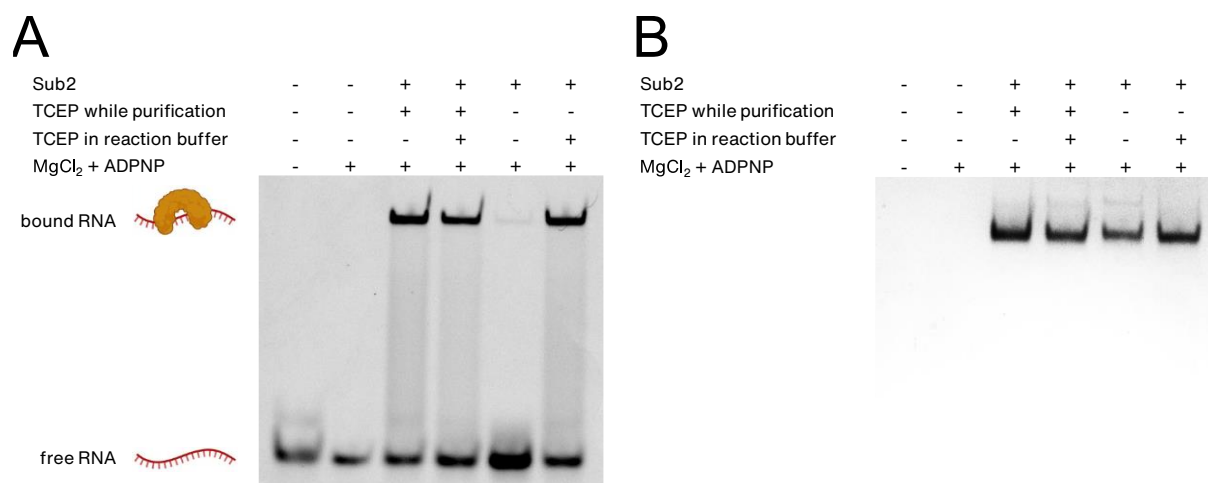


**Figure 11: Bioinformatical analysis of Sub2's cysteines.** (A) Crystal structure of Sub2 (5SUP) in complex with RNA (red), magnesium (yellow) and ADP-BEF (purple), highlighting the cysteines of Sub2 (cyan spheres). (B) Surface of the crystal structure of Sub2 (5SUP). Shown is the nucleotide binding pocket with cysteine 80 (cyan). (C) Crystal structure of Sub2 with focus on cysteine 217 and cysteine 258 forming a potential disulfide bridge. A-C created with PyMol. (D) Sequence logo of the 100 closest DECD-box RNA helicases to Sub2, highlighting the positions of Sub2's cysteines (for details see section Structure and amino acid sequence analysis). Created with WebLogo 3.

### Reducing agent is necessary to maintain Sub2's activity

Since reducing agents might have an influence on the corresponding bioconjugation and experiments on the cysteines, it was tested whether reducing agent is necessary to maintain the activity of Sub2 (Figure 12). Sub2 was purified in presence or absence of tris(2-carboxyethyl)phosphine (TCEP). A binding experiment was then carried out with these

purifications in the presence or absence of TCEP. The results show that a reducing agent while purification is necessary to maintain the activity. If purified without TCEP, the protein is no longer or hardly able to bind a 13 nt long single-stranded RNA. However, the activity can be recovered if the binding experiment is carried out in the presence of TCEP.



**Figure 12: Reducing agent is necessary to maintain Sub2's binding activity.** (A) EMSA of 1  $\mu$ M of wild-type Sub2, purified with or without TCEP and 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP 0.5 mM MgCl<sub>2</sub> and 2 mM TCEP, respectively without any reducing agent (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (B) Coomassie staining of EMSA in A.

## Introducing additional cysteines to Sub2

Since the aim was to specifically label Sub2 in order to perform FRET experiments, the existing cysteines need to be removed, but at the same time an additional cysteine has to be introduced on the surface at an unfunctional position. In this thesis, both approaches were taken simultaneously. On the one hand, additional cysteines were introduced into Sub2 wild-type and experiments were carried out, and on the other hand, cysteine deletion variants were produced and tested.

First of all, positions had to be chosen that were suitable for the introduction of a cysteine and the planned experiments. Table 15: Positions of additionally introduced shows an overview of the designed variants. On the one hand, positions in the different domains of Sub2 were selected in order to follow a conformational change of Sub2 via FRET experiments. Furthermore, a position close to the binding site of the C-box of Yra1 was chosen (Ren *et al.*, 2017, pdb: 5SUP) for tracking protein-protein interaction. Furthermore, a crystal structure was recently obtained showing the THO complex and Sub2 in a homodimer (Schuller *et al.*, 2020, pdb: 7AQO). The selected positions can also be used to track the dimerization of this complex. For details of the criteria used for the selection of the position shown, see section Structure and amino acid sequence analysis in methods.

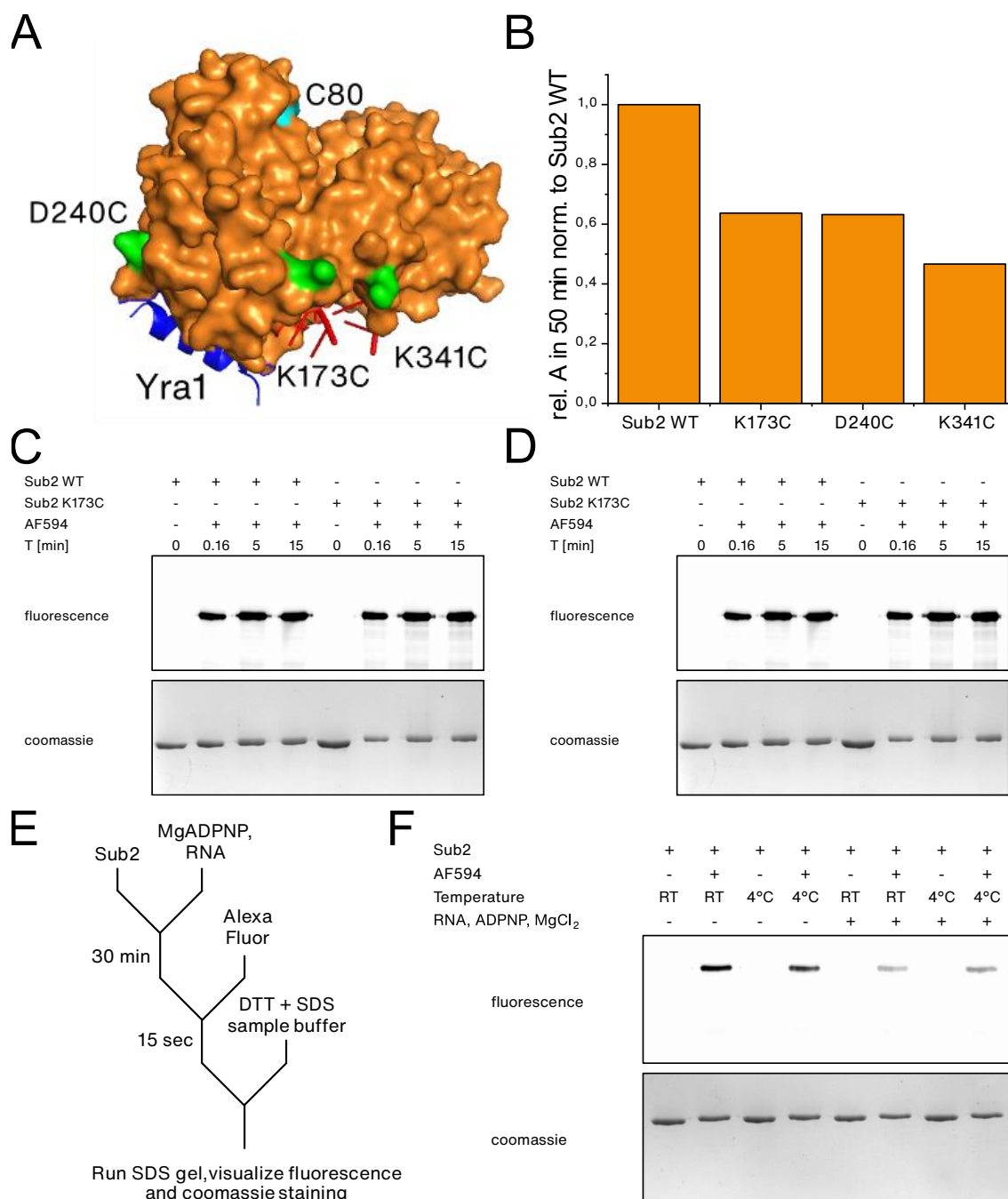
**Table 15:** Positions of additionally introduced cysteines

Residue	aa position	SA [Å <sup>2</sup> ] (5SUP)	Conservation	Conservation grade	Reason	Assay
LYS	173	53.21	20.60%	2	Proximity to RNA binding site (N-terminal RecA-like domain)	Conformational change, FRET THO-Sub2 dimer
LYS	341	49.74	12.10%	1	Proximity to RNA binding site (C-terminal RecA-like domain)	
ASP	240	51.22	20.60%	1	Proximity to Yra1	Protein-Protein interaction, negative control
LEU	177	54.47	9.30%	1	Proximity to RNA binding site (N-terminal RecA-like domain)	Conformational change,
ASN	320	42.82	10%	3	Proximity to RNA binding site (C-terminal RecA-like domain)	FRET THO-Sub2 dimer

Figure 13 A shows the examined variants in the structure. Variants K173C, D240C and K341C were selected for initial testing. These were tested for activity and then labelling experiments were carried out. The variants all show reduced activity in the helicase assay. Even though the activity is reduced, these variants can still be used, taking into consideration the reduced activity.

Next, labelling experiments were performed. For a first test, the wild type was labelled with K173C at different durations of incubation with the dye. The reaction of the thiols with the maleimides was very fast. Subsequently, the remaining variants were tested and potentially a stronger fluorescence is shown compared to the wild type, which can also be labelled (Figure 13 D). Since the fluorescence is nevertheless stronger than expected, the hypothesis arose that not only C80 on the surface can be labelled.

Therefore, another labelling experiment was carried out with different conditions. One was the temperature and the other one the presence of RNA and nucleotide. At lower temperatures, the reaction should be slower and more specific for cysteines on the surface. Furthermore, the change of conformation, upon binding nucleotide and RNA, could lead to a more compact structure and fewer cysteines would be accessible. The temperature does not seem to have any influence on the labelling efficiency. This has the advantage that the protein can be kept on ice permanently to prevent degradation. However, the fluorescence decreases significantly with prior incubation with RNA and nucleotide. Fewer cysteines seem to be accessible.



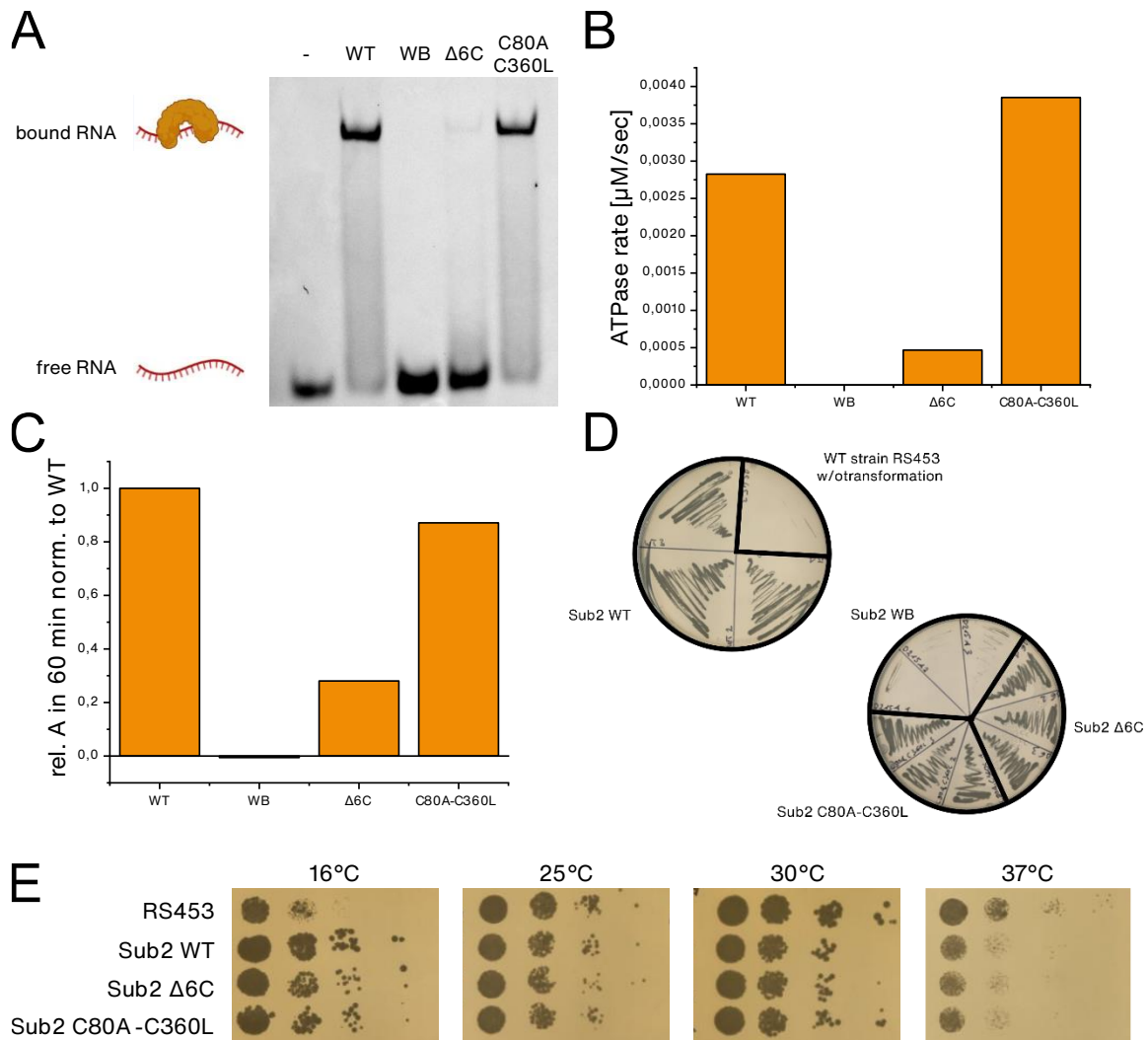
**Figure 13: Introducing additionally cysteines for future fluorescence-based assays.** (A) Crystal structure of Sub2 (5SUP) in complex with RNA (red) highlighting the introduced cysteines at amino acid position 173, 240 and 341 (green) and the existing cysteine 80 on the surface. The c-terminal box of Yra1 is shown in blue. Created with PyMol (B) Amplitude of the time course of the unwinding reaction of partial duplex Putnam RNA (for details see section Annealing of Substrates in methods) after 50 min normalized to wild-type Sub2 (Sub2 WT) in the presence of 2  $\mu$ M wild-type Sub2, Sub2 variant K173C, D240C or K341C, 1 mM  $MgCl_2$  and 2 mM ATP (see section Helicase assay for details). (C) Labelling Experiment of 2  $\mu$ M of wild-type Sub2 and variant K173C with 4  $\mu$ M Alexa fluor 594 (AF594) in binding buffer. Labelling reaction was stopped after the indicated times with 100 mM DTT and applied to a 10% SDS-Gel. After visualization of AF594 the gel was Coomassie stained (see section Bioconjugation for details). (D) Labelling Experiment of 2  $\mu$ M of wild-type Sub2 and variant K173C, D240C and K341C with 2  $\mu$ M Alexa fluor 594 (AF594) in binding buffer. Labelling reaction was stopped after 15 seconds with 100 mM DTT and applied to a 10% SDS-Gel. After visualization of AF594 the gel was Coomassie stained (see section Bioconjugation for details). (E) Diagram showing order-of-addition in the labelling assay of Sub2 with RNA, ADPNP and  $MgCl_2$  shown in F. (F) Labelling Experiment of 2  $\mu$ M of wild-type Sub2 with 2  $\mu$ M Alexa fluor 594 (AF594) in binding buffer. Before adding the dye, Sub2 incubated at room temperature (RT) or on ice (4°C) in the absence or presence of 1 mM ADPNP and 0.5 mM  $MgCl_2$ . Labelling reaction was stopped after 15 seconds with 100 mM DTT and applied to a 10% SDS-Gel. After visualization of AF594 the gel was Coomassie stained (see section Bioconjugation for details).

### **Deletion of all cysteines do not show an altered phenotype *in vivo***

As mentioned earlier, it would be desirable to have a cysteine-free protein in order to label it at a specific position. Since the function of the cysteines of Sub2, except for the one in the DECD motif, is unknown, different activities were tested and an *in vivo* survival assay was performed.

First, all six cysteines were deleted and second, the two cysteines specific for Sub2, C80 and C360, were deleted and tested *in vitro* (Figure 14). The Walker B variant of Sub2 was also tested as a negative control. If all cysteines are deleted, the binding activity is clearly decreased. Only a barely detectable part of the RNA is shifted. In contrast, the double mutant does not seem to have lost any activity. A similar pattern is shown in the ATPase and helicase assay. The activity of the  $\Delta 6C$  mutant has lost most of its activity, but is not completely inactive compared to the dead variant WB. The double mutant shows similar activity as the wild type.





**Figure 14: Substitution of all cysteines does not abolish Sub2's activity.** (A) EMSA of 1  $\mu$ M of wild-type Sub2 (WT), Walker B mutated (D215A, WB), all cysteines substituted ( $\Delta$ 6C) or C80A-C360L with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP and 0.5 mM  $MgCl_2$  (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details) (B) Initial velocities ( $\mu$ M/s) of ATPase in the presence of 2  $\mu$ M Sub2 wild-type, or the indicated variants, 1 mM  $MgCl_2$  and 1 mM ATP (see section ATPase assay in methods for details). (C) Amplitude of the time course of the unwinding reaction of partial duplex Putnam RNA (for details see section Annealing of Substrates in methods) after 60 min normalized to wild-type Sub2 (Sub2 WT) in the presence of 2  $\mu$ M wild-type Sub2, or the indicated variants, 1 mM  $MgCl_2$  and 2 mM ATP (see section Helicase assay for details) (D) FOA-plates to shuffle out the Sub2 wild-type plasmids. Lethal variants do not show any growth (see section Yeast viability assay (Dot spots)). (E) Ten-fold serial dilution of not transformed yeast strain RS453-Sub2 shuffle or same strain with shuffled Sub2 wild-type plasmid or the indicated variants of Sub2 on YPD plates at different temperatures. Cells were grown 2 (25°C, 30°C), 3 (37°C) or 7 days (16°C) (see section Yeast viability assay (Dot spots) for details).

Since it appears that the cysteines are not absolutely necessary for the enzymatic activity, the question arose whether these variants show a phenotype *in vivo* (Figure 14 D). The deletion of Sub2 leads to death of the yeast cells (Libri *et al.*, 2001). The WB variant was again chosen as a control. This variant dies upon transformation by shuffling out the wild type of Sub2. Surprisingly, neither  $\Delta$ 6C nor the double mutant show an obvious phenotype. Therefore, a dot spot assay was performed at different temperatures to expose the cells to additional stress (Figure 14 E). However, no difference is visible in contrast to the wild type. Deletion of all cysteines leads to less activity, but does not seem to play a primary role in the survival of the cell and *in vivo* function of Sub2.

## Chapter 3: Modulation of the activity of Sub2 by Tho1.

### Sub2 and Tho1 form a ternary complex on RNA

After investigating only Sub, the effect of Tho1 on Sub2 was analysed in different assays. First, the influence of Tho1 on the binding activity of Sub2 was investigated (Figure 15 A-C). As expected, Sub2 binds to single-stranded 13 nt long RNA. Tho1, on the other hand is not able to bind this short RNA efficiently under these reaction conditions. However, if Sub2 and Tho1 are incubated together with the RNA, a previously unknown band appears. This does not correspond to Sub2 or Tho1 alone. A ternary complex seems to form. The presence of Sub2 and Tho1 in this one band was confirmed by mass spectrometry<sup>1</sup>. The question is how this ternary complex is formed. Both proteins could bind cooperatively to the RNA or only one of the proteins binds directly to the RNA and the other protein interacts with the RNA-binding protein. To answer this question, binding-deficient variants of Sub2 and Tho1 were tested. In Figure 15 B, the Walker B variant of Sub2 was used and incubated together with Tho1. The formation of a ternary complex does not occur. The Walker B variant does not interfere with the formation of the complex. The assumption is that RNA-binding activity or nucleotide binding is necessary for the formation of the ternary complex. Furthermore, the wild type of Sub2 was tested with a binding-deficient variant of Tho1. In this variant, the C-terminal end is deleted. This variant is no longer able to bind any substrate. If Sub2 and Tho1  $\Delta$ CTE are incubated with RNA, a ternary complex is nevertheless formed. So far, for ternary complex formation the binding activity of Sub2, but not of Tho1, is required.

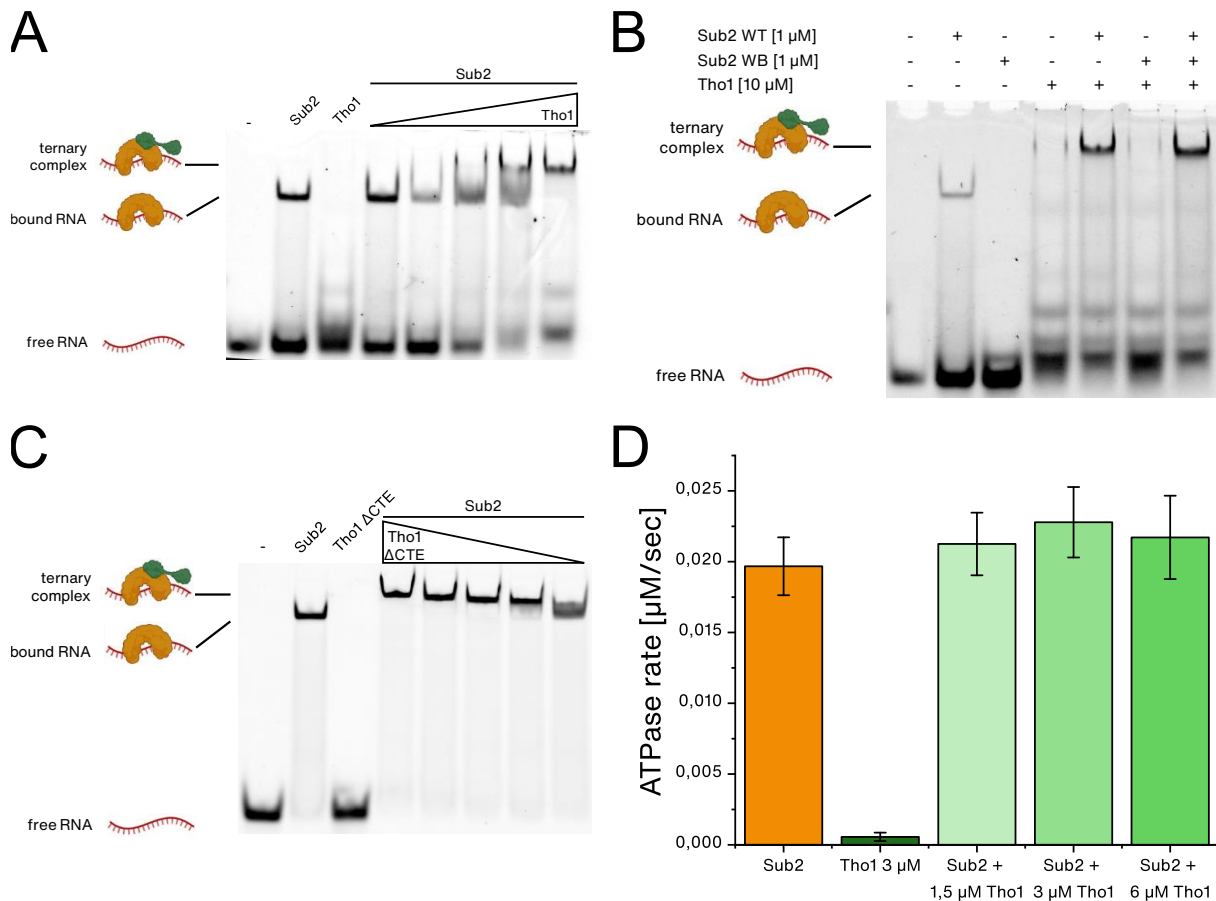
Furthermore, the influence of Tho1 on the ATPase activity of Sub2 was investigated (Figure 15 D). Tho1 has no influence on the ATPase activity of Sub2 under these reaction conditions. Also, different concentrations of Tho1 showed no change in the ATPase rate of Sub2.

The last activity to be investigated is the helicase activity (Figure 16). If Sub2 and Tho1 are incubated together in one reaction, a strong inhibition of the helicase activity of Sub2 is seen. To maintain further information of the mechanism of inhibition, dsRNA was pre-incubated with Tho1 and after a certain time Sub2 was added. This was also done *vice versa*.

---

<sup>1</sup> Prof. Dr. Lochnit, Institute of Biochemistry, Justus-Liebig-Universität Giessen, Germany, 2019



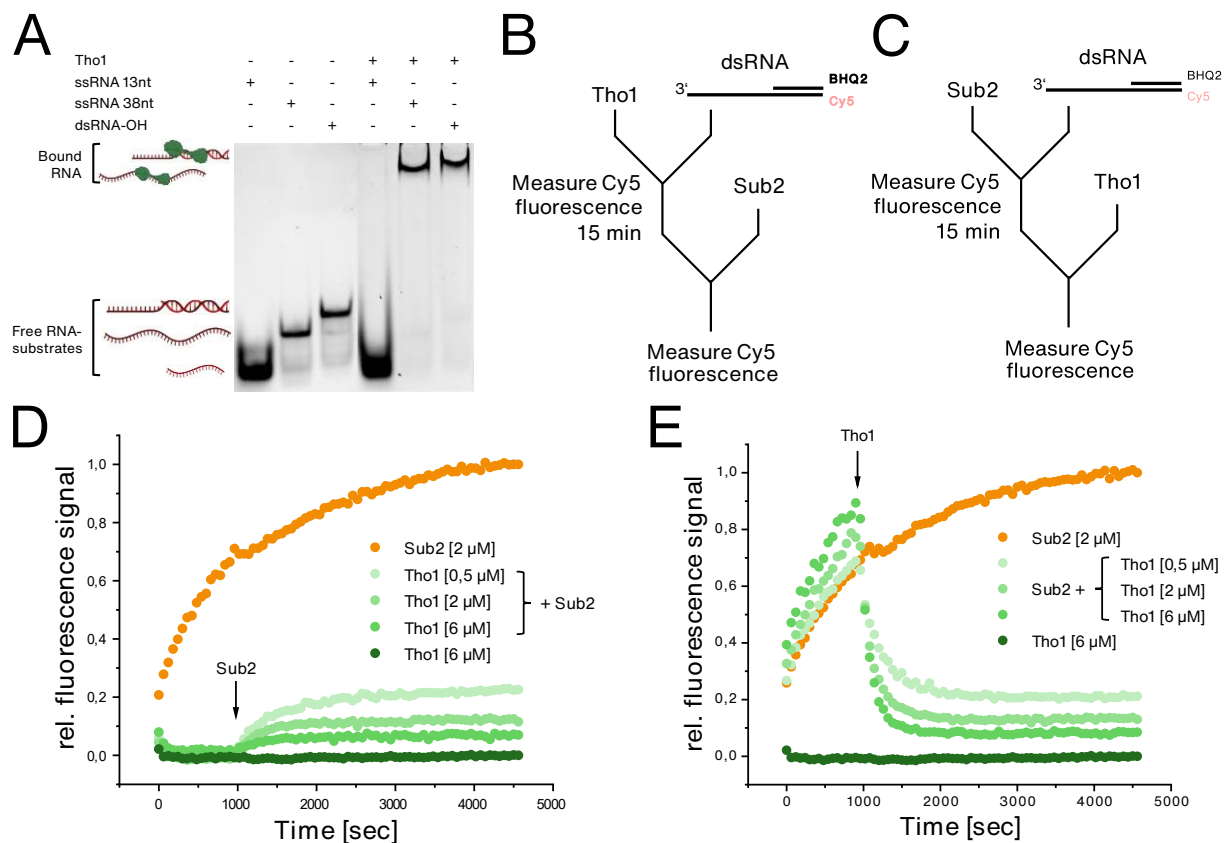


**Figure 15: Sub2 and Tho1 form a ternary complex on RNA, but Tho1 does not affect the ATPase activity of Sub2.** (A) EMSA of 1 μM of wild-type Sub2, 10 μM Tho1 or 1 μM Sub2 together with increasing concentrations of Tho1 (1.25, 2.5, 5, 10, 20 μM of Tho1) with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP 0.5 mM MgCl<sub>2</sub> (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (B) EMSA of only 1 μM of wild-type Sub2, Walker B (D215A, WB) or 10 μM Tho1 or in the indicated combinations with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP and 0.5 mM MgCl<sub>2</sub> (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (C) EMSA of 1 μM of wild-type Sub2, 6 μM Tho1 ΔCTE (see Figure 22 for details) or 1 μM Sub2 together with decreasing concentrations of Tho1 ΔCTE (12, 6, 3, 1.5, 0.75 μM of Tho1 ΔCTE) with 50 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP 0.5 mM MgCl<sub>2</sub> (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (D) Initial velocities (μM/s) of ATPase in the presence of 2 μM Sub2 wild-type, 3 μM Tho1 or 2 μM Sub2 together with the indicated concentrations of Tho1, 1 mM MgCl<sub>2</sub> and 1 mM ATP (see section ATPase assay in methods for details).

## Tho1 strongly inhibits unwinding of partial duplex RNA by Sub2

If Tho1 is pre-incubated and Sub2 is added afterwards, the signal increases but is significantly lower than the amplitude without Tho1. It is striking that Sub2 seems to be able to unwind more partial duplex RNA, the less Tho1 is in the reaction. This supports the assumption that enzymatic inhibition is taking place. If the experiment is carried out *vice versa*, i.e., the dsRNA is first incubated with Sub2 and then Tho1 is added at certain time point, the signal decreases rapidly, which means probably a reannealing of the substrate. It is also noticeable that the less Tho1 is present in the reaction, the higher the signal remains. After ruling out the possibility of a quench, the binding activity of Tho1 to the initial substrate and the products was tested (Figure 16 A). These are, on

the one hand, a dsRNA with a 3' overhang, a 13nt and a 38nt long RNA. Tho1 does not form a stable complex with short RNA but binds the dsRNA and the long ssRNA. Thus, Tho1 is able to bind to the initial substrate and also to the product. Since reannealing takes place while Sub2 unwinds the substrate, it is conceivable that Tho1 stabilises the double-stranded substrate and thus indirectly inhibits Sub2.



**Figure 16: Influence of Tho1 on the RNA unwinding activity of Sub2.** (A) EMSA of 10  $\mu$ M of wild-type Tho1 with 100 nM of 13 nt RNA 3'-labelled with Cy3 (ssRNA 13nt), 38 nt RNA 5'-labelled with Cy5 (ssRNA 38nt) or partial duplex RNA Putnam 5'-labelled with Cy5 (dsRNA-OH) (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (B) Diagram showing order-of-addition for real-time helicase assay of preincubated Tho1 with Sub2 and partial duplex RNA shown in D. (C) Diagram showing order-of-addition for real-time helicase assay of preincubated Sub2 with Tho1 and partial duplex RNA shown in E. (D) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 4600 sec) from partial duplex RNA (Putnam 10 nM, for details see section Annealing of Substrates in methods) in the presence of the indicated concentrations of Tho1 or 2  $\mu$ M wild-type Sub2, 1 mM MgCl<sub>2</sub> and 2 mM ATP. After 15 min 2  $\mu$ M Sub2 were added to the indicated reactions with Tho1 (see section Helicase assay in methods for details). (E) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 4600 sec) from partial duplex RNA (Putnam 10 nM, for details see section Annealing of Substrates in methods) in the presence of 2  $\mu$ M wild-type Sub2 or 6  $\mu$ M wild-type Tho1, 1 mM MgCl<sub>2</sub> and 2 mM ATP. After 15 min the indicated concentration of Tho1 were added to the indicated reactions with Sub2 (see section Helicase assay in methods for details).

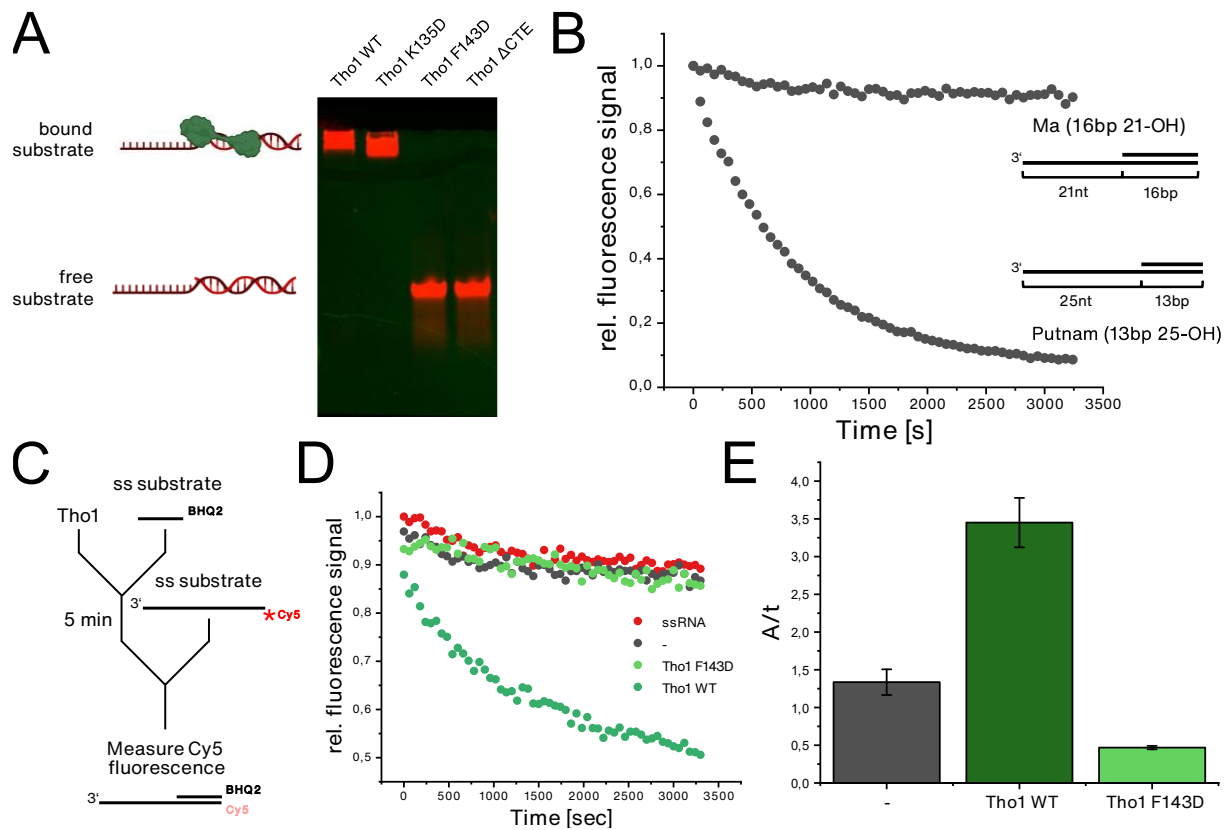
## Chapter 4: Tho1 has a novel annealing activity.

### Discovery of a novel annealing activity of Tho1

After the newly gained knowledge about the influence of Tho1 on the activities of Sub2, the question arose which mechanism takes place for inhibition of the helicase activity of Sub2. A new assay was introduced for this purpose: An annealing assay, which is based on the helicase assay. The first difference is that it does not work with a pre-annealed substrate, but starts the reaction with two single-stranded nucleic acids. Another difference is that instead of a substrate with a 13nt long double-stranded region and a 25nt long 3' overhang, there are two nucleic acids forming a double-stranded substrate with a 16nt long duplex with a 21nt long 3' overhang. The reason for this is that the Putnam substrates already anneal at room temperature at the given conditions and it proved to be difficult to measure the kinetics (Figure 17 B). To test whether the annealing is a reaction catalysed by Tho1, variants were produced and examined for their binding activity (Figure 17 A). The positions K135 and F143 were derived from *in vivo* RNA crosslink data (not published<sup>2</sup>). A non-binding variant (Tho1 F143D) was then tested together with the wild type of Tho1 in the annealing assay (Figure 17 D). Under the reaction conditions used and in the presence of wild-type Tho1, the two single strands anneal significantly faster, but not with the binding deficient variant F143D. To describe the curve as well as possible, the calculated amplitude was divided by the time constant (see section Annealing Assay in methods for details). This value is an indicator of the amount of annealed substrate and how fast it occurs.

---

<sup>2</sup> Dr. Philipp Keil, Institute of Biochemistry, Justus-Liebig-Universität, Giessen, Germany 2017

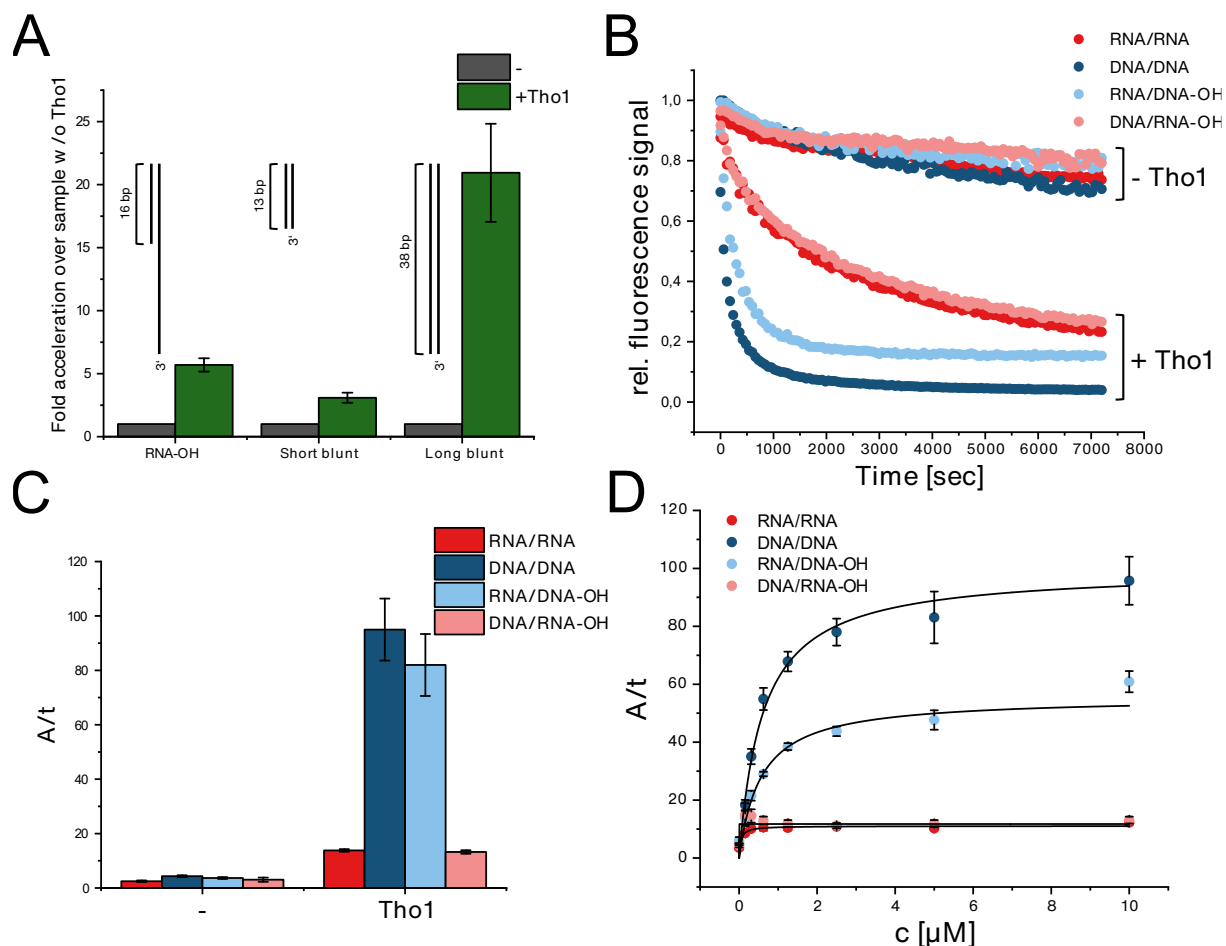


**Figure 17: Characterization of Tho1's annealing activity.** (A) EMSA of 6  $\mu$ M of wild-type Tho1 or the indicated variants with 100 nM partial duplex Putnam RNA 5'-labelled with Cy5 (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). (B) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 38 nt (Putnam) 37 nt (Ma) RNA (10 nM, for details see section Substrates in materials) in the presence of a complementary 13 nt (Putnam) or 16 nt (Ma) RNA labelled with BHQ2. (C) Diagram showing order-of-addition in annealing assay of Tho1 with two complementary strands of RNA (D) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 37 nt RNA (10 nM, for details see section Substrates in materials) in the presence of a complementary 16 nt RNA labelled with the quencher BHQ2 and 3  $\mu$ M Tho1 wild-type (WT) and the indicated variants of Tho1. (E) Ratio of calculated amplitude (A) to time constant (t) of the time course of an annealing reaction of 37 nt Ma RNA and the complementary 16 nt Ma RNA in the absence or presence of 3  $\mu$ M Tho1 wildtype (WT) or F143D (for details see section Annealing Assay in methods).

## Tho1 is a DNA annealing protein

After the annealing activity of Tho1 had been showed, the question arose whether Tho1 is specific for different substrates. The question was whether the overhang or the length of the duplex of the substrates used in these experiments plays a role. Therefore, the annealing assay was carried out with different RNA/RNA substrates. In addition to the standard Ma 16\_21 substrate, two other substrates without overhang were used. One double-stranded RNA with 13, and another with 38 base pairs. The speed of the annealing of the two strands in the absence or presence of Tho1 is examined. The result is that the acceleration of the 13 bp substrate is lower compared to the standard Ma substrate. However, annealing of the 38 bp duplex is more than 20-fold faster in the presence of Tho1.

Possible explanations are that the 13 bp duplex already anneals rapidly at room temperature and acceleration is hardly possible. On the other hand, the 38 bp duplex demonstrates that Tho1 is also able to remodel long and difficult substrates that might never anneal correctly without the presence of remodelling proteins.



**Figure 18: Tho1 anneals RNA, DNA and hybrid substrates.** (A) Ratio of calculated amplitude to time constant of the time course of an annealing reaction of indicated RNA substrates in the absence or presence of 5  $\mu$ M Tho1 wildtype (WT) normalized to samples without Tho1 (for details see section Annealing Assay in methods). Error bars are SE of  $n = 3$  independent experiments. (B) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 37 nt RNA or DNA (10 nM, for details see section Substrates in materials) in the presence of a complementary 16 nt RNA or DNA (10 nM) labelled with the quencher BHQ2 in the absence or presence of 5  $\mu$ M Tho1 wild-type (for details see section Annealing Assay in methods). (C) Quantification of time courses in B as ratio of calculated amplitude (A) and time constant ( $t$ ). Error bars are SE of  $n = 3$  independent experiments. (D) Ratio of calculated amplitude (A) to time constant ( $t$ ) of the time course of an annealing reaction of RNA, DNA and hybrid Ma substrate (for details see section Substrates in materials) in the presence of the indicated concentrations of Tho1 wild-type (for details see section Annealing Assay in methods). Error bars are SE of  $n = 3$  independent experiments.

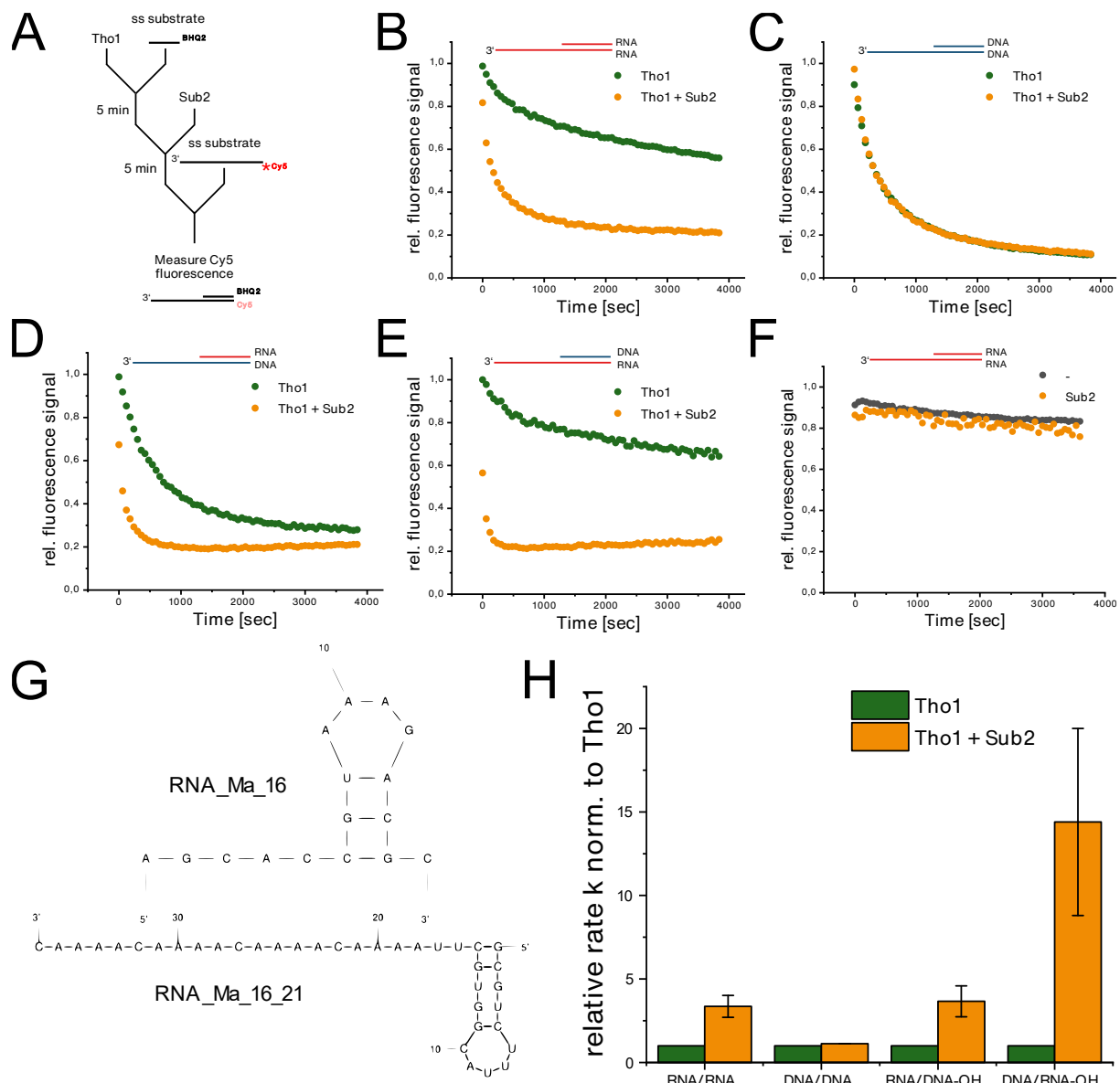
Furthermore, not only RNA was tested, but also DNA and hybrids. Sub2 already showed a higher activity on RNA/DNA hybrids. Moreover, Tho1 might have a role in mRNA export, and could thus also interact with DNA during transcription. The results of these experiments can be seen in Figure 18 B and C.

As already shown, the partial duplex RNA substrate with 16 bp anneals very slowly without Tho1, and annealing is accelerated in the presence of Tho1. DNA and hybrid substrates also anneal very slowly under the given reaction conditions without Tho1 (Figure 18 B). Tho1 processes not only RNA, but also DNA substrates, and the hybrid substrates are no exceptions. However, it can also be seen that although the substrates anneal at about the same rate without Tho1, the acceleration of the reaction of DNA/DNA and a hybrid with DNA overhang is much higher.

To provide further evidence for enzymatic activity and substrate specificity, a concentration series of Tho1 was made (Figure 18 D). This shows that the maximum acceleration for RNA/RNA and a hybrid with RNA overhang is already achieved at a concentration of less than 1  $\mu$ M. The maximum acceleration for DNA substrates is achieved at 10  $\mu$ M. This finding shows that Tho1 may be able to bring together even very long DNA strands which might form secondary structures when single stranded.

### **Sub2 enhances the annealing activity of Tho1**

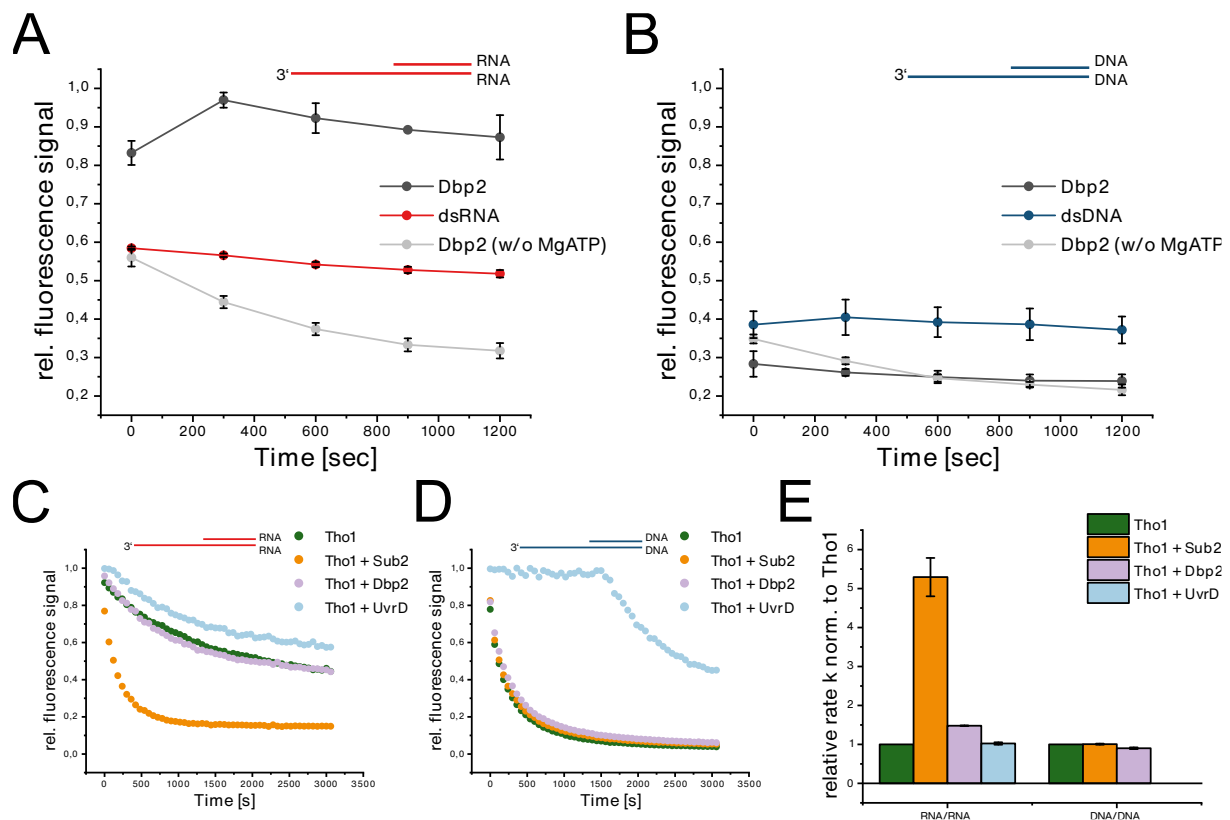
Since a previously unknown activity of Tho1 was found, and the influence of Tho1 on Sub2 was investigated beforehand, the next step was to investigate the influence of Sub2 on the annealing activity of Tho1 (Figure 19). As before, an annealing reaction of two single-stranded nucleic acids is observed in the presence of Tho1. In a further reaction, Sub2 is added. It can be observed that Sub2 enhances the annealing activity of Tho1. In the reactions in which at least one RNA strand was present (Figure 19 B, D, E), the nucleic acids anneal faster than with Tho1 alone. Sub2 has no influence on the annealing of the DNA/DNA substrates (Figure 19 C).



**Figure 19: Sub2 is enhancing the annealing activity of Tho1.** (A) Diagram showing order-of-addition in annealing assay with Tho1 and Sub2 with different substrates. (B-E) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 37 nt nucleic acid molecule (10 nM, for details see section Substrates in materials) in the presence of a complementary 16 nt nucleic acid molecule labelled with the quencher BHQ2 and 5  $\mu$ M Tho1 wild-type (WT) and the absence or presence of 2  $\mu$ M Sub2 (for details see section Annealing Assay in methods). B shows the annealing of RNA/RNA, C DNA/DNA, D RNA with DNA overhang and E DNA with RNA overhang. (F) Time course of CY5 fluorescence signal (normalized to initial signal intensity) from 37 nt RNA (10 nM, for details see section Substrates in materials) in the presence of a complementary 16 nt RNA labelled with the quencher BHQ2 and the absence or presence of 2  $\mu$ M Sub2. (G) Postulated secondary structure of the RNA nucleic acids used in the annealing experiments by UNAFold. (H) Calculated rates (k) of the time course of the annealing reactions shown in B-E normalized to reactions with only Tho1. Error bars are SE of n = 3 independent experiments.

To test whether the observed effect is a specific interaction between Tho1 and Sub2, these experiments were performed with other helicases (Figure 20). On the one hand, Dbp2 an DEAD-box RNA helicase from *S. pombe* was used and UvrD from *E. coli*. Dbp2 is another RNA DEAD-box helicase. UvrD is a translocating ssDNA helicase. After ensuring the activity of the helicases

under the reaction conditions used, the influence of these helicases on the annealing activity of Tho1 was investigated. Dbp2 has no significant influence on the reaction, neither on RNA/RNA nor on DNA/DNA annealing. UvrD shows no effect on RNA/RNA annealing by Tho1. However, in the reaction with UvrD, Tho1 and the two DNA strands, the signal remains at a maximum for a while, until after some time the signal decreases and the strands anneal accordingly.



**Figure 20: Annealing experiments with different helicases to test if the enhancing of the annealing activity of Tho1 by Sub2 is a specific interaction.** (A+B) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 1200 sec) from partial duplex RNA (A) or DNA (B) (Putnam 10 nM, for details see section Annealing of Substrates in methods) in the absence and presence of 500 nM Dbp2, 1 mM MgCl<sub>2</sub> and 2 mM ATP (see section Helicase assay in methods for details). Error bars SE of n = 2 independent experiments (C+D) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 37 nt RNA (C) or DNA (D) (Ma 10 nM, for details see section Substrates in materials) in the presence of a complementary 16 nt RNA (C) or DNA (D) labelled with the quencher BHQ2 and 5 μM Tho1 wild-type (WT) and the absence or presence of 500 nM Sub2, Dbp2 or UvrD (for details see section Annealing Assay in methods). (E) Calculated rates (k) of the time course of the annealing reactions shown in C-D normalized to reactions with only Tho1. Tho1 + UvrD on DNA/DNA is set to 0. Error bars are SE of n = 3 independent experiments.

## Annealing activity is conserved among Tho1's orthologues

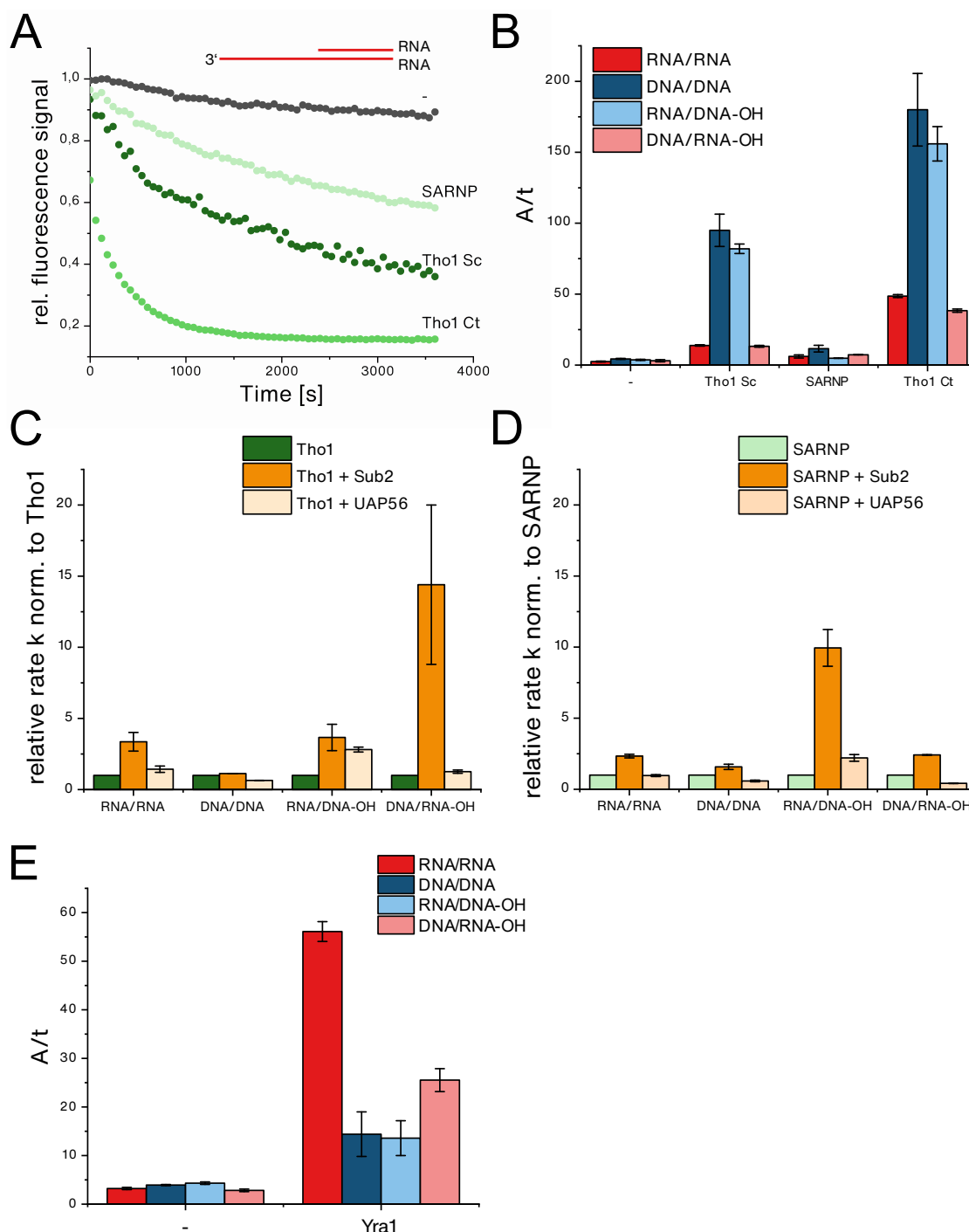
Now that the annealing activity of Tho1 has been described, the question arose whether this function is conserved within the Tho1 family. Therefore, not only Tho1 from *S. cerevisiae* was tested, but also Tho1 from the other fungus *Chaetomium thermophilum* and SARNP from *H. sapiens* (Figure 21 A-B). An annealing reaction was carried out with the orthologues, which is identical to that with



Tho1 from *S. cerevisiae*. It is shown that the orthologues of Tho1 from other organisms also accelerate the annealing of single stranded nucleic acids. A similar specificity was found for all proteins. Annealing reactions in which a long DNA strand was present, producing a DNA 3'-overhang in the product, were accelerated most by Tho1 and its orthologues. However, the overall activity was different. The protein from *C. thermophilum* showed higher activity than that from *S. cerevisiae*. Even though SARNP activity was seen, it was lower than that of the orthologues.

Sub2 is able to enhance to annealing activity of Tho1. After testing the activity of UAP56 (Figure 10) and SARNP the enhancing of the annealing activity was tested. It was not only tested in the same organism but also cross-tested to see whether the helicases can stimulate the annealing of the respective proteins (Figure 21 C+D). The annealing activity of Tho1 can be stimulated by UAP56 as well. However, the stimulation of activity is lower than with Sub2. This becomes very clear with a hybrid substrate with an RNA overhang. Under the reaction conditions shown, Sub2 enhances the activity of Tho1 more than tenfold, whereas it appears that UAP56 has no influence on the activity. If the effect were solely dependent on annealing and unwinding activity, one would expect the same results for both helicases, although not necessarily to the same intensity. Furthermore, SARNP was also tested with the helicases and, surprisingly, the annealing of substrates other than for Tho1 are stimulated by Sub2 and UAP56. Here, Sub2 does not enhance the hybrid with an RNA overhang the most, but surprisingly the one with a DNA overhang, which is also stimulated the most by UAP56. Furthermore, it can be seen that, while Sub2 enhances annealing by SARNP in the hybrid with RNA overhang, UAP56 has an inhibitory effect. In conclusion, it can be said that the enhancement of annealing activity is independent of the substrate, but rather depends on the proteins that are used.

To gain further insights about annealing activity, Yra1 (yeast annealing protein 1) was tested (Figure 21 E). It is already known that this protein has an RNA annealing activity (Portman *et al.*, 1997), but no DNA was tested. Since working with Yra1 from *S. cerevisiae* proved to be difficult (data not shown), its orthologue from *C. thermophilum* was used. So far, the orthologues from *C. thermophilum* always showed an identical pattern to the proteins from *S. cerevisiae*. It cannot be safely assumed that Yra1 from *C. thermophilum* behaves exactly the same as its orthologue from yeast. Nevertheless, Yra1 showed an opposite specificity to Tho1. This means that Yra1 could also anneal DNA/DNA and hybrids with DNA overhang, but showed more activity on RNA/RNA substrates and hybrids with RNA overhang.

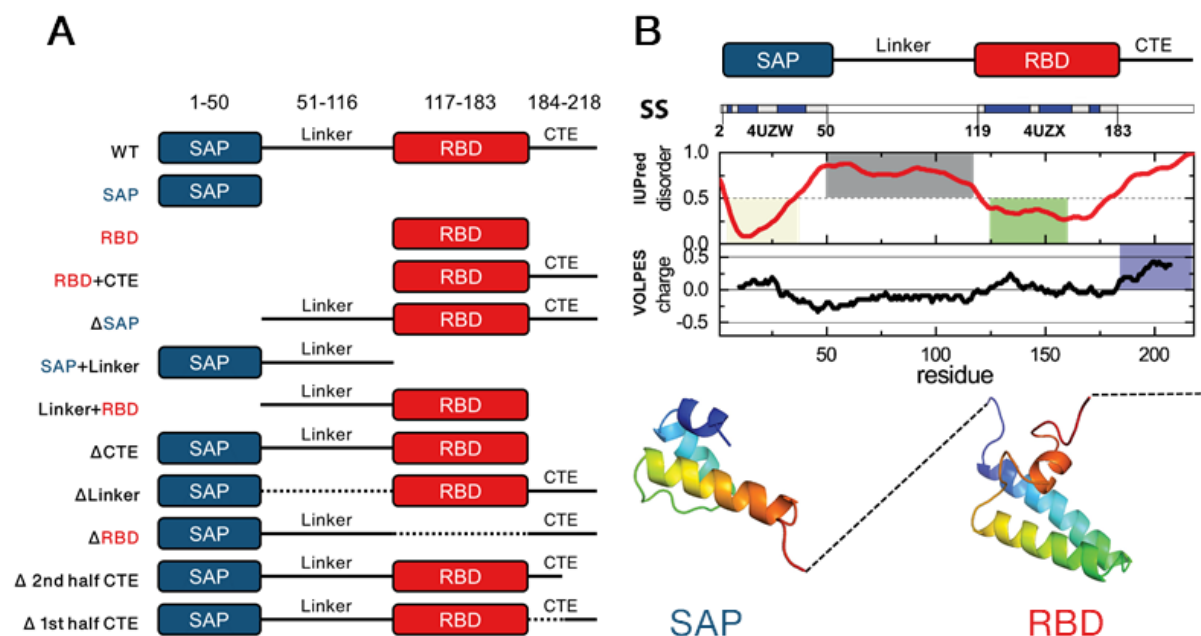


**Figure 21: Tho1's annealing activity is conserved among *Homo sapiens* and *Chaetomium thermophilum* and Yra1 is specific for RNA.** (A) Time course of Cy5 fluorescence signal (normalized to initial signal intensity) from 37 nt RNA (Ma 10 nM, for details see section Substrates in materials) in the presence of a complementary 16 nt RNA labelled with the quencher BHQ2 in the presence or absence of 5  $\mu$ M Tho1 from *S. cerevisiae* (Sc), from *C. thermophilum* (Ct) or SARNP from *H. sapiens* (for details see section Annealing Assay in methods). (B) Ratio of calculated amplitude (A) to time constant ( $t$ ) of the time course of an annealing reaction of 37 nt Ma nucleic acid and the complementary 16 nt Ma nucleic acid in the absence or presence of 5  $\mu$ M Tho1 from *S. cerevisiae* (Sc) or *C. thermophilum* (Ct) or human SARNP). (C) Calculated rates ( $k$ ) of the time course of the annealing reactions of Ma substrates in the presence of 5  $\mu$ M Tho1 and in the absence or presence of either 2  $\mu$ M Sub2 or UAP56 normalized to reactions with only Tho1. (D) Calculated rates ( $k$ ) of the time course of the annealing reactions of Ma substrates in the presence of 5  $\mu$ M SARNP and in the absence or presence of either 2  $\mu$ M Sub2 or UAP56 normalized to reactions with only SARNP. (E) Ratio of calculated amplitude (A) to time constant ( $t$ ) of the time course of an annealing reaction of 37 nt Ma nucleic acid and the complementary 16 nt Ma nucleic acid in the absence or presence of 100 nM Yra1 *C. thermophilum* (for details see section Annealing Assay in methods). Error bars are SE of  $n = 3$  independent experiments.

## Chapter 5: Tho1 domain analysis

### Domain analysis of Tho1 to discover their functionalities

Since annealing activity has not yet been described for Tho1, the question arose which domains of Tho1 are responsible for its activity and which for the interaction with Sub2. Therefore, different variants were cloned in order to describe the activity of Tho1 in detail (Figure 22 A). The variants were based on a previous study (Jacobsen *et al.*, 2016) and Tho1 was divided into four domains. These are the SAP domain, an unstructured linker that connects the SAP domain with the RNA-binding domain (RBD) and finally an unstructured C-terminal end (CTE). If taking the charge into account, it can be seen that the CTE is strongly positively charged (Figure 22 B). In contrast, the linker is rather negatively charged. The RBD is in the neutral range and the SAP domain also appears to be neutral with larger fluctuations. In addition to the current crystal structure of Tho1, a potential additional helix was predicted for the CTE (Jacobsen *et al.*, 2016). Therefore, two further variants were made in which the CTE was divided again. The annealing and binding activity and the influence on Sub2 were investigated. On the other hand, the formation of a ternary complex with RNA and the influence on the helicase activity of Sub2 were investigated.

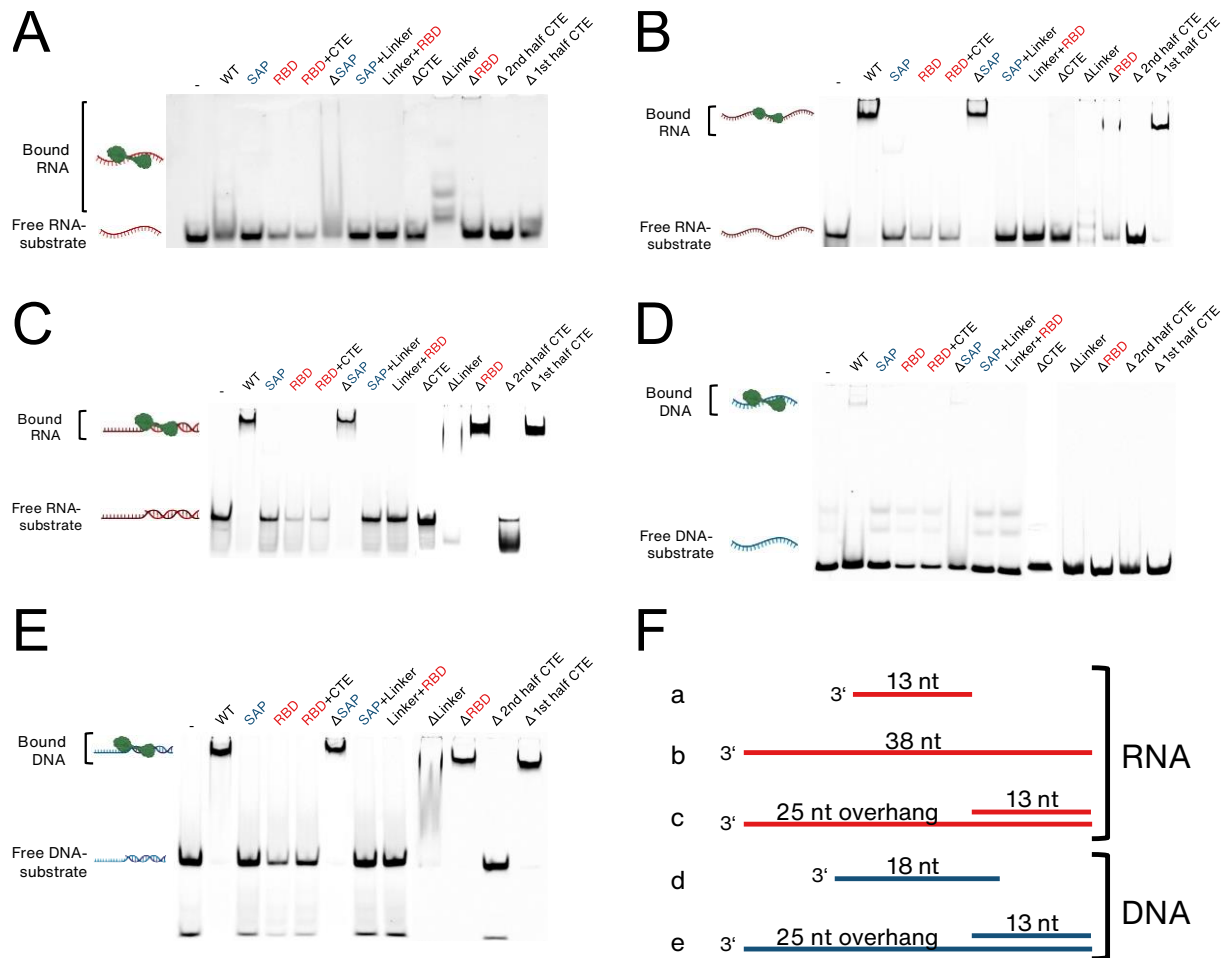


**Figure 22: Overview of analysed Tho1 domain variants and properties of Tho1.** (A) Overview of the analysed Tho1 domain variants. (B) Different properties of Tho1. Prognosed disorder were obtained by IUPred (Erdos *et al.*, 2021), and the charge profile were obtained by VOLPES-App (Bartonek and Zagrovic, 2019). At the bottom is the structure of the SAP (4UZW) and RNA-binding-domain (RBD, 4UZX) from (Jacobsen *et al.*, 2016).

## **The RBD and the C-terminal end are mainly responsible for binding of nucleic acids**

First, the binding of Tho1 variants to substrates is investigated (Figure 24). Although one might expect specificities of the domains to differ, no different specificity could be found among the variants. If RNA could be bound, DNA could also be bound.

The wild type shows the expected band. It forms a stable complex with long RNA and DNA under the reaction conditions. The SAP domain alone is not able to bind nucleic acids. The RNA-binding domain, on the other hand, does not show a defined band indicating a stable complex, but the fluorescence of the free RNA is reduced. The same observation can be made with the RBD in combination with the CTE. If the SAP domain is deleted ( $\Delta$ SAP), a defined band is seen as for the wild type. If one combines the SAP Domain with the linker (SAP+linker), the binding activity cannot be restored. Surprisingly, the linker even seems to have a negative effect, as the RBD+linker variant shows no evidence of binding activity. In comparison, as already mentioned, the RBD alone at least showed a reduction in the fluorescence intensity of the free RNA. Deletion of the CTE leads to loss of binding activity. If a closer look is taken at the CTE, we see that ultimately the deletion of the last 17 amino acids already leads to a loss of activity ( $\Delta$ 2nd half CTE). If, on the other hand, the area between RBD and the last 17 residues is deleted ( $\Delta$ 1st half CTE), a complex between the protein and RNA similar to the wild type is detectable. If the linker is deleted between the SAP and RNA binding domain, free RNA is hardly detectable. The RNA seems to be bound, but no stable complex is formed. The observation made when the RBD is deleted is unexpected. Here one sees that the fraction of free RNA decreases and part of the fluorescence is at the level of an expected stable complex. Although it was previously assumed that the RBD is mainly responsible for the binding activity, the protein can also bind substrates without this domain.

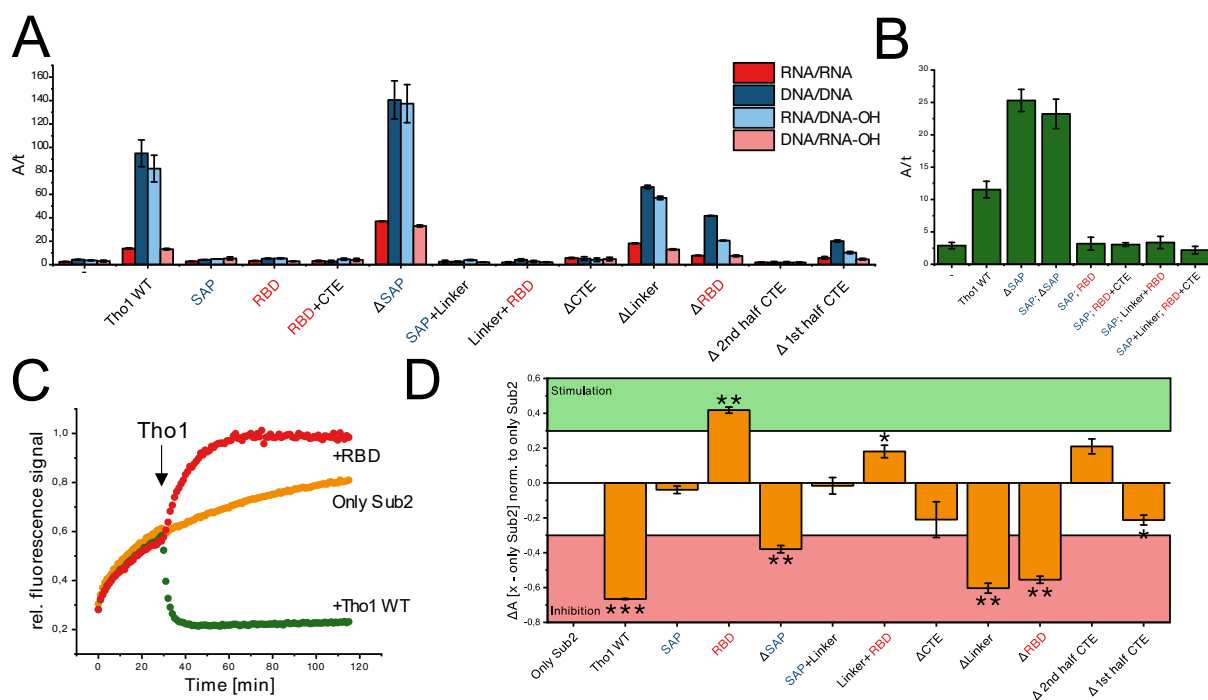


**Figure 23: Investigation of binding capability of Tho1 wild-type and domains.** (A-E) EMSA of 10 μM of wild-type Tho1 and the indicated domain variants with 100 nM of **A** 13nt ssRNA **B** 38nt ssRNA **C** partial duplex Putnam RNA **D** ssDNA (UvrD\_top) or **E** partial duplex Putnam DNA (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). All EMSA experiments (A-E) were cut together from different gels **(F)** Schematic overview of the substrates used in A-E (for details see Substrates in materials and Annealing of Substrates in methods).

## The C-terminal end plays an important role in annealing activity

Next, the variants of Tho1 were tested for their annealing activity (Figure 24 A). The first thing to notice is that there are no differences in substrate specificity. If the activity decreases or increases compared to the wild type, it does so for all substrates, regardless of whether RNA, DNA or hybrids. Neither the SAP domain nor the RBD are able to anneal partial double-stranded nucleic acids. The RBD together with the CTE also shows no activity. However, if only the SAP domain is deleted, nucleic acids are not only annealed in the presence of the protein, but this variant also shows a higher activity in comparison to WT. The linker in combination with the SAP domain or the RBD shows no annealing activity. If only the unstructured CTE is deleted, the protein also shows no activity. Because of this unexpected result, the CTE was again subdivided into two parts: the first and the second half. The first half is the already mentioned hypothetical helix, the second half is hypothetically unstructured. If the second strongly positive half is deleted, the protein already

loses all measurable annealing activity. In contrast, the variant in which the first half of the CTE has been deleted shows activity that is, however, much lower than that of the WT. Of all the variants that still had activity, this one had the lowest. The variant that no longer had the linker between the SAP domain and the RBD still showed activity, but this was also reduced, but not as other variants. If the RBD is deleted, the protein also loses activity, but is still able to anneal the substrates. Attempts were also made to restore lost annealing activity by combining complementary Tho1 variants that together would form a full length Tho1 (Figure 24 B). However, the activity was not restored.



**Figure 24: Analysis of annealing activity of Tho1 wild-type and domain variants.** (A) Ratio of calculated amplitude (A) to time constant (t) of the time course of an annealing reaction of 37 nt Ma nucleic acid and the complementary 16 nt Ma nucleic acid in the absence or presence of 5  $\mu$ M wild-type Tho1 (WT) or the indicated domain variants of Tho1. (B) Ratio of calculated amplitude (A) to time constant (t) of the time course of an annealing reaction of 37 nt Ma RNA and the complementary 16 nt Ma RNA in the absence or presence of 5  $\mu$ M wild-type Tho1 (WT),  $\Delta$ SAP variant or a combination of domain variants forming full-length Tho1 (for details see section Annealing Assay in methods). (C) Time course of Cy5 fluorescence signal (normalized to maximum signal intensity after 115 min) from partial duplex RNA (Putnam 10 nM, for details see section Annealing of Substrates in methods) in the presence of 2  $\mu$ M wild-type Sub2, 1 mM  $MgCl_2$  and 2 mM ATP. After 30 min 3  $\mu$ M of Tho1 wild-type or Tho1 RBD were added to the indicated reactions with Sub2 (see section Helicase assay in methods for details). (D) Influence of Tho1 wild-type and Tho1 domain variants on the helicase activity of Sub2 shown as change of amplitude after 120 min after addition of Tho1 after 30 min (like C). Reaction took place in the presence of partial duplex RNA (Putnam 10 nM, for details see section Annealing of Substrates in methods), 2  $\mu$ M Sub2, 1 mM ATP, 0.5 mM  $MgCl_2$  and the addition of 3  $\mu$ M of the indicated Tho1 variants (see section Helicase assay in methods for details). Error bars are SE of  $n = 3$  independent experiments. Significance tested via unpaired two-tailed t-test. P-value \*  $\leq 0.05$ , \*\*  $\leq 0.01$ , \*\*\*  $\leq 0.001$ .

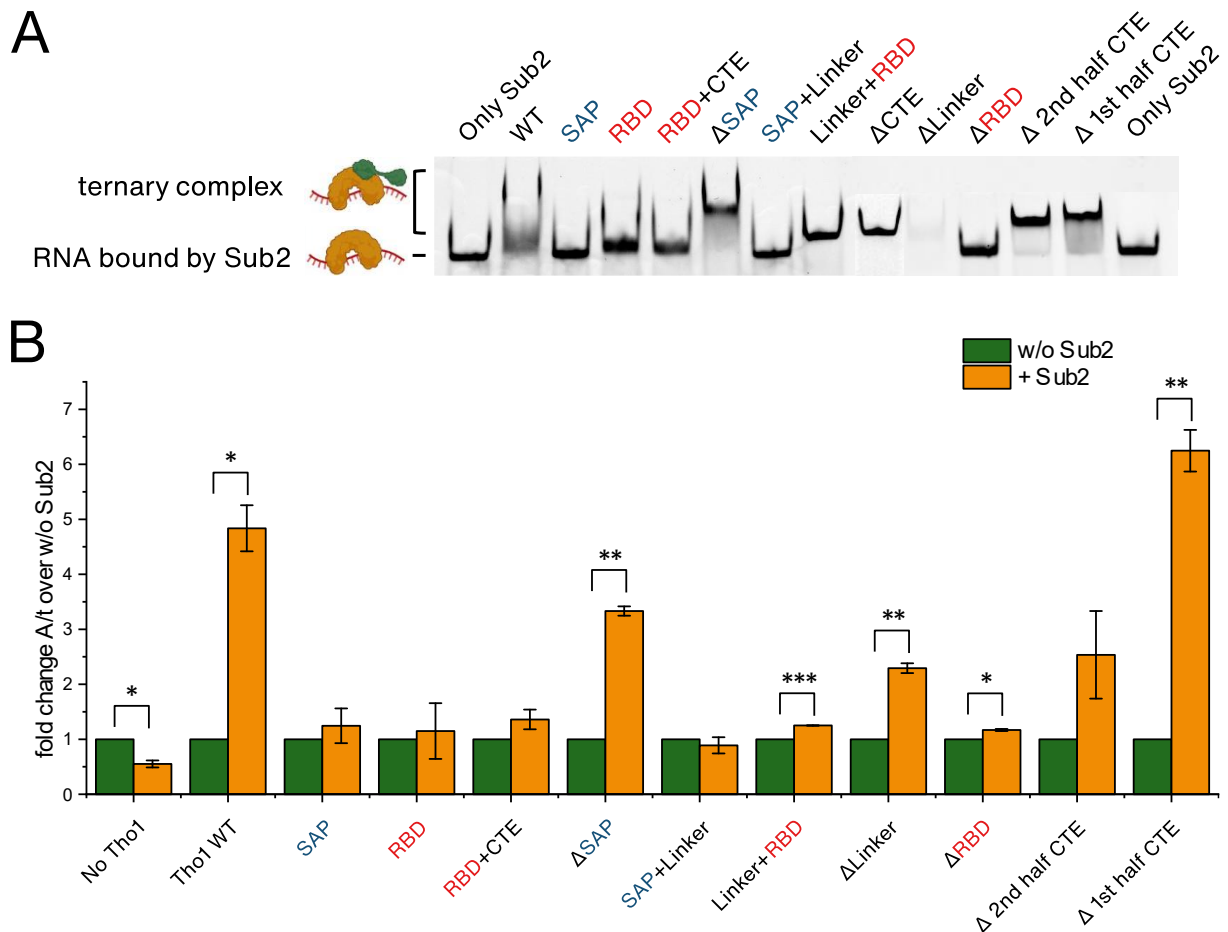
### **Annealing activity correlates with inhibition of Sub2's unwinding activity**

Next, the influence on the helicase activity of Sub2 was investigated to find out whether there is a direct correlation between annealing activity and inhibition mechanism (Figure 24 C+D). For this purpose, the substrate was pre-incubated with Sub2 and after a certain time Tho1 and its variants were added. In this independent experiment, the wild type of Tho1 shows inhibition of the helicase activity of Sub2, as expected. The SAP domain alone and in combination with the linker shows no inhibition. The variant without CTE also shows no effect. The linker in combination with the RBD also has no influence on the helicase activity of Sub2. If only the second half of the CTE is deleted, no inhibition is observed. These variants also showed no annealing activity. The variants that still have annealing activity also showed inhibition of Sub2 in this assay. However, surprisingly, a stronger inhibition by the WT than by the  $\Delta$ SAP variant is seen, although the measured annealing activity for  $\Delta$ SAP was higher than for the WT. Looking separately at the  $\Delta$ Linker,  $\Delta$ RBD and  $\Delta$ 1st half CTE variants, their inhibition correlates with their annealing activity. The  $\Delta$ linker variant already shows an inhibition similar to that of the WT, and is therefore even stronger than the  $\Delta$ SAP variant. The RBD alone shows a stimulation of helicase activity on an RNA/RNA substrate instead of an inhibition (Figure 24 C).

### **Tho1 interacts via its RNA-binding domain with Sub2**

To find out which domain of Tho1 is responsible for the interaction with Sub2, shift experiments with the domain variants were carried out with Sub2 (Figure 25 A). For the sake of clarity, only the influence of the Tho1 variants on the binding of RNA by Sub2 is shown. The free RNA is not shown. The SAP domain alone and also in combination with the linker do not form a ternary complex with Sub2. In contrast, all variants containing the RBD and the linker led to a clear shift (linker+RBD,  $\Delta$ CTE,  $\Delta$ 2nd half CTE,  $\Delta$ 1st half CTE,  $\Delta$ SAP). The RBD also seems to form a ternary complex with Sub2. Since it is a small protein (66 aa) that is also structurally compact, a large shift in a native gel would be unexpected. However, a difference to only Sub2 can certainly be observed. The effect of RBD+CTE cannot be clearly determined. However, since this Protein is larger than the RBD alone, one would expect a higher shift in the formation of a ternary complex with Sub2 and RNA than the RBD alone. But this is not the case. Deletion of the RBD leads to loss of the ability to form a ternary complex. If one deletes the complete CTE, a defined shift is seen. It does not matter whether only part of the CTE is deleted ( $\Delta$ 2nd half CTE,  $\Delta$ 1st half CTE). The deletion of the linker leads to an unexpected result. No free RNA is detectable and the band of Sub2 with bound RNA is also no longer visible. A light band can be seen, which indicates a

ternary complex. There is also no fluorescence in the pockets. It is unclear what happened to the substrate. Possibly this Tho1 variant is aggregated.



**Figure 25: Ternary complex formation of Sub2 and Tho1 domain variants and enhancing of annealing activity of these variants by Sub2. (A)** EMSA of 2  $\mu$ M wild-type Sub2 in the absence or presence of wild-type Tho1 or domain variants with 100 nM 13 nt RNA 3'-labelled with Cy3 in the presence of 1 mM ADPNP 0.5 mM  $MgCl_2$  (see section Electrophoretic Mobility Shift Assay (EMSA) in methods for details). The shown experiment was cut together from different gels **(B)** Ratio of calculated amplitude (A) to time constant (t) of the time course of an annealing reaction of 37 nt Ma RNA and the complementary 16 nt Ma RNA in the presence of Tho1 wild-type and Tho1 domain variants in the absence (w/o Sub2) or presence (+ Sub2) of Sub2 wild-type, normalized to samples without (w/o) Sub2. Error bars are SE of  $n = 3$  independent experiments. Significance tested via unpaired two-tailed t-test. P-value \*  $\leq 0.05$ , \*\*  $\leq 0.01$ , \*\*\*  $\leq 0.001$ .

As already shown, Sub2 can stimulate the annealing activity of Tho1 (see section Sub2 enhances the annealing activity of Tho1). To test whether Sub2 only stimulates active variants or can also activate dead variants, the annealing activity of Tho1 and its domain variants was again tested in the presence of Sub2 and dsRNA (Figure 25 B).

All proteins that show annealing activity alone can also be stimulated by Sub2. All of these variants show a significant increase in annealing activity in the presence of Sub2. However, a closer look



should be taken at Tho1  $\Delta$ RBD. Although the increase is significantly larger in the presence of Sub2, as the variance of the data obtained is small, one would expect a larger increase if the effect of enhancing is solely due to the secondary structure of the substrate.

The linker+RBD variant is in contrast to that finding of the  $\Delta$ RBD variant. This variant showed no activity in the annealing assay. However, it can be stimulated with high significance by Sub2 ( $p < 0.001$ ). Although the increase is small, it could be an indication that the increase in annealing activity is caused by an interaction with the RBD, especially when taking into account, that the increase of activity of the  $\Delta$ RBD is unexpected small. Nevertheless, further experiments are needed to confirm these findings.

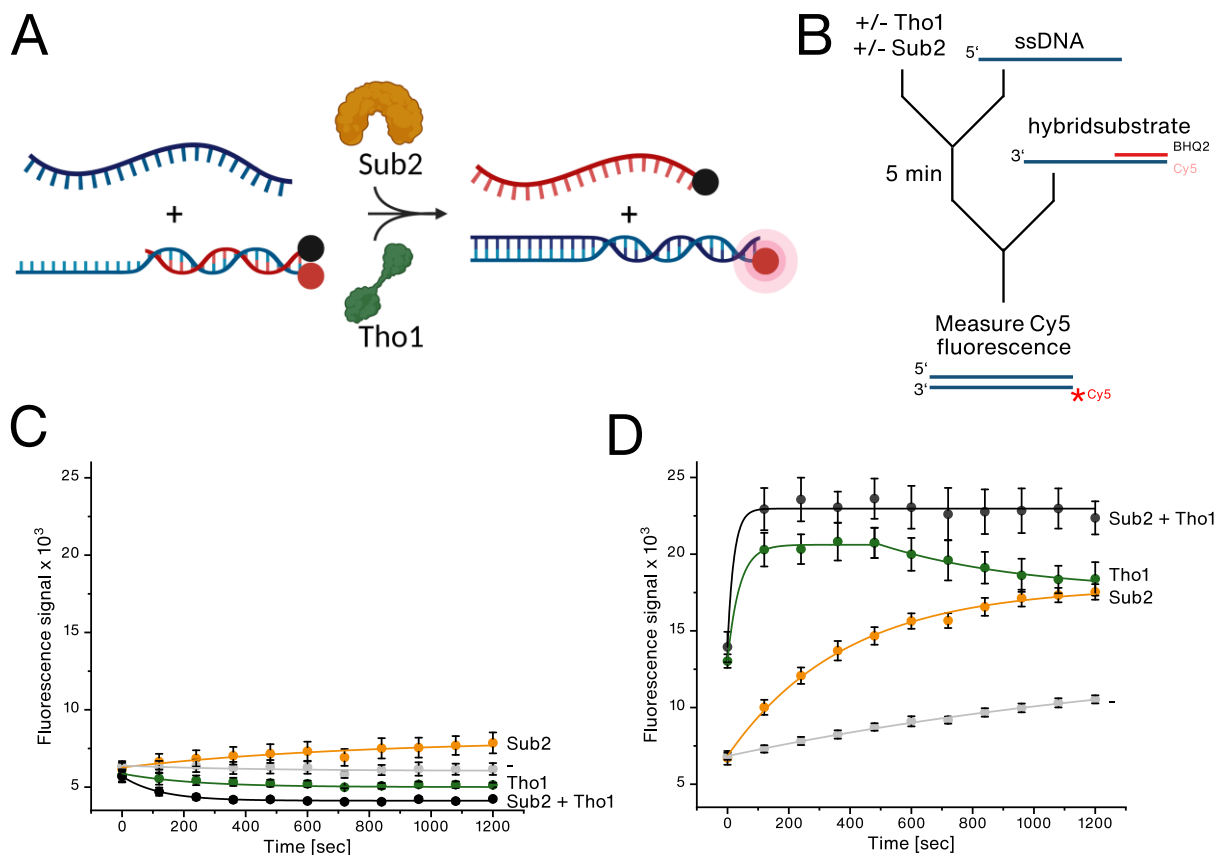
## Chapter 6: Sub2 and Tho1 might resolve R-Loops *in vitro* and *in vivo*

### **Tho1 is able to displace RNA-strands *in vitro***

It has already been shown that Sub2 and also its ortholog UAP56 might be able to resolve R-loops *in vivo* (Gomez-Gonzalez *et al.*, 2011, Perez-Calero *et al.*, 2020). For UAP56, *in vitro* experiments were carried out that showed it is possible for UAP56 to resolve R-loop like structures (Perez-Calero *et al.*, 2020). Tho1 is thought to have a role in the resolution of R-loops, but this has not yet been explicitly shown (Jimeno *et al.*, 2006). In this work, it will be shown that Tho1, although possessing annealing activity, is also capable of displacing an RNA strand from an RNA/DNA hybrid *in vitro*.

R-loops are structures of double stranded DNA, whereas one strand is annealed with a complementary RNA. The second DNA strand remains unpaired. In this work, the specificity of Tho1 for DNA has already been shown. In previous studies (Perez-Calero *et al.*, 2020), it was shown that UAP56 is better at unwinding RNA/DNA substrates. To confirm that the Sub2 used in this work can also unwind RNA/DNA hybrids and to better understand the mechanism of R-loop resolution, the activity of the helicases was tested on different substrates (Figure 9).

In another experiment, the effect of Sub2 and Tho1 on an R-loop-like substrate was tested. R-loops can be a hundred or more nucleotides long (Ginno *et al.*, 2012; Malig *et al.*, 2020; Sanz *et al.*, 2016), and thus the point of attack for resolving R-loops does not have to be at the branch point. In these cases, one has an RNA/DNA hybrid and a spatially separated single-stranded DNA strand. With this in mind, the following experiment was designed: A hybrid substrate with a DNA overhang was incubated together with a DNA perfectly complementary to the DNA strand of the substrate in the presence of Sub2, Tho1 or both proteins (Figure 26).



**Figure 26: Resolution of R-Loop like structure by Sub2 and Tho1 *in vitro*.** (A) Schematic model of the displacement assay used in D. Created with BioRender.com. (B) Diagram showing order-of-addition for displacement assay of Sub2 and Tho1 and partial duplex hybrid substrate in the presence of a perfect complementary DNA strand. (C) Time course of Cy5 fluorescence signal from partial duplex hybrid (Ma 10 nM, DNA overhang, for details see section Annealing of Substrates in methods) in the absence or presence of 2  $\mu$ M wild-type Sub2, 2.5  $\mu$ M wild-type Tho1, 1 mM MgCl<sub>2</sub> and 2 mM ATP. (D) Modified helicase assay into an RNA displacement assay. Time course of Cy5 fluorescence signal from partial duplex hybrid (Ma 10 nM, DNA overhang, (for details see section Annealing of Substrates in methods) in the presence of a perfect complementary DNA strand (Ma 10 nM), +/- 2  $\mu$ M wild-type Sub2, +/- 2.5  $\mu$ M wild-type Tho1, 1 mM MgCl<sub>2</sub> and 2 mM ATP. Curves are fitted with a single exponential function, whereas only Tho1 is fitted two times: From 0-480 sec and 480-1200 sec (see section Helicase assay in methods for details). Error bars are SE from n = 3 independent experiments. Curves are fitted with a single exponential function (see section Helicase assay in methods for details). Error bars are SE from n = 3 independent experiments.

The control experiments without the complementary strand are seen in Figure 26 C. Sub2 is able to partially unwind the substrate. Without any protein, the signal of the strand remains stable. However, if Tho1 is added to the reaction, the signal decreases, as Tho1 further anneals the free strands present in the equilibrium. With both proteins, the signal decreases further because, as already shown, Sub2 stimulates the annealing activity (Figure 19).

But when the complementary DNA strand is added, interesting results are obtained (Figure 26 D). The top strand alone leads to an increase in fluorescence, which means that it displaces the RNA strand. In the presence of Sub2, the signal increases even further. It is clearly stronger than without the top strand. In the presence of the complementary DNA, it acts as a trap that prevents reannealing. In the presence of the annealing protein Tho1 an initial burst, an increase in

fluorescence is seen. The RNA dissociates from the DNA. Presumably, Tho1 remodels the RNA/DNA hybrid by its annealing activity in the presence of a complementary DNA strand. After some time, the signal decreases again. Part of the RNA hybridises back onto the DNA and forms the initial substrate. An equilibrium is established. In the presence of both proteins, Sub2 and Tho1, an initial burst takes place, similar to Tho1 alone. However, compared to Tho1 alone, the signal remains stable and high. The RNA cannot rehybridize again.

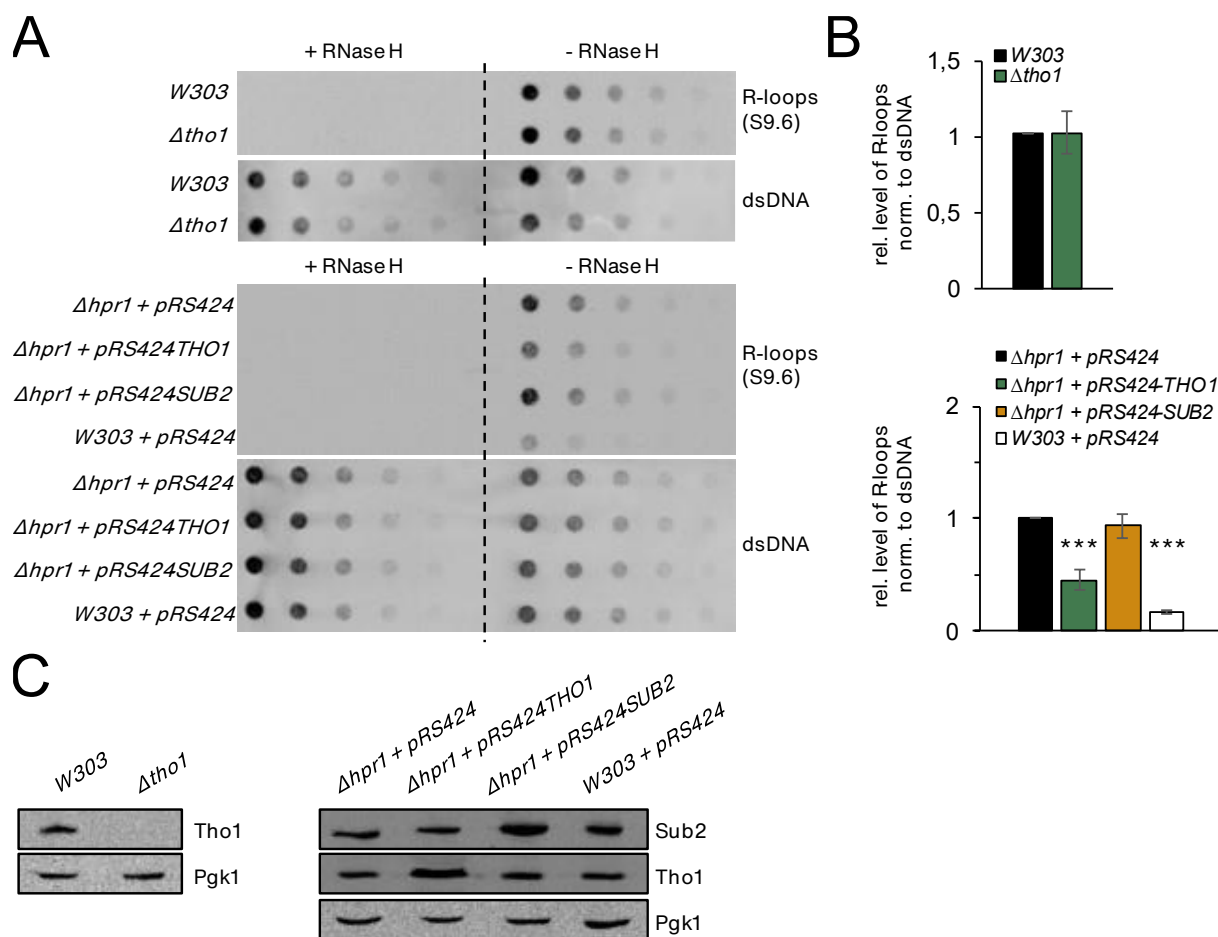
### **Tho1 prevents R-Loop formation *in vivo***

These results give a first indication that both Sub2 and Tho1 can resolve R-loops. Surprisingly, Tho1 alone is able to remodel an RNA/DNA hybrid to a dsDNA under these reaction conditions. To confirm the hypothesis that Tho1 is also able to unwind R-loops *in vivo*, experiments with an R-loop antibody in  $\Delta hpr1$  strains were performed (Figure 27).

In *S. cerevisiae* a  $\Delta hpr1$  mutant is associated with a missing THO complex. Defects in THO lead to mRNA export defects and an increased level of R-Loops. It is already known, that these phenotypes can be suppressed by overexpression of Sub2 or Tho1, but the mechanism behind that remains unclear.

With the new findings from this work and the already published results, different *in vivo* experiments were carried out. First, we investigated whether the deletion of Tho1 alters the amount of R-loops (Figure 27 B top). The result shows that a deletion of Tho1 does not lead to an altered level of R-loops.

Next,  $\Delta hpr1$  variants with overexpression of Tho1 and Sub2 were investigated with regard to the level of R-loops. Deletion of *hpr1* leads to an increased level of R-loops, and the other samples were normalized to it. The wild type, which was transformed with an empty plasmid, shows a significantly lower level of R-loops (Figure 27 B bottom, W303 + pRS424). It can be assumed that this is the wild type level of R-Loops and that the plasmid has no influence on the level of R-loops. To confirm the overexpression of Tho1 and Sub2, western blots were performed (Figure 27 C). Overexpression of Sub2 does not lead to a significant change in the amount of R-loops. However, overexpression of Tho1 leads to a significant decrease in the level of R-loops, but it does not reach wild-type level. Taken together *in vitro* and *in vivo* results, the findings suggest that Tho1 is involved in the resolution of R-loops.



**Figure 27: Tho1 overexpression is able to diminish the accumulation of R-loops in  $\Delta hpr1$ .** (A) R-loop dot blot assay. Relative levels of R-loops and dsDNA were determined using the anti-DNA-RNA Hybrid antibody (S9.6) and anti-dsDNA antibody. Specificity of the S9.6 antibody was confirmed by RNase H treatment. (B) Quantification of R-loop dot blots.  $n \geq 3$  (unpaired t-test, two-tailed \*\*\*  $p < 0.001$ ). (C) Western blots to confirm deletion of  $\Delta tho1$  and overexpression of Tho1/Sub2. Experiments carried out by Philipp Keil<sup>3</sup>.

<sup>3</sup> Dr. Philipp Keil, Institute of Biochemistry, Justus-Liebig-Universität, Giessen, Germany, 2021

# Discussion

## Chapter 1: Characterization of the DECD-Box RNA helicase Sub2 from *S. cerevisiae*

### Investigation of the activity of Sub2

At the beginning of this work, suitable assays had to be found to analyse the activities of the RNA helicase Sub2 as accurately as possible. For RNA-binding and ATPase activity, methods were chosen that are already published and used in the literature (see section Methods). For the helicase activity of Sub2 and also its ortholog UAP56, only discontinuous methods, such as gels, have been used so far (Perez-Calero *et al.*, 2020; Saguez *et al.*, 2013). However, as the helicase activity is a combination of binding, ATPase and conformational change, this method was insufficient for our needs. Therefore, a real-time assay was introduced that allows live observation of nucleic acid unwinding and can be easily modified to study different aspects of unwinding. Ultimately, this method led to the discovery of the annealing activity of Tho1 (see Chapter 4: Tho1 has a novel annealing activity in discussion).

To start *in vitro* experiments, the protein has to be purified. The purification used in this thesis was modified according to Saguez *et al.*, 2013. It is unusual to elute again with a low concentration of imidazole after the second Ni column, although the His-tag has already been cleaved off (Figure 3 A). However, during the optimisation of the purification, it was found that a considerable fraction of Sub2 remains on the column. After a further elution with a low concentration of imidazole and subsequent testing in the helicase assay, it was found that this fraction is much more active at the same concentration (data not shown). The non-specific binding of Sub2 to a Ni-column and its activity seem to correlate.

The electrophoretic mobility shift assay (EMSA) was chosen to investigate the RNA binding activity of Sub2. This was modified according to Saguez *et al.*, 2013. This method has advantages but also disadvantages. It is laborious, time-consuming and no real-time tracking of the activity is possible. One could also investigate the binding for example via anisotropy, but this method also has its drawbacks. The EMSA, on the other hand, even if not in solution or with the possibility to follow the binding in real time, is simple and clear in its entirety. Sub2 accordingly shows a clear shift which does not show with a nucleotide binding deficient variant (Figure 3 C). The wild type is also unable to bind RNA without an analogue of ATP (Figure 6 B).

To test the ATPase activity, a NADH-coupled assay was used (Ren *et al.*, 2017). In this assay, ADP is regenerated to ATP by the consumption of NADH. The absorption of NADH is measured, which decreases indirectly when ATP is hydrolysed. The advantage is that a negative influence of ADP or decreasing concentration of ATP is prevented. This allows a linear curve, which is analysed accordingly. As expected, the wild type of Sub2 shows ATPase activity, whereas the Walker B variant does not (Figure 3 E-F). The ATPase rates obtained are similar to those already reported in Ren *et al.*, 2017.

The last biochemical assay discussed is the helicase assay. This is based on the assay used in Ordabayev *et al.*, 2018. As expected, Sub2 unwinds the substrate, which have already been used in the literature to detect helicase activity of Sub2 (Putnam and Jankowsky, 2013; Saguez *et al.*, 2013). Sub2 wild type shows an exponential curve. Figure 17 B shows that the substrates (Putnam, partial duplex, 13 bp) used for unwinding already anneal at room temperature. This is probably the reason why Sub2 cannot unwind 100%, because at a certain concentration of single stranded RNAs, the speed of reannealing is the same as the unwinding and an equilibrium is reached. To prevent this, one could use a complementary RNA or DNA as a trap. However, this was deliberately not done in order to exclude additional influences and unintentional manipulation of the experiments. It can also be seen that the substrate, which has already been pre-annealed in a thermocycler, continues to anneal. It might be that after dilution of a stock solution, some strands fall apart due to the sudden change in conditions and reanneal over time. Due to this observation, a negative "unwound fraction" might occur, as the measured value is smaller than the initial value, resulting in a negative amplitude. This is observed with catalytic dead variants, such as the Walker B variant, or with DNA/DNA substrates which do not stimulate the activity of Sub2 (Figure 9). However, the result showed that the helicase assay used in this work is a reproducible assay which is suitable to investigate reliably the helicase activity of various helicases.

In the ATPase and helicase assay, different concentrations of Sub2 were used to ensure that it is an enzymatic activity of Sub2 (Figure 4). The curve of both experiments shows a slightly sigmoidal shape. The beginning of the curve is flatter than an expected hyperbolic function. From lower to higher concentrations, it changes to a steeper curve and ends in a plateau. Such a curve might indicate an oligomerisation of the protein. In such a model several Sub2 molecules have to act together in order to work effectively, a flatter curve can be observed at low concentrations on a concentration/activity plot. Oligomerisation has already been described for the DEAD-box RNA helicase Ded1 (Putnam *et al.*, 2015), and would therefore also be a possible mechanism for other helicases. A similar curve can be seen for the helicase activity of UAP56, the human orthologue to Sub2, but has not been described (Perez-Calero *et al.*, 2020). In the published data it is seen, that

only above a certain concentration the substrate will be unwound, or that at low concentrations the curve is flatter than an expected hyperbolic function. However, this has not been discussed at all by the authors. So far, no oligomerisation of Sub2 or its orthologues has been described, but there are indications of it in this work, and for UAP56 in previous studies (Perez-Calero *et al.*, 2020). These findings can be used as a basis for future studies. Especially in combination with FRET experiments.

### **Sub2 is a sensing protein**

The influence of magnesium and nucleotide on the activity of Sub2 was investigated (Figure 5). It turned out that Sub2 shows highest activity at an optimal ratio of nucleotide to magnesium. This was shown in all three activity assays. The optimum was always 2:1 (nucleotide : magnesium). It has already been shown that the activity of Sub2 decreases when this ratio is reduced (Putnam and Jankowsky, 2013). This has also been shown for Dbp5, a protein that is also involved in mRNA export. This might suggest that Sub2 and Dbp5 have some sensing function in the nucleus which leads to signalling a shortage of resources. It is more likely that a low concentration of ATP is measured, as this would have more fatal consequences than a magnesium concentration that is too high.

However, it is very interesting that Sub2 and Dbp5 are the only proteins tested in Putnam and Jankowsky, 2013 that had a complete loss of activity in the presence of ADP. In this work as well, the strong negative influence of ADP on Sub2 was quantified and confirmed (Figure 7). This underlines again the assumption that a shortage of ATP concentration can be measured and signalled by these proteins. If the cell lacks resources and energy, waste products accumulate, including ADP. The inhibition of Sub2 and Dbp5, and consequently the inhibition of mRNA export, might trigger survival mechanisms of the cell in which the production and export of mRNA is stopped in order to save resources.

### **Substrate specificity of Sub2**

The substrate specificity of Sub2 was tested, as no detailed studies for Sub2 are known to date. Sub2 can stably bind short single-stranded or partially double-stranded RNA substrates (Figure 8 A). In the presence of double-stranded RNA, two bands are seen. These are the double strand and the long 38 nt RNA strand, as the latter is labelled with Cy5. If the short unlabelled strand of the partial duplex was exchanged with DNA, a greatly reduced formation of protein-substrate complex was observed. The question is why Sub2 is not able to bind the part of the substrate that is single-



stranded and consists of RNA. A slight shifted band can be seen, but the binding efficiency is much lower than expected. Sub2 does not bind single-stranded DNA at all (Figure 8). However, it could be that Sub2 generally binds double-stranded structures with higher affinity, as they also occur in secondary structures of single-stranded nucleic acids. According to the prediction software UNAFold, the overhang of the used substrate does not form any secondary structures, which explains why the hybrid with RNA overhang cannot be stably bound. In addition, a blunt end substrate has been tested in the helicase assay and it looks like the overhang does not play a role (Figure 8 E). Furthermore, the hybrid substrate was also used in the helicase assay and shows that Sub2 unwinds this substrate much better than RNA/RNA.

In further experiments, it was shown that the best substrate to be unwound by Sub2 is a hybrid with RNA overhang, but nevertheless also a hybrid with DNA overhang is unwound at least as well as RNA/RNA (Figure 9). For UAP56 it already has been shown to unwind hybrids *in vitro* and to be involved in R-loop resolution (Perez-Calero *et al.*, 2020). The THO-Sub2 complex has also been claimed to be involved in R-loop prevention (Gomez-Gonzalez *et al.*, 2011), but this has never been shown for Sub2 alone, and so it may be that it is mainly the THO complex which is responsible for R-loop prevention. In this work, the ability of UAP56 to unwind hybrids was confirmed and additionally shown for the orthologue from *C. thermophilum*. It seems to be a conserved mechanism. The possibility of Sub2 being involved in R-loop resolution is discussed in detail below (see section Chapter 6: Sub2 and Tho1 might resolve R-Loops *in vitro* and *in vivo*).

## Chapter 2: Sub2 has several conserved cysteines with unknown function

### Introduction of additional cysteines

The introduction of the cysteines can be considered a success. Suitable positions were found to strategically introduce additional cysteines on the surface without losing the activity. The activity is reduced (Figure 13 B), but taking this reduced activity into account, *in vitro* experiments can be carried out. Furthermore, it was also possible to label the variants, and these showed increased fluorescence compared to the wild type (Figure 13 D). However, the wild type could also be labelled, and for the best specific results, the wild type cysteines need to be removed. Deletion of the cysteines showed no phenotype *in vivo* (Figure 14 D-E), and no complete loss of activity *in vitro* (Figure 14 A-C). It is questionable whether activity is still present after deletion of the wild-type cysteines and introduction of an additional cysteine. What is still pending is testing of the activity after labelling. Adding a dye could lead to steric hindrance or might have other possible effects.

After carrying out *in vitro* experiments, the variants shall be investigated *in vivo* as well. Since Sub2 is involved in different mechanisms, it should be checked whether the introduction of cysteines has an influence on the *in vivo* function of Sub2, especially in view of the fact that Sub2 may be involved in mechanisms which are a response to oxidative stress (see section Function of cysteines). If it turns out that these variants do not show any phenotype, these variants might be used in more complex systems, for example to follow the dimerization of the THO-Sub2 complex.

In a labelling experiment in the presence of Sub2 WT, ADPNP, magnesium and RNA, so similar to a binding experiment, it turned out that these proteins could no longer be labelled as before, which suggests that the cysteines already present in Sub2 are less accessible (Figure 13 F). On the one hand, it could be that the additional components in the reaction have an effect on the labelling efficiency. However, it is likely that due to a conformational change of the protein, it becomes more compact and some cysteines are no longer accessible. This could be made use of in the future, in case it is not possible to delete all cysteines and introducing a new one on the surface. In this way, one could nevertheless study conformational change under different conditions.

## Function of cysteines

Sub2 was purified with and without TCEP and then an EMSA was performed in the presence or absence of TCEP (Figure 12). The experiment shows that reducing agents, and thus reduced cysteines, are necessary for Sub2 to be functional. However, this experiment also shows that the activity can be restored if TCEP is added to the reaction. On the other hand, the deletion of cysteines does not lead to a complete loss of function (Figure 14). Therefore, it is possible that non-reduced cysteines form disulphide bridges between proteins and these aggregate or the protein is not properly folded.

Even if there is no complete loss of activity, the variant with no cysteine loses significantly activity (Figure 14 A-C). Therefore, at least one of the highly conserved cysteines has an important function because the other two cysteines, which are not highly conserved, can be deleted without loss of activity. It is unclear if one cysteine might be responsible of the loss of activity, or whether a combination of cysteines.

At first glance, the cysteines do not seem to play any role *in vivo* (Figure 14 D-E). At least the methods used in this thesis showed no phenotype of the  $\Delta 6C$  variant. Neither under optimal conditions nor under temperature stress, a difference to the wild type was detected. The variants are not temperature sensitive.

That cysteines have a special role among amino acids has already been described (see section Chapter 2: Sub2 has several conserved cysteines with unknown function. in results). Therefore, it is very likely that the highly conserved cysteines of Sub2 do have an important function. In order to describe the function of the cysteines, it is necessary to find a suitable assay. There is evidence that Sub2 may respond to oxidative stress (data not published). The experiment in the presence or absence of TCEP is a strong indication that Sub2 might play a role in the response to oxidative stress.

## Chapter 3: Modulation of the activity of Sub2 by Tho1

### Sub2 and Tho1 interact with each other

So far, only a yeast-two hybrid assay has shown that Sub2 and Tho1 interact with each other (Ito 2001). However, this method also shows a positive result for indirect interaction. The human orthologues UAP56 (Sub2), SARNP (Tho1) and AlyRef (Yra1) have been shown to form an ATP-dependent complex and interact with each other *in vitro* and *in vivo* (Chang *et al.*, 2013; Dufu *et al.*, 2010; Fujita *et al.*, 2020).

In this work, it is shown that Sub2 and Tho1 form a ternary complex with RNA (Figure 15 A). It can be seen that at a certain ratio of Sub2 to Tho1, another band appears. The presence of Sub2 and Tho1 in this band was verified by mass spectrometry. The formation of this complex depends on the binding activity of Sub2. An inactive variant of Sub2 and the Tho1 wild type do not show a ternary complex, but wild type Sub2 with an inactive variant of Tho1 does (Figure 15 B-C). Furthermore, it is shown that Tho1 wild-type does not bind the short 13 nt long single stranded RNA as well as longer 38 nt long RNA (Figure 16 A). Based on these findings, it can be assumed that Sub2 binds the RNA directly and Tho1 forms a protein-protein interaction with Sub2. An interaction of Sub2 with Tho1 has already been predicted. In a computational model, with the aid of AlphaFold, a structure was modelled in which Tho1 interacts directly with Sub2 via its RNA-binding domain (Humphreys *et al.*, 2021). The experimental evidence for this direct interaction is discussed further down below (see section Tho1 interacts via its RNA-binding domain with Sub2 in discussion).

## Chapter 4: Tho1 has a novel annealing activity

### Detailed investigation of the accelerated annealing of nucleic acids by Tho1

The influence of Tho1 on the helicase activity of Sub2 was investigated (Figure 16). It turned out that Tho1 strongly inhibits the activity of Sub2, but the mechanism of inhibition was unclear (Figure 16 D). One possible way of inhibition might be that Tho1 binds to the substrate and stabilises or blocks it for Sub2. Another possibility would be a direct protein-protein interaction and an inactivation of Sub2. To test Tho1's binding capabilities in the unwinding assay, the binding of the substrate but also of the products by Tho1 was investigated (Figure 16 A). Tho1 stably binds the substrate, but also a strand resulting from unwinding.

Since it was shown that Tho1 binds to product and substrate, but also very likely interacts with Sub2, a further experiment was carried out. Sub2 was pre-incubated with the substrate, which was unwound as expected. Then Tho1 was added. The signal could have continued to increase, plateaued or decreased. When Tho1 is added, the signal drops rapidly. This could be a protein-protein interaction where Tho1 binds to Sub2 and leads to a disassociation of Sub2 from the RNA. After dissociation of Sub2 from the RNA and preventing it from rebinding, the RNA can reanneal. Although, that it was shown that Sub2 and Tho1 form a stable ternary complex, those experiments were done with ADPNP and the use of ATP in the helicase assay may have a different effect.

To investigate the effects seen previously in detail, an annealing assay was developed based on the helicase assay but starting with two single strands. Before testing for annealing, point mutations in conserved amino acids were introduced, to find a variant with loss of binding activity. This variant may serve as a negative control. Two highly conserved amino acids in the RNA binding domain were selected (Supplementary figure 1). F143 showed a loss of binding ability and was used as a negative control (Figure 17 A). Furthermore, the Putnam (partial duplex 13 bp) substrate was found to anneal rapidly at room temperature. Therefore, a substrate with a 3 nt longer duplex, that reportedly showed no spontaneous annealing (Ma *et al.*, 2013) was used (Figure 17 B).

Tho1 WT and F143D were tested in the annealing assay. In the presence of Tho1 F143D, no change in the signal is detected. However, when the wild-type is added to the two single stranded RNAs, the signal decreases. Thus, a novel annealing activity of Tho1 was found.

In addition to the substrate with overhang, blunt-end substrates were also tested (Figure 18 A). For the annealing of substrates by Tho1, the overhang plays no role. An acceleration of annealing can be seen even without overhang. Furthermore, duplexes of different lengths were tested. One was a 13 bp and the other a 38 bp long substrate. The 13 bp long duplex is not accelerated as much as

the standard substrate. This could be due to the fact that the substrate already anneals relatively quickly at room temperature without Tho1 and that further acceleration does not show a great effect. The influence on the 38 bp long substrate, on the other hand, is very distinct. The acceleration of annealing is about 20 times faster than without Tho1. Tho1 is thus able to anneal short but also longer substrates and does not need a single stranded overhang for that purpose. A more detailed analysis is required to understand the mechanism of annealing by Tho1. This should include shorter substrates and nucleic acids with different extent of secondary structures.

Since Tho1 has an SAP domain that is also supposed to bind DNA, new substrates were used that were not limited to RNA. Furthermore, Tho1 seems to play an active role in transcription, where not only RNA but also DNA plays a role. In addition to RNA/RNA and DNA/DNA substrates, hybrids with a DNA or RNA overhang were also tested (Figure 18 B-D).

Although Tho1 can anneal RNA, it turned out that annealing of two DNA strands is accelerated the most by Tho1 (Figure 18 D). However, a substrate with a DNA overhang is also annealed with a similar speed. That observation might depend on the degree of secondary structure. The long RNA strand (37 nt) has a more stable fold (NUPACK) compared to the corresponding DNA strand.

The question arises for what purpose the annealing activity is needed. Tho1 can anneal any combination of RNA and DNA. This means it could support the formation of harmful R-loops. But *in vivo*, many different proteins are involved in, for example, transcription, so the situation is much more complex than under isolated *in vitro* conditions. On the other hand, Tho1 might be involved in the remodelling of RNA. It could form specific secondary structures by annealing, binding and stabilising. However, the results suggest that Tho1 is mainly a DNA/DNA annealing protein. Thus, Tho1 could help to re-anneal chromosomal DNA after transcription. And not only that, but it could also remove harmful secondary structures of the DNA and thus contribute to the proper annealing of the DNA. The role of Tho1 in the resolution of R-loops and the annealing of DNA is discussed further below in Chapter 6: Sub2 and Tho1 might resolve R-Loops *in vitro* and *in vivo*.

The orthologues of Tho1 from *H. sapiens* and *C. thermophilum* were tested (Figure 21 A). The annealing activity appears to be conserved. Even though the activity of SARNP is significantly lower, an acceleration of annealing can be seen. The question remains unanswered why SARNP shows a significantly lower activity. It could be that more mechanisms have evolved in *H. sapiens* which make the annealing activity of SARNP redundant, and consequently adopted another major function. *C. thermophilum*, on the other hand, shows almost twice the activity of the yeast orthologue.

*C. thermophilum*'s optimum temperature is 50°C. A temperature at which strands are already separated locally or it is more difficult for strands to anneal again. Therefore, a higher activity is needed. All orthologues show the highest activity on DNA.

To check whether the annealing activity of Tho1 is specific for DNA or whether DNA anneals faster in general, Yra1 (Yeast RNA annealing protein 1) from *C. thermophilum* was purified and tested (Figure 21 D). It has already been reported (Portman *et al.*, 1997) that Yra1 anneals RNA, but no other substrates, such as DNA or hybrids, have been tested. Experiments showed that Yra1 from *C. thermophilum* is much more active than Tho1. It should be mentioned that in the experiments with Yra1, only 100 nM of the protein was used. For comparison: for the experiments with Tho1 it was 5 µM. Besides the high activity, the experiments showed as well that Yra1 anneals DNA substrates, but shows most activity on an RNA/RNA substrate.

During transcription, the THO complex recruits Sub2, and Sub2 in turn recruits Yra1. Considering that Sub2 may specifically recognise and bind double-stranded regions, an interaction between Sub2 and Yra1 could occur on secondary structures of mRNA, leading to remodelling of the mRNA and subsequent packaging into mRNPs and export from the cell.

### **Sub2 enhances the annealing activity of Tho1**

After the influence of Tho1 on Sub2 was investigated, experiments were also carried out to investigate the influence of Sub2 on the annealing activity of Tho1 (Figure 19). The results show that Sub2 stimulates the annealing activity of Tho1. However, the requirement for this is that at least one of the substrate strands is RNA. In fact, the stimulation is strongest on a hybrid with an RNA overhang. Sub2 has no effect on DNA. A possible explanation would be that Sub2 stimulates the activity of Tho1 via protein-protein interaction. But this would not explain why Sub2 has no influence on DNA/DNA. However, it could also be that the interaction is so specific that only the RNA annealing activity is stimulated. Another possibility is that Sub2 dissolves secondary structures that appear in the single strands. This might explain why Sub2 has no effect on DNA annealing. Furthermore, it would also explain why the hybrid with an RNA overhang is stimulated the most. The longer RNA forms larger secondary structures (Figure 19 G). And since there is no short RNA for Sub2 to inhibit the activity on the long strand, the effect is the greatest when only the long RNA strand is present.

To investigate the stimulation in more detail, and to test whether this is a specific Sub2-Tho1 interaction or dependent on the substrates, other helicases were tested (Figure 20). One was

another DEAD-box RNA helicase, Dbp2 from *S. pombe*, which is also involved in mRNA export, and UvrD, a DNA helicase and translocase from *E. coli*.

UvrD has no influence on RNA annealing. If, on the other hand, a DNA substrate is used, UvrD keeps the strands separate until all of the ATP is used up and the substrates can be annealed by Tho1. Since UvrD is a translocase, it can be assumed that secondary structures are resolved, but UvrD is always able to bind and unwind again and again and therefore has a negative effect on the annealing activity of Tho1. Unwinding by a DEAD-box helicase, however, allows a more specific mechanism.

First, the activity of Dbp2 was tested in the helicase assay (Figure 20 A-B). The activity is significantly lower than the activity of Sub2. However, one would expect that even a small effect would be visible if the stimulation effect would be substrate specific. However, Dbp2 shows no effect on the annealing activity, neither with DNA nor RNA. The findings suggest a mechanism based on a protein-protein interaction between Tho1 and Sub2. On the other hand, due to the lack of control experiments, such as substrates free of secondary structures, the possibility of an effect independent of protein interaction, such as the removal of those secondary structures, cannot be completely excluded. Therefore, further experiments were carried out with the orthologues from *H. sapiens* and these were cross tested with the yeast proteins (Figure 21 B-C).

In previous experiments, UAP56 showed a similar activity to Sub2. However, UAP56 does not stimulate the annealing activity of Tho1 in the way that Sub2 does. UAP56 stimulates activity on a hybrid with DNA overhang similar to Sub2, but has little effect on a hybrid with RNA overhang. Furthermore, UAP56 seems to have little effect on the annealing activity of SARNP. It does accelerate the annealing of a substrate with DNA overhang, but otherwise UAP56 has no effect. In comparison, Sub2 has a much stronger effect on the hybrid with DNA overhang. In contrast, Sub2 has the strongest effect on the annealing by Tho1 of a hybrid with RNA overhang.

Sub2 has a different effect on Tho1 than on SARNP. UAP56 has little effect on either annealing protein despite helicase activity. These different effects of the helicases on the annealing proteins suggest that it is eventually a specific effect which depends on a protein-protein interaction. Further experiments should be carried out to investigate these findings. The orthologues from *C. thermophilum* could be used for further cross experiments or new substrates which, for example, do not have secondary structures or cannot be processed by Sub2/UAP56.



## Chapter 5: Tho1 domain analysis

### **The C-terminal end is needed for binding activity, the role of the RNA-binding domain remains unclear**

If the highly basic C-terminal end is deleted, Tho1 is no longer able to bind nucleic acids (Figure 23). It is already sufficient if only the last 17 amino acids are deleted. ( $\Delta$ 2nd half CTE). Even the deletion of the RBD (according to Jacobsen *et al.*, 2016), shows no loss of binding activity. In contrast, RBD+CTE alone shows no clear shift in the EMSA experiment. However, the fluorescence of the free RNA clearly decreases. It can therefore be assumed that the substrate is bound in some form, but it remains unclear whether the variants cannot enter the gel due to a lack of charge, or whether the variant may be aggregated and bind the substrate non-specifically. Furthermore, even though the linker and SAP are not directly involved in the binding activity, it could be that these domains have a stabilising effect on the overall structure of the protein. If the linker is deleted, there is no loss of activity, but a no defined band forms and a smear is seen. The role of the RNA binding domain remains unclear. In the presence of only RBD, the fluorescence of free RNA decreases, but no clear shift can be seen, similar to RBD+CTE. The role of the SAP domain is unclear as well, as deletion of SAP shows no change in activity and the SAP domain alone shows no indication of binding activity.

The CTE has a strong positive charge and is suitable for binding negatively charged nucleic acids. Even though it was previously assumed to be a disordered region, the CTE seems to play a very important role in the activity of Tho1.

### **The C-terminal end is crucial for the annealing activity**

Binding activity is a requirement for the annealing of two single stranded nucleic acids. However, assuming that the RBD alone and RBD+CTE can bind nucleic acids, then substrate binding does not necessarily lead to annealing (Figure 24 A). The CTE already shows an important function in binding activity, but it is also becoming apparent that this part of Tho1 plays an extraordinary role in annealing activity. Deletion of the CTE also leads to a loss of activity in the annealing assay. And as for the binding activity, the deletion of the last 17 amino acids leads already to a loss of function. Deleting the first part of the CTE leads to a strong reduction of the annealing activity, but not to a complete loss. Since annealing activity is an active process, possibly associated with a conformational change, it could be that deletion of the first half of the CTE leads to a steric conflict that is harmful for the annealing function.

The RBD does not seem to play a primary role in annealing activity. If it is deleted, the activity decreases, but not all activity is lost. It is remarkable that in the binding experiments no difference between  $\Delta$ RBD and the wild type could be detected, but in the annealing assay a clear reduction in annealing activity can be observed.

If the SAP domain is deleted the annealing activity increases. It is still unclear what function the SAP domain provides in Tho1, but it seems to possibly have a regulatory function or serves an interaction that has not yet been described. It has been shown *in vivo* that deletion of the SAP domain leads to better recovery of the  $\Delta$ hpr1 phenotype (Jimeno *et al.*, 2006). It may therefore be that the annealing function is directly related to the ability to rescue this phenotype.

### **Tho1 interacts via its RNA-binding domain with Sub2**

The binding activity of Tho1 is not required to form a ternary complex with Sub2 and RNA and is further confirmed by the domain analyses (Figure 25 A). The CTE plays no role in the formation of the complex, and the deletion of SAP still leads to an interaction. Deletion of the linker leads to a very weak band that would correspond to a ternary complex. But most of the RNA does not run into the gel at all. Considering the binding experiments, the linker seems to have an important structural function.

If Tho1 consists only of the RNA-binding domain, a slight shift is visible which is higher than Sub2 alone. This is particularly evident from the raised edges of the band. When the RBD is deleted ( $\Delta$ RBD, SAP+linker, SAP), the formation of a ternary complex can no longer take place. The RNA binding domain of Tho1 is an interaction domain with an interface for Sub2.

The experimental results are supported by computational models of Sub2 and Tho1 (Humphreys *et al.*, 2021). It was shown that one helix in the RNA-binding domain might potentially be an interface for interaction with Sub2 (Figure 28).

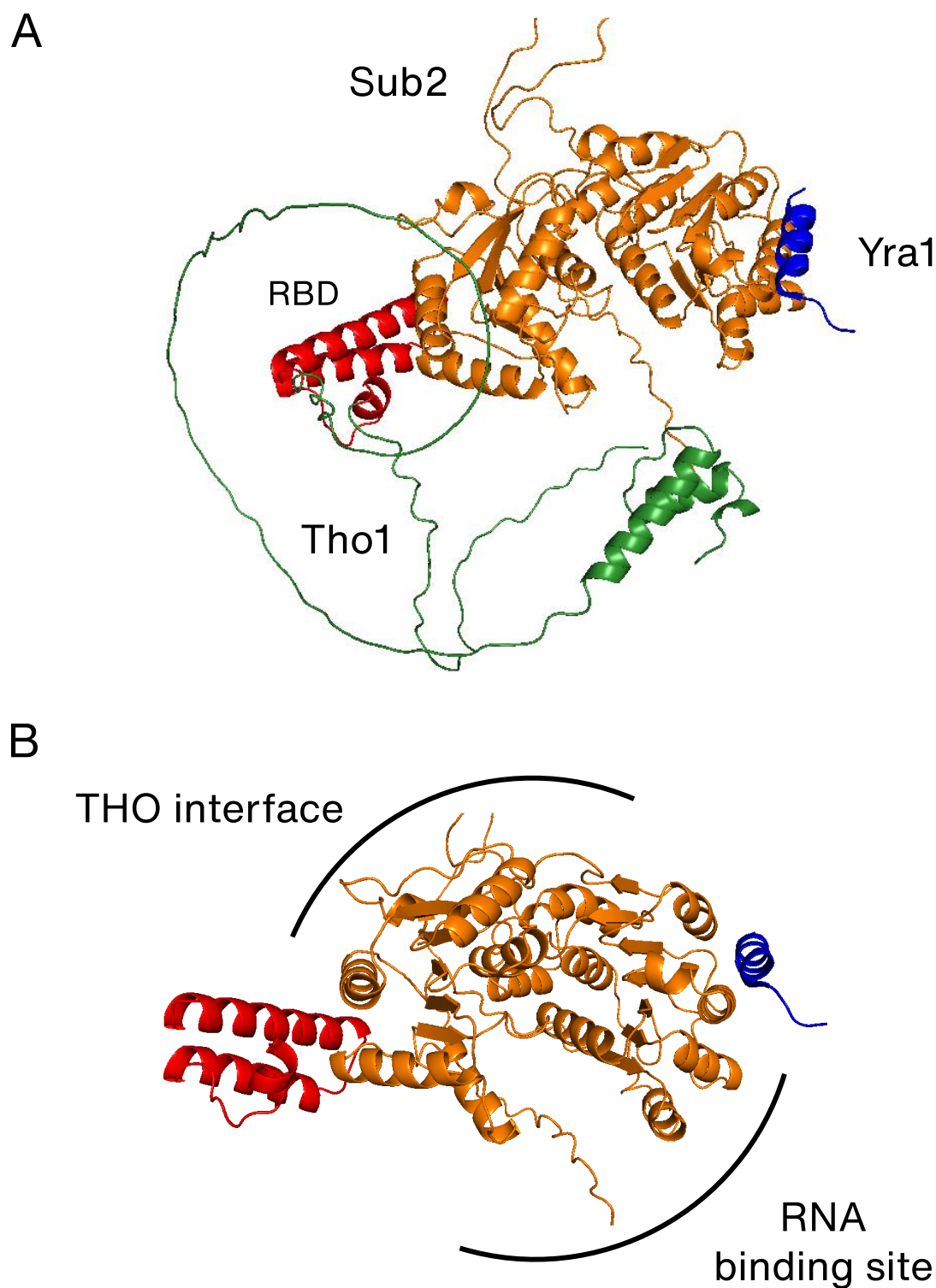
Interestingly, two residues, K135 and F143, have been found to crosslink to RNA *in vivo* (not published<sup>4</sup>). The variant K135D and F143D were used as controls (Figure 17) and F143 lost its ability to bind to RNA and to promote RNA-annealing.

However, since neither of these activities are required for ternary complex formation with Sub2 and RNA, it was interesting to note that Tho1 K135D was able to form a ternary complex with

---

<sup>4</sup> Dr. Philipp Keil, Institute of Biochemistry, Justus-Liebig-Universität, Giessen, Germany 2017

Sub2 and RNA, whereas Tho1 F143D did not (data not shown). In the structural models F143 but not K135 is a key residue of the protein-protein interface between Tho1 and Sub2.



**Figure 28: Modelled structure of Sub2 in complex with Tho1 and Yra1.** (A) Yra1 (blue) interacts with Sub2's (orange) n-terminal RecA-like domain (pdb: 5SUP, (Ren *et al.*, 2017)). Tho1 (green) interact with its RNA-Binding domain (red) with Sub2's c-terminal (pdb: n/a, supplementary material (Humphreys *et al.*, 2021)). (B) Same structure as A from different angle without full Tho1. When Sub2 interacts simultaneously with Yra1 and Tho1's RBD, the interface for the THO complex (Schuller *et al.*, 2020) and the RNA-binding site are still accessible. Created with PyMol.

### **Inhibition of Sub2 correlates with annealing activity**

The effect of the domains of Tho1 on the unwinding of partial duplex RNA by Sub2 was investigated (Figure 24 C-D). In general, the annealing function of Tho1 corresponds to the ability to inhibit Sub2 (WT,  $\Delta$ SAP,  $\Delta$ linker,  $\Delta$ RBD,  $\Delta$ 1st half CTE). However, it is noticeable that the intensity of annealing activity does not correlate completely with the intensity of inhibition of Sub2.

In addition to inhibition, stimulation of helicase activity was observed as well. The effect is particularly strong with the RBD. However, linker+RBD also shows significant stimulation. But the mechanism of stimulation is not known. Two mechanisms are possible: On the one hand, the single-stranded product could be bound and thus prevent reannealing of the substrates and would thus lead to a shift in the equilibrium. However, this is contradicted by the fact that the linker+RBD variant shows no signs of binding activity, whereas the RBD alone does (Figure 23). The second possibility would therefore be a direct protein-protein interaction that leads to an increased activity of Sub2. The interaction could interfere with different steps of the unwinding cycle. For example, the ADP release and thus the regeneration of Sub2 could be stimulated, or the conformational change to destabilise the double strand could be supported. Since evidence is growing for a direct interaction between Sub2 and the RBD of Tho1, the latter possibility is the more likely.

### **The RNA-binding domain of Tho1 is necessary for stimulation by Sub2**

In general, it is seen that a variant which still has annealing activity can also be stimulated by Sub2 (Figure 25 B). First assumption is a modulation of the RNA, independent of a possible interaction between Sub2 and Tho1.

The  $\Delta$ RBD variant showed annealing activity (Figure 24 A). However, the effect of stimulation is unexpectedly small, assuming that stimulation is solely dependent on the substrates used. On the other hand, the linker+RBD variant showed no annealing activity, but shows a very significant stimulation by Sub2. It may be that this variant has minimal, undetectable annealing activity, and this is then stimulated by Sub2 and made measurable. Another possibility is that this inactive variant can be activated by interaction with Sub2. However, one must be careful with significance tests, as they can only support a statement. Findings must be confirmed by additional experiments.

In this case, it is only safe to say that a greater stimulation by Sub2 of the  $\Delta$ RBD variant was expected, assuming it depends on the substrate. In addition, the experiments with the human orthologues (see section Sub2 enhances the annealing activity of Tho1 in discussion) suggest that stimulation is based on specific protein interaction. To confirm with certainty that enhancing is based on a protein-protein interaction, further experiments need to be performed. For example,

experiments could be performed at higher salt concentrations to disrupt interactions, or Tho1 F143D could be used to disrupt a potential interaction with Sub2. The computer model from Humphreys *et al.*, 2021 might serve as a basis for more point mutants.

## Summary of the domain analysis

This part summarises the discussed functions of the domains of Tho1.

There is little evidence on the function of the SAP domain. The literature suggests that the SAP domain is a specific DNA binding domain that binds AT rich regions (Aravind and Koonin, 2000). In addition, the SAP domain often occurs in context with DNA repair associated proteins (Aravind and Koonin, 2000). It has also been reported that SAR-A can still bind to DNA after deletion of the SAP domain, similar to Tho1 (Gohring *et al.*, 1997). It has also been suggested that the SAP domain is a dimerization domain (Aravind and Koonin, 2000). In any case, it can be said that without the SAP domain, the activity of Tho1 is increased and the protein does not lose any other activity. It is possible that the SAP domain fulfils a specific function *in vivo*. The results of this work suggest that the SAP domain plays a role in the stability or regulation of Tho1. Interestingly, the *A. thaliana* Tho1 orthologue MOS11 does not contain a SAP-domain, but functions in mRNA-export and TREX complex formation (Sørensen *et al.*, 2017).

The linker may be responsible for supporting the structural integrity of Tho1. On the one hand, it gives Tho1 a certain flexibility, but also the potential possibility that the SAP and RNA-binding domains can act independently in relative distance.

The region of Tho1 previously referred to as the RNA-binding domain appears, according to the results of this work, to be an interaction domain for Sub2. This domain may also have a role in binding and annealing. However, this appears to be secondary.

The C-terminal end has not received much attention in the past. However, it seems to play a very large role in the activity of Tho1. It is highly basic but also predicted to be disordered. Which does not mean that this region cannot undergo a conformational change. The CTE is absolutely necessary for binding and annealing, but plays no role in the interaction with Sub2.

The RBD and CTE may need to be redefined. Recent *in silico* findings show defined repetitive conserved motifs within Tho1 and its orthologues (not published, see appendix Supplementary figure 1). These conserved "boxes" should be further investigated in the future, as they are promising motifs that appear not only in Tho1 from yeast, but also in several homologues.

## Chapter 6: Sub2 and Tho1 might resolve R-Loops *in vitro* and *in vivo*

R-loops are structures of three nucleic acids consisting of a partially unwound double-stranded DNA where one strand is associated with a complementary RNA. These structures occur most often during transcription, where the transcribed RNA anneals with the still unwound DNA. This region can be several hundred base pairs long. The point of interest for resolving R-loops does not necessarily have to be at the "fork" of these structures, but can also be in the middle of the R-loop, where the complementary DNA strand is distant, but nevertheless available to reanneal. This assumption is the basis of the way the R-Loops resolution assay is designed in this work. Therefore, an R-loop like substrate consisting of a partial duplex hybrid and a free, perfectly complementary DNA (see section Tho1 is able to displace RNA-strands *in vitro* in results for details) was used in this assay (Figure 26).

In the presence of the complementary DNA strand, Sub2 is able to unwind the hybrid, as the complementary strand acts as a trap, taking the single strand DNA out of the equilibrium. A surprising result is seen with Tho1. While in the absence of the complementary DNA strand nothing happens, in the presence of the complementary DNA strand, the RNA strand is displaced from its associated DNA strand. It can be assumed that this effect is based on the annealing activity of Tho1. Tho1 is able to anneal the overhang of the hybrid with the complementary DNA strand, and subsequently displace the RNA strand by further annealing. After an initial burst, the fluorescence decreases again. This leads to a new formation of the RNA/DNA hybrid. Tho1 can not only anneal DNA but also these hybrids. After annealing the DNA, the reaction runs in the opposite direction again until an equilibrium is reached. But this changes when Sub2 and Tho1 can act together. A burst can be seen and the immediate establishment of an equilibrium which remains constantly high. Tho1 displaces the RNA and Sub2 ensures that the RNA is no longer available for Tho1 to reanneal it again to the DNA. *In vitro*, Sub2 and Tho1 show cooperative activity to resolve R-loops.

*In vivo*, there is no increase of the level of R-loops when Tho1 is deleted. However, overexpression of Tho1 not only rescues the  $\Delta hpr1$  phenotype, but also leads to a decrease in R-loop levels. Even if overexpression of Sub2 rescues the same phenotype, no reduction in R-loops is seen.

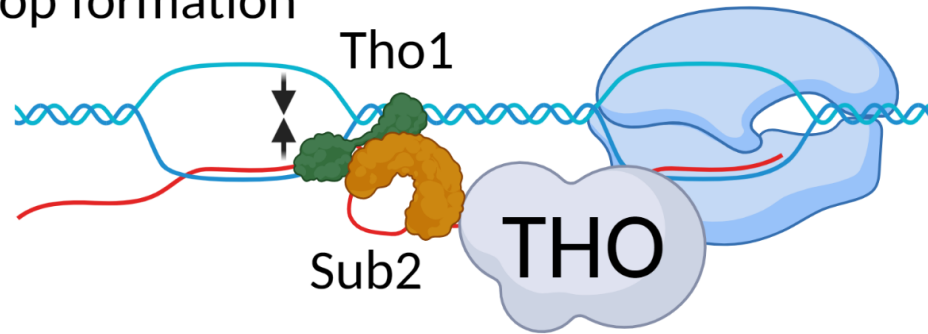
It has already been shown that Sub2-THO is involved in R-loop prevention (Gomez-Gonzalez *et al.*, 2011). Since prevention can also be a passive process, it may be that overexpression of Sub2 has no effect on R-loop resolution. Although it has been shown that Sub2 and also its orthologues can unwind hybrids, this activity could be only supportive. As it has been shown in the *in vitro* experiment, Tho1 has a much greater potential to resolve R-loops, and Sub2 has a supporting role. This might be the mechanism of R-loop resolution *in vivo*.

### **A model for R-Loop resolution by Sub2 and Tho1**

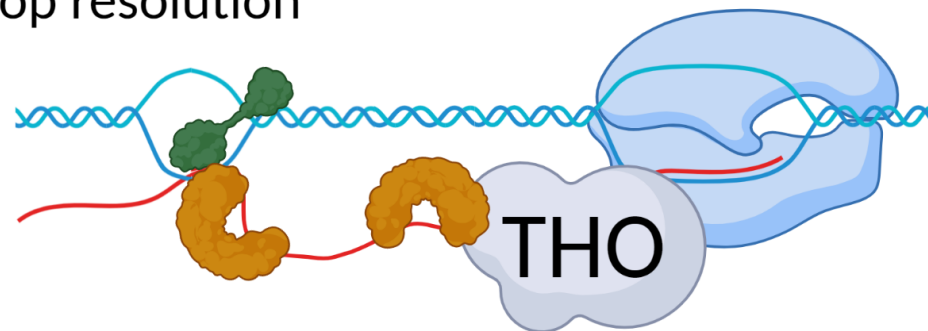
Sub2 and Tho1 cooperate to resolve R-loops. The THO-Sub2 complex already associates with the mRNA during transcription. This is already the first step to prevent R-loops, as the RNA strand is bound and it becomes more difficult for it to hybridise with the DNA. Since this is only a preventive mechanism, overexpression of Sub2 does not help to decrease the level of R-loops. However, overexpression can nevertheless rescue the  $\Delta hpr1$  phenotype. Overexpression of Sub2 compensates for the relatively low RNA binding affinity of Sub2 (Jimeno *et al.*, 2006). This bypasses the requirement for the THO complex to be present. Accumulation of Sub2 on the mRNA, leads to recruitment of Yra1, and subsequently to recruitment of other export factors, and ultimately to mRNA export itself.

Nevertheless, should an R-loop form, Sub2 is able to recruit Tho1 via protein-protein interaction. Interaction with Tho1 does not exclude interaction with Yra1, as the interface for the two proteins is in different RecA-like domains of Sub2 (Figure 28, Humphreys *et al.*, 2021). Recruitment of Tho1 leads to R-loop resolution by an active annealing mechanism of the DNA strands, which displaces the RNA. Sub2 can then support the removal of the R-loop with its binding and unwinding activity. Yra1 is then recruited and mRNA export takes place. Since this is an active process through Tho1, the overexpression will lead to a reduction of R-loops. The formation of the R-loops leads to stalling of the polymerase and mRNA, which prevents export. Tho1 resolves this and the mRNA can then be exported via Sub2 and Yra1, even without a functioning THO complex.

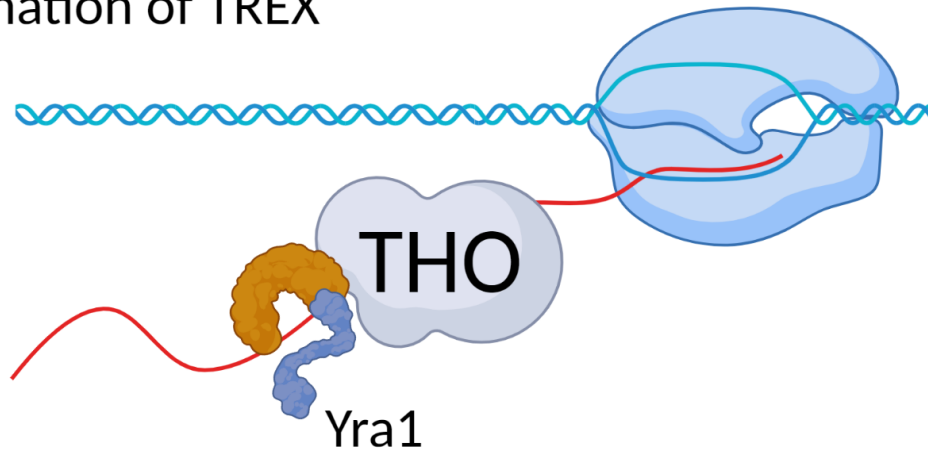
### R-Loop formation



### R-Loop resolution



### Formation of TREX



mRNA Export

**Figure 29: Proposed model for the resolution of R-loops by Sub2 and Tho1.** When R-loops form and the mRNA is stalled, Tho1 can be recruited by protein-protein interaction with Sub2, which has already bound the RNA together with the THO complex. By Tho1's annealing activity, the DNA rehybridizes and the mRNA is displaced from the DNA. To prevent the mRNA from rehybridizing, Sub2 assists in keeping the RNA single-stranded and separated from the DNA. After the R-loop is removed, Yra1 is recruited to form the TREX complex which leads to the export of mRNA. Created with BioRender.com.



## Conclusions

Sub2 was studied in detail with regard to buffer conditions, substrate and nucleotide specificity. It was shown that Sub2 has reduced activity with increasing concentrations of magnesium. But more interesting is the fact that Sub2 is strongly inhibited by ADP and thus might be a sensing protein for cellular stress situations. However, these findings need further investigation to show that inhibition by ADP serves a biological role.

In addition, it has been shown that Sub2 unwinds RNA/DNA hybrids better than RNA/RNA. This was also shown for the Sub2 orthologues from *H. sapiens* and *C. thermophilum*. The question is whether this is a specific effect or whether all DEAD-box helicases show the same properties with regard to substrate specificity. For this, further helicases might be tested in the future.

Cysteines could be introduced, but the deletion of the existing cysteines might be critical. Nevertheless, the function of the conserved and Sub2 specific cysteines is still completely unknown. They might have a role in a signalling pathway, as it was shown in this work that Sub2 could be a sensing protein. However, the right assays need to be found to proof that. Unfortunately, this project was not continued because there was not enough capacity and Tho1 was prioritized.

While studying the influence of Tho1 on Sub2, a strong inhibition of unwinding activity was shown. It turned out that Tho1 has an annealing activity. This activity was analysed and additionally the functions of the domains of Tho1 were investigated in detail. An *in silico* postulated interaction between Tho1 and Sub2 was experimentally confirmed. Furthermore, it was shown that the C-terminal end plays an essential role in the binding and annealing activity of Tho1. The function of the SAP domain remains unclear. It may have an *in vivo* function that could not be determined with *in vitro* experiments used in this work. For the characterisation of the SAP domain, extended *in vivo* experiments should be carried out in the future.

Finally, it was shown that Tho1 can resolve R-loop-like structures *in vitro*. RNA is removed from a DNA strand via its annealing activity. It was also shown *in vivo* that overexpression of Tho1 leads to a reduction of the level of R-loops. Taken together, a model was designed in which Sub2 and Tho1 cooperatively resolve R-loops *in vitro* and *in vivo*.

# References

- Aravind L, Koonin EV. 2000.** SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends in Biochemical Sciences* **25**:112–114. doi: 10.1016/s0968-0004(99)01537-6.
- Aregger R, Klostermeier D. 2009.** The DEAD box helicase YxiN maintains a closed conformation during ATP hydrolysis. *Biochemistry* **48**:10679–10681. doi: 10.1021/bi901278p.
- Bartonek L, Zagrovic B. 2019.** VOLPES: an interactive web-based tool for visualizing and comparing physicochemical properties of biological sequences. *Nucleic Acids Research* **47**:W632–W635. doi: 10.1093/nar/gkz407.
- Bjork P, Wieslander L. 2014.** Mechanisms of mRNA export. *Seminars in Cell & Developmental Biology* **32**:47–54. doi: 10.1016/j.semcdb.2014.04.027.
- Bjork P, Wieslander L. 2017.** Integration of mRNP formation and export. *Cell Mol Life Sci* **74**:2875–2897. doi: 10.1007/s00018-017-2503-3.
- Chang C-T, Hautbergue GM, Walsh MJ, Viphakone N, van Dijk TB, Philipsen S, Wilson SA. 2013.** Chtop is a component of the dynamic TREX mRNA export complex. *Embo Journal* **32**:473–486. doi: 10.1038/emboj.2012.342.
- Chavez S, Aguilera A. 1997.** The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes & Development* **11**:3459–3470. doi: 10.1101/gad.11.24.3459.
- Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. 2000.** A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *Embo Journal* **19**:5824–5834. doi: 10.1093/emboj/19.21.5824.
- Chavez S, Garcia-Rubio M, Prado F, Aguilera A. 2001.** Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol* **21**:7054–7064. doi: 10.1128/MCB.21.20.7054-7064.2001.
- Chen C, Tan M, Wu ZF, Zhang YB, He FY, Lu YJ, Li SB, Cao M, Li GH, Wu J, Cheng H, Lei M. 2021.** Structural and functional insights into R-loop prevention and mRNA export by budding yeast THO-Sub2 complex. *Science Bulletin* **66**:2347–2352. doi: 10.1016/j.scib.2021.08.004.
- Collins R, Karlberg T, Lehtio L, Schutz P, van den Berg S, Dahlgren LG, Hammarstrom M, Weigelt J, Schuler H. 2009.** The DEXD/H-box RNA Helicase DDX19 Is Regulated by an alpha-Helical Switch. *Journal of Biological Chemistry* **284**:10296–10300. doi: 10.1074/jbc.C900018200.

- Dodson CA**, Ferguson N, Rutherford TJ, Johnson CM, Fersht AR. **2010**. Engineering a two-helix bundle protein for folding studies. *Protein Eng Des Sel* **23**:357–364. doi: 10.1093/protein/gzp080.
- Dominguez-Sanchez MS**, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. **2011**. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7**:e1002386. doi: 10.1371/journal.pgen.1002386.
- Dufu K**, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, Reed R. **2010**. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes & Development* **24**:2043–2053. doi: 10.1101/gad.1898610.
- Erdos G**, Pajkos M, Dosztanyi Z. **2021**. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res* **49**:W297–W303. doi: 10.1093/nar/gkab408.
- Fan HY**, Merker RJ, Klein HL. **2001**. High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. *Mol Cell Biol* **21**:5459–5470. doi: 10.1128/MCB.21.16.5459-5470.2001.
- Fleckner J**, Zhang M, Valcarcel J, Green MR. **1997**. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* **11**:1864–1872.
- Folco EG**, Lee CS, Dufu K, Yamazaki T, Reed R. **2012**. The proteins PDIP3 and ZC11A associate with the human TREX complex in an ATP-dependent manner and function in mRNA export. *PLoS One* **7**:e43804. doi: 10.1371/journal.pone.0043804.
- Fujita KI**, Yamazaki T, Harada K, Seno S, Matsuda H, Masuda S. **2020**. URH49 exports mRNA by remodeling complex formation and mediating the NXF1-dependent pathway. *Biochim Biophys Acta Gene Regul Mech* **1863**:194480. doi: 10.1016/j.bbagr.2020.194480.
- Gasteiger E**, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. **2005**. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ.
- Gatfield D**, Le Hir H, Schmitt C, Braun IC, Kocher T, Wilm M, Izaurralde E. **2001**. The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila*. *Curr Biol* **11**:1716–1721.
- Gibson DG**, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. **2009**. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**:343–345. doi: 10.1038/nmeth.1318.
- Ginno PA**, Lott PL, Christensen HC, Korf I, Chedin F. **2012**. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Molecular Cell* **45**:814–825. doi: 10.1016/j.molcel.2012.01.017.

- Gohring F**, Schwab BL, Nicotera P, Leist M, Fackelmayer FO. **1997**. The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. *Embo Journal* **16**:7361–7371. doi: 10.1093/emboj/16.24.7361.
- Gomez-Gonzalez B**, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A, Foiani M, Aguilera A. **2011**. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* **30**:3106–3119. doi: 10.1038/emboj.2011.206.
- Gwizdek C**, Hobeika M, Kus B, Ossareh-Nazari B, Dargemont C, Rodriguez MS. **2005**. The mRNA nuclear export factor Hpr1 is regulated by Rsp5-mediated ubiquitylation. *The Journal of biological chemistry* **280**:13401–13405. doi: 10.1074/jbc.C500040200.
- Gwizdek C**, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, Stutz F, Dargemont C. **2006**. Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proc Natl Acad Sci U S A* **103**:16376–16381. doi: 10.1073/pnas.0607941103.
- Heath CG**, Viphakone N, Wilson SA. **2016**. The role of TREX in gene expression and disease. *Biochem J* **473**:2911–2935. doi: 10.1042/BCJ20160010.
- Hermanson GT**. **2013**. *Bioconjugate Techniques*. Academic Press.
- Hetzer MW**, Wente SR. **2009**. Border control at the nucleus: Biogenesis and organization of the nuclear membrane and pore complexes. *Developmental cell* **17**:606–616. doi: 10.1016/j.devcel.2009.10.007.
- Humphreys IR**, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR, Stancheva VG, Li XH, Liu K, Zheng Z, Barrero DJ, Roy U, Kuper J, Fernandez IS, Szakal B, Branzei D, Rizo J, Kisker C, Greene EC, Biggins S, Keeney S, Miller EA, Fromme JC, Hendrickson TL, Cong Q, Baker D. **2021**. Computed structures of core eukaryotic protein complexes. *Science* eabm4805. doi: 10.1126/science.abm4805.
- Hurt E**, Luo MJ, Rother S, Reed R, Strasser K. **2004**. Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proceedings of the National Academy of Sciences of the United States of America* **101**:1858–1862. doi: 10.1073/pnas.0308663100.
- Ito T**, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. **2001**. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98**:4569–4574. doi: 10.1073/pnas.061034498.
- Jacobsen JO**, Allen MD, Freund SM, Bycroft M. **2016**. High-resolution NMR structures of the domains of *Saccharomyces cerevisiae* Tho1. *Acta Crystallogr F Struct Biol Commun* **72**:500–506. doi: 10.1107/S2053230X16007597.

- Jimeno S**, Luna R, Garcia-Rubio M, Aguilera A. **2006**. Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. *Mol Cell Biol* **26**:4387–4398. doi: 10.1128/MCB.00234-06.
- Jimeno S**, Rondon AG, Luna R, Aguilera A. **2002**. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J* **21**:3526–3535. doi: 10.1093/emboj/cdf335.
- Kapadia F**, Pryor A, Chang TH, Johnson LF. **2006**. Nuclear localization of poly(A)<sup>+</sup> mRNA following siRNA reduction of expression of the mammalian RNA helicases UAP56 and URH49. *Gene* **384**:37–44. doi: 10.1016/j.gene.2006.07.010.
- Katahira J**. **2012**. mRNA export and the TREX complex. *Biochim Biophys Acta* **1819**:507–513. doi: 10.1016/j.bbagr.2011.12.001.
- Katahira J**, Straesser K, Saiwaki T, Yoneda Y, Hurt E. **2002**. Complex formation between Tap and p15 affects binding to FG-repeat nucleoporins and nucleocytoplasmic shuttling. *Journal of Biological Chemistry* **277**:9242–9246. doi: 10.1074/jbc.M110007200.
- Kipp M**, Gohring F, Ostendorp T, van Drunen CM, van Driel R, Przybylski M, Fackelmayer FO. **2000**. SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol Cell Biol* **20**:7480–7489. doi: 10.1128/MCB.20.20.7480-7489.2000.
- Kistler AL**, Guthrie C. **2001**. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev* **15**:42–49.
- Libri D**, Graziani N, Saguez C, Boulay J. **2001**. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev* **15**:36–41.
- Luna R**, Rondon AG, Perez-Calero C, Salas-Armenteros I, Aguilera A. **2019**. The THO Complex as a Paradigm for the Prevention of Cotranscriptional R-Loops. *Cold Spring Harb Symp Quant Biol* **84**:105–114. doi: 10.1101/sqb.2019.84.039594.
- Ma WK**, Cloutier SC, Tran EJ. **2013**. The DEAD-box protein Dbp2 functions with the RNA-binding protein Yra1 to promote mRNP assembly. *J Mol Biol* **425**:3824–3838. doi: 10.1016/j.jmb.2013.05.016.
- MacMorris M**, Brocker C, Blumenthal T. **2003**. UAP56 levels affect viability and mRNA export in *Caenorhabditis elegans*. *RNA* **9**:847–857.
- Malig M**, Hartono SR, Giafaglione JM, Sanz LA, Chedin F. **2020**. Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. *Journal of Molecular Biology* **432**:2271–2288. doi: 10.1016/j.jmb.2020.02.014.
- Marino SM**, Gladyshev VN. **2008**. Functional Diversity of Cysteine Residues in Proteins and Unique Features of Catalytic Redox-active Cysteines in Thiol Oxidoreductases. *Molecules and Cells* **26**:228–235.

- Marino SM**, Gladyshev VN. **2009**. A Structure-Based Approach for Detection of Thiol Oxidoreductases and Their Catalytic Redox-Active Cysteine Residues. *Plos Computational Biology* **5**:e1000383. doi: 10.1371/journal.pcbi.1000383.
- Marino SM**, Gladyshev VN. **2010**. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *Journal of Molecular Biology* **404**:902–916. doi: 10.1016/j.jmb.2010.09.027.
- Ordabayev YA**, Nguyen B, Niedziela-Majka A, Lohman TM. **2018**. Regulation of UvrD Helicase Activity by MutL. *J Mol Biol* **430**:4260–4274. doi: 10.1016/j.jmb.2018.08.022.
- Paulsen CE**, Carroll KS. **2013**. Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. *Chemical reviews* **113**:4633–4679. doi: 10.1021/cr300163e.
- Perez-Calero C**, Bayona-Feliu A, Xue X, Barroso SI, Munoz S, Gonzalez-Basallote VM, Sung P, Aguilera A. **2020**. UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide. *Genes Dev* **34**:898–912. doi: 10.1101/gad.336024.119.
- Piruat JI**, Aguilera A. **1998**. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J* **17**:4859–4872. doi: 10.1093/emboj/17.16.4859.
- Poole LB**. **2015**. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic Biol Med* **80**:148–157. doi: 10.1016/j.freeradbiomed.2014.11.013.
- Portman DS**, O'Connor JP, Dreyfuss G. **1997**. YRA1, an essential *Saccharomyces cerevisiae* gene, encodes a novel nuclear protein with RNA annealing activity. *RNA* **3**:527–537.
- Putnam A**, Jankowsky E. **2012**. Analysis of duplex unwinding by RNA helicases using stopped-flow fluorescence spectroscopy. *Methods in enzymology* **511**:1–27. doi: 10.1016/B978-0-12-396546-2.00001-2.
- Putnam AA**, Gao Z, Liu F, Jia H, Yang Q, Jankowsky E. **2015**. Division of Labor in an Oligomer of the DEAD-Box RNA Helicase Ded1p. *Molecular Cell* **59**:541–552. doi: 10.1016/j.molcel.2015.06.030.
- Putnam AA**, Jankowsky E. **2013**. AMP Sensing by DEAD-Box RNA Helicases. *Journal of Molecular Biology* **425**:3839–3845. doi: 10.1016/j.jmb.2013.05.006.
- Ren Y**, Schmiede P, Blobel G. **2017**. Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. *Elife* **6**. doi: 10.7554/eLife.20070.
- Rother S**, Straesser K. **2009**. mRNA export-an integrative component of gene expression. Ralph Kehlenbach (Ed.), *Nuclear Transport*, Landes Bioscience, Austin.
- Rudolph MG**, Klostermeier D. **2015**. When core competence is not enough: functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biological Chemistry* **396**:849–865. doi: 10.1515/hsz-2014-0277.

- Saguez C**, Gonzales FA, Schmid M, Boggild A, Latrick CM, Malagon F, Putnam A, Sanderson L, Jankowsky E, Brodersen DE, Jensen TH. **2013**. Mutational analysis of the yeast RNA helicase Sub2p reveals conserved domains required for growth, mRNA export, and genomic stability. *RNA* **19**:1363–1371. doi: 10.1261/rna.040048.113.
- Sanz LA**, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, Chédin F. **2016**. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular Cell* **63**:167–178. doi: 10.1016/j.molcel.2016.05.032.
- Schuller SK**, Schuller JM, Prabu JR, Baumgartner M, Bonneau F, Basquin J, Conti E. **2020**. Structural insights into the nucleic acid remodeling mechanisms of the yeast THO-Sub2 complex. *Elife* **9**. doi: 10.7554/eLife.61467.
- Sengoku T**, Nureki O, Nakamura A, Kobayashi S, Yokoyama S. **2006**. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* **125**:287–300. doi: 10.1016/j.cell.2006.01.054.
- Shi H**, Cordin O, Minder CM, Linder P, Xu RM. **2004**. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc Natl Acad Sci U S A* **101**:17628–17633. doi: 10.1073/pnas.0408172101.
- Sørensen BB**, Ehrnsberger HF, Esposito S, Pfab A, Bruckmann A, Hauptmann J, Meister G, Merkl R, Schubert T, Längst G, Melzer M, Grasser M, Grasser KD. **2017**. The Arabidopsis THO/TREX component TEX1 functionally interacts with MOS11 and modulates mRNA export and alternative splicing events. *Plant molecular biology* **93**:283–298. doi: 10.1007/s11103-016-0561-9.
- Steimer L**, Klostermeier D. **2012**. RNA helicases in infection and disease. *Rna Biology* **9**:751–771. doi: 10.4161/rna.20090.
- Strasser K**, Bassler J, Hurt E. **2000**. Binding of the Mex67p/Mtr2p heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. *The Journal of cell biology* **150**:695–706.
- Strasser K**, Hurt E. **2001**. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**:648–652. doi: 10.1038/35098113.
- Strasser K**, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E. **2002**. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**:304–308. doi: 10.1038/nature746.
- Sugiura T**, Sakurai K, Nagano Y. **2007**. Intracellular characterization of DDX39, a novel growth-associated RNA helicase. *Exp Cell Res* **313**:782–790. doi: 10.1016/j.yexcr.2006.11.014.
- Talavera MA**, La Cruz EM de. **2005**. Equilibrium and kinetic analysis of nucleotide binding to the DEAD-box RNA helicase DbpA. *Biochemistry* **44**:959–970. doi: 10.1021/bi048253i.

- Taniguchi I**, Ohno M. **2008**. ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56. *Mol Cell Biol* **28**:601–608. doi: 10.1128/MCB.01341-07.
- Tanner N**, Cordin O, Banroques J, Doère M, Linder P. **2003**. The Q Motif. *Molecular cell* **11**:127–138. doi: 10.1016/S1097-2765(03)00006-6.
- Wende W**, Friedhoff P, Strasser K. **2019**. Mechanism and Regulation of Co-transcriptional mRNP Assembly and Nuclear mRNA Export. In: Oeffinger M, Zenklusen D, eds. *Biology of Mrna: Structure and Function*.
- Wickramasinghe VO**, Laskey RA. **2015**. Control of mammalian gene expression by selective mRNA export. *Nat Rev Mol Cell Biol* **16**:431–442. doi: 10.1038/nrm4010.
- Xie Y**, Clarke BP, Kim YJ, Ivey AL, Hill PS, Shi Y, Ren Y. **2021**. Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2. *Elife* **10**. doi: 10.7554/eLife.65699.
- Yang Q**, Jankowsky E. **2005**. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**:13591–13601. doi: 10.1021/bi0508946.
- Younis S**, Kamel W, Falkeborn T, Wang H, **Di Yu**, Daniels R, Essand M, Hinkula J, Akusjärvi G, Andersson L. **2018**. Multiple nuclear-replicating viruses require the stress-induced protein ZC3H11A for efficient growth. *Proceedings of the National Academy of Sciences* **115**:E3808–E3816. doi: 10.1073/pnas.1722333115.
- Zhang M**, Green MR. **2001**. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev* **15**:30–35.



# List of figures

Figure 1: Overexpression of Sub2 or Tho1 rescue the $\Delta$ THO phenotype. ....	9
Figure 2: Annealed substrates used in this thesis.....	37
Figure 3: Purification and biochemical analysis (RNA-binding, ATPase and helicase) of Sub2 and variants thereof used in this thesis. ....	42
Figure 4: Concentration dependency of Sub2 wild-type in ATPase and helicase assay.....	43
Figure 5: The ratio of magnesium ions to nucleotide is crucial for the activity of Sub2. ....	44
Figure 6: Impact of different nucleotides on the activity of Sub2. ....	45
Figure 7: The ATPase Sub2 is inhibited by ADP.....	47
Figure 8: Investigation of substrate specificity of Sub2.....	49
Figure 9: Sub2 unwinds RNA/DNA hybrids. ....	50
Figure 10: Unwinding of RNA/DNA hybrids is conserved among <i>Homo sapiens</i> and <i>Chaetomium thermophilum</i> . ....	51
Figure 11: Bioinformatical analysis of Sub2's cysteines.....	53
Figure 12: Reducing agent is necessary to maintain Sub2's binding activity. ....	54
Figure 13: Introducing additionally cysteines for future fluorescence-based assays. ....	56
Figure 14: Substitution of all cysteines does not abolish Sub2's activity.....	58
Figure 15: Sub2 and Tho1 form a ternary complex on RNA, but Tho1 does not affect the ATPase activity of Sub2.. ....	60
Figure 16: Influence of Tho1 on the RNA unwinding activity of Sub2. ....	61
Figure 17: Characterization of Tho1's annealing activity. ....	63
Figure 18: Tho1 anneals RNA, DNA and hybrid substrates.....	64
Figure 19: Sub2 is enhancing the annealing activity of Tho1. ....	66
Figure 20: Annealing experiments with different helicases to test if the enhancing of the annealing activity of Tho1 by Sub2 is a specific interaction. ....	67
Figure 21: Tho1's annealing activity is conserved among <i>Homo sapiens</i> and <i>Chaetomium thermophilum</i> and Yra1 is specific for RNA. ....	69
Figure 22: Overview of analysed Tho1 domain variants and properties of Tho1.....	70
Figure 23: Investigation of binding capability of Tho1 wild-type and domains. ....	72
Figure 24: Analysis of annealing activity of Tho1 wild-type and domain variants. ....	73
Figure 25: Ternary complex formation of Sub2 and Tho1 domain variants and enhancing of annealing activity of these variants by Sub2.. ....	75
Figure 26: Resolution of R-Loop like structure by Sub2 and Tho1 <i>in vitro</i> . ....	78
Figure 27: Tho1 overexpression is able to diminish the accumulation of R-loops in $\Delta$ <i>hpr1</i> . ....	80

Figure 28: Modelled structure of Sub2 in complex with Tho1 and Yra1.. .....	94
Figure 29: Proposed model for the resolution of R-loops by Sub2 and Tho1.. .....	99

# List of tables

Table 1: Chemicals and consumables .....	11
Table 2: Equipment and devices .....	13
Table 3: Kits .....	19
Table 4: Enzymes .....	19
Table 5: Antibodies .....	19
Table 6: Markers .....	20
Table 7: Softwares .....	20
Table 8: <i>E. coli</i> strains .....	20
Table 9: Yeast strains .....	21
Table 10: Plasmids .....	21
Table 11: Primers .....	23
Table 12: Substrates .....	26
Table 13: Thermocycler program for PCR with Phusion polymerase .....	31
Table 14: Thermocycler program for PCR with Taq polymerase .....	32
Table 15: Positions of additionally introduced cysteins .....	55

# Abbreviations

Abbreviation	Description
°C	degree Celsius
μ	micro
μL	microlitre
μM	micromolar
2-ME	2-Mercaptoethanol
5-FOA	5-Fluoroorotic acid
A	amplitude
ADP	adenosindiphosphate
ADPNP	adenylyl-imidodiphosphat
ATP	adenosintriphosphate
BHQ	black hole quencher
BisTris	Bis(2-hydroxyethyl) amino-tris(hydroxymethyl)methane
bp	base pair
c	concentration
C	cysteine
<i>C. thermophilum</i>	<i>Chaetomium thermophilum</i>
ca	circa
CBB	Coomassie Brilliant Blue
CTE	c-terminal end
Da	Dalton
DC	Duty Cycle
dest.	destillata
DNA	desoxyribonucleic acid
ds	double strand
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
etc	et cetera

EtOH	ethanol
FRET	Förster resonance energy transfer
<i>H. sapiens</i>	<i>Homo Sapiens</i>
HCl	hydrochloric acid
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonacid
His	histidine
IGEPAL	octylphenoxypolyethoxyethanol
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranosid
KCl	potassium chloride
kDa	Kilodalton
L	litre
LB-Medium	Lysogeny Broth Media
LDH	lactate dehydrogenase
mA	Milliampere
MES	2-(N-Morpholino)ethansulpho acid
min	minute
mL	millilitre
mM	millimolar
MOPS	3-(N-Morpholino) propansulfon acid
mRNA	messenger ribonucleic acid
mRNP	messenger ribonucleoprotein particle
NADH	Nicotinamide adenine dinucleotide
nm	nanometre
nM	nanomolar
NP-40	nonylphenoxypolyethoxylethanol
nt	nucleotide
NTE	n-terminal end
OC	Output Control
OCDase	orotidine 5'-phosphate decarboxylase
OD <sub>600</sub>	optical density at 600 nanometres
PAGE	polyacrylamide gel electrophoresis

PCR	polymerase chain reaction
PDB	protein databank
PEP	phosphoenolpyruvate
pH	$\log_{10} \alpha (H^+)$
PK	pyruvate kinase
PMSF	phenylmethylsulfonylfluoride
r av	average radius
r max	maximum radius
RBD	RNA binding domain
RNA	ribonucleic acid
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SAP	SAF-A/B, Acinus and PIAS associated domain
SE	Standard error
sec	seconds
SOB	super optimal broth
SOC	super Optimal broth with Catabolite repression media
t	time
TCEP	tris(2-carboxyethyl) phosphine
TEV	Tobacco Etch Virus
TREX	transport export complex
Tris	Tris(hydroxymethyl)-aminomethan
U	unit
WA	Walker A
WB	Walker B
WT	wildtype
YPD	Yeast extract, peptone, dextrose

# Publications

**Tho1 interacts with the RNA-helicase Sub2 via its RNA-binding domain and stimulates R-Loop resolution**

Matthias Bastian Miosga, Francisca Amoah, Phillip Keil, Katja Sträßer, and Peter Friedhoff

2022 (manuscript in process)

# Eidesstattliche Erklärung

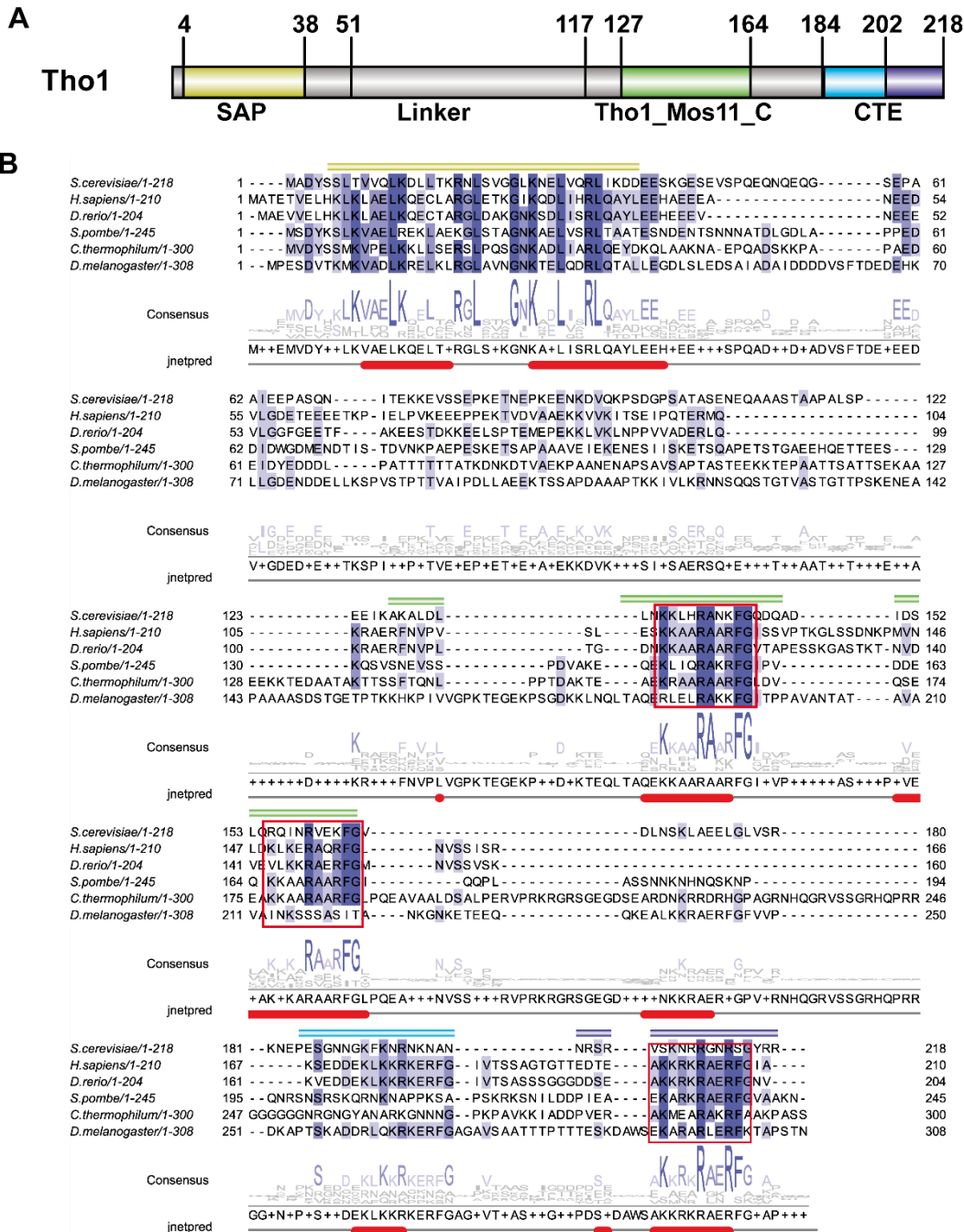
Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiats-Software zu. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe Ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

Gießen, den: .....      Unterschrift: .....

(Vorname, Nachname)



# Appendix



**Supplementary figure 1: Tho1/SARNP domain organization and sequence alignment.** (A) Schematic representation of the *S. cerevisiae* Tho1 is shown on top. Pfam domains are SAP (aa 4-38) and Tho1\_MOS11\_C (127-164) connected by a non-conserved, disordered linker. (B). Sequence alignment of Tho1 orthologues from *Saccharomyces cerevisiae* (P40400), *Homo sapiens* (P82979), *Danio rerio* (F1QT63), *Schizosaccharomyces pombe* (O74871), *Chaetomium thermophilum* (G0SH02) and *Drosophila melanogaster* (Q9VHC8). The domain of ScTho1 is shown and colored as in (A). The sequence alignment was JalView (Version: 2.11.1.7) (Waterhouse, 2009 Jalview) using MUSCLE (Edgar, 2004 Muscle). Conserved residues are highlighted with color, with darker shades of slate-blue indicating high conservation. The consensus sequence is shown below. Secondary structure prediction for *S. cerevisiae* Tho1 using the alignment was performed by Jpred (Drozdetskiy, 2015). Helices are marked as red tubes. Note, in addition to the conserved two helical motifs (boxed in red) of the Tho1\_Mos11\_C domain (Jacobsen, 2016), the C-terminal extension contains a third box.

# Quickguide

**Helicase reaction buffer**  
(helicase and ATPase assay)  
20 mM MES pH 6.5  
4 mM TCEP

**Binding buffer (EMSA)**  
10 mM MES pH 6.5  
2 mM TCEP  
5% glycerin

Standard substrates

Binding:

**Putnam 13 RNA**

ATPase:

**Putnam 13\_25 rev**

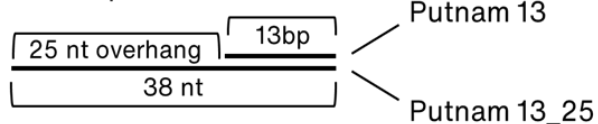
Helicase:

**Partial duplex Putnam**

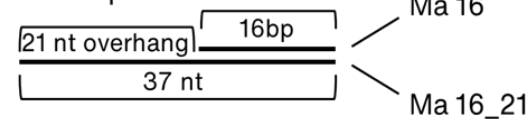
Annealing:

2x single strands Ma forming **partial duplex Ma**

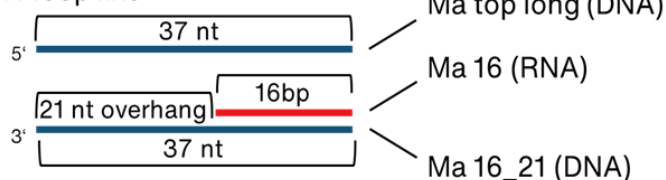
**Partial duplex Putnam**



**Partial duplex Ma**



**R-loop like**

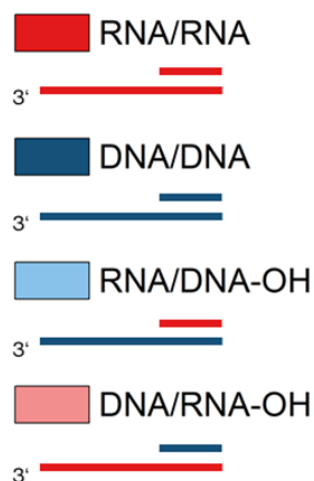



















**Sub2**



**Tho1**

Colorcode



Overview domain analysis	Binding activity (EMSA)	Annealing activity	Effect on Sub2 Helicase	Get enhanced by Sub2	Ternary complex formation
WT  Linker  CTE			Inhibition		
SAP 					
RBD 			Stimulation		
RBD+CTE  CTE					
ΔSAP Linker  CTE			Inhibition		
SAP+Linker  Linker					
Linker+RBD Linker 					
ΔCTE  Linker 					
ΔLinker  Linker  CTE			Inhibition		
ΔRBD  Linker CTE			Inhibition		
Δ 2nd half CTE  Linker  CTE					
Δ 1st half CTE  Linker  CTE					