JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Structural and functional insights into TREX complex interactions

---

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
– Dr. rer. nat. –
am Fachbereich 08 für Biologie und Chemie
der Justus-Liebig-Universität Gießen

vorgelegt von
Carina Kern

Gießen, 2025

Die vorliegende Arbeit wurde am Institut für Biochemie (Fachbereich 08: Biologie und Chemie) der Justus-Liebig-Universität Gießen unter der Leitung von Prof. Dr. Katja Sträßer erstellt.

Dissertation eingereicht: Februar 2025

Erstgutachter: Prof. Dr. Katja Sträßer
Fachbereich 08: Biologie und Chemie
Institut für Biochemie
Justus-Liebig-Universität Gießen

Zweitgutachter: Apl. Prof. Dr. Elena Evguenieva-Hackenberg
Fachbereich 08: Biologie und Chemie
Institut für Mikrobiologie und Molekularbiologie
Justus-Liebig-Universität Gießen

Table of Contents

Abstract VII

Abstract (German version) VIII

List of Figures X

List of Tables XII

List of Abbreviations XIII

1. Introduction 1

1.1. Gene expression 1

1.2. Biogenesis of mRNA 1

1.2.1. The synthesis of mRNA 2

1.2.2. Processing of transcribed mRNA 6

1.2.3. Assembly of mRNPs 9

1.2.4. Export of mRNA into the cytoplasm 12

1.3. The TREX complex 15

1.3.1. TREX homologs across selected species 15

1.3.2. Functional relevance of the TREX components 17

1.3.3. Recruitment of the TREX complex to active genes 19

1.3.4. Functions of TREX in mRNA export via Mex67-Mtr2 21

1.3.5. Structural characteristics of the TREX complex 23

1.4. The TREX interacting factor Tho1 24

1.5. Aims and objectives 25

2. Material 27

2.1. Instruments 27

2.2 Chemicals 28

2.3. Buffers and solutions 31

2.4. DNA and protein standards 34

2.5. Oligonucleotides 34

2.6. Kits 37

2.7. Enzymes 37

2.8. Antibodies 38

2.9. Plasmids 38

2.10. Strains 41
2.11. Media 42

3. Methods 43

3.1. Microbiological techniques used for *E. coli* 43
3.1.1. Preparation of Z-competent *E. coli* cells 43
3.1.2. Transformation of *E. coli* cells 43
3.1.3. Preparation of plasmid DNA 43
3.1.4. Protein expression in *E. coli* 44
3.2. Microbiological techniques used for *S. cerevisiae* 44
3.2.1. Preparation of competent *S. cerevisiae* cells 44
3.2.2. Transformation of *S. cerevisiae* cells 44
3.2.3. Treatment of *S. cerevisiae* cells with thiolutin 45
3.2.4. Dot spot assay 45
3.3. Basic molecular biology techniques 46
3.3.1. Polymerase chain reaction (PCR) 46
3.3.2. Screening of *E. coli* clones 46
3.3.3. Screening of *S. cerevisiae* clones 46
3.3.4. Nucleic acid electrophoresis 47
3.3.5. Plasmid cloning 47
3.3.6. Sequencing of plasmid DNA 48
3.4. Basic biochemical techniques 48
3.4.1. SDS-PAGE and Coomassie G-250 staining 48
3.4.2. Western blot 48
3.4.3. Protein identification by mass spectrometry 49
3.4.4. Alkaline lysis of yeast cells 49
3.4.5. Precipitation of proteins 49
3.4.6. Precipitation of DNA 49
3.4.7. Isolation of RNA 49
3.4.8. Quantitative reverse transcription PCR (RT-qPCR) 50
3.5. Protein purification techniques 51
3.5.1. Affinity purification of recombinant proteins 51
3.5.2. Ion exchange chromatography of recombinant proteins 51
3.5.3. Size exclusion chromatography of recombinant proteins 52
3.5.4. Purification of co-expressed S-Y-T complex 53
3.5.5. Purification of *in vitro* reconstituted protein complexes 53
3.5.6. RNA pull-down 54
3.5.7. Tandem Affinity Purification (TAP) of yeast TREX complex 55
3.5.8. Cross-linking of purified yeast TREX complex 55

3.6. Structural studies of purified proteins and complexes 56
3.6.1. Circular Dichroism Spectroscopy 56
3.6.2. Limited proteolysis with trypsin 56
3.6.3. Crystallization of recombinant proteins 56
3.6.4. Cross-linking mass spectrometry (XL-MS) of yeast TREX complex 57
3.6.5. Structure determination by negative stain EM 57
3.7. Functional studies of purified proteins and complexes 57
3.7.1. Electrophoretic Mobility Shift Assay (EMSA) 57
3.7.2. Fluorescence Resonance Energy Transfer (FRET) assay 58

4. Results 59

4.1. The TREX complex in *S. cerevisiae* 59
4.1.1. Preparation and cross-linking of the TREX complex 59
4.1.2. Cross-linking mass spectrometry of the yeast TREX complex 60
4.1.2.1. Compatibility of cross-links with recombinant yeast TREX structures 62
4.1.2.2. XL-MS data comparison of endogenous and recombinant THO/TREX 63
4.1.3. Electron microscopy of the yeast TREX complex 64
4.1.3.1. Structure determination of BS^3 cross-linked TREX complex 64
4.1.3.2. EM of ADH-DMTMM cross-linked TREX complex 66
4.1.3.3. EM of uncross-linked TREX complex 67
4.1.4. The TREX subunit Tho2 68
4.1.4.1. The unstructured C-terminal domain of Tho2 69
4.1.4.2. Role of Tho2 in mRNA stability 70
4.2. The Sub2-Yra1-Tho1 complex from *C. thermophilum* 71
4.2.1. The TREX subunit Sub2 in *C. thermophilum* (*ct*Sub2) 71
4.2.1.1. Purification of *ct*Sub2 72
4.2.1.2. Crystal structure of *ct*Sub2 73
4.2.2. The Tho1 protein in *C. thermophilum* (*ct*Tho1) 75
4.2.2.1. Purification of *ct*Tho1 75
4.2.2.2. Structural analysis of *ct*Tho1 76
4.2.2.3. Generation and purification of *ct*Tho1 deletion mutants 77
4.2.2.4. Secondary structure analysis of *ct*Tho1 mutants 79
4.2.2.5. DNA- and RNA-binding of *ct*Tho1 mutants 81
4.2.2.6. Annealing activity of *ct*Tho1 mutants 82
4.2.2.7. Interaction of *ct*Tho1 mutants with *ct*Sub2 83
4.2.2.8. Overview of functional analyses in *ct*Tho1 mutants 85
4.2.3. The reconstituted *ct*Sub2-*ct*Tho1 complex 85

4.2.4. The RNA annealing protein Yra1 in *C. thermophilum* (*ct*Yra1) 87
4.2.4.1. Purification of *ct*Yra1 88
4.2.4.2. *In vitro* reconstitutions with *ct*Yra1 89
4.2.5. Structure of the *ct*Sub2-*ct*Yra1-*ct*Tho1 (S-Y-T) complex 90
4.2.5.1. Co-expression and purification of different S-Y-T complexes 90
4.2.5.2. Nuclease treatment of the S-Y-T complex 92
4.2.5.3. Binding of the S-Y-T complex to a specific RNA 93
4.2.5.4. Stability of the S-Y-T complex 94
4.2.5.5. Limited proteolysis of the S-Y-T complex 97
4.2.5.6. Structural analysis of the S-Y-T complex by cryo-EM 97

5. Discussion 99

5.1. The endogenous yeast TREX complex 99
5.1.1. Successful purification of the yeast TREX complex 99
5.1.2. Structural insights into the yeast TREX complex by XL-MS 99
5.1.3. Compatibility of cross-links with recombinant yeast TREX complexes 101
5.1.4. Electron microscopy of the yeast TREX complex 102
5.1.5. Functional analysis of flexible regions in TREX components 103
5.1.6. Role of Tho2 in mRNA stability 104
5.2. The Sub2-Yra1-Tho1 (S-Y-T) complex from *C. thermophilum* 105
5.2.1. The structure of Sub2 from *C. thermophilum* (*ct*Sub2) 105
5.2.2. The role of *ct*Tho1 in mRNA biogenesis 106
5.2.3. *In vitro* interactions of *ct*Sub2 with *ct*Tho1 and *ct*Yra1 108
5.2.4. Structural analysis of the *ct*Sub2-*ct*Yra1-*ct*Tho1 (S-Y-T) complex 109
5.3. Conclusions and outlook 111

References 114

Publications 138

Acknowledgments 139

Appendix 140

Eigenständigkeitserklärung 144

Abstract

Gene expression in eukaryotes involves mRNA synthesis, processing, and export from the nucleus to the cytoplasm. The highly conserved TREX (transcription/export) complex plays a crucial role in coupling these processes, ensuring the efficient transition of mRNA from transcription to translation (Sträßer et al. 2002). Since only limited research was available on the structural organization of the TREX complex, the first objective was to investigate the endogenous TREX complex structure from *Saccharomyces cerevisiae*. The second objective was to analyze the structure of the Sub2-Yra1-Tho1 subcomplex from *Chaetomium thermophilum*. Tho1 has been suggested to function in transcription, splicing, export, and/or translation (Aravind & Koonin 2000; Hashii et al. 2004; Leaw et al. 2004). In addition to the structural objectives, this study also aimed to gain information of Tho1's functions in *C. thermophilum*.

This research showed the predominantly dimeric arrangement of the complete endogenous *S. cerevisiae* TREX complex, determined using XL-MS and EM. While the flexible C-terminus of yeast Tho2 is not essential for the formation of TREX, it is critical for cellular functions, particularly under stress conditions. Co-expressing the *C. thermophilum* proteins *ct*Sub2, *ct*Tho1, and *ct*Yra1 in *Escherichia coli* enabled the purification of a stable *ct*Sub2-*ct*Tho1-*ct*Yra1 (S-Y-T) complex. Despite nuclease treatment during the purification, this complex retained its association with nucleic acids. Cryo-EM images of the nuclease-treated S-Y-T complex revealed larger complexes, approximately 20 nm in size, tightly bound to RNA. Especially the interaction between *ct*Sub2 and *ct*Tho1 was highly stable, which allowed the successful *in vitro* reconstitution of this protein-protein complex. Further *in vitro* reconstitution experiments with various *ct*Tho1 deletion mutants revealed that the C-terminal domain (CTD) of *ct*Tho1, which includes two DIM motifs, is crucial for interacting with *ct*Sub2. Although the C-terminal end (CTE) of *ct*Tho1, which contains an additional DIM motif, is not essential, it enhances the *ct*Sub2 interaction. Moreover, the flexible CTE carries out important functions of *ct*Tho1, such as binding and annealing different RNA and DNA oligos.

In conclusion, this study shows the predominantly dimeric organization of the endogenous yeast TREX complex and highlights the importance of structurally unresolved, flexible protein regions – such as the yeast Tho2 C-terminus or the ctTho1 CTE – in maintaining cellular functionality. Given that *ct*Tho1 has three DIM

motifs, the S-Y-T complex in *C. thermophilum* likely contains multiple *ct*Sub2 molecules. By interacting with the alpha-helical N- and C-box regions of *ct*Yra1, the *ct*Sub2 proteins are possibly compacted by several *ct*Yra1 proteins, resulting in a multiprotein S-Y-T complex.

Abstract (German version)

Die Genexpression in Eukaryonten umfasst die Synthese, die Prozessierung und den Export von mRNA aus dem Zellkern in das Zytoplasma. Der hochkonservierte TREX-Komplex (Transkription/Export) spielt eine entscheidende Rolle bei der Kopplung dieser Prozesse und gewährleistet den effizienten Übergang der mRNA von der Transkription zur Translation (Sträßer et al. 2002). Da die strukturelle Organisation des TREX-Komplexes nicht vollständig erforscht ist, bestand das erste Ziel darin, die Struktur des endogenen TREX-Komplexes aus *Saccharomyces cerevisiae* zu untersuchen. Das zweite Ziel war die Analyse der Struktur des Sub2-Yra1-Tho1 Subkomplexes aus *Chaetomium thermophilum*. Es wird angenommen, dass Tho1 bei der Transkription, dem Spleißen, dem Export und/oder der Translation eine Rolle spielt (Aravind & Koonin 2000; Hashii et al. 2004; Leaw et al. 2004). Zusätzlich zu den strukturellen Zielen sollten auch Informationen über die Funktionen von Tho1 in *C. thermophilum* gewonnen werden.

Es konnte die überwiegend dimere Anordnung des vollständigen endogenen TREX-Komplexes von *S. cerevisiae*, die mit XL-MS und EM bestimmt wurde, konnte gezeigt werden. Während der flexible C-Terminus von *S. cerevisiae* Tho2 für die Bildung von TREX nicht wesentlich ist, ist er für zelluläre Funktionen, insbesondere unter Stressbedingungen, entscheidend. Die Koexpression der *C. thermophilum* Proteine *ct*Sub2, *ct*Tho1 und *ct*Yra1 in *Escherichia coli* ermöglichte die Aufreinigung eines stabilen *ct*Sub2-*ct*Tho1-*ct*Yra1 (S-Y-T) Komplexes. Trotz Nuklease-Behandlung während der Reinigung behielt dieser Komplex seine Assoziation mit Nukleinsäuren bei. Kryo-EM-Bilder des Nuklease-behandelten S-Y-T-Komplexes zeigten größere Komplexe von etwa 20 nm, die eng an RNA gebunden sind. Insbesondere die Interaktion zwischen *ct*Sub2 und *ct*Tho1 war sehr stabil, was die erfolgreiche *In-vitro*-Rekonstitution dieses Protein-Protein-Komplexes ermöglichte. Weitere *In-vitro*-Rekonstitutionsexperimente mit verschiedenen *ct*Tho1-Deletionsmutanten zeigten, dass die C-terminale Domäne (CTD) von *ct*Tho1, die zwei DIM-Motive enthält, für die

Wechselwirkung mit *ct*Sub2 entscheidend ist. Obwohl das C-terminale Ende (CTE) von *ct*Tho1, das ein zusätzliches DIM-Motiv enthält, nicht entscheidend ist, verstärkt es die Wechselwirkung mit *ct*Sub2. Darüber hinaus erfüllt das flexible CTE wichtige Funktionen von *ct*Tho1, wie die Bindung und das Annealen verschiedener RNA- und DNA-Oligos.

Zusammenfassend zeigt diese Studie die überwiegend dimere Organisation des endogenen *S. cerevisiae* TREX-Komplexes und unterstreicht die Bedeutung strukturell ungelöster, flexibler Proteinregionen – wie dem C-Terminus von *S. cerevisiae* Tho2 oder dem CTE von *ct*Tho1 – für die Aufrechterhaltung der zellulären Funktionalität. Da *ct*Tho1 drei DIM-Motive besitzt, enthält der S-Y-T-Komplex in *C. thermophilum* wahrscheinlich mehrere *ct*Sub2-Moleküle. Durch die Interaktion mit den alpha-helicalen N- und C-Box-Regionen von *ct*Yra1 werden die *ct*Sub2-Proteine möglicherweise durch mehrere *ct*Yra1-Proteine zusammengepackt, was zu einem Multiprotein-S-Y-T-Komplex führt.

List of Figures

Figure 1: Schematic representation of gene expression based on the central dogma of molecular biology 1
Figure 2: After transcription, processing, and association with RNPs, the mRNP is exported to the cytoplasm 2
Figure 3: Phosphorylation of the CTD plays a crucial role in aiding the transcription cycle of RNAPII 4
Figure 4: Mechanism of RNAPII pausing and transition to productive elongation 4
Figure 5: The m7G cap structure at the 5’ end of RNA is generated in several enzymatic reactions 6
Figure 6: The spliceosome catalyzes the excision of introns and the ligation of exons in a two-step transesterification process 8
Figure 7: CPSF/CPF binds in the 3’ UTR of pre-mRNAs and facilitates 3’ end processing 9
Figure 8: The nuclear pore complex (NPC) facilitates the mRNA export from the nucleus to the cytoplasm 12
Figure 9: Recruitment of mRNA export factors and passage through the NPC in yeast 14
Figure 10: The recruitment of TREX to the active gene 20
Figure 11: TREX facilitates mRNA export by interacting with Mex67-Mtr2 22
Figure 12: High-resolution structures of the yeast and human THO-Sub2/UAP56 complex 23
Figure 13: Preparation of the endogenous TREX complex from *S. cerevisiae* 60
Figure 14: BS^3 and ADH-DMTMM cross-links of the TREX complex were mainly observed between the THO components 61
Figure 15: Overview of all observed BS^3 and ADH-DMTMM cross-links of endogenous yeast TREX complex 63
Figure 16: Comparison of inter-subunit cross-links 64
Figure 17: Negative-stain EM structure of the BS^3 cross-linked endogenous TREX complex 65
Figure 18: Negative-stain EM of the ADH-DMTMM cross-linked endogenous TREX complex 67
Figure 19: Negative-stain EM of the uncross-linked endogenous TREX complex 68
Figure 20: The unstructured C-terminal domain of Tho2 is not necessary for TREX complex formation but for growth at low and high temperatures 69

Figure 21: Overexpression of Tho2 leads to an increased half-life of transcripts 70
Figure 22: Sequence alignment of *H. sapiens* UAP56/DDX39B, *S. cerevisiae* Sub2 and *C. thermophilum* Sub2 72
Figure 23: The purification of *ct*Sub2 yielded high amounts of pure protein 72
Figure 24: The crystal structure of *ct*Sub2 displays the protein in the open conformation 74
Figure 25: Sequence alignment of *H. sapiens* SARNP, *S. cerevisiae* Tho1 and *C. thermophilum* Tho1 75
Figure 26: The purification of *ct*Tho1 yielded high amounts of pure protein 76
Figure 27: Structural comparison of the Tho1 proteins from *S. cerevisiae*, *H. sapiens* and *C. thermophilum* 77
Figure 28: Various *ct*Tho1 mutants were generated based on the structure prediction 78
Figure 29: The *ct*Tho1 mutants were purified in reasonably high and pure quantities 78
Figure 30: The secondary structure content of all *ct*Tho1 mutants aligns with the secondary structure prediction 80
Figure 31: The CTE is the most essential region of *ct*Tho1 for RNA- and DNA-binding 81
Figure 32: The CTE of *ct*Tho1 is necessary for annealing nucleic acids 82
Figure 33: The CTD of *ct*Tho1 is primarily required for the interaction with *ct*Sub2 84
Figure 34: The S-Y complex displays a stoichiometry where *ct*Sub2 is approximately twice as abundant as *ct*Tho1 87
Figure 35: Sequence alignment of *H. sapiens* ALYREF, *S. cerevisiae* Yra1 and *C. thermophilum* Yra1 88
Figure 36: The *ct*Yra1 proteins and mutants purified in reasonably high and pure quantities 89
Figure 37: The *in vitro* reconstitution of the *ct*Sub2-*ct*Yra1-(*ct*Tho1) complex is not efficient 90
Figure 38: Deletions in *ct*Tho1 result in alterations in S-Y-T stoichiometry and nucleic acid content during co-expression 91
Figure 39: Nuclease treatment of S-Y-T improves protein yield without compromising the stoichiometry or integrity of the complex 93
Figure 40: The nuclease treated S-Y-T complex can bind to a specific RNA introduced to it 94
Figure 41: The S-Y-T complex remains stable during short-term storage at 4°C and long-term storage at -80°C 95
Figure 42: The S-Y-T complex retains its stability until *ct*Yra1 dissociates at a salt concentration of 250 mM 96

Figure 43: Limited proteolysis of S-Y-T shows that *ct*Sub2 is the most stable and structured component of the complex 97
Figure 44: The cryo-EM analysis of S-Y-T reveals RNA bound complexes with a size of 20 nm 98
Figure 45: The positive charge of the *ct*Tho1 CTE most likely facilitates the interaction with nucleic acids 107
Figure 46: The yeast Tho1 CTD forms a complex with two UAP56/DDX39 molecules 108
Figure 47: Model illustrating the potential role of the *ct*Sub2-*ct*Yra1-*ct*Tho1 complex in mRNA assembly 113
Figure S1: Data processing workflow showing the image processing procedure with the number of particles and the resolution of the 3D reconstruction 140
Figure S2: Verification of protein expression by Western blot 140
Figure S3: Overexpression of Tho2, Npl3 and Nab2 leads to an increased half-life of transcripts 141
Figure S4: *In vitro* reconstitutions and pull-down experiments of the remaining *ct*Tho1 mutants 142
Figure S5: Negative-stain EM of the recombinant *ct*THO complex 143

List of Tables

Table 1: Components of the *S. cerevisiae* TREX complex and their homologs in *H. sapiens* and *C. thermophilum* 16
Table 2: Used instruments and the respective manufacturer 27
Table 3: Chemicals and their respective supplier 28
Table 4: Used buffers and solutions sorted by their purpose 31
Table 5: Oligonucleotides used for colony PCR and/or sequencing 34
Table 6: Oligonucleotides used for cloning 34
Table 7: Oligonucleotides used for qPCR 36
Table 8: Oligonucleotides used for EMSA, FRET annealing assay and RNA pull-down 36
Table 9: Used kits and their suppliers 37
Table 10: All enzymes and the purpose they were used for 37
Table 11: Primary antibodies used for Western blot 38
Table 12: Secondary antibodies used for Western blot 38

Table 13: Plasmids employed for expressing *ct*Sub2 recombinantly in *E. coli* 39
Table 14: Plasmids employed for expressing *ct*Tho1 recombinantly in *E. coli* 39
Table 15: Plasmids employed for expressing *ct*Yra1 recombinantly in *E. coli* 39
Table 16: Plasmids employed for co-expressing different complexes recombinantly in *E. coli* 40
Table 17: Plasmids employed for expressing or cloning different genes in *S. cerevisiae* 41
Table 18: *Escherichia coli* strains used for general cloning and recombinant protein expression 41
Table 19: *Saccharomyces cerevisiae* strains used in this study 41
Table 20: Used media and their composition 42
Table 21: The isoelectric points of all purified recombinant proteins 52
Table 22: Summary of the functional characteristics of *ct*Tho1 mutants 85

List of Abbreviations

APS	ammonium persulfate
C-	carboxy-terminal
ct	*Chaetomium thermophilum*
CTD	carboxy-terminal domain
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
E	elution
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
f1 ori	f1 bacteriophage origin of replication
FT	flow through
FTpA	FLAG-TEV-2xProteinA-tag
GSH	glutathione
GST	glutathione S-transferase
His_7	7x histidine-tag

HTpA	His_7-TEV-2xProteinA-tag
IgG	immunoglobulin G
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISO	isothermal reaction buffer
KanR	kanamycin resistance
LB	Luria-Bertani medium
m7G	7-methylguanosine
mRNP	messenger ribonucleoprotein ≥
N-	amino-terminal
NAD^+	β-Nicotinamide adenine dinucleotide
NMD	nonsense-mediated mRNA decay
NPC	nuclear pore complex
OD	optical density
ori	origin of replication
PAGE	polyacrylamide gel electrophoresis
PAS	polyadenylation signal
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
poly(A)	polyadenosine
(pre-)mRNA	(precursor) messenger ribonucleic acid
ProtA	Protein A
RBS	ribosome binding site
RNA	ribonucleic acid
RNAPII	RNA polymerase II
RRM	RNA recognition motif
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
SR	serine–arginine-rich
STE	sodium chloride-Tris-EDTA buffer
S-Y-T	*ct*Sub2-*ct*Yra1-*ct*Tho1
TAE	Tris-acetate-EDTA buffer
TAP	tandem affinity purification
TCA	trichloroacetic acid
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus protease recognition and cleavage site

TPR	tetratrico peptide repeat
TREX	transcription/export
UBA	ubiquitin associated
XL-MS	cross-linking mass-spectrometry

1. Introduction

1.1. Gene expression

Essentially, the central dogma of molecular biology describes the process of gene expression (Crick 1970; Fig. 1). This highly regulated, multi-step process begins with the replication of DNA in the nucleus, where the genetic material is duplicated to ensure the accurate transmission of genetic information during cell division (O'Donnell et al. 2013). Following DNA replication, specific genes are transcribed into precursor messenger RNAs (pre-mRNAs) by RNA polymerase II (RNAPII). Transcription is the first step in converting genetic information into functional molecules, such as proteins (Clancy & Brown 2008). Pre-mRNA is processed co-transcriptionally and by interacting with several nuclear RNA binding proteins (RBPs) a mature messenger ribonucleoprotein particle (mRNP) is assembled which is then exported to the cytoplasm for translation (Meinel & Strasser 2015; Singh et al. 2015; Wegener & Müller-McNicoll 2019; Wende et al. 2019; Khong & Parker 2020). During translation, the mRNA sequence is read by ribosomes, and the corresponding amino acids are assembled into proteins with various cellular functions, ranging from structural roles to enzymatic activities (Clancy & Brown 2008).

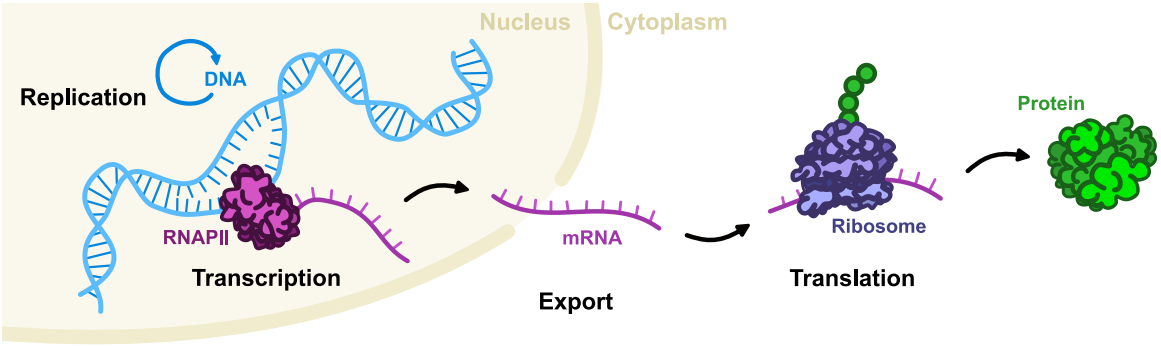


Figure 1: Schematic representation of gene expression based on the central dogma of molecular biology. The process begins with DNA replication in the nucleus, where the genetic material is duplicated. RNA polymerase II (RNAPII) then transcribes specific genes from DNA into mRNA. The mRNA is exported from the nucleus to the cytoplasm. In the cytoplasm, the ribosome translates the mRNA into a functional protein.

1.2. Biogenesis of mRNA

The biogenesis of mRNA is a fundamental process in gene expression. In eukaryotic cells it begins in the nucleus with the synthesis of pre-mRNA by RNAPII. Pre-mRNA processing takes place co-transcriptionally and includes capping, splicing, and the

formation of the 3' end. Simultaneously, the mRNA interacts with several nuclear RBPs resulting in the assembly of an mRNP (Meinel & Strasser 2015; Singh et al. 2015; Wegener & Müller-McNicoll 2019; Wende et al. 2019; Khong & Parker 2020). Finally, the mature mRNP is exported to the cytoplasm (Fig. 2).

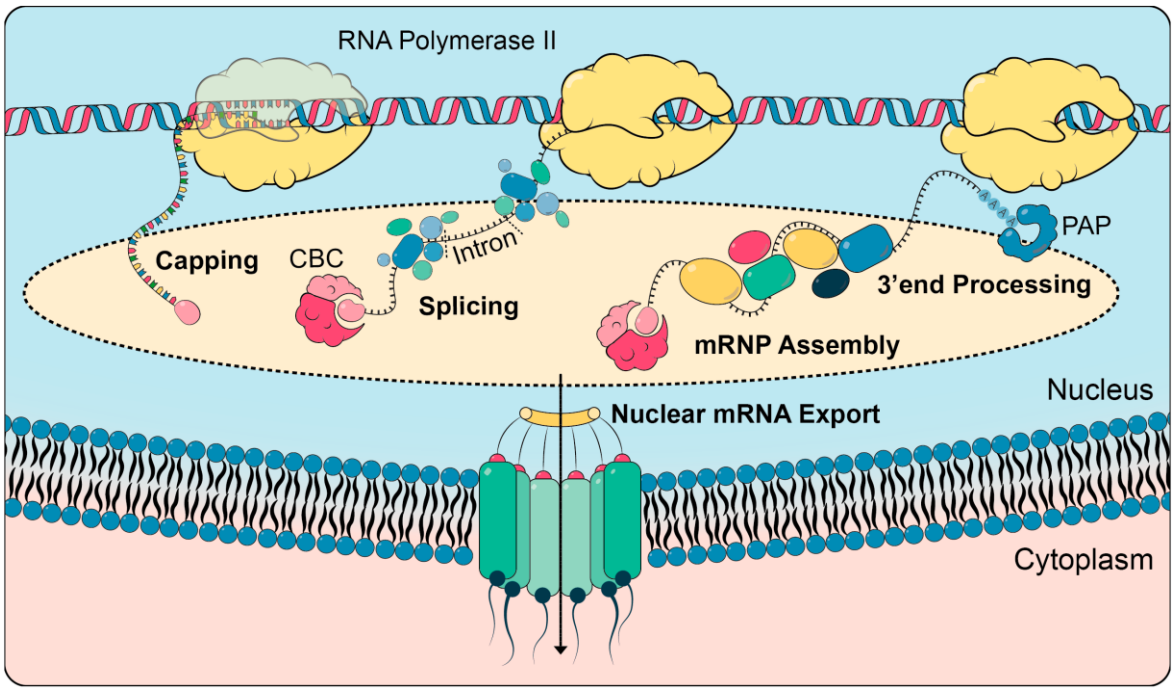


Figure 2: After transcription, processing, and assembly with RNPs, the mRNP is exported to the cytoplasm. RNA polymerase II is responsible for the transcription of protein-coding genes and the synthesis of mRNA. Subsequently, the mRNA undergoes a series of modifications: (1) a cap is added to its 5' end for binding the cap-binding complex (CBC), (2) the spliceosome removes introns, and (3) at the 3' end, the mRNA is cleaved, and a poly(A) tail is attached by poly(A) polymerase (PAP). Simultaneously, nuclear mRNA binding proteins bind to the mRNA, forming an mRNP complex, which is finally exported from the nucleus to the cytoplasm. Figure by Zarnack et al. 2020, licensed under the Creative Commons Attribution (CC BY) license.

Throughout the biogenesis of mRNA, multiple quality control mechanisms ensure that only correctly processed and functional mRNA molecules are translated. For instance, the nuclear exosome can degrade aberrant pre-mRNAs with incomplete splicing. These and other quality control processes maintain the integrity of gene expression and prevent the synthesis of faulty or harmful proteins (Hilleren & Parker, 2003; Mitchell & Parker 2014).

1.2.1. The synthesis of mRNA

The expression of protein-coding genes begins when genetic information encoded in DNA is transcribed into mRNA by RNAPII. Transcription is a highly regulated process that involves distinct stages – initiation, elongation, and termination. Each

of these stages is regulated by a complex network of transcription factors and regulatory elements (Casamassimi & Ciccodicola 2019).

For transcription initiation, the preinitiation complex (PIC) is assembled at the core promoter region of a gene. The PIC consists of RNAPII, along with the general transcription factors (GTFs) TBP, TFIIB, TFIIE, TFIIF, and TFIIH (Hampsey 1998). Transcription starts in a specific region at the 5' end of a gene, called the transcription start site (TSS). The TSS is situated within the core promoter region, a brief sequence spanning approximately 50 base pairs upstream and 50 base pairs downstream of the TSS (Haberle & Stark 2018). This region typically contains key sequence elements, which serve as binding sites for GTFs and RNAPII, and regulatory elements, which can either enhance or inhibit transcription in a gene-specific fashion. Eukaryotic core promoters comprise various DNA sequence elements, such as the TATA box, initiator (Inr) elements, and the downstream core promoter element (DPE). However, these elements are not universally present across all promoters. For instance, while many promoters contain a TATA box, a significant number are TATA-less (Hampsey 1998; Chen & Xu 2022).

The C-terminal domain (CTD) of Rpb1, the largest subunit of RNAPII, plays a crucial role in the transcription cycle and co-transcriptional mRNA processing. The CTD consists of multiple repeats of a heptapeptide sequence (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) and acts as a platform for the coordinated recruitment of the transcriptional machinery and mRNA processing factors. Prior to transcription initiation, the CTD is largely hypo-phosphorylated, allowing for the stable association of RNAPII with the promoter. After PIC assembly, the DNA is unwound around the TSS. Initial transcription then results in the abortive synthesis of a short 2–3 nucleotide RNA. Promoter escape occurs soon after transcription initiation and is marked by the release of most GTFs and TFIIH-mediated phosphorylation of Serine 5 (S_5) and Serine 7 (S_7) in the CTD of Rpb1 (Wade & Struhl 2008; Rodríguez-Molina et al. 2023; Fig. 3).

In some eukaryotes, RNAPII is paused after synthesizing 20–50 nucleotides of RNA (Strobl & Eick 1992; Kwak & Lis 2013; Gajos et al. 2021). Promoter-proximal pausing of RNAPII is a common regulatory mechanism that, for instance, helps facilitate efficient capping of nascent transcripts. The pause occurs after RNAPII clears the promoter and recruits negative elongation factors (Whelan & Pelchat

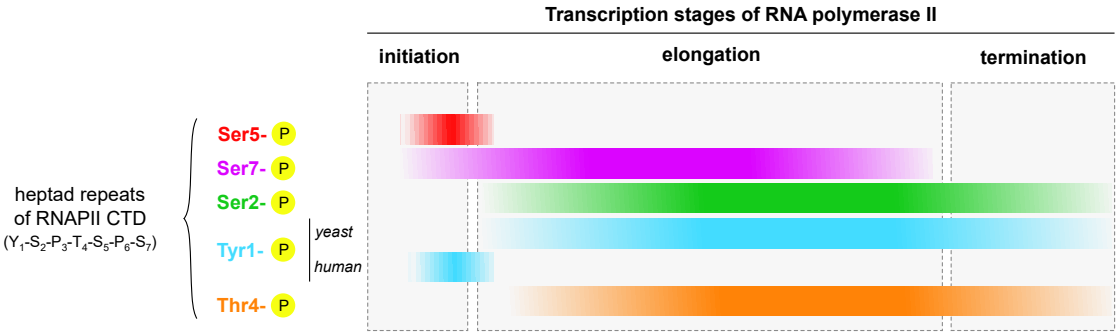


Figure 3: Phosphorylation of the CTD plays a crucial role in aiding the transcription cycle of RNAPII. Transcription progresses through three distinct stages: initiation, elongation, and termination. The C-terminal domain (CTD) of the largest RNAPII subunit, Rpb1, undergoes varying phosphorylation states during the transcription cycle to enable the recruitment and replacement of stage-specific factors. The patterns of phosphorylation are represented by colored bars. Figure adapted from Rodríguez-Molina et al. 2023, licensed under the Creative Commons CC BY license.

2022). The DRB Sensitivity-Inducing Factor (DSIF), composed of the subunits Spt4 and Spt5, functions as a positive transcription elongation factor and interacts with the Negative Elongation Factor (NELF) (Yamaguchi et al. 2002). The RNAPII processivity is slowed through the interaction with DSIF, while full pausing occurs when NELF (consisting of NELF-A, -B, -C/D, and -E) binds to both RNAPII and the nascent transcript (Whelan & Pelchat 2022; Fig. 4). The positive transcription elongation factor b (P-TEFb), initially held in an inactive state within the 7SK snRNP complex, becomes activated through its interaction with either BRD4 or SEC. Once activated, P-TEFb is recruited to the paused RNAPII, where its kinase subunit CDK9

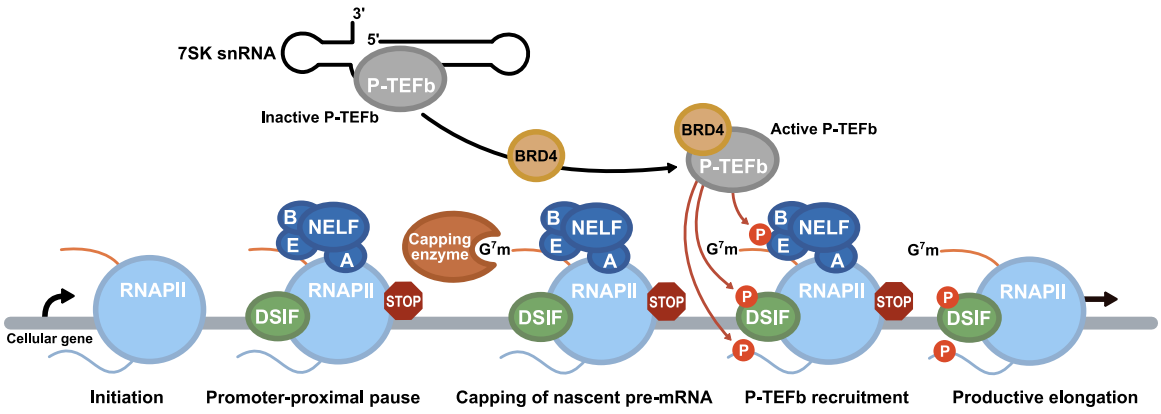


Figure 4: Mechanism of RNAPII pausing and transition to productive elongation. After transcription initiation, RNAPII pauses as DSIF and NELF bind. P-TEFb activation via BRD4 or SEC releases it from the 7SK snRNP complex. The active P-TEFb complex targets RNAPII, where CDK9 phosphorylates Spt5, NELF-E, and S2 of RNAPII CTD. This phosphorylation dissociates NELF, converts DSIF into a positive elongation factor, and enables productive elongation. Figure taken from Whelan & Pelchat 2022, licensed under the Creative Commons Attribution (CC BY) license.

phosphorylates Spt5, NELF-E, and the RNAPII CTD at Serine 2 (S_2). This triggers the release of NELF, converts DSIF into a positive elongation factor, and allows transcription to resume (Fujinaga et al. 2023; Whelan & Pelchat 2022; Fig. 4).

At the transition from RNAPII pausing to productive elongation, S_5 of the RNAPII CTD is dephosphorylated by the Ssu72 phosphatase (Rodríguez-Molina et al. 2023; Fig. 3). As transcription elongation progresses, the phosphorylated S_2 facilitates the recruitment of splicing factors and several elongation factors like Paf1, Spt4, and Dst1 (Harlen et al. 2016; Couvillion et al. 2022). When RNAPII approaches the 3' end of the gene, phosphorylated S_2 recruits termination factors and the machinery needed for 3' end processing (Rodríguez-Molina et al. 2023). As RNAPII transcribes the Polyadenylation Signal (PAS) in yeast, Y_1 is dephosphorylated by Glc7 phosphatase and its regulatory subunit Ref2. This allows phosphorylated S_2 to facilitate the recruitment of processing and termination factors (Rodríguez-Molina et al. 2023).

Initially, two main models were proposed to explain transcription termination and RNAPII release. The allosteric model (Logan et al. 1987) proposes that transcription of cleavage and termination signals leads to structural changes in RNAPII, which cause its detachment from the DNA. This process is likely supported by termination factors that weaken the interaction between RNAPII and the DNA. In contrast, the torpedo model (Connelly & Manley 1988) proposes that the cleavage of the nascent transcript generates a 5’ end that is then degraded by a 5’-3’ exoribonuclease. When the exoribonuclease encounters RNAPII, it induces the release of RNAPII from the DNA. A more recent model merges the allosteric and torpedo models, emphasizing the combined roles of termination factors, specific DNA sequences in the 3’ end regions, and RNAPII modifications in promoting the release of RNAPII from DNA (Luo et al. 2006; Eaton et al. 2020; Lopez Martinez & Svejstrup 2024).

In conclusion, mRNA synthesis is a coordinated process that includes initiation, elongation, and termination, governed by various protein complexes and transcription factors. The interactions between RNAPII, its CTD, and regulatory elements throughout these stages ensure accurate mRNA processing and the production of mature transcripts.

1.2.2. Processing of transcribed mRNA

Pre-mRNAs undergo three processing events that occur either simultaneously with or after transcription. These processes include capping, splicing, and polyadenylation and are essential for converting pre-mRNA into mature mRNA, which can then be translated into protein. Each of these steps are important in ensuring the stability and functionality of the final mRNA product (Carmody & Wente 2009).

The first processing step occurs when the pre-mRNA is just 20-30 nucleotides long. During this phase, a 7-methylguanosine (m7G) cap structure is attached to the 5' end (Shatkin & Manley 2000; Decroly et al. 2011; Fig. 5), which protects the mRNA against 5'-3' degradation (Bousquet-Antonelli et al. 2000). The capping process begins with RNA triphosphatase (RTPase) removing the γ-phosphate from

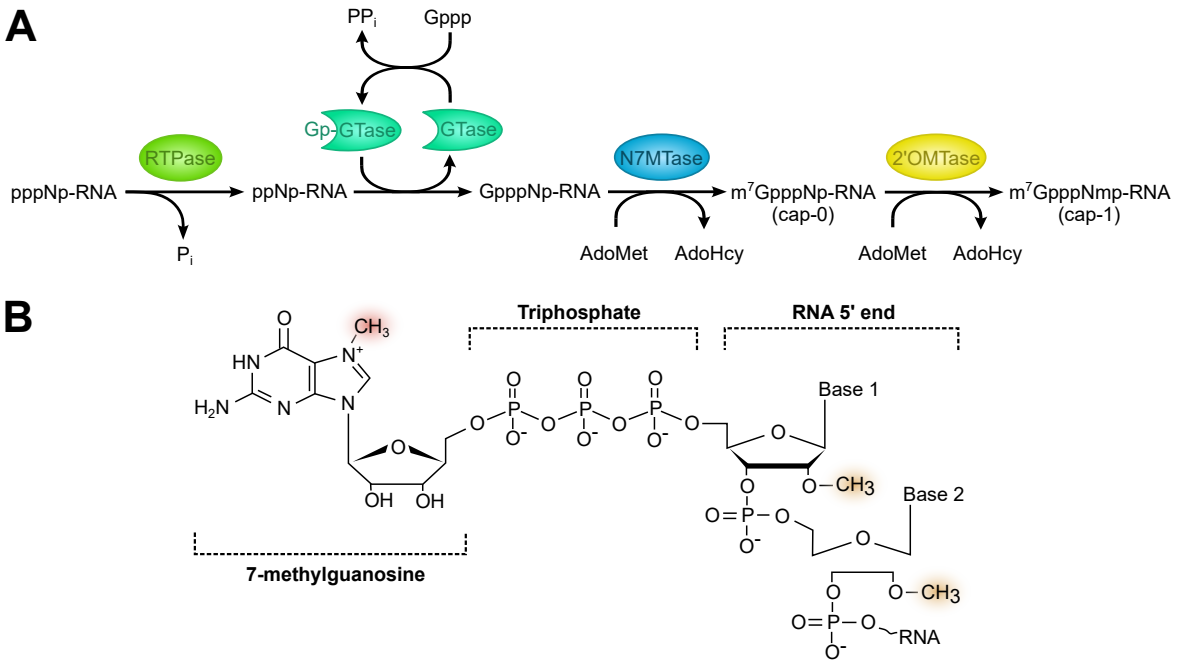


Figure 5: The m7G cap structure at the 5' end of RNA is generated in several enzymatic reactions. *(A)* RNA triphosphatase (RTPase) initiates the removal of the γ-phosphate from the 5' triphosphate, leading to the formation of a diphosphate 5' end and the release of inorganic phosphate. By using a GTP (Gppp) molecule, the guanylyltransferase (GTase) creates a covalent enzyme–guanylate intermediate (Gp–GTase). The GTase transfers a 5'-phosphoguanosine (GMP) to the 5' diphosphate of the RNA. Then, (guanine-N7)-methyltransferase (N7MTase) transfers the methyl group from S-adenosyl-L-methionine (AdoMet) to the cap guanine, resulting in the cap-0 structure, 7-methyl-GpppNp (m7GpppNp), and releasing S-adenosyl-L-homocysteine (AdoHcy) as a by-product. The capping process concludes with the methylation of the ribose-2'-O position of the initial nucleotide by the AdoMet-dependent (nucleoside-2'-O)-methyltransferase (2'OMTase), resulting in the formation of the cap-1 structure (m7GpppNm2'-Op). *(B)* The mRNA cap comprises a 7-methylguanosine connected to the 5' nucleoside of the mRNA chain via a 5'–5' triphosphate bridge. The red shading indicates the methyl group at the N7 position of guanosine, while the orange shading represents the 2'-O-methyl groups of the first and second nucleotide residues, forming the cap-1 and cap-2 structures, respectively. Adapted from Decroly et al. 2011, reproduced with permission from Springer Nature (license number: 5963580244100).

the 5’ triphosphate, resulting in a diphosphate 5’ end. Guanylyltransferase (GTase) then utilizes a GTP molecule to add a guanylate cap to the RNA, forming a covalent enzyme–guanylate intermediate. Finally, methylation occurs in two steps: first by (guanine-N7)-methyltransferase (N7MTase) to create the cap-0 structure, followed by (nucleoside-2’-O)-methyltransferase (2’OMTase) methylating the ribose-2’-O position of the initial nucleotide to produce the cap-1 structure (Decroly et al. 2011; Fig. 5). Once the m7G cap has been synthesized, it is bound by the nuclear cap-binding complex (CBC). The complex is conserved from yeast to humans, comprising Cbp20 and Cbp80 in *Saccharomyces cerevisiae*, and NCBP1 and NCBP2 in humans (Zarnack et al. 2020).

Following 5' capping, the transcript undergoes splicing, where non-coding introns are removed, and protein-coding exons are joined together. Splicing involves a two-step transesterification process, catalyzed by the spliceosome (Fig. 6). This dynamic ribonucleoprotein (RNP) complex consists of five snRNAs (U1, U2, U4, U5, and U6), along with numerous associated proteins (Will & Lührmann 2011; Herzel et al. 2017). The assembly of the spliceosome begins with U1 binding to the 5' splice site, along with the Msl5-Mud2 (or SF1-U2AF in humans) heterodimer binding to the branch point. This process leads to the formation of the commitment complex (or early/E complex in mammalians) (Colot et al. 1996). The interaction between Msl5 and the branch point is subsequently substituted with U2, resulting in the formation of the prespliceosome (A-complex). Then, the tri-snRNP containing U4, U5, and U6, joins and the precatalytic B complex is formed. Following multiple structural changes, protein exchanges, and the elimination of the U1 and U4 snRNPs, the active spliceosome (B* complex) becomes capable of initiating the first transesterification splicing reaction (Wahl et al. 2009). In this initial step of the splicing process, the 2'-OH group of the branch adenosine residue initiates a nucleophilic attack on the phosphodiester bond at the 5' splice site. This reaction releases the 5' exon and results in the creation of a lariat structure consisting of the intron and exon 2, connected through a 2'-5'-phosphodiester linkage (Ruskin et al. 1984). This reaction leads to the formation of the C complex, which then facilitates the second transesterification splicing reaction, resulting in the removal of the lariat intron structure and the joining of the exons through ligation (Konarska et al. 1985; Tseng & Cheng 2013; Fig. 6).

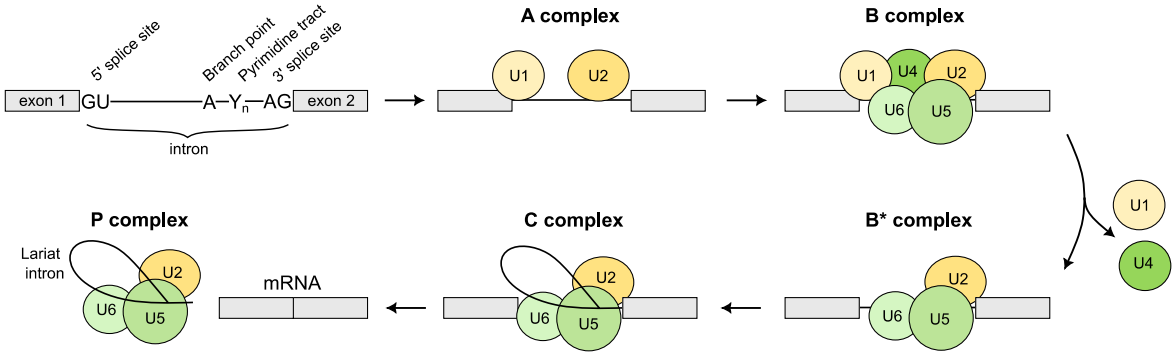


Figure 6: The spliceosome catalyzes the excision of introns and the ligation of exons in a two-step transesterification process. The A complex is established when U1 binds to the 5' splice site, and U2 attaches to the branch point. The A complex combines with U4, U5, and U6 to create the B complex. Eliminating U1 and U4 triggers the activation of the spliceosome (B* complex). Upon concluding the initial splicing step, the C complex takes forms. Following the second splicing step, the fully spliced mRNA product is liberated from the post-splicing (P) complex. Figure adapted from Gehring & Roignant 2021, licensed under the Creative Commons CC-BY-NC-ND license.

Once accurate splicing has been accomplished, the final stage of processing for the nascent mRNA involves the 3' end processing (Hocine et al. 2010; Fig. 7). Pre-mRNA 3' end processing requires only cleavage and polyadenylation as enzymatic activities. In the case of the mammalian 3' end processing machinery, over 14 proteins have been recognized, while the yeast 3' end processing machinery is associated with more than 20 proteins (Mandel et al. 2008). This machinery comprises the highly conserved cleavage and polyadenylation factor, referred to as CPF in yeast and CPSF in metazoans. CPSF/CPF interacts with RNAPII during transcription and consists of distinct subunits, including the endonuclease Ysh1/CPSF73, which is responsible for pre-mRNA cleavage, and the Pap1/PAP polymerase that adds the poly(A) tail. The poly(A) tail plays crucial roles in nuclear export, mRNA stability, and translation regulation. Additionally, CPSF/CPF contains two protein phosphatases, Ssu72/SSU72 and Glc7/PP1, that regulate transcription and 3' end processing (Casañal et al. 2017; Fig. 7). The 3' end cleavage and polyadenylation process are guided by specific sequence elements located within the 3’ untranslated region (3’ UTR) of the pre-mRNA. These cis elements are found in almost all eukaryotic pre-mRNAs that undergo polyadenylation. However, histone pre-mRNAs, although they undergo cleavage, lack the sequence elements required for polyadenylation (Mandel et al. 2008). The polyadenylation signal (PAS) sequence marks the sites for 3’ end processing and is located 10-30 nucleotides upstream of the cleavage site, often accompanied by auxiliary RNA motifs that

recruit cleavage factors. The PAS is bound directly by CPSF/CPF, with the consensus sequence AAUAAA showing the highest affinity for the processing complex in most eukaryotes (Kumar et al. 2019; Fig. 7). However, in humans, only around 50% of PAS sites have this canonical AAUAAA sequence. Noncanonical PAS sequences are bound less efficiently by CPSF/CPF, which likely influences the regulation of 3' end processing (Boreikaitė & Passmore 2023).

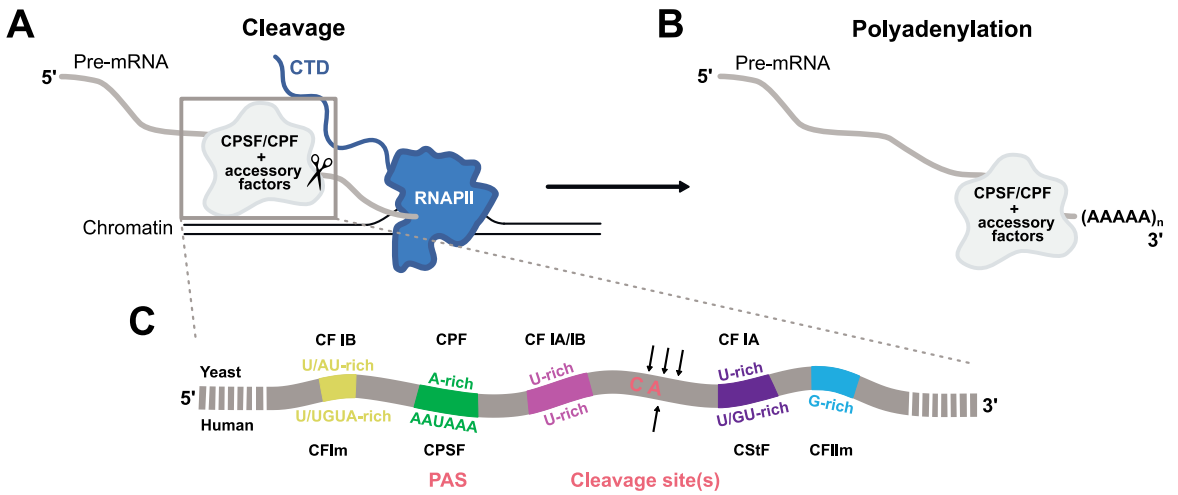


Figure 7: CPSF/CPF binds in the 3' UTR of pre-mRNAs and facilitates 3' end processing. CPSF/CPF contains the endonuclease Ysh1/CPSF73 for (A) pre-mRNA cleavage and Pap1/PAP polymerase for (B) adding the poly(A) tail, as well as two protein phosphatases, Ssu72/SSU72 and Glc7/PP1, that regulate transcription and 3' end processing. (C) The 3' end cleavage and polyadenylation are guided by specific sequence elements in the 3' untranslated region (3' UTR) of pre-mRNA. The polyadenylation signal (PAS) is located 10-30 nucleotides upstream of the cleavage site and indicates the sites for 3' end processing. CPSF/CPF directly binds the PAS, with the AAUAAA consensus sequence having the highest affinity for the processing complex in most eukaryotes. The protein factors that interact with each element in yeast (top) and humans (bottom) are shown. Figure taken from Boreikaitė & Passmore 2023, licensed under a Creative Commons Attribution 4.0 International License.

In summary, the processing of transcribed mRNA involves capping, splicing, and polyadenylation, which together transform pre-mRNA into a mature and functional mRNA. Processing not only enhances mRNA stability and translation efficiency but also is tightly regulated by specific protein complexes that ensure the accuracy and timing of each processing step.

1.2.3. Assembly of mRNPs

Already during synthesis and processing, mRNA is bound by various RBPs, resulting in the assembly of mRNPs. The assembly of an mRNP is a critical stage in the gene expression process as it stabilizes the mRNA and facilitates its export

from the nucleus to the cytoplasm. Moreover, the composition of mRNPs can influence subsequent stages of the mRNA life cycle, including mRNA localization, translation, and degradation within the cytoplasm (Müller-McNicoll & Neugebauer 2013; Gehring et al. 2017; Wende et al. 2019).

The initial binding of proteins to mRNAs begins as soon as the pre-mRNA is long enough to emerge from RNAPII (Gehring et al. 2017). The CBC is the first stable protein complex to associate with nascent mRNA once the m7G cap is added to its 5' end (see chapter 1.1.2.; Zarnack et al. 2020; Mars et al. 2021). The small CBC subunit Cbp20/NCBP1 directly associates with the cap structure and then proceeds to recruit the large subunit Cbp80/NCBP2 (Müller-McNicoll & Neugebauer 2014; Rambout & Maquat 2020). Together, they create the CBC heterodimer that functions as a central hub for various factors regulating nearly all aspects of gene expression. The CBC protects the mRNA from degradation and facilitates transcription elongation, splicing, 3' end formation, and nuclear export (Izaurralde et al. 1994; Flaherty et al. 1997; Hosoda et al. 2005; Cheng et al. 2006; Zarnack et al. 2020).

During splicing, additional proteins bind stably to the mRNA in a splicing-dependent manner. For instance, the exon junction complex (EJC) assembles on the mRNA approximately 20–24 nucleotides upstream of the spliced exon-exon junction. The DEAD-box RNA helicase eIF4A3 (DDX48) as well as the proteins MAGOH, RBM8A (Y14), and BTZ (also referred to as MLN51 or CASC3) are the core components of the EJC. This central core serves as a platform for binding additional factors associated with splicing, mRNA transport, translation, and the process of nonsense-mediated decay (NMD) (Saulière et al. 2012; Singh et al. 2012; Le Hir et al. 2016; Schlautmann & Gehring 2020; Zarnack et al. 2020).

In metazoans, the TREX complex is also likely recruited during splicing (Masuda et al. 2005). Originally, TREX was first identified in *S. cerevisiae*, where it couples transcription to mRNA export (see chapter 1.2.; Sträßer et al. 2002). The conserved TREX complex is recruited to the mRNP through various protein-protein and mRNA interactions and is essential in nuclear mRNP assembly and other stages of gene expression (Zarnack et al. 2020).

After 3' end processing with mRNA cleavage and polyadenylation (see chapter 1.1.2.), the poly(A) tail is bound by PABPN in mammals and Pab1 in yeast. These

proteins protect the poly(A) tail, control its length, and influence various downstream processes. Since Pab1 is primarily localized in the cytoplasm, Nab2 serves as the main nuclear poly(A)-binding protein in *S. cerevisiae* (Brambilla et al. 2019; Soucek et al. 2012; Zarnack et al. 2020).

Other proteins that contribute to mRNP assembly include Tho1 and its mammalian homolog SARNP. For the recruitment of these proteins, interaction with the TREX complex, as well as splicing and capping are needed (see chapters 1.2. and 1.3.; Jimeno et al. 2006; Dufu et al. 2010). Moreover, mRNPs can also include several SR proteins. For instance, Npl3 is an SR-like protein in yeast that binds to mRNA co-transcriptionally and plays a role in transcription elongation, splicing, 3' end formation, and nuclear mRNA export (Lee et al. 1996; Bucheli & Buratowski 2005; Dermody et al. 2008; Kress et al. 2008; Meinel et al. 2013; Zarnack et al. 2020). Another SR-like protein is the nuclear poly(A)-binding protein Nab2. Besides functions in 3’ end processing and mRNP assembly, Nab2 is involved in mRNA export (Batisse et al. 2009; Zarnack et al. 2020).

Accurate mRNA processing and mRNP assembly are required for mRNA export to the cytoplasm (see chapter 1.1.4.). Several nuclear mRNP components serve as export adaptors, recruiting the mRNA exporters Mex67-Mtr2 in *S. cerevisiae* and NXF1-NXT1 (Tap-p15) in mammals. In yeast, export adaptors include the TREX complex components Hpr1, Yra1, Gbp2, and Hrb1 (see chapter 1.2.), as well as Npl3 and Nab2 (Gilbert & Guthrie 2004; Gwizdek et al. 2006; Iglesias et al. 2010; Luo et al. 2001; Hackmann et al. 2014; Zarnack et al. 2020).

Quality control systems monitor the maturation of mRNAs and lead to the degradation of mRNAs that have undergone incorrect processing or assembly. Defective mRNAs can be eliminated via two pathways: either by the nuclear exosome or through a process referred to as nonsense-mediated mRNA decay (NMD) in the cytoplasm (Kervestin & Jacobson 2012; Schoenberg & Maquat 2012; Shoemaker & Green 2012; Lykke-Andersen & Bennett 2014; Singh et al. 2018).

In summary, the assembly of mRNPs includes various RBPs and protein complexes that are essential for mRNA maturation and export. Key mRNP components like the CBC, EJC, and TREX play crucial roles in mRNA processing, transcription elongation, and the recruitment of export adaptors. Accurate mRNA processing and

the assembly of a mature mRNP are required for the successful transport of mRNA to the cytoplasm for translation.

1.2.4. Export of mRNA into the cytoplasm

Because eukaryotic transcription and translation are compartmentalized by the nuclear envelope, the processed mRNAs need to be transported from the nucleus into the cytoplasm. This transport is a critical step in gene expression, ensuring that mRNAs reach the ribosomes in the cytoplasm for translation. The translocation of mRNA across the nuclear envelope is tightly regulated and involves interactions of transport proteins with nuclear pore complexes (NPCs) (Katahira 2012; Stewart 2019).

To enable the export of mRNA, it must travel through the nuclear interior until it reaches an NPC (Fig. 8). The NPC is a complex with a molecular weight of about 110 kDa comprising around 30 distinct nucleoporins (also referred to as Nups). While the precise number and arrangement of nucleoporins may differ among species, the fundamental structural framework of the NPC is consistent across all eukaryotic organisms (Fabre & Hurt 1997; Rout et al. 2000). Typically, NPCs feature the nuclear basket on the nucleoplasmic side and the cytoplasmic ring with associated filaments on the cytoplasmic side. The central channel, the inner ring, the outer ring, and the transmembrane ring link these two components (Fig. 8).

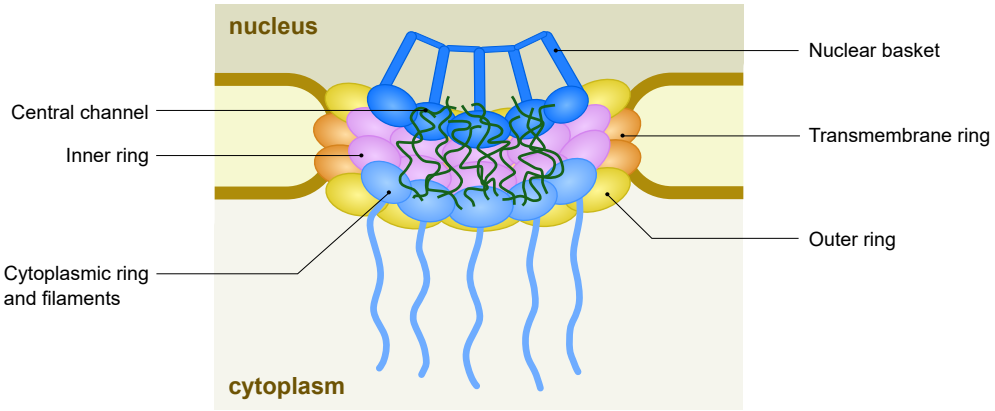


Figure 8: The nuclear pore complex (NPC) facilitates the mRNA export from the nucleus to the cytoplasm. The NPC comprises several primary subcomplexes, including the cytoplasmic ring and filaments, the outer ring, the inner ring, the transmembrane ring, the central channel, and the nuclear basket. Each of these subcomplexes is composed of numerous nucleoporins (Nups). Adapted from Coyne & Rothstein 2022, reproduced with permission from Springer Nature (license number: 5963590266191).

Around one-third of nucleoporins are inherently disordered proteins referred to as FG-nucleoporins (FG-Nups). FG-Nups, characterized by abundant phenylalanine-glycine repetitions, form a discerning permeability barrier responsible for controlling the transport of large molecules (Peyro et al. 2021; Coyne & Rothstein 2022; Yu et al. 2023).

In general, the transport through NPCs requires so called karyopherins or importin-β family members. These nuclear transport receptors detect nuclear localization signals (NLS) or nuclear export signals (NES), short peptide sequences that determine whether a cargo protein is imported or exported. Additionally, karyopherins detect nucleotide motifs in RNA cargoes, enabling the export of RNAs. In terms of their functions, there are three types of karyopherins: importins that mediate import, exportins that handle export, and carriers that can both import certain cargoes and export others. Their activity is regulated by the small GTPase Ran, which exists in its GTP-bound state in the nucleus and GDP-bound state in the cytoplasm. This RanGTP gradient, maintained by RanGEF (guanine nucleotide exchange factor) in the nucleus and RanGAP (GTPase-activating protein) in the cytoplasm, drives directional transport. Importins bind cargo in the cytoplasm and release it in the nucleus upon RanGTP binding, while exportins bind cargo in the nucleus with RanGTP and release it in the cytoplasm after RanGTP hydrolysis. This mechanism governs the export of various RNAs, including tRNA, miRNA, snRNA, and rRNA, but general mRNA export follows a different pathway, using unrelated transport receptors and not directly depending on the Ran gradient (Cole & Hammell 1998; Macara 2001; Köhler & Hurt 2007; Carmody & Wente 2009; Güttler & Görlich 2011; Jovanovic-Talisman & Zilman 2017).

The nascent mRNA transcript is already prepared for its export during transcription and processing. The THO complex is the core component of the TREX complex (see chapter 1.2.). Shortly after RNAPII initiates transcription, the THO complex binds to the newly synthesized transcript and its component Hpr1 is thought to recruit the DEAD-box RNA helicase Sub2 (UAP56/DDX39 in humans; Fig. 9), another component of the TREX complex (see chapter 1.2.1.; Jimeno et al. 2002; Sträßer et al. 2002; Zenklusen et al. 2002; Hurt et al. 2004). DEAD-box RNA helicases contain two RecA-like domains that cooperatively bind ATP and RNA in a sequence-independent way. While their ATP-dependent helicase activity is weakly

processive, these enzymes mainly facilitate local RNA:RNA rearrangements, modulate RNA:DNA and DNA:DNA interactions, and remodel RNP complexes (Taschuk & Cherry 2020).

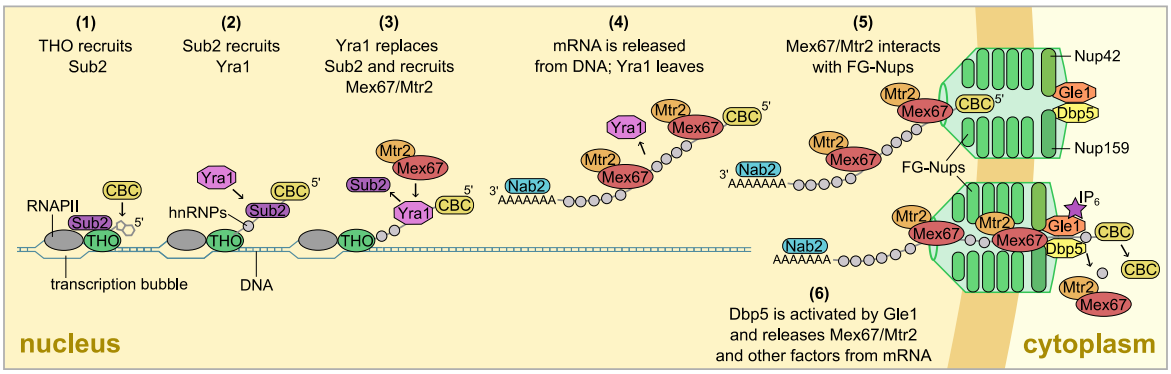


Figure 9: Recruitment of mRNA export factors and passage through the NPC in yeast. *(1)* Soon after RNAPII starts transcription, the newly synthesized transcript is bound by the THO complex, which recruits the TREX component Sub2. *(2)* Sub2 recruits another TREX component, Yra1. *(3)* After Sub2 is displaced by Yra1, Yra1 facilitates the recruitment of the export factor Mex67-Mtr2. *(4)* When the mRNA is released from the DNA, Yra1 dissociates and Mex67-Mtr2 brings the mRNA to the NPC, *(5)* by interacting with the FG-Nups. *(6)* Dbp5 is activated as a result of its interaction with Gle1, which, in turn, associates with inositol hexakisphosphate (IP_6). Dbp5 initiates the release of Mex67-Mtr2 and various other factors. Adapted from Carmody & Wente 2009, reproduced with permission from Journal of cell science (license number: 1575708-1).

Following the recruitment of Sub2 to the transcript, Sub2 is believed to subsequently recruit the next TREX component, Yra1 (ALYREF in human; see chapter 1.2.1.; Luo et al. 2001; Sträßer & Hurt 2001). It was proposed that Sub2 is then likely replaced by the export factor Mex67-Mtr2 (NXF1-NXT1 in humans; Sträßer & Hurt 2001; Fig. 9). When the mRNA is released from the DNA, Mex67-Mtr2 brings the mRNP to the NPC by interacting with the FG-Nups (Segref et al. 1997; Sträßer et al. 2000; Magistris 2001; Fig. 9). As the mRNP traverses the NPC, it eventually reaches the cytoplasmic filaments. In this location, the yeast Nup159, or in humans Nup214, interacts with the DEAD-box protein Dbp5 (DDX19 in humans) (Schmitt et al. 1999; Napetschnig et al. 2007; Moeller et al. 2009). Additionally, the neighboring yeast Nup42, or in humans hCG1, binds to Gle1 (Strahm et al. 1999; Kendirgi et al. 2005; Fig. 9). Dbp5 is activated as a result of its interaction with Gle1, which, in turn, associates with inositol hexaphosphate (IP_6) to enable the activation of the ATPase activity of Dbp5 (Alcázar-Román et al. 2006). Dbp5 initiates the release of Mex67-Mtr2 and various other factors, including the poly(A)-binding protein Nab2 (Lund & Guthrie 2005; Tran et al. 2007; Adams & Wente 2020; Fig. 9). This remodeling of

the mRNP at the NPC exit contributes to determining the direction of mRNA transport (Carmody & Wente 2009). Within the cytoplasm, mRNA undergoes translation to synthesize proteins. Gle1 is involved in translation initiation (Rayala et al. 2004), and, additionally, both Gle1 and Dbp5 have functions in translation termination (Gross et al. 2007; Bolger et al. 2008).

In summary, the export of processed mRNAs from the nucleus to the cytoplasm is crucial for gene expression, allowing for translation at the ribosomes. This transport through the NPC involves a series of coordinated interactions, particularly between components of the TREX complex and the export factor Mex67-Mtr2.

1.3. The TREX complex

TREX (transcription/export) is a multi-functional complex that couples transcription of protein-coding genes (see chapter 1.1.1.) to mRNA export into the cytoplasm (see chapter 1.1.4.). Having crucial roles in mRNA biogenesis, the TREX complex is found in diverse organisms, ranging from *S. cerevisiae* to more advanced eukaryotes like humans (Jimeno et al. 2002; Sträßer et al. 2002; Zenklusen et al. 2002). In *S. cerevisiae*, the TREX complex consists of the pentameric THO complex, comprising Tho2, Hpr1, Mft1, Thp2, and Tex1, along with the ATP-dependent DEAD-box RNA helicase Sub2, the nuclear mRNA export adapter Yra1, and the SR-like proteins Gpb2 and Hrb1 (Strasser et al. 2002; Hurt et al. 2004).

1.3.1. TREX homologs across selected species

Since TREX is a highly conserved complex, it has been studied and characterized in a range of different species. Besides the model organism *S. cerevisiae*, TREX has also been studied in the human system (Katahira 2012). As can be seen in table 1, humans contain several homologs of the yeast TREX complex proteins. For instance, the yeast THO components Hpr1, Tho2, Tex1 and Mft1 are homologous to THOC1, THOC2, THOC3 and THOC7 in humans, respectively (Heath et al. 2016). The THO component Thp2 is specific to *S. cerevisiae* and cannot be found in humans. However, higher eukaryotes contain THOC5, which is structurally homologous to Thp2 (Pühringer et al. 2020). The human TREX component THOC6 does not have homologs in *S. cerevisiae* (Katahira 2012). The yeast DEAD-box RNA helicase Sub2 is homologous to human UAP56/DDX39 and yeast Yra1 to

ALYREF/THOC4 in humans (Jimeno et al. 2002; Sträßer et al. 2002; Masuda et al. 2005). The yeast SR-like proteins Gbp2 and Hrb1 are not identified as components. Additionally, the human TREX complex encompasses various interacting partners such as SARNP (equivalent to Tho1 in yeast), ZC3H11A, IUF, LUZP4, POLDIP3, and CHTOP (Dufu et al. 2010; Heath et al. 2016). Proteins derived from thermophilic organisms exhibit higher stability in comparison to proteins originating from mesophilic organisms (Jaenicke 1991; Jaenicke & Böhm 1998). Due to this characteristic, thermophiles are frequently employed to acquire three-dimensional structures of individual proteins or protein complexes (Bock et al. 2014). In an effort to examine the entire structure of the TREX complex, the thermophilic fungus *Chaetomium thermophilum* could serve as a promising source of stable proteins. Based on the genome published by Amlacher et al. (2011), Dr. Wolfgang Wende (Institute for Biochemistry, JLU Gießen) identified components of *ct*TREX (table 1). Homologs of yeast proteins Hpr1, Tho2, Mft1, Tex1, Sub2, and Yra1 are present in *C. thermophilum*. Unlike in yeast where the SR-like proteins Gbp2 and Hrb1 are encoded by separate genes, there is a single gene coding for *ct*Gbp2/Hrb1 in *C. thermophilum*. The THO component Thp2 is absent in *C. thermophilum*; however, the *ct*TREX complex includes the homolog of the human THOC5 (Amlacher et al. 2011; Bock et al. 2014; table 1).

Table 1: Components of the *S. cerevisiae* TREX complex and their homologs in *H. sapiens* and *C. thermophilum* (adapted from Katahira 2012).

	S. cerevisiae	***Homo sapiens***	***C. thermophilum***
THO components	Hpr1	THOC1	*ct*Hpr1
	Tho2	THOC2	*ct*Tho2
	Thp2	-	-
	Mft1	THOC7	*ct*Mft1
	-	THOC5	*ct*Thoc5
	-	THOC6	-
	Tex1	THOC3	*ct*Tex1
DEAD-box type helicase	Sub2	UAP56/DDX39	*ct*Sub2
Adaptor mRNA binding protein	Yra1	ALYREF/THOC4	*ct*Yra1
SR-like proteins	Hrb1		*ct*Hrb1/Gbp2
	Gbp2		

In summary, the TREX complex has a high degree of conservation across species. While several homologs of yeast TREX components are present in humans, distinct differences, such as the absence of Thp2 in humans and or the absence of THOC6 in yeast, are evident. Homology with thermophilic organisms like *C. thermophilum* can allow deeper insights into the TREX structure, as their more stable proteins are better suited for structural analysis.

1.3.2. Functional relevance of the TREX components

Observing the effects on an organism when a gene is absent is one of the most direct approaches to understanding its function. Therefore, the use of mutations and deletions of genes, due to their potential to disrupt cellular processes, frequently serve as valuable tools in unraveling gene functions (Alberts et al. 2002). The strategy of testing and analyzing various mutants has also been employed to elucidate the functional significance of individual THO/TREX components.

In general, null mutations in Tho2, Hpr1, Mft1, and Thp2 exhibit very similar effects, including deficiencies in transcription and RNA export, along with hyper-recombination phenotypes. For instance, it was shown that yeast *Δhpr1* mutants experience significant constraints in transcriptional elongation (Chávez & Aguilera 1997). Moreover, *Δhpr1* mutants generate DNA:RNA hybrids between the DNA template and the RNA transcript. This structure, comprising the DNA:RNA hybrid and the non-template single-stranded DNA, is referred to as an R-loop (Aguilera & García-Muse 2012). The notable increase in R-loops observed in *Δhpr1* mutants, which substantially hinders transcription, suggests the crucial role of Hpr1 in promoting efficient transcription elongation (Huertas & Aguilera 2003). Tho2 constitutes the largest subunit of the pentameric THO complex. Investigations of *Δtho2* mutants indicate that Tho2 shares a comparable significance with Hpr1 in transcription elongation. Additionally, the overexpression of Tho2 is able to partially suppress the transcriptional defects observed in *Δhpr1* mutants (Piruat & Aguilera 1998). Studies on *Δmft1* and *Δthp2* mutants reveal that the absence of either Mft1 or Thp2 leads to comparable transcription and recombination phenotypes as observed in *Δhpr1* and *Δtho2* mutants (Chavez et al. 2000; Jimeno et al. 2002). Apart from mutants affecting the four canonical THO components, the *Δtex1* mutation has virtually no discernible phenotypic effects in yeast (Peña et al. 2012).

In summary, these studies suggest that the THO complex plays crucial roles in transcription elongation (Rondón et al. 2003). Consistent with this role, genome-wide studies have shown that THO is present on virtually all actively transcribed genes. However, it is noteworthy that THO is not uniformly distributed across the entire length of an actively transcribed gene like RNAPII. Instead, THO exhibits a 5'-3' gradient, increasing in distribution along the gene, and is notably absent from promoter regions (Gómez-González et al. 2011; Meinel et al. 2013).

As the THO complex travels with the transcription machinery, its subunits interact with various factors. For instance, the THO complex component Hpr1 was shown to directly interact with Sub2 (Zenklusen et al. 2002). Therefore, it is believed that THO recruits Sub2 to the mRNA during transcription (Sträßer et al. 2002; Zenklusen et al. 2002; Hurt et al. 2004). The yeast Sub2 protein is a member of the DEAD-box RNA helicase superfamily (La Cruz et al. 1999). Various studies in *S. cerevisiae* have implicated Sub2 in pre-mRNA splicing (Kistler & Guthrie 2001; Libri et al. 2001; Zhang & Green 2001). Additionally, it was observed that thermosensitive *Δsub2* mutants exhibit a strong mRNA export defect (Sträßer & Hurt 2001). This finding indicates that Sub2 plays a critical role in mRNA export, which is further supported by its direct interaction with the mRNA export adapter Yra1 (Luo et al. 2001), which directly binds the export factor Mex67-Mtr2 (Sträßer & Hurt 2000).

Yra1 is a member of the REF (RNA and export factor binding proteins) family of hnRNP-like proteins. The REF proteins have a nonspecific affinity for RNA and share an RNP-motif RNA binding domain (RBD) as a common structural feature (Portman et al. 1997). Additionally, REF proteins contain two highly conserved N- and C-terminal boxes (N-Box and C-Box), which are interspersed by two more variable domains (Stutz et al. 2000). Depletion of Yra1 as well as studies on thermosensitive *YRA1* mutants revealed the nuclear accumulation of poly(A)$^+$ RNA in the absence of Yra1 (Sträßer & Hurt 2000). In *S. cerevisiae*, Yra2 is a nonessential homolog of Yra1. While *Δyra2* mutants do not display a growth phenotype, the overexpression of Yra2 can rescue the otherwise lethal *Δyra1* mutants. This implies that Yra2 has functions that are overlapping with those of Yra1 (Zenklusen et al. 2001).

In yeast, the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 are identified as components of the TREX complex. These two proteins are recruited to the nascent

RNA transcript during transcription and stably bind to TREX (Hurt et al. 2004). Gbp2 and Hrb1 each possess three RNA recognition motifs: RRM1, RRM2, and RRM3. While RRM1 and RRM2 are essential for binding to RNA and DNA, the RRM3 of Gbp2 is specifically necessary for interaction with THO (Martínez-Lumbreras et al. 2016). Mutants lacking Gbp2 (*Δgbp2*) and Hrb1 (*Δhrb1*) are viable and do not exhibit any defects in mRNA export, unlike other mutants lacking components of the THO/TREX complex (Hurt et al. 2004; Sträßer et al. 2002).

In summary, gene mutations and deletions were crucial in understanding the functions of the different TREX complex components. Null mutations in key components such as Tho2, Hpr1, Mft1, and Thp2 particularly lead to significant transcription and mRNA export deficiencies. Studies on Sub2 and Yra1 highlighted their roles as essential mRNA export factors. Collectively, these findings demonstrate how the TREX complex effectively links transcription to mRNA export.

1.3.3. Recruitment of the TREX complex to active genes

In *S. cerevisiae*, the TREX complex is recruited to active genes already during transcription elongation (Sträßer et al. 2002; Zenklusen et al. 2002; Reed & Cheng 2005). It was observed that TREX can directly associate with RNAPII. Specifically, this interaction is facilitated by the phosphorylation of the RNAPII CTD at positions S_2 and S_5 (Meinel et al. 2013; Fig. 10). The significance of the S_2 phosphorylation aligns with the fact that Ctk1, the kinase responsible for the phosphorylation at this position, interacts with TREX (Hurt et al. 2004). In addition to the S_2 and S_5 phosphorylation, the recruitment of TREX also necessitates the presence of RNA (Meinel et al. 2013; Fig. 10). Notably, the C-terminus of Tho2 interacts with nucleic acids and is essential for the presence of THO at transcribed genes (Peña et al. 2012).

The Prp19 complex (Prp19C) is another crucial factor for the effective recruitment of TREX to the actively transcribed mRNA. Prp19C comprises eight core proteins and up to 19 associated proteins in *S. cerevisiae* (Chanarat & Sträßer 2013). Prp19C functions in many different cellular processes, including splicing (Chan & Cheng 2005; Hogg et al. 2010; Chanarat & Sträßer 2013), genome maintenance (Henriques et al. 1989; Mahajan & Mitchell 2003; Chanarat & Sträßer 2013), and protein degradation by the proteasome (Löscher et al. 2005; Sihn et al. 2007; Chanarat & Sträßer 2013). In addition to these functions, it was demonstrated that

the Prp19C component Syf1 is required for TREX occupancy at transcribed genes (Fig. 10). Specifically, when the C-terminal domain of Syf1, which contains tetratricopeptide repeat (TPR) motifs, is mutated, Prp19C cannot efficiently bind to RNAPII anymore. Additionally, in these *Δsyf1* mutants it was observed that the binding of TREX to active genes is reduced. Hence, it is suggested that Prp19C plays a crucial role in transcription elongation and contributes to a more stable recruitment of TREX to genes being transcribed by RNAPII (Chanarat et al. 2011). An important factor for recruiting Prp19C and subsequently TREX to the transcribed gene is the Mud2 protein (Minocha et al. 2018). The sequence of Mud2 is similar to the metazoan splicing factor U2AF65. Mud2 participates in early pre-mRNA splicing and is known to interact with the Msl5p/BBP splicing factor and Sub2 (Abovich et al. 1994; Kistler & Guthrie 2001). While Prp19C interacts with RNAPII through the C-terminus of Syf1 (Chanarat et al. 2011), Mud2 directly interacts with the phosphorylated S_2 of the CTD. After binding to the CTD, Mud2 facilitates the recruitment of Prp19C and TREX to intron-containing and intronless genes (Minocha et al. 2018; Fig. 10).

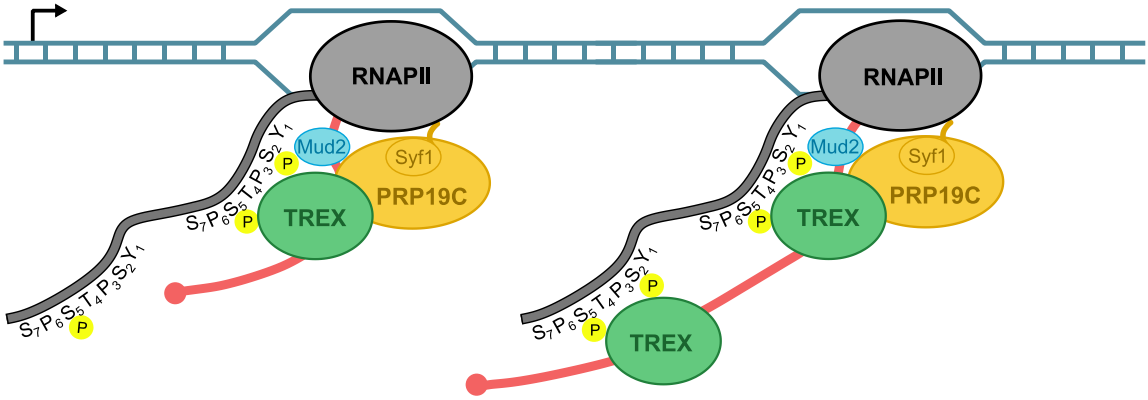


Figure 10: The recruitment of TREX to the active gene. The TREX complex is recruited directly by interacting with the phosphorylated S_2 and S_5 of the CTD of RNAPII. Interaction with RNA is also required for TREX recruitment. The S_2 phosphorylation recruits Mud2, which, in turn, is involved in the recruitment of Prp19C. Additionally to Mud2, Prp19C can directly interact with RNAPII with the C-terminal end of its subunit Syf1. Mud2 and Prp19C contribute to a more stable recruitment of TREX. With transcription elongation, it is possible that additional TREX complexes may be required to maintain the nascent mRNA in proximity to the CTD. Adapted from Meinel et al. 2013, licensed under the Creative Commons Attribution License, and Minocha et al. 2018.

Unlike in yeast, the metazoan TREX complex is recruited in a splicing-dependent manner, with ALYREF likely facilitating this process through its interaction with the EJC component eIF4A3 (Masuda et al. 2005; Gromadzka et al. 2016). Additionally, the metazoan TREX is recruited to the 5' end of mRNA via ALYREF's interaction

with the CBC, and to the 3' end through its interaction with PABPN1 (Cheng et al. 2006; Shi et al. 2017; Zarnack et al. 2020). The recruitment of ALYREF to the 3' end is consistent with the interaction between the yeast ALYREF homolog Yra1 and Pcf11. In yeast, Pcf11, along with Clp1, plays a role in pre-mRNA cleavage at the 3' end. After Pcf11 is recruited to RNAPII via the CTD, it can then recruit Yra1 to the 3' end (Johnson et al. 2009). Finally, the displacement of Yra1 with Clp1 is achieved by regulating the accessibility of Clp1 to Pcf11 (Johnson et al. 2011).

In conclusion, the recruitment of the TREX complex to actively transcribed genes in *S. cerevisiae* depends on crucial interactions with the phosphorylated CTD of RNAPII, along with key factors such as Mud2 and Prp19C. In metazoans, however, TREX is recruited in a splicing-dependent manner. Regardless of the organism, TREX recruitment is a key early step in ensuring accurate mRNA biogenesis and efficient gene expression.

1.3.4. Functions of TREX in mRNA export via Mex67-Mtr2

The TREX complex is crucial for the export of mRNAs via the export factor Mex67-Mtr2 (NXF1-NXT1 in humans). By interaction of the heterodimer Mex67-Mtr2 with the FG-Nups of the NPC (see chapter 1.1.4.), the complex is able to export mRNA into the cytoplasm (Segref et al. 1997; Santos-Rosa et al. 1998; Sträßer et al. 2000; Terry & Wente 2007). Moreover, Mex67-Mtr2 interacts with various protein factors to facilitate the export of mRNAs (Chanarat et al. 2012; Fig. 11).

It has been shown that two components of the TREX complex interact with the export factor Mex67-Mtr2. Firstly, there is a direct interaction between Hpr1 and the ubiquitin-associated (UBA) domain of Mex67 (Gwizdek et al. 2006). Secondly, as previously mentioned, Yra1 has the capability to recruit Mex67-Mtr2 to the transcript (Sträßer & Hurt 2000). Moreover, it was shown that Yra1 can associate with the poly(A)-binding protein Nab2 (Kashyap et al. 2005; Batisse et al. 2009). Conversely, Nab2 is known to directly interact not only with Yra1 but also with Mex67. Overexpressing either Nab2 or Mex67 was observed to suppress mRNA export defects in *Δyra1* mutants. Additionally, Yra1 has the capability to enhance the interaction between Nab2 and Mex67 (Iglesias et al. 2010). Similar to Yra1, Hpr1, and Nab2, Npl3 has the ability to interact with Mex67-Mtr2, and thereby contributing to efficient mRNA export (Köhler & Hurt 2007). However, affinity purifications using

TAP-tagged Npl3 revealed that Npl3 does not have a direct association with TREX. Therefore, Npl3 is not recruited to the nascent transcript by THO/TREX but rather by other structures, such as the CBC (Hurt et al. 2004).

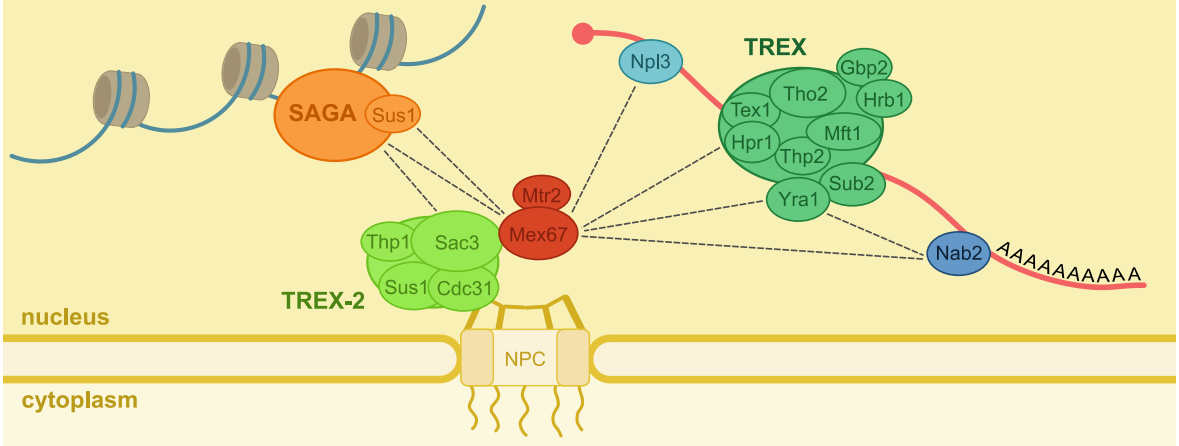


Figure 11: TREX facilitates mRNA export by interacting with Mex67-Mtr2. The TREX components Yra1 and Hpr1 directly interact with Mex67-Mtr2. Other factors that associate with Mex67-Mtr2 are Nab2 and Npl3. The TREX-2 complex is bound to the NPC by its subunit Cdc31. The TREX-2 component Sac3 is essential for mRNA export by interacting with Mex67-Mtr2. The TREX-2 component Sus1 is also part of the SAGA complex. TREX-2 functionally links SAGA-dependent gene expression to mRNA export. Adapted from Köhler & Hurt 2007, reproduced with permission from Springer Nature (license number: 5963591402041), and Chanarat et al. 2012, with permission by Taylor & Francis.

Sac3, another crucial mRNA export adaptor in yeast, has been demonstrated to interact with Yra1 and facilitate the recruitment of the Mex67-Mtr2 export receptor. Together with Thp1, Sus1, and Cdc31, Sac3 constitutes the TREX-2 complex. This complex is bound to the NPC by Cdc31 interacting with the nuclear basket proteins Nup1 and Nup60 (Fischer et al. 2002; Köhler & Hurt 2007; Rondón et al. 2010). The TREX-2 component Sus1 is also part of the SAGA complex that is involved in transcription initiation, facilitating histone acetylation and deubiquitylation (Rodríguez-Navarro et al. 2004; Köhler et al. 2006). Therefore, TREX-2 may link SAGA-dependent gene expression to mRNA export (Köhler & Hurt 2007).

In conclusion, the TREX complex is essential for mRNA export though its interaction with the Mex67-Mtr2 export factor, facilitating mRNA transport across the NPC. Key TREX components, such as Yra1 and Hpr1, interact directly with Mex67-Mtr2, while additional factors like Nab2 and Npl3 further support this process. The TREX-2 complex connects mRNA export to transcription through the SAGA complex. These interactions ensure efficient gene expression by tightly coordinating mRNA export and transcription.

1.3.5. Structural characteristics of the TREX complex

During this study, structures of human and yeast THO-Sub2/UAP56 complexes, which were expressed recombinantly, were published (Pühringer et al. 2020; Schuller et al. 2020; Chen et al. 2021; Xie et al. 2021). The structure of the endogenous human TREX complex was recently published (Pacheco-Fiallos et al. 2023), and it is generally in line with the structure of the recombinant complex (Pühringer et al. 2020). Overall, the monomeric arrangement of the THO components and Sub2 is conserved between the yeast and human TREX complexes (Fig. 12). In general, Tho2 together with Hpr1 forms a platform to which the other complex components attach. For instance, Mft1 and Thp2 create an elongated coiled-coil structure that stretches diagonally from the one end of the

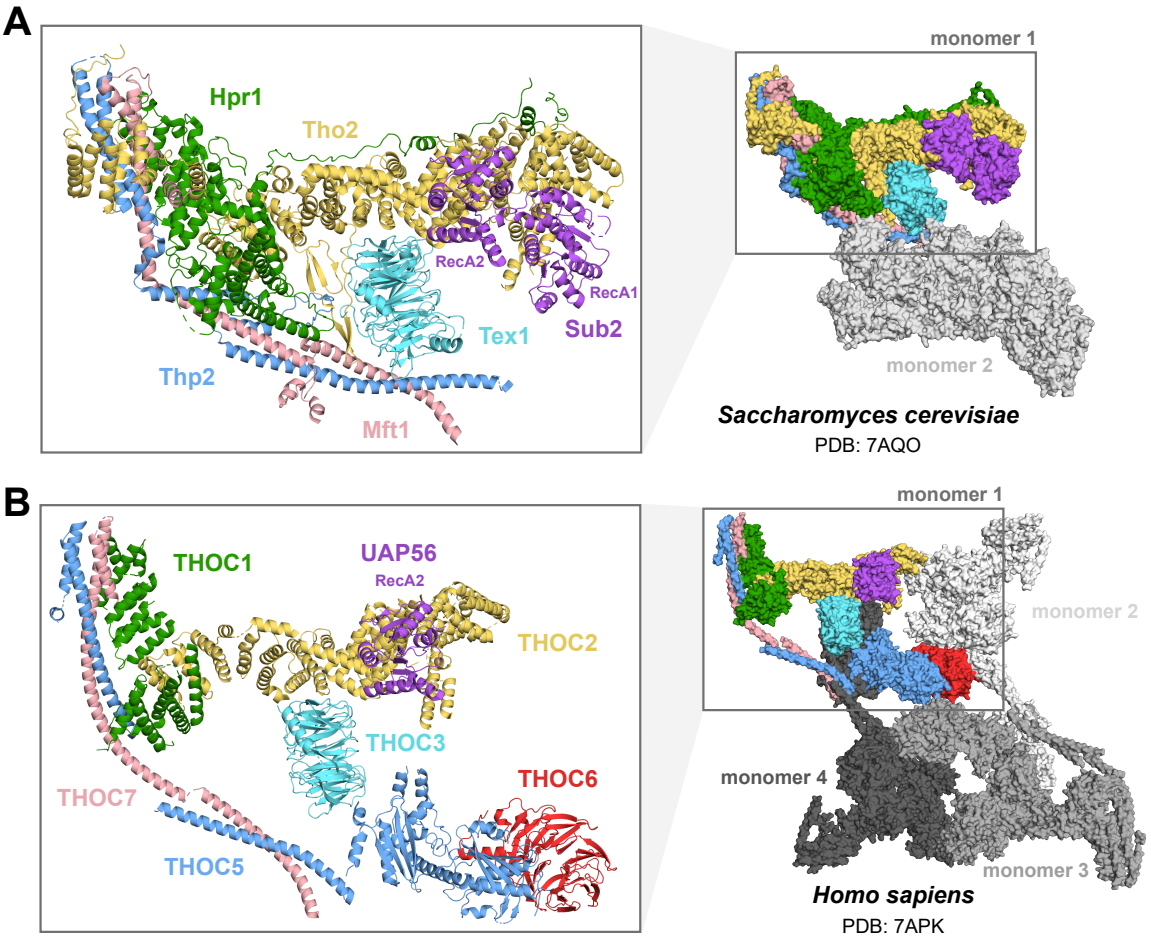


Figure 12: High-resolution structures of the yeast and human THO-Sub2/UAP56 complex. *(A)* In *S. cerevisiae* the recombinant THO-Sub2 complex (PDB: 7AQO, Schuller et al. 2020) mainly assembles into a dimer (shown in surface representation). The single complex components are indicated for one monomer shown in cartoon representation. *(B)* In *H. sapiens* the recombinant THO-UAP56 complex (PDB: 7APK, Pühringer et al. 2020) mainly assembles into a tetramer (shown in surface representation). The single complex components are indicated for one monomer shown in cartoon representation.

platform, whereas Sub2 binds to the opposite end of the platform. Tex1, which mainly consists of β-sheets is associated to the middle of the Hpr1-Tho1 platform (Fig. 12). Although the overall arrangement is similar between yeast and human, the preferred oligomeric states seem to differ. While the *S. cerevisiae* complex seems to preferentially assemble into a dimer, the *H. sapiens* mainly assembles into a tetramer (Fig. 12). Nevertheless, other oligomeric states, ranging from dimeric to octameric were observed for the recombinant yeast and human THO-Sub2/UAP56 complexes (Pühringer et al. 2020; Schuller et al. 2020; Chen et al. 2021; Xie et al. 2021).

1.4. The TREX interacting factor Tho1

The TREX complex recruits various protein factors to the transcript. One of these factors in *S. cerevisiae* is Tho1, which is recruited in a THO and RNA-dependent manner (Jimeno et al. 2006). Initially, Tho1 was identified as a suppressor of the transcriptional defect of Hpr1 by overexpression (Piruat & Aguilera 1998). Later it was reported, that the overexpression of Tho1 not only suppresses the mRNA export defect and hyperrecombination phenotypes seen in *Δhpr1* but in all null mutants of the THO complex (specifically *Δhpr1*, *Δtho2*, *Δmft1*, *Δthp2*). However, the overexpression of Tho1 cannot suppress the observed phenotypes in *Δsub2* or *Δthp1* (Jimeno et al. 2006). Notably, the *Δtho1* mutant itself does not have a known phenotype (Piruat & Aguilera 1998).

Based on sequence alignment, the human SARNP protein was identified as a homolog of the yeast Tho1 (Jimeno et al. 2006). SARNP, also known as CIP29, has been implicated in various cancers, suggesting its functional significance in humans (Fukuda et al. 2002; Hashii et al. 2004). Both the yeast Tho1 and the human SARNP have an SAP domain in common, which contains a potential DNA-binding motif (Aravind & Koonin 2000). Tho1/SARNP has been suggested to play a role in transcription, splicing, export, and/or translation (Aravind & Koonin 2000; Hashii et al. 2004; Leaw et al. 2004). The human SARNP was identified as a primary interacting partner of UAP56/DDX39 (Sub2 in yeast). Furthermore, it was demonstrated that SARNP has the capability to bind RNA and enhance the RNA unwinding activity of DDX39 (Sugiura et al. 2007). The interaction between SARNP and DDX39 appears to be conserved, as the yeast proteins Tho1 and Sub2 have

also been reported to interact (Jimeno et al. 2006). This implies the importance of the interaction between the two proteins. Nevertheless, the full physiological relevance of this interaction remains not entirely understood (Jimeno et al. 2006; Sugiura et al. 2007). Notably, it was reported that DDX39, ALYREF (Yra1 in yeast), and SARNP form a trimeric complex. This complex formation was described as ATP-dependent. While DDX39 directly interacts with SARNP and ALYREF, there is no direct interaction between SARNP and ALYREF. There is speculation that the trimeric complex might act as a helicase, unwinding RNA secondary structures at the 5' end of mRNA. It is plausible that the unwinding at the 5' end creates a platform, potentially facilitating processes such as the recruitment of the THO complex or the export through the NPC (Dufu et al. 2010).

1.5. Aims and objectives

The TREX complex plays a vital role in mRNA biogenesis by linking transcription to mRNA export (Sträßer et al. 2002). Gaining insights into the structure of the TREX complex is essential for clarifying its diverse functions, which are crucial for efficient gene expression and cellular processes. A comprehensive understanding of the structure of the native TREX complex will improve the knowledge of its regulatory mechanisms across different organisms, ranging from *S. cerevisiae* to humans. Therefore, the primary aim of the first part of this study is to elucidate the structural organization of the complete TREX complex by purifying it from its native environment in *S. cerevisiae*, followed by analysis using cross-linking mass spectrometry (XL-MS) and negative-stain electron microscopy (EM). This will provide a more accurate representation of its physiological state, allowing for a better understanding of the multiple complex conformations during its various functions in mRNA processing and export. Furthermore, this study will specifically focus on the TREX component Tho2 from *S. cerevisiae*, particularly on its flexible C-terminal region. By analyzing a C-terminally truncated Tho2 mutant, this study aims to gain insights into its functional significance for the cell and into its contribution to the formation of the TREX complex. Additionally, overexpressing Tho2 in *S. cerevisiae* will be used to evaluate its impact on mRNA stability.

The second part of this study investigates the structural and functional interactions within the dimeric Sub2-Tho1 complex and the trimeric Sub2-Tho1-Yra1 complex.

The conserved interactions, particularly between Tho1 and Sub2, suggest the importance of this complex, though its full physiological role remains unclear (Jimeno et al. 2006). Using recombinantly expressed proteins from *C. thermophilum*, the aim is to resolve the structural organization of these complexes, as understanding their structure and how they interact with DNA and RNA is crucial for revealing their roles in mRNA biogenesis and gene expression regulation. In addition to crystallization and cryo-EM, various biochemical assays will provide further insights into these complexes. Furthermore, this study particularly focuses on *ct*Tho1. By testing various deletion mutants of *ct*Tho1 in different functional assays, the aim is to identify regions of *ct*Tho1 required for its interaction with *ct*Sub2, as well as with nucleic acids.

In summary, this study aims to enhance the understanding of mRNA biogenesis, particularly regarding the TREX complex. Several TREX components, including Tho2, Sub2, and Yra1, along with the TREX interacting partner Tho1, are analyzed in more detail. By employing XL-MS and EM, and through various biochemical approaches, the aim is to provide insights into the structural organization and functions of these complexes and their components.

2. Material

2.1. Instruments

The instruments that were used during this study are listed in table 2.

Table 2: Used instruments and the respective manufacturer.

Name	Manufacturer
6870D Freezer/Mill	SPEX SamplePrep, Stanmore, United Kingdom
ÄKTA™ start System	GE Healthcare, Chicago, USA
ÄKTA™ purifier System	GE Healthcare, Chicago, USA
Bio Dancer	New Brunswick Scientific, Edison, USA
BioDoc Analyze GelDoc	Biometra, Göttingen, Germany
Centrifuge 5424 R	Eppendorf, Hamburg, Germany
Centrifuge 5430	Eppendorf, Hamburg, Germany
Centrifuge 5430 R	Eppendorf, Hamburg, Germany
Centrifuge Avanti JXN-26	Beckman Coulter, Krefeld, Germany
Centrifuge Heraeus Megafuge 40R	Thermo Fisher Scientific, Dreieich, Germany
Centrifuge Heraeus Pico 17	Thermo Fisher Scientific, Dreieich, Germany
ChemoCam Imager	Intas Science Imaging Instruments, Göttingen, Germany
Duetta Spectrometer	Horiba Scientific, Oberursel, Germany
Electrophoresis Power Supply – EPS 301	GE Healthcare, Chicago, USA
Epson Perfection 3200 Photo	Seiko Epson Corporation, Suwa, Japan
Gel iX Imager	Intas Science Imaging Instruments, Göttingen, Germany
GSTrap HP 1-ml affinity column	Cytiva, Marlborough, USA
HiTrap Q HP 1 ml anion exchange column	Cytiva, Marlborough, USA
HiTrap SP 1 ml cation exchange column	Cytiva, Marlborough, USA
HiTrap TALON Crude 1 mL column	Cytiva, Marlborough, USA
Incubator BM 400	Memmert, Schwabach, Germany
Incubator IKA KS 4000 ic control	IKA-Werke, Staufen, Germany
Incubator Incucell 55	MMM Medcenter Einrichtungen, Gräfelfing, Germany
Incubator Innova 40	New Brunswick Scientific, Edison, USA
Incubator Innova 4000	New Brunswick Scientific, Edison, USA
Incubator Multitron Pro	Infors, Einsbach, Germany
J-710 Circular Dichroism Spectrophotometer	Jasco, Easton, USA

Mini-PROTEAN® Tetra Vertical Electrophoresis Cell	Bio-Rad Laboratories, Vienna, Austria
Mobicol "classic" column	MoBiTec, Göttingen, Germany
neoLab Rotator	neoLab Migge, Heidelberg, Germany
Owl EasyCast B1A electrophoresis system	Thermo Fisher Scientific, Dreieich, Germany
pH meter pH 1000L	VWR International, Darmstadt, Germany
Quant Studio 3, Real Time PCR System	Thermo Fisher Scientific, Dreieich, Germany Scientific
Rotational Vacuum Concentrator System	Christ, Osterode, Germany
Sonifier 250	Branson Ultrasonics Corporation, Danbury, USA
Strep-Tactin XT 4Flow high capacity 1 mL column	IBA Lifesciences, Göttingen, Germany
Superdex 75 Increase 10/300 GL	Cytiva, Marlborough, USA
Superdex 200 Increase 10/300 GL	Cytiva, Marlborough, USA
Superose 6 Increase 10/300 GL	Cytiva, Marlborough, USA
Techne Dri-Block Heater DB-3	Bibby Scientific, Staffordshire, United Kingdom
Thermocycler peqSTAR XS	VWR International, Darmstadt, Germany
Thermocycler TPersonal 48	Biometra, Jena, Germany
Thermomixer 5436	Eppendorf, Hamburg, Germany
Trans-Blot Turbo Transfer System	Bio-Rad Laboratories, Vienna, Austria
Typhoon FLA 9500	GE Healthcare, Chicago, USA
Ultracentrifuge Optima XN-80	Beckman Coulter, Krefeld, Germany
WPA CO8000 Cell Density Meter	Biochrom, Cambridge, United Kingdom

2.2. Chemicals

All chemicals that were used during this study are listed in table 3.

Table 3: Chemicals and their respective supplier.

Name	**Supplier**
4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium (DMTMM) chloride	Sigma-Aldrich Chemie, Taufkirchen, Germany
Acrylamide – Solution (40 %) – Mix 29 : 1	AppliChem, Darmstadt, Germany
Adenine hemisulfate salt	Sigma-Aldrich Chemie, Taufkirchen, Germany
Adipic dihydrazide (ADH-d_0/d_8)	Sigma-Aldrich Chemie, Taufkirchen, Germany
Agarose Basic	AppliChem, Darmstadt, Germany
Ammonium bicarbonate (NH_4HCO_3)	Sigma-Aldrich Chemie, Taufkirchen, Germany
Ammonium persulfate (APS)	Sigma-Aldrich Chemie, Taufkirchen, Germany

Ammonium sulfate	Merck, Darmstadt, Germany
Ampicillin	AppliChem, Darmstadt, Germany
ANTI-FLAG M2 Affinity Gel	Sigma-Aldrich Chemie, Taufkirchen, Germany
Bacto Peptone	Becton, Dickinson and Company (BD), Erembodegem, Belgium
Bacto Yeast Extract	Becton, Dickinson and Company (BD), Erembodegem, Belgium
Benzamidine HCl	MP Biomedicals, Irvine, USA
bis(sulfosuccinimidyl)suberate (BS^3-d_0/d_{12})	Creative Molecules, San Francisco, USA
β-cyclodextrin	Tokyo Chemical Industry, Tokyo, Japan
β-Mercaptoethanol	Merck, Darmstadt, Germany
β-Nicotinamide adenine dinucleotide (NAD^+)	Fluka Chemie, Buchs, Switzerland
Boric acid	Merck, Darmstadt, Germany
Bromophenol blue	AppliChem, Darmstadt, Germany
Coomassie Brilliant Blue G-250	AppliChem, Darmstadt, Germany
CheLuminate-HRP ECL solution	Intas Science Imaging Instruments, Göttingen, Germany
Chloroform	Merck, Darmstadt, Germany
D-(+)-Biotin	Alfa Aesar, Thermo Fisher Scientific, Waltham, USA
D-(+)-Glucose monohydrate	Sigma-Aldrich Chemie, Taufkirchen, Germany
D-(+)-Sucrose	AppliChem, Darmstadt, Germany
Deoxyadenosine triphosphate (dATP)	Thermo Fisher Scientific, Waltham, USA
Deoxycytidine triphosphate (dCTP)	Thermo Fisher Scientific, Waltham, USA
Deoxyguanosine triphosphate (dGTP)	Thermo Fisher Scientific, Waltham, USA
Deoxythymidine triphosphate (dTTP)	Thermo Fisher Scientific, Waltham, USA
Dextrin from potato starch	Sigma-Aldrich Chemie, Taufkirchen, Germany
Diethyl pyrocarbonate (DEPC)	Carl Roth, Karlsruhe, Germany
Dimethyl sulfoxide (DMSO)	Grüssing, Filsum, Germany
di-Potassium hydrogen phosphate (K_2HPO_4)	Merck, Darmstadt, Germany
di-Sodium hydrogen phosphate heptahydrate (Na_2HPO_4 * 7 H_2O)	Carl Roth, Karlsruhe, Germany
DTT BioChemica	AppliChem, Darmstadt, Germany
Dynabeads M-280 Streptavidin	Thermo Fisher Scientific, Waltham, USA
Ethanol ≥99,5 %	Thermo Fisher Scientific, Waltham, USA
Ethidium bromide solution (1%)	Carl Roth, Karlsruhe, Germany
Ethylenediaminetetraacetic acid disodium salt dehydrate (EDTA)	Sigma-Aldrich Chemie, Taufkirchen, Germany

FLAG Peptide	Carl Roth, Karlsruhe, Germany
Genetecin (G418)	Thermo Fisher Scientific, Waltham, USA
Glutathione (reduced; GSH)	Merck, Darmstadt, Germany
Glutathione Sepharose 4B	GE Healthcare, Chicago, USA
Glycerin ROTIPURAN® ≥99,5 %, p.a.	Carl Roth, Karlsruhe, Germany
HDGreen™ Plus DNA stain	Intas Science Imaging Instruments, Göttingen, Germany
HEPES	Carl Roth, Karlsruhe, Germany
Hydrochloric acid (HCl)	Carl Roth, Karlsruhe, Germany
Herring Sperm DNA	Thermo Fisher Scientific, Waltham, USA
IGEPAL CA-630	Sigma-Aldrich Chemie, Taufkirchen, Germany
IgG Sepharose 6 Fast Flow	Cytiva, Marlborough, USA
Imidazole	Merck, Darmstadt, Germany
Isopropanol (2-Propanol)	Carl Roth, Karlsruhe, Germany
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Carl Roth, Karlsruhe, Germany
Kanamycin	AppliChem, Darmstadt, Germany
Leupeptin	Merck, Darmstadt, Germany
L-Histidine	Sigma-Aldrich Chemie, Taufkirchen, Germany
Lithium chloride (LiCl)	Merck, Darmstadt, Germany
L-Leucine	Sigma-Aldrich Chemie, Taufkirchen, Germany
L-Tryptophan	Sigma-Aldrich Chemie, Taufkirchen, Germany
Magnesium chloride hexahydrate ($MgCl_2$ * 6 H_2O)	Merck, Darmstadt, Germany
MES	Carl Roth, Karlsruhe, Germany
Methanol	Merck, Darmstadt, Germany
Oligo(dt) primer	NEB, Frankfurt/Main, Germany
Pepstatin A	Merck, Darmstadt, Germany
Phenylmethylsulfonyl fluoride (PMSF)	AppliChem, Darmstadt, Germany
Polythylene glycol 4000 (PEG-4000)	Fluka Chemie, Buchs, Switzerland
Potassium chloride (KCl)	ORG Laborchemie, Bunde, Germany
Potassium dihydrogen phosphate (KH_2PO_4)	Merck, Darmstadt, Germany
Potassium hydroxide (KOH)	Merck, Darmstadt, Germany
Powdered milk, fat free, blotting grade	Carl Roth, Karlsruhe, Germany
Protino Ni-NTA Agarose	Macherey-Nagel, Düren, Germany

SDS ultrapure	AppliChem, Darmstadt, Germany
Sodium acetate (NaOAc)	Merck, Darmstadt, Germany
Sodium chloride (NaCl)	Merck, Darmstadt, Germany
Sodium dihydrogen phosphate monohydrate (NaH_2PO_4 * H_2O)	Carl Roth, Karlsruhe, Germany
Sodium hydroxide (NaOH)	Merck, Darmstadt, Germany
Strep-Tactin XT 4Flow resin	IBA Lifesciences, Göttingen, Germany
PowerUp SYBR Green Master Mix	Applied Biosystems, Waltham, USA
Random Hexamers	Jena Bioscience, Jena, Germany
Tetramethylethylenediamine (TEMED)	Carl Roth, Karlsruhe, Germany
Thiolutin	Biomol, Hamburg, Germany
Trichloroacetic acid ≥99 %, p.a. (TCA)	Carl Roth, Karlsruhe, Germany
Tris(2-carboxyethyl)phosphin (TCEP)	Sigma-Aldrich Chemie, Taufkirchen, Germany
Tris ultrapure	AppliChem, Darmstadt, Germany
Triton X-100	AppliChem, Darmstadt, Germany
TRIzol reagent	Thermo Fisher Scientific, Waltham, USA
Tryptone BioChemica	AppliChem, Darmstadt, Germany
Tween 20	Sigma-Aldrich Chemie, Taufkirchen, Germany
Uracil	Sigma-Aldrich Chemie, Taufkirchen, Germany

2.3. Buffers and solutions

Table 4 provides a listing of the compositions of all used buffers and solutions. Typically, self-prepared buffers and solutions were prepared with MilliQ water (Merck, Darmstadt, Germany). Stock solutions, in general, underwent autoclaving at 120°C for 20 minutes. For solutions and buffers sensitive to heat, sterile filtration through a 0.22 µm filter was employed.

Table 4: Used buffers and solutions sorted by their purpose. For self-made buffers the composition is specified. For purchased buffers the supplier is listed.

Purpose	Name	Composition/Supplier
Affinity purification of GST-tagged proteins	Binding/wash buffer	50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT
	Elution buffer	50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 10 mM reduced glutathione

Affinity purification of His-tagged proteins	Binding/wash buffer	50 mM sodium phosphate (pH 7.4), 100 mM NaCl
	Elution buffer	50 mM sodium phosphate (pH 7.4), 100 mM NaCl, 250 mM imidazole
Affinity purification of Strep-tagged proteins	Binding/wash buffer	100 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM $MgCl_2$, 0.1 % (v/v) Tween 20
	Elution buffer	100 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM $MgCl_2$, 0.1 % (v/v) Tween 20, 50 mM biotin
Agarose gel electrophoresis	6x Gel Loading Dye, Purple	NEB, Frankfurt/Main, Germany
	50x Tris-acetate-EDTA buffer (TAE)	2 M Tris (pH 8.0), 1 M Na-OAc, 50 mM EDTA
Alkaline lysis	Pre-treatment solution	7.5% β-mercaptoethanol, 1.85 M NaOH
CD spectroscopy	CD buffer	10 mM Na-Phosphate (pH 7.5)
DNase I Treatment	10x reaction buffer for DNase I + $MgCl_2$	Thermo Fisher Scientific, Waltham, USA
	50 mM EDTA	Thermo Fisher Scientific, Waltham, USA
E. coli harvest	1x Sodium Chloride-Tris-EDTA buffer (STE)	10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1 mM EDTA
EMSA	2x EMSA buffer	20 mM MES (pH 6.5), 5% glycerol, 2 mM TCEP
FRET annealing assay	Annealing buffer	20 mM MES (pH 7.0)
Gibson Assembly	5x Isothermal reaction buffer (ISO)	500 mM Tris-HCl (pH 7.5), 50 mM $MgCl_2 \cdot 7H_2O$, 1 mM dATP, 1 mM dTTP, 1 mM dCTP, 1 mM dGTP, 50 mM DTT, 25% (w/v) PEG-8000, 5 mM NAD^+
In vitro reconstitution	Reconstitution/wash buffer	50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5 mM DTT
	Elution buffer	50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5 mM DTT, 10 mM reduced glutathione or 50 mM biotin or 250 mM imidazole
Ion (anion) exchange chromatography	Binding buffer	20 mM Tris-HCl (pH 8.0)
	Elution buffer	20 mM Tris-HCl (pH 8.0), 1 M NaCl
Ion (cation) exchange chromatography	Binding buffer	50 mM MES (pH 6.0)
	Elution buffer	50 mM MES (pH 6.0), 1 M NaCl

Native PAGE	10x TBE buffer	1 M Tris-HCl (pH 8.3), 1 M boric acid, 25 mM EDTA
Reverse Transcription	M-MuLV Reverse Transcriptase Reaction Buffer	NEB, Frankfurt/Main, Germany
PCR	5x Q5® Reaction Buffer	NEB, Frankfurt/Main, Germany
	10x *Taq* buffer	100 mM Tris-HCl (pH 9.0), 0.5 M KCl, 1% (w/v) Triton X-100, 15 mM $MgCl_2$
Protease Inhibitor	100x Protease Inhibitor	8 ng/ml Leupeptin, 137 ng/ml Pepstatin A, 17 ng/ml PMSF, 0.33 mg/ml Benzamidine, solved in 100% EtOH
RNA pull-down	Equilibration buffer for beads	20 mM Tris-HCl (pH 7.5)
	RNA capture buffer	20 mM Tris-HCl (pH 7.5), 1 M NaCl, 1 mM EDTA
	Protein-RNA-Binding buffer	20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2 mM $MgCl_2$, 0.1% Tween 20
	Wash buffer	20 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2 mM $MgCl_2$, 0.1% Tween 20
	Elution buffer	20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2 mM $MgCl_2$
SDS-PAGE	4x Stacking gel buffer	0.5 M Tris-HCl (pH 6.8), 8 mM EDTA, 0.4% (w/v) SDS
	4x Separating gel buffer	1.5 M Tris-HCl (pH 8.8), 8 mM EDTA, 0.4% (w/v) SDS
	4x Laemmli SDS sample buffer	250 mM Tris-HCl (pH 6.8), 9% (w/v) SDS, 40% (v/v) glycerol, 0.2% (w/v) bromphenol blue, 100 mM DTT
	5x SDS-PAGE running buffer	125 mM Tris-HCl, 0.95 M glycine, 0.5% (w/v) SDS
	Coomassie G-250 staining solution	15% (v/v) EtOH, 85% (v/v) *Milli-Q* water, 4% (w/v) ammonium sulfate, 1.5% (w/v) β-cyclodextrin, 0.015% (w/v) Coomassie G-250
Size exclusion chromatography	SEC buffer	50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5 mM DTT
Tandem affinity purification (TAP)	TAP buffer	50 mM HEPES (pH 7.5), 200 mM KCl, 1.5 mM $MgCl_2$, 0.15% IGEPAL
	Buffer A	20 mM HEPES (pH 8.4), 150 mM KCl
	Buffer B	20 mM HEPES (pH 7.0), 150 mM KCl
Transformation of *S. cerevisiae*	Solution I	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM LiOAc
	Solution II	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM LiOAc, 40% PEG 4000
Western blot	1x TBS-T	50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1 % (v/v) Tween 20
	Semi-dry blotting buffer	25 mM Tris, 192 mM glycine, 20% (v/v) methanol

2.4. DNA and protein standards

The 1 kb Plus DNA Ladder from NEB served as the DNA standard for agarose gel electrophoresis. For SDS-PAGE protein standards, either the *Unstained Protein Standard, Broad Range (10-200 kDa)*, or the *Color Prestained Protein Standard, Broad Range (11-245 kDa)*, both from NEB, were utilized.

2.5. Oligonucleotides

All oligonucleotides shown in tables 5 to 7 were synthesized by and purchased from *Biolegio B.V.* (Nijmegen, Netherlands). The oligonucleotides employed in EMSAs, FRET annealing assays, and RNA pull-down experiments are listed in table 8 and were synthesized and acquired from *Eurogentec* (Seraing, Belgium).

Table 5: Oligonucleotides used for colony PCR and/or sequencing.

Name	Sequence (5'-3')
036	CCCGCGAAATTAATACGACTC
037	TAGAGGCCCCAAGGGGTTAT
232	AATTAACCCTCACTAAAGGG
562	GTCAGGGGGGCGGAGCCTATGGAAAAAC
867	AGGCTATCGAGAAGCGGTTCG
868	CCTGTCAAGAAGACCCTTGAAG
883	CTCCGACGGATGTTCGAAAC
1020	CTTCTTAACCTGACCGACAGACTGCTGAAAGTACTC
1407	GGGAGGAAGGTTCTGTGTGAAGGAC
PK218	GATAATATGGTTGGGGTTCAGG
PK300	GTCCCAAAACCTTCTCAAGC
PK301	GAGCAGCTTAGAGGATATCC
PK302	AGAAGTAGGGCACTTAGACC
PK303	ACTTTCTCCCTGTTGCAGAG

Table 6: Oligonucleotides used for cloning.

Name	Sequence (5'-3')
1013	AATTCGAGCTCCGTCGACAAGCTTGCGGCCG
1023	GCTAGCGCCCTGAAAATACAGGTTTTCG
1039	AATTAATAGACTGGATGGAGGCGG

1042	CCGCCTCCATCCAGTCTATTAATT
1051	CGGACACTTAAGCGACGCATAATTTTGTTTAACTTTAAGAAGGA
1052	GCGTCGCTTAAGTGTCCGGTCGACGGAGCTCGAATTC
1053	AGCGACGGCGCGCCAGACATAATTTTGTTTAACTTTAAGAAGGA
1054	GTCTGGCGCGCCGTCGCTGTCGACGGAGCTCGAATTC
1055	GACCGACCTGCAGGACGAATAATTTTGTTTAACTTTAAGAAGGA
1056	TCGTCCTGCAGGTCGGTCGTCGACGGAGCTCGAATTC
1057	ACGGCACGCCGGCGACGAATAATTTTGTTTAACTTTAAGAAGGA
1058	TCGTCGCCGGCGTGCCGTGTCGACGGAGCTCGAATTC
1059	CGCGACGTTAATTAAGCGACTGAGATCCGGCTGCTAACAAAGC
1060	GTCGCTTAATTAACGTCGCGGTCGACGGAGCTCGAATTC
1084	CTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTG
1091	CACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAG
1187	GCGCCCTGAAAATACAGGTTTTCTCCGGACTTGTACAGCTCGTCCATGC
1223	GAATTAATTGAAGGTGCAGAGTTTAGC
1252	AGAGCCGTGATGGTGGTGATGGTGATGAGAGCCTAGTTCACCTTGAAAATATAAATTTTC
1257	CTCATCACCATCACCACCATCACGGCTCTGCAGAACAGACGCTACTTTCC
1297	GCTAAACTCTGCACCTTCAATTAATTC
1301	GAAAACCTGTATTTTCAGGGCGCTAGCCACCACCACCACCACCACTGAAATTCGAGCTCCGTCGAC
1302	GAAAACCTGTATTTTCAGGGCGCTAGCTGGAGCCACCCGCAGTTCG
1304	CCCTGAAAATACAGGTTTTCGGTCAAGACATCCTCCTCCATCGGGGC
1378	CTTGTCGACGGAGCTCGAATTCAACTAGTCAGAGCACTGTCGAGAGCGGC
1379	GAAAACCTGTATTTTCAGGGCGCTAGCACCGACGCGAAGACGGAAGC
1380	GACGGAGCTCGAATTCAACTAGTGGAGTCGGCCTGCGGCTCCGCATTTTTG
1384	GCTTGTCGACGGAGCTCGAATTCAACTAGTAAGGTTCTGTGTGAAGGACG
1386	GCCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC
1387	GAAGGAGATATACCATGGGCGATAAGAAGGGCTCCTATGTCGGC
1388	CGAAAACCTGTATTTTCAGGGCGCTAGCCCCGAGCGTGTTCCTCGTAAAC
1389	CGAAAACCTGTATTTTCAGGGCGCTAGCAAGAAGCCTGCTCCTGCTGAG
1406	GTCCTTCACACAGAACCTTCCTCCCCAGGAGGCAGTGGCCGCTCTCGAC
1407	GGGAGGAAGGTTCTGTGTGAAGGAC
1408	GGTTTGGATGTCCAGTCGGAAG
1409	CTTCCGACTGGACATCCAAACCTTCGGCTTCCGTCTTCGCGTCG

1410	CTTGTCGACGGAGCTCGAATTCAACTAGTAGAAGAAGCCGGATCATCCGCGATTTTC
1416	CCTGTATTTTCAGGGCGCTAGCGGTCGCAGCAACCGCCCTGTC
1517	CTAAATTAATAGGATCCATGCTACCAAAAAAAAAAAGAAAAGTTACCGG
1518	ACCTCCGCTTACATCAACAC
1519	GCTACCAAAAAAAAAAAGAAAAGTTACCGGTATGTCTCAAGAACAGTACACAG
1520	CCTCTAGGATCAGCGGGTTTAAACGTTCATTTCCGTATCTTGTTC
1521	CTGTGTACTGTTCTTGAGACATACCGGTAACTTTTCTTTTTTTTTTTGGTAGC
1522	CTGTGTACTGTTCTTGAGACATACCGGTAACTTTTCTTTTTTTTTTTGGTAGC
1532	ACCGGTTGCATAATCCGGAACATCATACGGATAGCTAGCGCCCTGAAAATACAGGTTTTC
1533	CTATCCGTATGATGTTCCGGATTATGCAACCGGTTCGGCCGAAGAGGATCTGATCGAC
1534	CTATCCGTATGATGTTCCGGATTATGCAACCGGTTCCGGCAAGCTCGACCAGACTCTTG
1535	CTATCCGTATGATGTTCCGGATTATGCAACCGGTGTCGACTACAGCTCTATGAAAGTTC
BK169	GTGTTGATGTAAGCGGAGGT
PK291	CCCAAAAAAAAAAAGAAAAGTTACCGGTATGTCTGAAGCTCAAGAAAC
PK292	CCCTCTAGGATCAGCGGGTTTAAACTTACCTGGTTGGTGATCTTTCACG
PK293	CCCAAAAAAAAAAAGAAAAGTTACCGGTATGGCAGAACAGACGCTACTTTCC
PK294	CCCTCTAGGATCAGCGGGTTTAAACTTACCTCTGGTACCTACTGACG
PK295	GCCGTTAAAATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGAC
PK296	TAGCATGGATCCTATTAATTTAGTGTGTGTATTTGTGTTTGCGTG
PK297	ACACTAAATTAATAGGATCCATGCTAATGTCTCAAGAACAGTACACAG
PK298	GTAAAGTTAGAAGCCATTTTAACGGCGTTCATTTCCGTATCTTGTTC

Table 7: Oligonucleotides used for qPCR.

Name	Sequence (5'-3')
RPL4A qPCR fw	TTGCAATTGGCTCCAGGTGCTC
RPL4A qPCR rev	TGGGATGGCAAAGTGTAGCCGAC
RPS1A qPCR fw	TCGCCTTCACCAGAAAGCAAGC
RPS1A qPCR rev	GGCCAAGGTAGATCCTTGAACTTCC

Table 8: Oligonucleotides used for EMSA, FRET annealing assay and RNA pull-down.

Name	Sequence (5'-3')
DNA_Ma-16-Cy3	AGCACCGTAAAGACGC-Cy3
DNA_Ma-16-21-Cy5	Cy5-GCGTCTTTACGGTGCTTAAAACAAAACAAAACAAAAC
DNA_Ma-Long-top-Cy3	GTTTTGTTTTGTTTTGTTTTAAGCACCGTAAAGACGC-Cy3
RNA_Ma-16-Cy3	AGCACCGUAAAGACGC-Cy3

RNA_Ma-16-21-Cy5	Cy5–GCGUCUUUACGGUGCUUAAAACAAAACAAAACAAAAC
Biotin_Ma-16-21-Cy5	Biotin-TEG–GCGUCUUUACGGUGCUUAAAACAAAACAAAACAAAAC

2.6. Kits

Table 9 provides a list of the kits employed for the preparation of Z-competent DH5α and Rosetta 2 (DE3) cells, as well as for PCR clean-up and plasmid preparation.

Table 9: Used kits and their suppliers.

Kit	Supplier
Mix & Go *E. coli* Transformation Kit & Buffer Set	Zymo Reasearch Corp., Irvine, USA
NucleoSpin Plasmid (NoLid)	Macherey-Nagel, Düren, Germany
NucleoSpin Gel and PCR Clean-up kit	Macherey-Nagel, Düren, Germany

2.7. Enzymes

All enzymes used in this study are listed in table 10.

Table 10: All enzymes and the purpose they were used for.

Enzyme	Supplier
DNase I	Thermo Fisher Scientific, Waltham, USA
DpnI (recognition site: Gm6A^TC)	NEB, Frankfurt/Main, Germany
M-MuLV Reverse Transcriptase	NEB, Frankfurt/Main, Germany
Phusion High Fidelity DNA polymerase	NEB, Frankfurt/Main, Germany
Q5 High-Fidelity DNA polymerase	NEB, Frankfurt/Main, Germany
RNase inhibitor RiboLock	Thermo Fisher Scientific, Waltham, USA
RNase I	Thermo Fisher Scientific, Waltham, USA
RNase III	Thermo Fisher Scientific, Waltham, USA
RNase A	Thermo Fisher Scientific, Waltham, USA
RNase H	NEB, Frankfurt/Main, Germany
RNase T1	Thermo Fisher Scientific, Waltham, USA
T5 exonuclease	NEB, Frankfurt/Main, Germany
Taq DNA polymerase	Self-made
Taq DNA ligase	NEB, Frankfurt/Main, Germany
TEV protease	Self-made
Trypsin, Proteomics Grade	Sigma-Aldrich, Taufkirchen, Germany

Zymolyase 100T	Carl Roth, Karlsruhe, Germany
Zymolyase 20T	Carl Roth, Karlsruhe, Germany

2.8. Antibodies

Table 11 to 12 list the primary and secondary antibodies utilized for immunoblotting.

Table 11: Primary antibodies used for Western blot.

Name	Source	Dilution	Supplier
anti-HA	rabbit, monoclonal	1:1000	R&D Systems, Minneapolis, USA
anti-MCP	rabbit	1:5000	Merck, Darmstadt, Germany
anti-Nab2	mouse	1:5000	Swanson lab (3F2)
anti-Npl3	rabbit, polyclonal	1:5000	Guthrie lab, San Francisco, USA
anti-Pgk1	mouse, monoclonal	1:5000	Invitrogen, Waltham, USA
Peroxidase-anti-peroxidase (PAP)	rabbit, polyclonal	1:5000	Sigma-Aldrich, Taufkirchen, Germany

Table 12: Secondary antibodies used for Western blot.

Name	Source	Dilution	Supplier
anti-rabbit-HRPO	goat, monoclonal	1:3000	Bio-Rad, Hercules, USA
anti-mouse-HRPO	goat, monoclonal	1:3000	Bio-Rad, Hercules, USA

2.9. Plasmids

The plasmids created for the recombinant expression of individual genes such as *ct*Sub2 (table 13), *ct*Tho1 (table 14), and *ct*Yra1 (table 15), as well as for the co-expression of various complexes (table 16), are derived from the pET24d vector (Novagen). This vector features an f1 bacteriophage origin of replication (f1 ori) and a high-copy-number ColE1/pMB1/pBR322/pUC origin of replication (ori). Additionally, it includes a T7 promoter and terminator essential for the expression of recombinant proteins. Given the presence of a lacI gene in the vector, recombinant protein expression is triggered only upon induction with IPTG. Kanamycin serves as the selectable marker in the vector. Polycistronic plasmids carried a maximum of four genes, each preceded by a ribosome binding site. The T7 promoter was located

upstream of the initial gene, while the T7 terminator was positioned downstream of the last gene.

Table 13: Plasmids employed for expressing *ct*Sub2 recombinantly in *E. coli*. The inserted gene is implicated in the name of the plasmid. Abbreviations: GST = glutathione S-transferase, His_7 = 7x histidine-tag, TEV = tobacco etch virus protease recognition and cleavage site.

Plasmid	Source
pET24d-GST-TEV-*ct*Sub2	W. Wende
pET24d-His_7-TEV-*ct*Sub2	W. Wende
pET24d-His_7-TEV-HA-*ct*Sub2	this study
pET24d-His_7-TEV-*ct*Sub2-ΔN	this study

Table 14: Plasmids employed for expressing *ct*Tho1 recombinantly in *E. coli*. The inserted gene is implicated in the name of the plasmid. Abbreviations: His_7 = 7x histidine-tag, TEV = tobacco etch virus protease recognition and cleavage site.

Plasmid	Source
pET24d-His_7-TEV-*ct*Tho1	W. Wende
pET24d-His_7-TEV-HA-*ct*Tho1	this study
pET24d-Strep-TEV-*ct*Tho1	this study
pET24d-Strep-TEV-ΔCTE	this study
pET24d-Strep-TEV-ΔSAP-ΔLinker	this study
pET24d-Strep-TEV-ΔCTD-ΔCTE	this study
pET24d-Strep-TEV-CTD	this study
pET24d-Strep-TEV-SAP	this study
pET24d-Strep-TEV-CTE	this study
pET24d-Strep-TEV-Linker	this study
pET24d-Strep-TEV-ΔSAP-ΔCTE	this study
pET24d-Strep-TEV-ΔCTD	this study
pET24d-Strep-TEV-$\Delta\alpha_{CTE}$	this study
pET24d-Strep-TEV-$\Delta\alpha_{1\text{-}CTD}$	this study
pET24d-Strep-TEV-ΔαCTD-$\Delta\alpha_{CTE}$	this study

Table 15: Plasmids employed for expressing *ct*Yra1 recombinantly in *E. coli*. The inserted gene is implicated in the name of the plasmid. Abbreviations: His_7 = 7x histidine-tag, TEV = tobacco etch virus protease recognition and cleavage site.

Plasmid	Source
pET24d-*ct*Yra1-TEV-His_7	this study

pET24d-*ct*Yra1-TEV-Strep	this study
pET24d-His_7-mCherry-TEV-C-Box	this study
pET24d-His_7-TEV-*ct*Yra1	this study
pET24d-His_7-TEV-HA-*ct*Yra1	this study
pET24d-N-Box-TEV-mCherry-His_7	this study

Table 16: Plasmids employed for co-expressing different complexes recombinantly in *E. coli*. The inserted gene is implicated in the name of the plasmid. Abbreviations: GST = glutathione S-transferase, His_7 = 7x histidine-tag, TEV = tobacco etch virus protease recognition and cleavage site.

Plasmid	**Source**
pET24d-(His_7-TEV-*ct*Sub2)-(His_7-TEV-*ct*Yra1)-(Strep-TEV-*ct*Tho1)	this study
pET24d-(His_7-TEV-*ct*Sub2)-(His_7-TEV-*ct*Yra1)-(Strep-TEV-ΔCTE)	this study
pET24d-(His_7-TEV-*ct*Sub2)-(His_7-TEV-*ct*Yra1)-(Strep-TEV-ΔSAP-ΔLinker)	this study
pET24d-(His_7-TEV-*ct*Sub2)-(Strep-TEV-*ct*Yra1)-(His_7-TEV-*ct*Tho1)	this study
pET24d-(His_7-TEV-*ct*Tex1)-(His_7-TEV-*ct*Thoc5)-(GST-TEV-*ct*Mft1)-(His_7-TEV-*ct*Hpr1)	this study
pET24d-His_7-TEV-*ct*Tho2-ΔC(1566)	W. Wende
pET24d-(Strep-TEV-HA-*ct*Tho1)-(His_7-TEV-HA-*ct*Tho1)-(His_7-TEV-HA-*ct*Sub2)	this study
pET24d-(*ct*Sub2)-(His_7-TEV-*ct*Tho1)-(*ct*Yra1)	this study
pET24d-(*ct*Sub2)-(Strep-TEV-*ct*Tho1)-(*ct*Yra1)	this study

The plasmids designed for the expression in *S. cerevisiae* are constructed using the pET296-YcpLac111 (Addgene #53249) or the pRS315 (Sikorski & Hieter 1989) vector as a foundation. To replicate the plasmid in *E. coli*, both vectors are equipped with a bacterial origin of replication (ori) and utilize ampicillin as the selectable marker. To facilitate the correct transmission and replication of the plasmid in yeast, the vectors incorporate the yeast CEN centromeric sequence and the yeast ARS origin of replication region. To facilitate selection in yeast, both vectors carry the gene for 3-isopropylmalate dehydrogenase, essential for leucine biosynthesis (LEU2 marker). The plasmid pET264-pUC 24xMS2V6 Loxp KANr Loxp (Addgene #104393) was employed for the genomic insertion of 24xMS2V6 into the 3'UTR of RPS1A. The plasmid pSH47 was utilized for expressing the Cre recombinase and excising the KANr marker following the insertion of the 24xMS2V6 loops. Table 17 lists all yeast plasmids utilized in this study.

Table 17: Plasmids employed for expressing or cloning different genes in *S. cerevisiae*. The inserted gene is implicated in the name of the plasmid. Abbreviations: MCP = MS2-coat protein.

Plasmid	Source
pET264-pUC-24xMS2V6	Addgene #104393
pET296-YcpLac111 (empty vector)	Addgene #53249
pET296-YcpLac111-MCP-GFP	Addgene #104394
pET296-YcpLac111-MCP	P. Keil
pET296-YcpLac111-MCP-Npl3	this study
pET296-YcpLac111-MCP-Tho2	this study
pET296-YcpLac111-Nab2	this study
pET296-YcpLac111-Nab2-MCP	this study
pET296-YcpLac111-Npl3	this study
pET296-YcpLac111-Tho2	this study
pRS315	Sikorski & Hieter 1989
pRS315-PTH-Tho2	this study
pRS315-PTH-Tho2-ΔC	this study
pSH47	Euroscarf

2.10. Strains

The DH5α *E. coli* strain was utilized for routine cloning, while the Rosetta 2 (DE3) *E. coli* strain was employed for the expression of recombinant proteins (table 18). The *S. cerevisiae* strains used in this study are listed in table 19.

Table 18: *Escherichia coli* strains used for general cloning and recombinant protein expression.

Strain	Genotype
DH5α	*F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 Φ80dlacZΔM15 Δ(lacZYA-argF) U169, hsdR17(r_K^- m_K^+), λ–*
Rosetta 2 (DE3)	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)*

Table 19: *Saccharomyces cerevisiae* strains used in this study.

Strain	Genotype	Reference
BY4743	*MAT a; his3Δ0; leu2Δ0; LYS2Δ0; ura3Δ0*	Euroscarf
RPS1A-MS2 (BY4743)	*MAT a; his3Δ0; leu2Δ0; LYS2Δ0; ura3Δ0; RPS1A-MS2*	this study
Δ*tho2* Hpr1-FTpA (W303)	*MAT α; YNL139C::kanMX Hpr1-FTpA::HIS*	B. Keil
W303	*MAT a; ura3-1; trp1-1; his3-11,15; leu2-3,112; ade2-1; can1-100; GAL+*	Thomas & Rothstein 1989

2.11. Media

E. coli cells (DH5α, Rosetta 2) were grown in LB medium. Z-competent cells were cultivated in SOB medium. SOC medium was used for the recovery of transformed *E. coli* competent cells. *S. cerevisiae* cells were grown in YPD or SDC medium. All used media and their composition are listed in table 20. The media were prepared with *MilliQ* water (Merck, Darmstadt, Germany) and autoclaved (120°C, 20 min) before use.

Table 20: Used media and their composition.

Name	Composition	Reference
LB	1% (w/v) peptone, 0.5% (w/v) yeast extract, 85.6 mM NaCl, pH 7.2	Bertani 1951
SDC	0.67% (w/v) yeast nitrogen base; 0.06% (w/v) complete, synthetic mix of aa; drop out as required; 2% (w/v) glucose; (2% (w/v) agar for plates)	Werner-Washburne et al. 1996
SOB	2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, pH 6-7	Hanahan 1983
SOC	2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose, pH 6-7	Hanahan 1983
YPD	2% (w/v) peptone; 2% (w/v) glucose; 1% (w/v) yeast extract; (2% (w/v) agar for plates)	Sherman 1991

3. Methods

Most of the following methods are derived from or based on the collected edition *Current Protocols in Molecular Biology* from Ausubel et al. (1994).

3.1. Microbiological techniques used for *E. coli*

If not stated otherwise, *E. coli* cells were grown in LB medium or on LB-agar plates. Depending on which resistance gene the plasmid carried, the corresponding antibiotic was added to the media with a final concentration of 25 µg/ml for selection.

3.1.1. Preparation of Z-competent *E. coli* cells

For the preparation of Z-competent DH5α and Rosetta 2 (DE3) cells, the *Mix & Go E. coli Transformation Kit & Buffer Set* (Zymo Reasearch) was used according to manufacturer's instructions.

3.1.2. Transformation of *E. coli* cells

For each transformation reaction, a 50 µl aliquot of Z-competent cells was thawed on ice. Then, 100 ng of plasmid DNA was added and gently mixed. The cells were incubated for 20 min on ice, before 400 µl of SOC medium was added to each transformation. The cells were incubated for at least 1 h at 37°C and 200 rpm, so that positive clones can begin expressing antibiotic resistance genes that are present on the transformed plasmids. After incubation, the cells were concentrated by centrifugation at 710×g for 1 min. Most of the supernatant was decanted and the pellet was gently resuspended in the remaining medium (~100 µl). Finally, the cells were plated onto pre-warmed agar plates and grown over night at 37°C.

3.1.3. Preparation of plasmid DNA

Clones that were tested positive by colony PCR (chapter 3.3.2.) were used to inoculate a 3 ml culture. The cultures were grown over night at 37°C and 200 rpm. For extraction of plasmid DNA, the kit *NucleoSpin Plasmid (NoLid)* (Macherey-Nagel) was used according to manufacturer's instructions. The concentration of extracted plasmid DNA was measured with a NanoDrop 1000 Spectrophotometer (Thermo Scientific).

3.1.4. Protein expression in *E. coli*

For protein expression, Rosetta 2 (DE3) cells were transformed (chapter 3.1.2.). The transformed cells were then used for inoculation of a starter culture, which was grown over night at 37°C and 200 rpm. The next day a main culture was inoculated with 1:100 of starter culture. The culture was grown in a baffled flask at 37°C and ~160 rpm for 2-3 h until an optical density of OD_{600nm} = 0.6 was reached. For protein expression, cells were induced with 1 mM IPTG. Cells were grown overnight at 20°C and 130 rpm. The culture was harvested by centrifugation at 5,000×g for 5 min. The pellet was resuspended in 1x STE buffer and centrifuged again at 3,000×g and 4°C for 10 min. The supernatant was decanted and the pellet either stored at -20°C until further use or processed immediately.

3.2. Microbiological techniques used for *S. cerevisiae*

If not stated otherwise, *S. cerevisiae* cells were grown in YPD or the appropriate SDC medium.

3.2.1. Preparation of competent *S. cerevisiae* cells

For five to six transformations of *S. cerevisiae*, an overnight pre-culture was used to start a 50 mL main culture at an OD_{600nm} = 0.2. After incubation at 30°C and 170 rpm, the cells were harvested at 2,800×g for 5 min when they reached mid-log phase at an OD_{600nm} = 0.6 to 0.8. Then, the cell pellet was washed once with 10 mL water and centrifuged again at 2,800×g for 5 min. Afterwards, the cell pellet was washed with 500 µl of solution I and transferred to a 2 mL tube. After centrifugation at 2,800×g for 3 min, the cells were finally resuspended in 250 µl of solution I.

3.2.2. Transformation of *S. cerevisiae* cells

For each transformation, 50 µl of competent *S. cerevisiae* cells were used. The cells were mixed with 5 µl single-stranded carrier DNA (2 mg/mL) and 300 µL solution II. Then, either 500 ng of plasmid DNA was added or linearized DNA from a 300 µl PCR reaction, that was precipitated with ethanol (chapter 3.4.6.), was added for the integration into the yeast genome. As a negative control a transformation without DNA was performed. After the cell mixture incubated for 30 min on a turning wheel at room temperature, the cells were heat shocked for 10 min at 42°C and

immediately afterwards incubated on ice for 3 min. Prior to centrifugation of the cell suspension at 1,200×g for 3 min, 1 ml of water was added to dilute solution II. For more difficult transformations, like the integration of linearized DNA into the genome, the cells were incubated in 1 ml YPD medium for 2 hours at 30°C and 170 rpm. After recovery, the cells were centrifuged at 1,200×g for 3 min. The cell pellet was resuspended in 100 µl water and plated out on selective SDC plates. When plasmid DNA was used for transformation, cells were directly plated out without recovery. The cells were incubated for two to four days at 30°C. Finally, single colonies were picked, re-streaked on a fresh SDC plate and incubated over night at 30°C.

3.2.3. Treatment of *S. cerevisiae* cells with thiolutin

An over-night pre-culture was used to start a 150 ml main culture at an OD_{600nm} of 0.2. The culture incubated at 30°C and 170 rpm. When the culture reached an OD_{600nm} of 0.5, 8 ml of the culture were harvested by centrifuging at 1,200×g and 4°C for 3 min. The pellet was washed once with H_2O and centrifuged again at 1,200×g and 4°C for 3 min. After discarding the supernatant, the pellet was flash frozen in liquid nitrogen and stored at -80°C. To stop the transcription, the remaining culture was treated with 15 µg/ml thiolutin. The culture was then continued to be incubated at 30°C and 170 rpm. After 10, 30 and 40 min additional 8 ml of culture were harvested as described above.

3.2.4. Dot spot assay

Single colonies of yeast cells, that were freshly transformed (chapter 3.2.2.), were collected using a small loop and suspended in 1 ml of water. To standardize the cell count for all spotted strains, the OD_{600nm} was measured and diluted to 0.2. Three consecutive 10-fold serial dilutions were prepared for each strain. Subsequently, 7.5 µl of each strain and its dilutions were spotted onto the respective media plates, air-dried, and then incubated at 16°C, 25°C, 30°C, and 37°C for a maximum of 7 days. The assay was repeated with three biological replicates of each strain.

3.3. Basic molecular biology techniques

3.3.1. Polymerase chain reaction (PCR)

PCR was performed in a peqSTAR XS (VWR) or a TPersonal 48 (Biometra) thermocycler. The annealing temperature was set according to melting temperature of the used primers minus 5°C and the selected time for elongation was dependent on the length of the expected PCR product. In general, Q5 High-Fidelity DNA polymerase (NEB) was used for PCR. The protocol *PCR Using Q5 High-Fidelity DNA Polymerase (M0491)* (NEB 2023) was used for setting up the PCR reaction and defining the thermocycling conditions. The PCR products were analyzed by agarose gel electrophoresis (chapter 3.3.4.).

3.3.2. Screening of *E. coli* clones

To screen for recombinant clones after transformation, a colony PCR using *Taq* DNA polymerase was performed. For this purpose, up to 10 colonies from each transformation plate were picked with a sterile toothpick, transferred to 20 µl H_2O and subsequently streaked on a new agar plate. The colonies were then screened by PCR using vector-specific forward and reverse primers that amplified over the inserted gene. Occasionally, insert-specific primers were additionally used for screening. The protocol *PCR Protocol for Taq DNA Polymerase with Standard Taq Buffer (M0273)* (NEB 2023) was used for setting up the PCR reaction and defining the thermocycling conditions.

3.3.3. Screening of *S. cerevisiae* clones

After transformation of a PCR product and re-streaking single colonies on a fresh plate, the *S. cerevisiae* clones were screened for correct genomic integration of the linearized DNA. Before the colony PCR was performed, the yeast cells were lysed using zymolase. For lysis, a small amount of yeast colony was scraped from the plate und resuspended in 15 µl of 2.5 mg/ml zymolase solution. After incubation for 20 min at room temperature, the cells were put into a thermocycler and incubated for 5 min at 37°C followed by incubation for 5 min at 95°C. Finally, 60 µl of H_2O were added to the cell suspension and 5 µl were used as template in each PCR reaction. The colonies were screened by using forward and reverse primers that amplified over the inserted DNA sequence. The protocol *PCR Protocol for Taq DNA*

Polymerase with Standard Taq Buffer (M0273) (NEB 2023) was used for setting up the PCR reaction and defining the thermocycling conditions.

3.3.4. Nucleic acid electrophoresis

In general, agarose gel electrophoresis was used for analysis of PCR products. For preparation of an agarose gel, 1% (w/v) agarose was dissolved in 1x TAE buffer and subsequently heated in a microwave. Before the lukewarm gel was poured into a gel tray, Intas HDGreen™ Plus DNA stain was added. After the gel had completely solidified, it was covered with 1x TAE buffer. Prior to loading, DNA samples were mixed with 6x Purple Gel Loading Dye (NEB). As a molecular weight marker the 1 kb Plus DNA Ladder from NEB was used. After loading the gel, a voltage between 80 and 120 V was applied to separate the DNA fragments according to their size. The DNA bands were finally visualized in an *Intas Gel iX* Imager by UV light.

To analyze the nucleic acid content in a protein purification, native acrylamide gel electrophoresis was performed. For running a native 6% TBE gel, a Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad) filled with 1x TBE buffer was used. Prior to loading, the samples were mixed with 6x Purple Gel Loading Dye (NEB). The gels were run for 80 min at a voltage of 100 V and an electric current of 100 mA. As a molecular weight marker the 1 kb Plus DNA Ladder from NEB was used. For staining, the gel was incubated for 10 min in ethidium bromide solution. The gel was then destained with tab water for 10 min and finally analyzed with the gel documentation system *BioDocAnalyze* (Biometra).

3.3.5. Plasmid cloning

Gibson Assembly was employed to clone plasmids harboring a single gene, as well as to perform cloning of polycistronic plasmids. In the case of polycistronic plasmids, a ribosome binding site was inserted downstream of each individual gene. The insert and backbone fragments were amplified using Q5-PCR (chapter 3.3.1) and then subjected to DpnI digestion at 37°C for one hour. The purification of PCR products was carried out using the *NucleoSpin Gel and PCR Clean-up* kit (Macherey-Nagel), following the instructions provided by the manufacturer. Gibson Assembly was performed according to *Gibson Assembly Protocol (E5510)* (NEB 2023). The recombinant plasmids were finally transformed into DH5α cells (chapter 3.1.2.).

3.3.6. Sequencing of plasmid DNA

After colony PCR, plasmid DNA was further analyzed by sequencing. Therefore, DNA template solution containing 40-100 ng/µl plasmid were mixed with 3 µl 20 µM sequencing primer solution in a total volume of 15 µl. For sequencing either a vector-specific (forward or reverse primer from colony PCR) or an insert-specific primer was used. Sequencing was done by the sequencing service *Microsynth Seqlab* (Göttingen, Germany).

3.4. Basic biochemical techniques

3.4.1. SDS-PAGE and Coomassie G-250 staining

For SDS-PAGE a Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad) filled with 1x SDS-PAGE running buffer was used. Prior to loading, the samples were mixed with Laemmli SDS sample buffer, incubated for 5-10 min at 95°C and centrifuged at max. speed for 1 min. The gels were run at maximum voltage of 300 V and an electric current of 50 mA per gel until the dye front reached the bottom of the gel. As a marker either the *Unstained Protein Standard, Broad Range (10-200 kDa)* or the *Color Prestained Protein Standard, Broad Range (10-250 kDa)* from NEB was used. For staining of the proteins, the gel was incubated for at least 5 min in Coomassie G-250 solution. The gel was then destained with tab water.

3.4.2. Western blot

Proteins were separated using SDS-PAGE, as outlined in the preceding chapter. Subsequently, the proteins were transferred from the SDS-PAGE gel to a nitrocellulose membrane (Porablot NCL, Macherey-Nagel) using the semi-dry blotting device *Trans-Blot Turbo Transfer System* (Biorad). Blotting of the proteins was performed at 25 V for 45 min. After blocking the membrane with 5% milk powder dissolved in TBST buffer, the membrane was incubated over night with the specific primary antibody at 4°C. On the following day, the membrane underwent three 10-minute washes with TBST buffer. Subsequently, the membrane was incubated with the appropriate secondary antibody for 1 hour at room temperature. Following a series of three 10-minute washes with TBST buffer, the western blot was detected using *CheLuminate-HRP ECL* solution (Applichem). The *ChemoCam* Imager (Intas) was used to image the chemiluminescence signals.

3.4.3. Protein identification by mass spectrometry

For the identification of proteins, the samples were first analyzed by SDS-PAGE and Coomassie G-250 staining (chapter 3.4.1.). Subsequently, individual protein bands were excised from the gel. The analysis by mass spectrometry was done by *Protein Analytics* (Prof Dr. G. Lochnit, Gießen, Germany).

3.4.4. Alkaline lysis of yeast cells

To analyze protein amounts in yeast cells, 10 ml of yeast culture were harvested when the culture reached an OD_{600nm} of 0.6. The cell pellet was then resuspended in 500 µl cold ddH_2O and subsequently mixed with 150 µl of pre-treatment solution which contained 7.5% β-mercaptoethanol and 1.85 M NaOH. After incubation on ice for 20 min, proteins were precipitated (chapter 3.4.5.).

3.4.5. Precipitation of proteins

To precipitate one volume of protein solution, 1/10 volume of 100% trichloroacetic acid (TCA) was added. The mixture was thoroughly vortexed, before incubation for at least 20 min on ice. To pelletize the proteins, the solution was centrifuged at 10,000×g for 20 min at 4°C. The supernatant was removed and the pellet resuspended in Laemmli SDS sample buffer. The pH was adapted with 2 M Tris. The sample was incubated at 95°C for 2-5 min and centrifuged at maximum speed for 1 min, before it was analyzed by SDS-PAGE (chapter 3.4.1.) or by western blot (chapter 3.4.2.).

3.4.6. Precipitation of DNA

To precipitate one volume of DNA, three volumes of 100% ethanol and 1/10 volume of 3 M NaOAc (pH 5.2) were added to the sample. After incubation for at least 20 min at -20°C, the solution was centrifuged at maximum speed and 4°C for 15 min. After washing the pellet once with 70% ethanol and centrifuging again at maximum speed and 4°C for 10 min, the pellet was air dried and dissolved in 20 µl H_2O.

3.4.7. Isolation of RNA

To isolate total RNA from *S. cerevisiae*, an over-night pre-culture was used to start a main culture at an OD_{600nm} = 0.2. When the cells reached an OD_{600nm} of 0.5, 10 ml of culture were harvested. Then, 1 ml of TRIzol reagent was added to one cell pellet. After vortexing and incubation for 5 min on ice, 200 µl chloroform were added. The

sample was vortexed again and incubated for 2-3 min at room temperature. For separation of proteins and nucleic acids, the sample was centrifuged at 12,000×g and 4°C for 15 min. The upper aqueous phase, which contains the RNA, was transferred to a new tube. To precipitate the RNA, 500 µl of ice cold isopropanol were added. After vortexing, the sample was incubated for 10 min at room temperature. To obtain a RNA pellet, the sample was centrifuged at 12,000×g and 4°C for 10 min. The pellet was washed once with 1 ml of ice cold 75% ethanol, followed by centrifugation at 7,500×g and 4°C for 5 min. The supernatant was discarded and the remaining RNA pellet was air dried. The RNA was dissolved in the appropriate volume of DEPC-treated RNase-free water or buffer. Contaminating DNA was removed by digestion with DNase I at 37°C for 30 min. The digestion was stopped by adding 2.5 mM EDTA and incubation for 10 min at 65°C. The concentration of extracted total RNA was measured with a NanoDrop 1000 Spectrophotometer (Thermo Scientific).

3.4.8. Quantitative reverse transcription PCR (RT-qPCR)

For quantification of RNA in a sample, the RNA template had to be converted into cDNA first. Therefore, 100-200 ng of RNA were mixed with random hexamer as well as oligo(dt) primers and dNTPs. After annealing the sample for 5 min at 65°C, M-MuLV Reverse Transcriptase (NEB) was added. The mixture was incubated for 5 min at room temperature, followed by an incubation for 1 h at 42°C. The M-MuLV enzyme was inactivated by incubating the sample for 10 min at 65°C. The generated cDNA was then precipitated by adding 100% ethanol. After centrifuging for 20 min at maximum speed and 4°C, the cDNA pellet was washed once with 80% ethanol. The sample was centrifuged again at maximum speed and 4°C for 15 min. Then, the pellet was dried in a CHRIST evaporation machine at 42°C and 1750 rpm for 10 min. The cDNA was dissolved in 50 µl H_2O and a 10x dilution was used for qPCR. For one qPCR reaction, 2.5 µl of sample were mixed with 5 µl *PowerUp SYBR Green Mastermix* (ThermoFisher Scientific), 2.3 µl H_2O, 1 µM forward primer and 1 µM reverse primer. The *QuantStudio 3 cycler* (Applied Biosystems) was used to perform the qPCR. Every sample was measured as a technical triplicate. For verification of primer specificity, the primer melting curve was analyzed. Calculation

of Ct values was done by using the *QuantStudio Design and Analysis Software v1.4* (Applied Biosystems).

3.5. Protein purification techniques

3.5.1. Affinity purification of recombinant proteins

In general, recombinant proteins were purified with the *ÄKTA start system* (GE Healthcare). Depending on the specific affinity tag of the protein, the system was equipped with either a GSTrap HP 1-mL affinity column (Cytiva), a Strep-Tactin XT 4Flow high capacity 1 mL column (iba) or a HiTrap TALON Crude 1 mL column (Cytiva). The pellet that was harvested after protein expression (chapter 3.1.4.) was resuspended in the lysis buffer suitable for the specific column and supplemented with protease inhibitor. Then, the cells were lysed with a Branson Sonifier 250 ultrasound disruptor. Afterwards, the lysate was pre-cleared by centrifugation at 7,000×g for 10 min at 4°C and finally cleared by ultra-centrifugation at 36,700×g for 20 min at 4°C. Before starting the purification, the run parameters were set with the ÄKTA start software *UNICORN start* (GE Healthcare). First, the column was equilibrated with the appropriate binding buffer. Then the sample was applied using the pump at a flow rate of 1 ml/min. The column was washed with 8 to 10 CVs of binding buffer, before an isocratic elution with 10 to 20 CVs was performed. GST-tagged proteins were eluted with buffer containing 10 mM reduced glutathione, Strep-tagged proteins with buffer containing 50 mM biotin and His-tagged proteins with buffer containing 250 mM imidazole. Elution fractions of 1 ml were automatically collected with the fraction collector Frac30. At last, the column was equilibrated with 3 CV of binding buffer at a flow rate of 1 ml/min. The purified proteins were analyzed by SDS-PAGE (chapter 3.4.1.).

3.5.2. Ion exchange chromatography of recombinant proteins

Depending on the purity of the protein, the affinity purification was followed up by a second step of purification using an ion exchange column with the *ÄKTA start system* (GE Healthcare). Based on the charge of the particular protein, either a HiTrap SP 1 ml cation exchange column (Cytiva) or a HiTrap Q HP 1 ml anion exchange column (Cytiva) was employed. Because ion exchange chromatography relies on the interactions between charged protein molecules and oppositely

charged stationary phase materials, the distinct isoelectric points (pI) of each protein were taken into account (table 21). The pH of the buffer solution was then adjusted accordingly.

Table 21: The isoelectric points of all purified recombinant proteins. The isoelectric point was calculated with *SnapGene*.

Protein	Isoelectric point
GST-*ct*Sub2	5.83
His-*ct*Sub2	6.12
His-*ct*Sub2(ΔN)	6.75
His-*ct*Tho1	9.67
Strep-*ct*Tho1	9.84
Strep-*ct*Tho1(ΔCTE)	4.58
Strep-*ct*Tho1(ΔSAP-ΔLinker)	11.30
Strep-*ct*Tho1(ΔCTD-ΔCTE)	4.35
Strep-*ct*Tho1-CTD	7.16
Strep-*ct*Tho1-SAP	5.83
Strep-*ct*Tho1-CTE	11.70

Protein	Isoelectric point
Strep-*ct*Tho1-Linker	4.15
Strep-*ct*Tho1(ΔSAP-ΔCTE)	4.45
Strep-*ct*Tho1(ΔCTD)	9.62
Strep-*ct*Tho1($\Delta\alpha_{CTE}$)	9.46
Strep-*ct*Tho1($\Delta\alpha_{1\text{-}CTE}$)	9.44
Strep-*ct*Tho1(ΔCTD-$\Delta\alpha_{CTE}$)	8.38
***ct*Yra1-His**	10.40
***ct*Yra1-Strep**	10.37
N-Box-mCherry-His	6.44
His-mCherry-C-Box	5.62

Prior to applying the sample to an ion exchange column, the sample was desalted. After equilibrating the appropriate ion exchange column with buffer containing no salt, the sample was applied using the pump at a flow rate of 1 ml/min. The column was washed with 5 CV of buffer, before a salt gradient elution was performed. The gradient was generated by using the ÄKTA start mixer, which forms the gradient by mixing a no salt buffer with a buffer containing 1 M NaCl. Elution fractions of 1 ml were automatically collected with the fraction collector Frac30. At last, the column was equilibrated with 3 CV of no salt buffer at a flow rate of 1 ml/min. The purified elution fractions were analyzed by SDS-PAGE (chapter 3.4.1.).

3.5.3. Size exclusion chromatography of recombinant proteins

Following affinity purification and potentially ion exchange chromatography, size exclusion chromatography (SEC) was utilized as the concluding purification step. The ÄKTA purifier system (GE Healthcare) was used to perform SEC. Depending on the molecular weight of the applied sample, either a Superdex 75 Increase 10/300 GL (Cytiva), a Superdex 200 Increase 10/300 GL (Cytiva), or a Superose 6 Increase 10/300 GL (Cytiva) size exclusion column was selected. After the column

was equilibrated with the appropriate SEC buffer, the sample was applied using the sample loop. The SEC run was performed for 1.5 column volumes at a flow rate of 0.5 ml/min. Elution fractions of 1 ml were automatically collected with the fraction collector Frac-950. The elution was tracked by measuring the UV absorption at 280 nm in a chromatogram. The fractions corresponding to a peak at 280 nm and those adjacent to the peak were subjected to analysis using SDS-PAGE (chapter 3.4.1.).

3.5.4. Purification of co-expressed S-Y-T complex

After the co-expression of the trimeric S-Y-T complex containing His-*ct*Sub2, His-*ct*Yra1, and Strep-*ct*Tho1 (chapter 3.1.4.), the harvested pellet was resuspended in lysis buffer supplemented with protease inhibitor, lysed and subsequently cleared by ultracentrifugation (chapter 3.5.1.). When the objective was to eliminate nucleic acids from the lysate, the supernatant underwent treatment with a mixture of RNases (RNase A, RNase T1, RNase III, RNase H, RNase I) and DNase I for 30 min at 4°C on a turning wheel. Subsequently, the lysate was incubated with equilibrated Strep-TactinXT 4Flow resin (iba) for 1 hour at 4°C on a rotating wheel. After incubation, the resin was pooled by centrifugation at 500×g and 4°C for 5 min. The supernatant (flow-through) was removed, and the resin was transferred to a Mobicol "classic" column (MoBiTec). The resin was subjected to three 10-minute washes with lysis buffer at 4°C on a rotating wheel. Following each washing step, the wash was collected by centrifugation at 500×g and 4°C for 1 min. The proteins were ultimately eluted by incubating with lysis buffer containing 50 mM biotin for 30 min at 4°C on a rotating wheel. After collecting the initial elution through centrifugation at 500×g and 4°C for 1 min, a second elution step was conducted. SDS-PAGE and Coomassie G-250 staining were employed for the analysis of lysate, flow-through, wash steps, and eluates (chapter 3.4.1.). Following affinity purification, the complex was subsequently subjected SEC using the Superdex 200 Increase 10/300 GL (Cytiva) column for further purification (chapter 3.5.3.).

3.5.5. Purification of *in vitro* reconstituted protein complexes

In vitro reconstitution studies were carried out using two to three distinct recombinant proteins. Each of the individual proteins employed in the *in vitro* reconstitution was purified separately through affinity purification (chapter 3.5.1) and

ion exchange chromatography (chapter 3.5.2.). The proteins intended for assessment of their ability to form a complex *in vitro* were mixed with buffer in equimolar quantities or in two to threefold excess over stoichiometric amounts. After incubation in a Mobicol "classic" column (MoBiTec) for 1 h at 4°C on a rotating wheel, the appropriate resin was added to pull-down the reconstituted complex. The resin underwent three 10-minute washes with reconstitution buffer at 4°C on a rotating wheel. After each washing step, the wash was collected by centrifugation at 500×g and 4°C for 1 minute. In a GST pull-down, elution was carried out using reconstitution buffer containing 10 mM reduced glutathione, for Strep pull-down, elution was conducted with reconstitution buffer containing 50 mM biotin, and for His pull-down, elution was achieved using reconstitution buffer containing 250 mM imidazole. The proteins were incubated in elution buffer for 30 minutes at 4°C on a rotating wheel. Following the initial elution, which was collected by centrifugation at 500×g and 4°C for 1 minute, a second elution step was performed. Input, flow through, washing steps and eluates were finally analyzed by SDS-PAGE and Coomassie G-250 staining (chapter 3.4.1.).

3.5.6. RNA pull-down

The RNA pull-down is based on the instructions for the *Pierce Magnetic RNA-Protein Pull-Down Kit* (Thermo Fisher Scientific). A 37-nucleotide long RNA, labeled with biotin at the 5' end, was employed for the pull-down. First, 200 pmol of RNA was immobilized onto 50 µl of equilibrated Streptavidin Dynabeads (Thermo Fisher Scientific) in RNA capture buffer. Following a 20-minute incubation at room temperature on a rotating wheel, the supernatant was removed. Subsequently, 100 µl of Protein-RNA-Binding buffer containing RNase inhibitor was used for equilibration of the beads. A master mix was prepared, consisting of 10 µl of 10x Protein-RNA-Binding buffer, 30 µl of 50% glycerol, 15 to 20 µl of protein, and 30 to 45 µl of DEPC water. The Protein-RNA-Binding buffer was then removed from the beads, and 100 µl of the master mix was added to the beads. Following a 1-hour incubation at 4°C on a rotating wheel, the supernatant was removed. The beads underwent three washes with 100 µl of wash buffer. In the end, the complex was eluted using 50 µl of a buffer containing RNases. The elution was performed for

30 min at 4°C on a rotating wheel. Input, flow through, washing steps and eluate were finally analyzed by Western blot (chapter 3.4.2.).

3.5.7. Tandem Affinity Purification (TAP) of yeast TREX complex

A sequence coding for a FLAG-TEV-2xProtein-A (FTpA) tandem affinity tag was inserted C-terminally into the endogenous *HPR1* locus into the *S. cerevisiae* strain RS453 (MATa) using standard yeast genetic techniques (chapter 3.2.2.). Yeast were grown in YPD to an OD_{600nm} of 3.5, harvested by centrifugation and frozen in liquid nitrogen. Frozen yeast pellets were lysed with a cryo-mill (SPEX) and stored at -80°C. The grindate was resuspended in TAP buffer containing additional protease inhibitor mix. The lysate was pre-cleared at 4,200×g for 10 min at 4°C, followed by ultracentrifugation of the supernatant at 132,600×g for 1 h at 4°C. The clear supernatant was transferred to a 50 ml conical centrifuge tube, which was subsequently rotated with 600 µl IgG beads (Cytiva) for 1 h or for 16-18 h at 4°C. Beads were separated from the lysate by centrifugation at 700×g for 3 min at 4°C and washed twice with TAP buffer containing additional 1 mM DTT. For elution, beads were rotated with TEV protease in TAP buffer for 1 h at 16°C. Following centrifugation, the resulting supernatant (TEV eluate) was rotated with 450 µl FLAG beads (Sigma-Aldrich) in buffer A or buffer B (20 mM HEPES, pH 7.0; 150 mM KCl) for 1 h at 4°C. Beads were washed once with buffer A or B containing additional 0.15% IGEPAL. For elution, the sample was rotated with 100 µg/ml FLAG peptide (Sigma-Aldrich) for 15 min at room temperature.

3.5.8. Crosslinking of purified yeast TREX complex

The purified *S.c.* TREX complex was concentrated and subsequently cross-linked with either bis(sulfosuccinimidyl)suberate (BS^3-d_0/d_{12}, Creative Molecules) in buffer A or with adipic dihydrazide (ADH-d_0/d_8, Sigma-Aldrich) and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium (DMTMM) chloride (Sigma-Aldrich) in buffer B for 1 h at 37°C and 500 rpm using a ThermoMixer (Eppendorf). The BS^3 cross-linking reaction was stopped by incubation with 50 mM ammonium bicarbonate (NH_4HCO_3) for 20 min at 37°C at 500 rpm in a ThermoMixer. The ADH/DMTMM cross-link reaction was stopped by desalting with a Zeba Spin Desalting Column (ThermoFisher Scientific). The cross-linked samples were flash frozen in liquid nitrogen and stored at -80°C until further processing.

3.6. Structural studies of purified proteins and complexes

3.6.1. Circular Dichroism Spectroscopy

Circular Dichroism (CD) spectroscopy was utilized to assess the secondary structure content of the purified recombinant proteins. Each protein was measured in 10 mM Na-Phosphate solution at pH 7.5 using a cuvette with a 0.5 mm pathlength. Ultimately, the far-UV CD spectra were monitored using a J-710 Circular Dichroism Spectrophotometer (Jasco). The machine units represented in millidegrees of ellipticity were transformed into mean residue molar ellipticity using the following equation (where n is the number of peptide bonds and $\mathrm{Ellipticity}$ is the raw data from the instrument):

$$\theta(deg^1 \cdot cm^2 \cdot dmol^{-1}) = \frac{\mathrm{Ellipticity(mdeg)} \cdot 10^6}{\mathrm{Pathlength(mm)} \cdot [Protein](\mu M) \cdot n}$$

3.6.2. Limited proteolysis with trypsin

The co-expressed S-Y-T complex was analyzed after purification (chapter 3.5.4.) by limited proteolysis. Approximately 10 µg of complex were either treated with 1 µg or with 5 µg of trypsin at room temperature. The trypsin digestion was stopped after 1, 2, 5, 10, 30, 60, 90, and 120 minutes by adding protease inhibitor. The samples from each time point were analyzed by SDS-PAGE and Coomassie G-250 staining (chapter 3.4.1.). The sample after 120 minutes of incubation with 5 µg trypsin was analyzed by Liquid chromatography coupled with mass spectrometry (LC-MS) for intact protein quantification (Prof. Dr. Robert Tampé, Goethe University Frankfurt).

3.6.3. Crystallization of recombinant proteins

Following affinity purification (chapter 3.5.1), ion exchange chromatography (chapter 3.5.2), and size exclusion chromatography (chapter 3.5.3), the individual recombinant His-*ct*Sub2 and His-*ct*Tho1 were subjected to crystallization testing. Furthermore, the *in vitro* reconstituted His-*ct*Sub2/Strep-*ct*Tho1 and His-*ct*Sub2(ΔN)/Strep-*ct*Tho1(ΔCTE) complexes were also assessed for crystallization. Therefore, an automated robotic system was used to screen various crystallization conditions (Michaela Stumpf & Karin Fritz-Wolf, Justus-Liebig University Gießen).

3.6.4. Cross-linking mass spectrometry (XL-MS) of yeast TREX complex

For the cross-linking mass spectrometry (XL-MS) analyses, 40 µg of purified yeast TREX complex each were cross-linked in three independent experiments with (1) 0.5 mM BS^3, (2) 1 mM ADH and 1 mM DMTMM or (3) 2.5 mM ADH and 2.5 mM DMTMM (chapter 3.5.7. and 3.5.8.). The BS^3 and ADH-DMTMM cross-linked endogenous TREX complex was processed for cross-linking mass spectrometry (XL-MS) essentially as described in Kern et al. (2023). The sample processing, liquid chromatography-tandem mass spectrometry, and the identification of cross-linked peptides were performed by Dr. Alexander Leitner (ETH Zurich).

3.6.5. Structure determination by negative stain EM

For negative stain EM, 20 µg of uncross-linked, endogenous yeast TREX complex was used (chapter 3.5.7.). Additionally, 20 µg of BS^3 cross-linked and 20 µg of ADH-DMTMM cross-linked TREX complex (chapter 3.5.7. and 3.5.8.) were analyzed by negative stain EM. The processing of the samples and data analysis is essentially described in Kern et al. (2023) and was performed by Dr. Christin Radon (University of Potsdam).

3.7. Functional studies on purified proteins and complexes

3.7.1. Electrophoretic Mobility Shift Assay (EMSA)

The Electrophoretic Mobility Shift Assay (EMSA) was used to detect the interaction between a protein and nucleic acids. The nucleic acids were labeled with either a Cy3 fluorescent dye at the 3’ end or a Cy5 fluorescent dye at the 5’ end. In general, 5 µM of protein was incubated with 50 nM nucleic acid in buffer containing 10 mM MES pH 6.5, 5% glycerol and 2 mM TCEP. After incubating at room temperature for 20 min, 10 µl of the sample was loaded onto a pre-cooled native 10% TBE gel. The gel was run in 1x TBE buffer at an electric current of 100 mA and at voltage of 100 V for 80 min in the cold room (chapter 3.3.4.). After washing the gel briefly with water, the fluorescent signals were imaged with the laser scanner Typhoon FLA 9500 (GE Healthcare).

3.7.2. Fluorescence Resonance Energy Transfer (FRET) assay

To observe the DNA/DNA-, DNA/RNA- and RNA/RNA-annealing activity of different *ct*Tho1 mutants, a fluorescence resonance energy transfer (FRET) assay was performed. For this assay, two complementary nucleic acids were used. One nucleic acid was labeled with a Cy3 fluorescent dye at the 3' end, while the other one was labeled with a Cy5 fluorescent dye at the 5' end. The Tho1-promoted annealing of the two CyDye-labeled nucleic acids leads to a FRET signal. When the Cy3 donor and Cy5 acceptor are close enough for FRET to occur, it leads to an emission peak of Cy5 at around 680 nm. The more duplexes form, the higher is the increase in Cy5 emission. For this assay, an excitation wavelength of 515 nm was used. The emission spectrum ranging from 530 to 750 nm was recorded with the spectrofluorometer *Duetta* (Horiba Scientific). One reaction contained 10 nM of Cy3 labeled RNA or DNA, 10 nM of Cy5 labeled RNA or DNA and 250 nM of protein in 20 mM MES pH 7 buffer. The sample was measured without adding protein and directly after protein was added.

4. Results

This study is divided into two parts. The first focuses on the structural analysis of the TREX complex from *S. cerevisiae*, with some results published by Kern et al. (2023). The second part investigates the *ct*Sub2-*ct*Yra1-*ct*Tho1 complex from *C. thermophilum*, with particular focus on the role of *ct*Tho1.

4.1. The TREX complex in *S. cerevisiae*

By coupling transcription to mRNA export, TREX is essential for accurate and efficient gene expression and overall cellular function (Sträßer et al. 2002). Gaining structural insights into TREX is essential for understanding its diverse functions. To achieve this, the complex was purified from *S. cerevisiae* and then analyzed using a combination of cross-linking mass spectrometry (XL-MS) and negative-stain electron microscopy (EM).

4.1.1. Preparation and cross-linking of the TREX complex

For purification, the TREX subunit Hpr1 was tagged with a C-terminal FLAG-TEV-Protein A tag, and the native complex was isolated from *S. cerevisiae* using tandem affinity purification (TAP). As shown in Figure 13A, the THO subunits (Tho2, Mft1, Thp2, Tex1) co-purified with Hpr1, along with Sub2, Yra1, Gbp2, and Hrb1, with Yra1 being notably more abundant. Mass spectrometry confirmed the identity of all components. Figure 13B shows the domain structure of the nine TREX components, highlighting the unstructured C-terminal regions in Tho2 and Mft1, as well as in Hpr1, Tex1, and Thp2 to a lesser extent.

After purifying the complex, it was cross-linked either with bissulfosuccinimidyl suberate (BS^3) or with a combination of adipic dihydrazide (ADH-d_0/d_8) and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium (DMTMM) chloride. BS^3 covalently links primary amino-groups (e.g. lysines) with a distance of 11.4 Å (Fig. 13C). By using ADH-DMTMM, cross-linking with DMTMM results in zero-length cross-links (ZLXL; Fig. 13D) and cross-linking with ADH-DMTMM leads to carboxyl-carboxyl cross-linking reactions with a distance of 11.1 Å (Fig. 13E; Leitner et al. 2014).

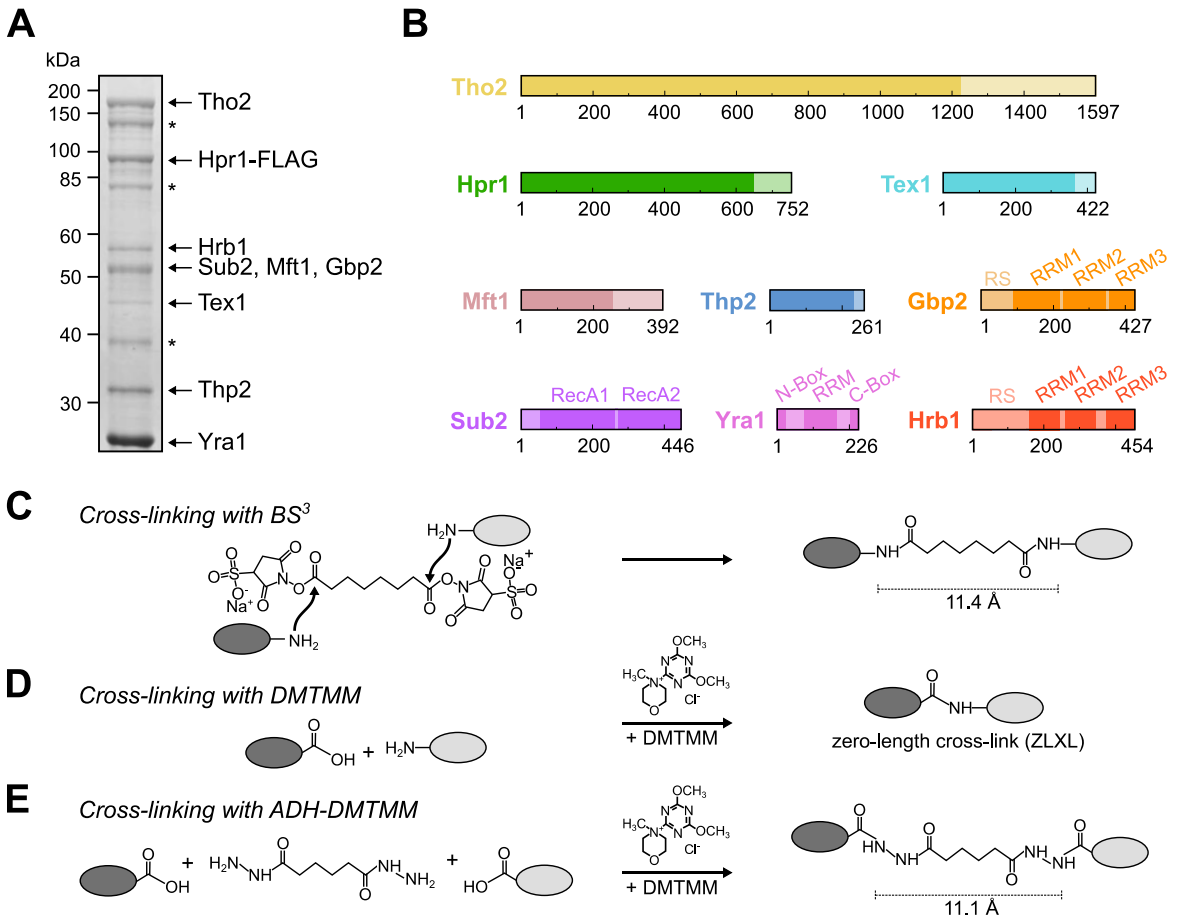


Figure 13: Preparation of the endogenous TREX complex from *S. cerevisiae*. *(A)* Representative purification of the native TREX complex by tandem affinity purification. The TREX subunit Hpr1 was endogenously tagged with a C-terminal FLAG-TEV-Protein A tag. The FLAG eluate was analyzed on a 10% SDS-PAGE gel stained with Coomassie Brilliant Blue. Subunits corresponding to Tho2, Hpr1-FLAG, Hrb1, Sub2, Mft1, Gbp2, Tex1, Thp2 and Yra1 were identified by mass spectrometry and are indicated accordingly. The asterisks indicate degradation bands. Molecular weight marker bands are indicated in kDa. *(B)* Scheme of the domain structures of all nine TREX subunits. Flexible domains that are not resolved in published TREX structures are depicted in a lighter color. *(C-E)* Used cross-linking reagents and their reaction mechanisms. *(C)* Cross-linking with BS^3 covalently links primary amino-groups (e.g. lysines) with a distance of 11.4 Å. *(D)* Cross-linking with DMTMM, a carboxyl-amine reactive cross-linker, leads to zero-length cross-links (ZLXL). *(E)* Cross-linking with ADH combined with DMTMM results in carboxyl-carboxyl cross-linking reactions with a distance of 11.1 Å. *(C-E)* Schema was modified after Leitner et al. (2014).

4.1.2. Cross-linking mass spectrometry of the yeast TREX complex

The structural architecture of the endogenous TREX complex was analyzed by cross-linking mass spectrometry (XL-MS), which was performed by Dr. Alexander Leitner at ETH Zurich. The TREX preparation, where the complex was cross-linked with BS^3 (performed by Manuel Koschitza), yielded many cross-links in the flexible regions of TREX components (Fig. 14A). In an attempt to obtain more cross-links in other regions of the TREX components, ADH-DMTMM was used as a second cross-linking chemistry (Fig. 14B).

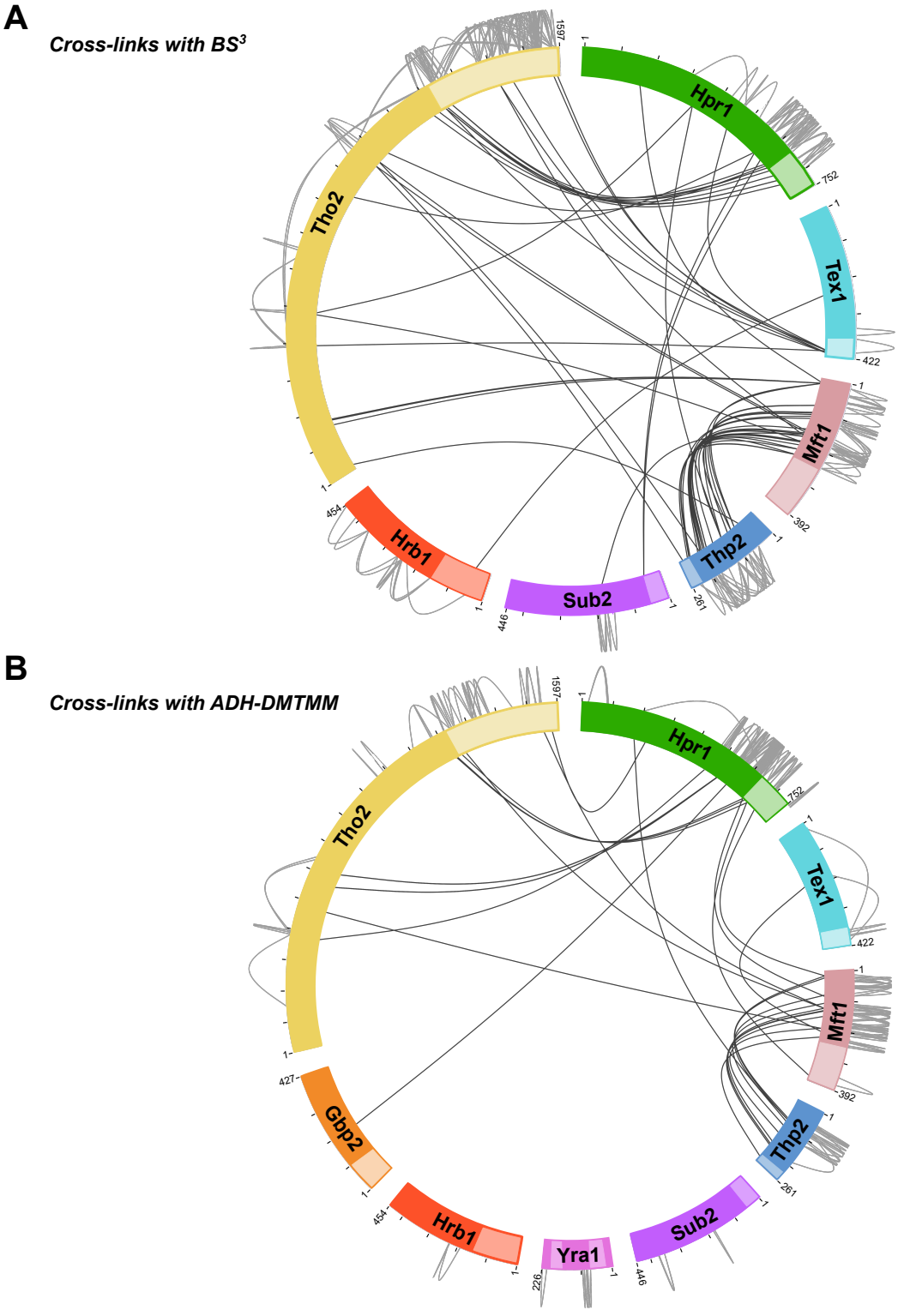


Figure 14: BS^3 and ADH-DMTMM cross-links of the TREX complex were mainly observed between the THO components. Cross-links observed between components of the *(A)* BS^3 and *(B)* ADH-DMTMM cross-linked TREX complex. Intra-molecular cross-links are shown in light grey lines outside of the circle, while inter-molecular cross-links are shown in dark grey lines inside the circle.

By employing both cross-linking chemistries, over 550 cross-linked peptide pairs were identified. Out of these, more than 445 were non-redundant site pairs involving interactions between THO components and Sub2 (Fig. 15). Most cross-links were contributed by BS^3 and DMTMM, while only few ADH-DMTMM cross-links were detected. Although Gbp2 and Yra1 were present in the BS^3 cross-linked sample, no intra- or inter-molecular cross-links were detected for these proteins (Fig. 14A). In the ADH-DMTMM cross-linked TREX sample, inter-molecular cross-links were found among all components except Hrb1, Yra1, and Sub2. Intra-molecular cross-links were present in all components except Gbp2 (Fig. 14B).

Overall, the BS^3 and ADH-DMTMM cross-links show a heterogeneous distribution across the individual TREX subunits, with most located in the C-terminal regions of Tho2 and Hpr1, as well as in the first 250 residues of Mft1 and Thp2. Cross-links between residues within the same protein can either occur within a single copy of the protein or between two separate copies of the same protein. Most cross-links connecting different subunits were found between Tho2 and Hpr1, as well as between Mft1 and Thp2 (Fig. 14).

4.1.2.1. Compatibility of cross-links with recombinant yeast TREX structures

To assess the alignment between the endogenous and the recombinant TREX complex, the BS^3 and ADH-DMTMM cross-links in Figure 14 were mapped onto the 3.4 Å cryo-EM structure by Schuller et al. (2020; PDB ID: 7APX).

Less than 40% of the cross-links (171 out of 445; Fig. 15) had both residues present in the structure. For the remaining cross-links, one or both residues were missing in the structure. This was especially apparent in the C-terminal region of Tho2, which was missing over 350 residues, and in Hpr1, where about 150 C-terminal residues were absent (Fig. 14). The extensive cross-linking in the flexible regions made it difficult to model the full-length proteins, as these regions have few connections to more structured regions or other subunits.

Of the mapped cross-links, 78% (133 out of 171) fitted the monomeric unit, while only three aligned with dimer contacts. The 35 remaining site pairs did neither fit the monomer nor dimer structure, mainly involving intra-molecular cross-links in Thp2 and Mft1 or inter-molecular connections between them.

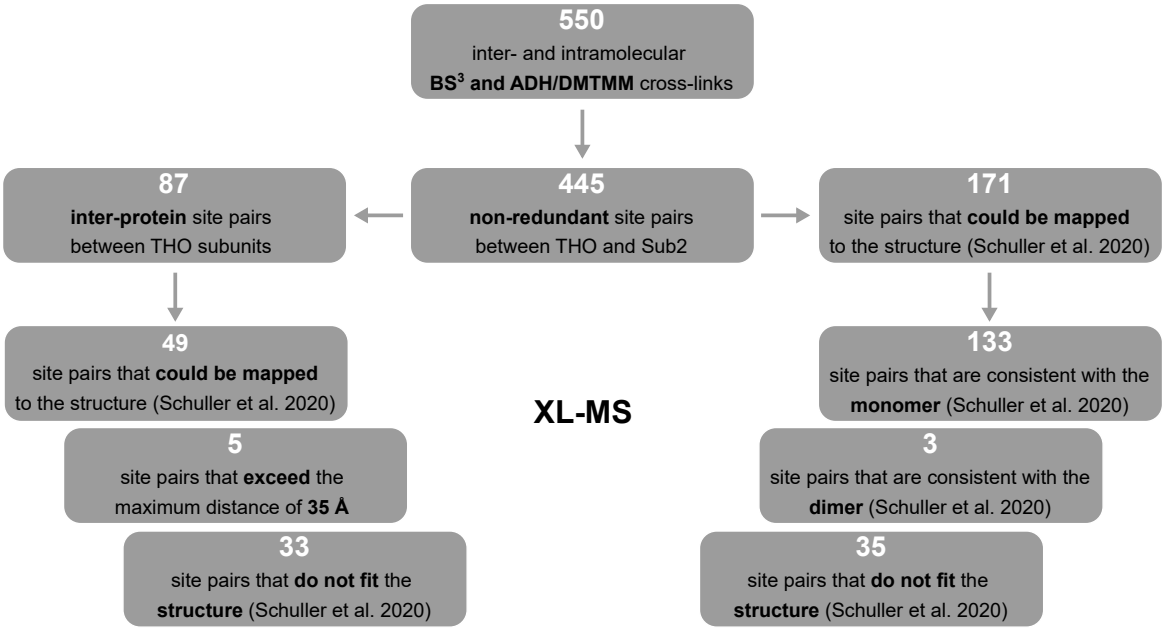


Figure 15: Overview of all observed BS^3 and ADH-DMTMM cross-links of endogenous yeast TREX complex. Among 550 cross-links occurring within and between molecules, 445 correspond to unique site-pairs without redundancy. A total of 171 of these site-pairs were matched to the structure of the recombinant yeast THO-Sub2 complex (Schuller et al. 2020). Altogether, 87 distinct non-redundant inter-protein site pairs were detected among the five THO components.

4.1.2.2. XL-MS data comparison of endogenous and recombinant THO/TREX

The BS^3/ADH-DMTMM cross-linking dataset was compared to the DSS/EDC results from Xie et al. (2021), who used XL-MS to analyze interactions in a reconstituted yeast THO-Sub2-Gbp2 complex. Xie et al. (2021) used disuccinimidyl suberate (DSS), a non-sulfonated variant of BS^3, along with the zero-length coupling reagent N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide (EDC). Figure 16 compares the inter-protein cross-links identified in both studies, showing a high degree of similarity in frequently observed subunit contacts.

A total of 87 inter-protein site pairs were identified in the BS^3/ADH-DMTMM cross-linked endogenous THO complex (Fig. 15), compared to 113 pairs in the DSS/EDC cross-linked recombinant complex. In the BS^3/ADH-DMTMM dataset, most cross-links were between Thp2 and Mft1 (40 out of 87) and Tho2 and Mft1 (10 out of 87). Xie et al. (2021) found the most frequent DSS/EDC cross-links between Tho2 and Hpr1 (34 out of 113), Tho2 and Mft1 (28 out of 113), and Thp2 and Mft1 (17 out of 113; Fig. 16). Approximately 40% (49 out of 87) of the BS^3/ADH-DMTMM cross-links mapped to the recombinant TREX complex structure (Schuller et al. 2020). Among the remaining cross-links, five exceeded the 35 Å distance limit, and

33 could not be assigned to the structure (Figs. 15 and 16). In contrast, only about 20% (21 out of 113) of the DSS/EDC cross-links were mapped successfully. Unlike the natively purified complex, all amino acid contacts from Xie et al. were compatible with the 35 Å threshold (Fig. 16).

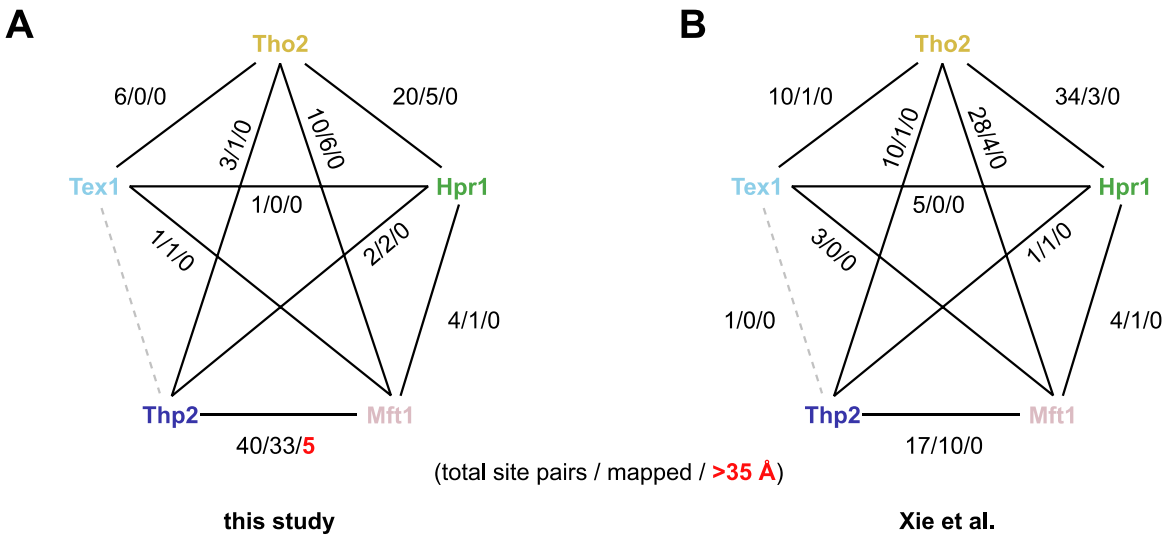


Figure 16: Comparison of inter-subunit cross-links. *(A)* BS^3/ADH-DMTMM cross-links, excluding contacts with Sub2, and *(B)* DSS/EDC cross-links observed in the recombinant complex by Xie et al. (2021). Displayed are the counts of unique cross-linking site pairs observed for each subunit pair, along with the count of contacts that could be correlated with PDB ID: 7APX, thus linking structured regions as determined by cryo-EM. Additionally, those that surpass a distance of 35 Å are highlighted in red. Figure taken from Kern et al. 2023, licensed under a Creative Commons License (Attribution-NonCommercial 4.0 International).

4.1.3. Electron microscopy of the yeast TREX complex

Over 60% (274 of 445) of the non-redundant cross-links identified in the XL-MS analysis fell within unresolved regions in the recent THO–Sub2 structures (e.g. Schuller et al. 2020). To assess whether the assembly of the endogenous TREX complex differed from the recombinant THO–Sub2 complex, freshly prepared BS^3 cross-linked, ADH-DMTMM cross-linked, and uncross-linked complexes were analyzed by negative-stain electron microscopy (UANS EM) with single-particle analysis (performed by Dr. Christin Radon, University of Potsdam) (Supplementary Fig. S1).

4.1.3.1. Structure determination of BS^3 cross-linked TREX complex

On the negative-stain grid, the freshly purified and BS^3 cross-linked TREX complex exhibited a varied and heterogeneous formation (Fig. 17A). Nevertheless, specific dimeric particles that resembled previously published 2D classes could be observed

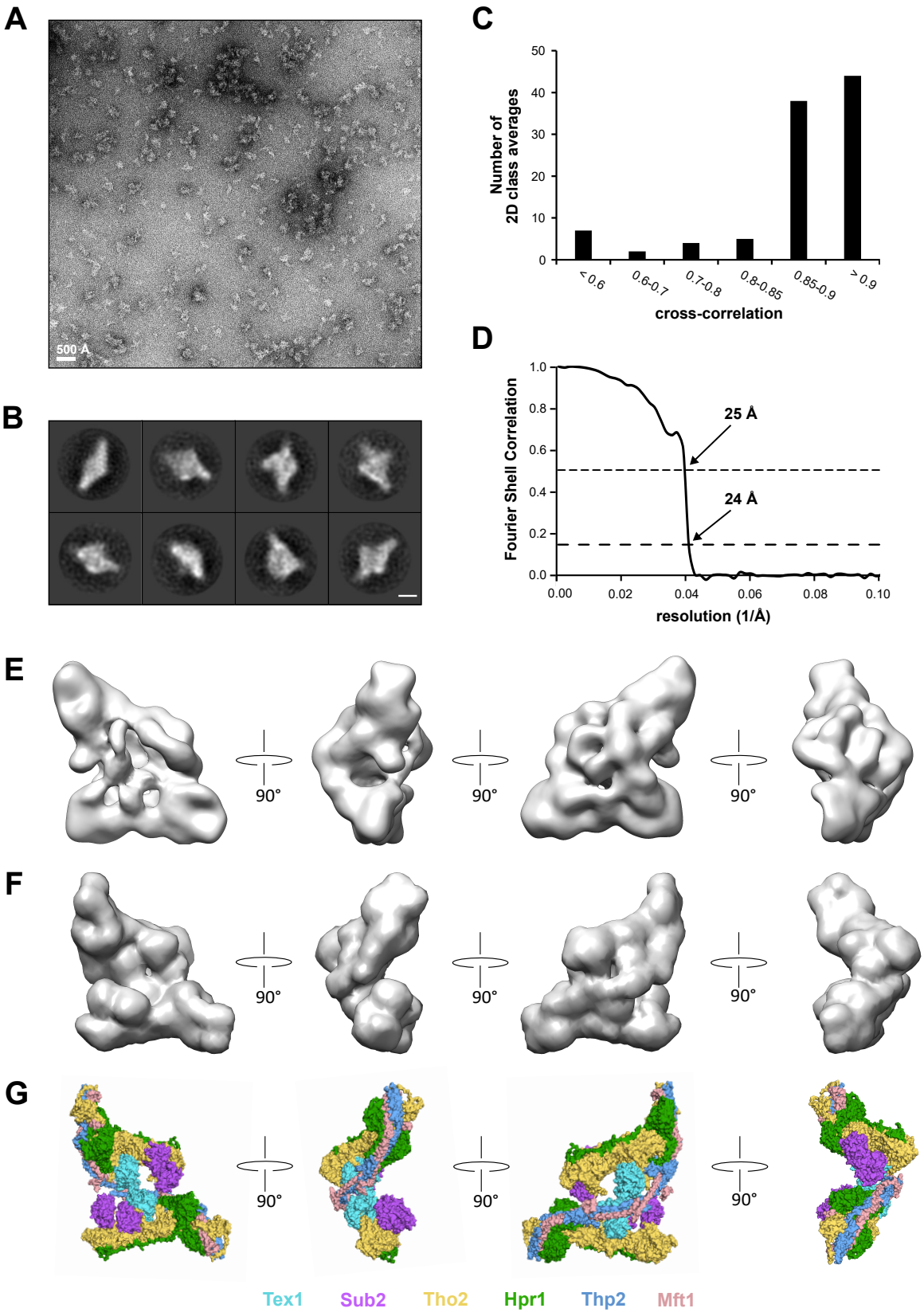


Figure 17: Negative-stain EM structure of the BS^3 cross-linked endogenous TREX complex. *(A)* A representative negative-stain micrograph of the BS^3 cross-linked *S. cerevisiae* TREX complex. Scale bar represents 500 Å. *(B)* Representative reference-free 2D class averages of the TREX complex. Scale bar represents 100 Å. *(C)* Plot showing cross-correlation coefficient of 100 2D class

averages from 42,784 particles obtained by projection matching. *(D)* Fourier shell correlation (FSC) plot for the negative-stain reconstruction of the TREX complex. The resolution at FSC = 0.143 (dashed line) and FSC = 0.5 (dotted line) are indicated. *(E)* Three-dimensional reconstruction of the TREX complex resolved to 26 Å. *(F)* Representative views of the Tho–Sub2 dimer structure (PDB ID: 7AQO) in surface representation lowpass-filtered to 25 Å. *(G)* Color-coded THO–Sub2 dimer PDB model depicted in the same views as in *(F)*. The single-particle analysis was performed by Dr. Christin Radon (University of Potsdam). Figure adapted from Kern et al. 2023, licensed under a Creative Commons License (Attribution-NonCommercial 4.0 International).

(Fig. 17B; Pena et al. 2012; Schuller et al. 2020; Xie et al. 2021). To determine the fraction of particles resembling dimers within the UANS dataset, the reference-free 2D class averages were compared with the dimeric map previously published for the yeast THO–Sub2 complex by Schuller et al. employing image cross-correlation techniques. When selecting a cross-correlation coefficient threshold that incorporates distinct dimeric views into the dataset, the fraction of particles that correlate with dimers within the UANS dataset was 50%, while an additional 27% fell within the coefficient range of 0.897 to 0.85 (Fig. 17C). Therefore, most of the particles within the dataset exhibited a dimeric resemblance. Following this, the dimeric fraction was used to generate a 3D reconstruction without applying any symmetry, resulting in an overall resolution of 26 Å (Fig. 17D and E).

The 3D reconstruction and reference-free class averages align closely with the published THO–Sub2 dimer (Schuller et al., 2020), and the yeast TREX sample resembles the *in vitro* structure in shape and organization, with minor domain orientation differences (Fig. 17E-G).

4.1.3.2. EM of ADH-DMTMM cross-linked TREX complex

The UANS EM analysis of the ADH-DMTMM cross-linked TREX complex captured 531 micrographs with a focus range of -3 to -1 µm. The sample showed some aggregation and heterogeneous particles of various sizes (Fig. 18A). A total of 1,091 particles were manually selected and boxed at 248 pixels with a 500 Å mask. Autopicking was performed on a set of 60,586 particles and three 2D class averages were used as templates. The particles were grouped into three 2D classes based on size (Fig. 18B-D). The first class contained 5,016 particles (320-350 Å; Fig. 18B), the second included 29,660 particles (250-300 Å; Fig. 18C), and the third had 14,444 particles (170-230 Å; Fig. 18D). Due to the dataset's diversity, matching 2D images across classes was challenging.

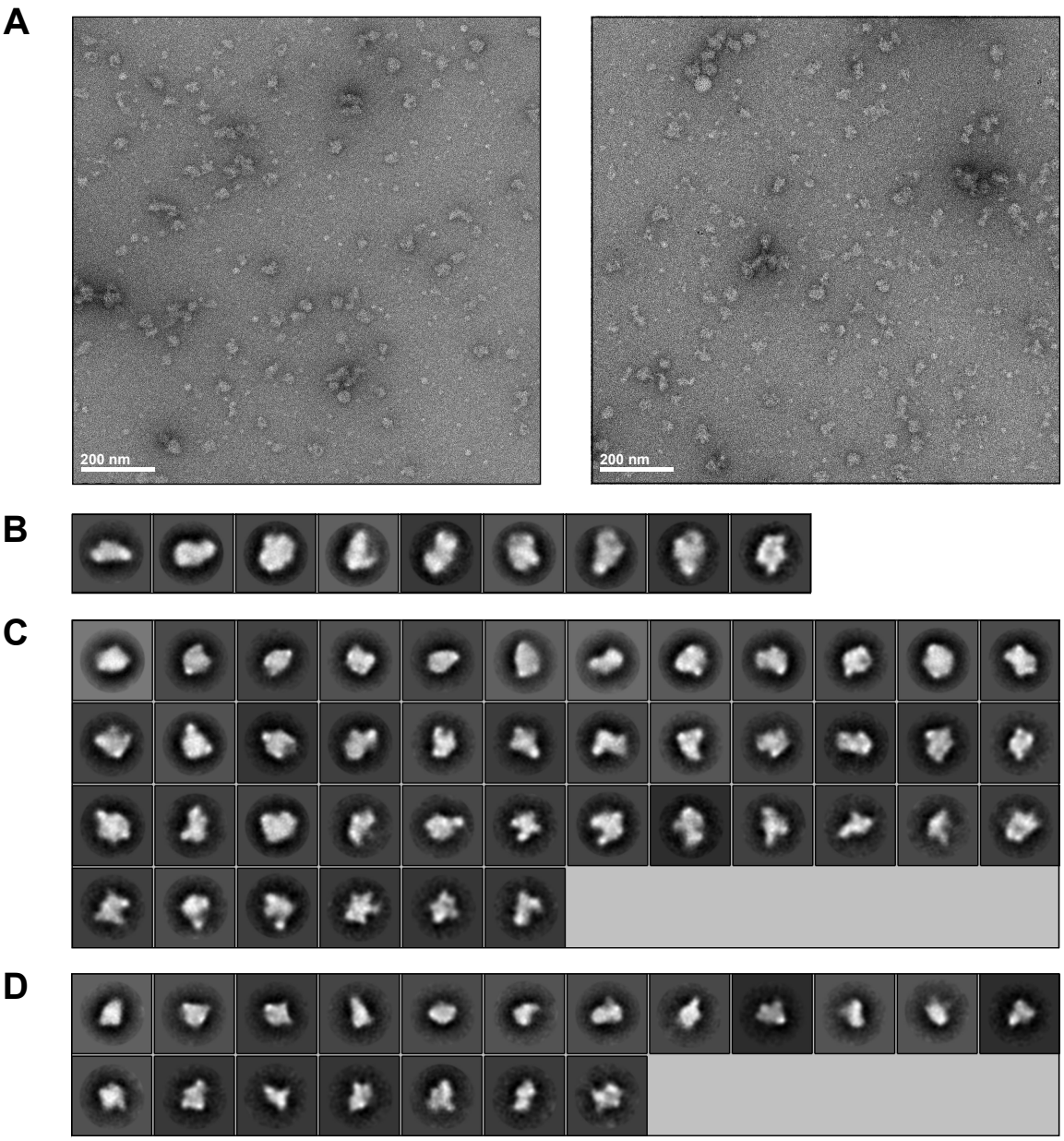


Figure 18: Negative-stain EM of the ADH-DMTMM cross-linked endogenous TREX complex. *(A)* Two representative negative-stain micrographs of the ADH-DMTMM cross-linked *S. cerevisiae* TREX complex. Scale bar represents 200 nm. *(B-C)* Three representative reference-free 2D class averages of the TREX complex. The black circle indicates 500 Å, the box size is 630 Å. 2D class averages of *(B)* 5,016 particles with a size between 320 and 350 Å, of *(C)* 29,660 particles with a size between 250 and 300 Å, and of *(D)* 14,444 particles with a size between 170 and 230 Å. The single-particle analysis was performed by Dr. Christin Radon (University of Potsdam).

4.1.3.3. EM of uncross-linked TREX complex

In addition to the cross-linked TREX complexes, a freshly purified uncross-linked complex was analyzed by UANS EM (Fig. 19). In total, 142 micrographs were recorded with a focus range of -3 to -1.5 µm. The sample showed some aggregation and particle heterogeneity similar to ADH-DMTMM cross-linked particles (Fig. 19A).

A total of 1,220 particles were manually boxed (248 pixels, 500 Å mask), and autopicking was applied to 22,806 particles using three 2D class averages as templates. After 2D classification, the best averages included 18,158 particles sized 170-350 Å (Fig. 19B).

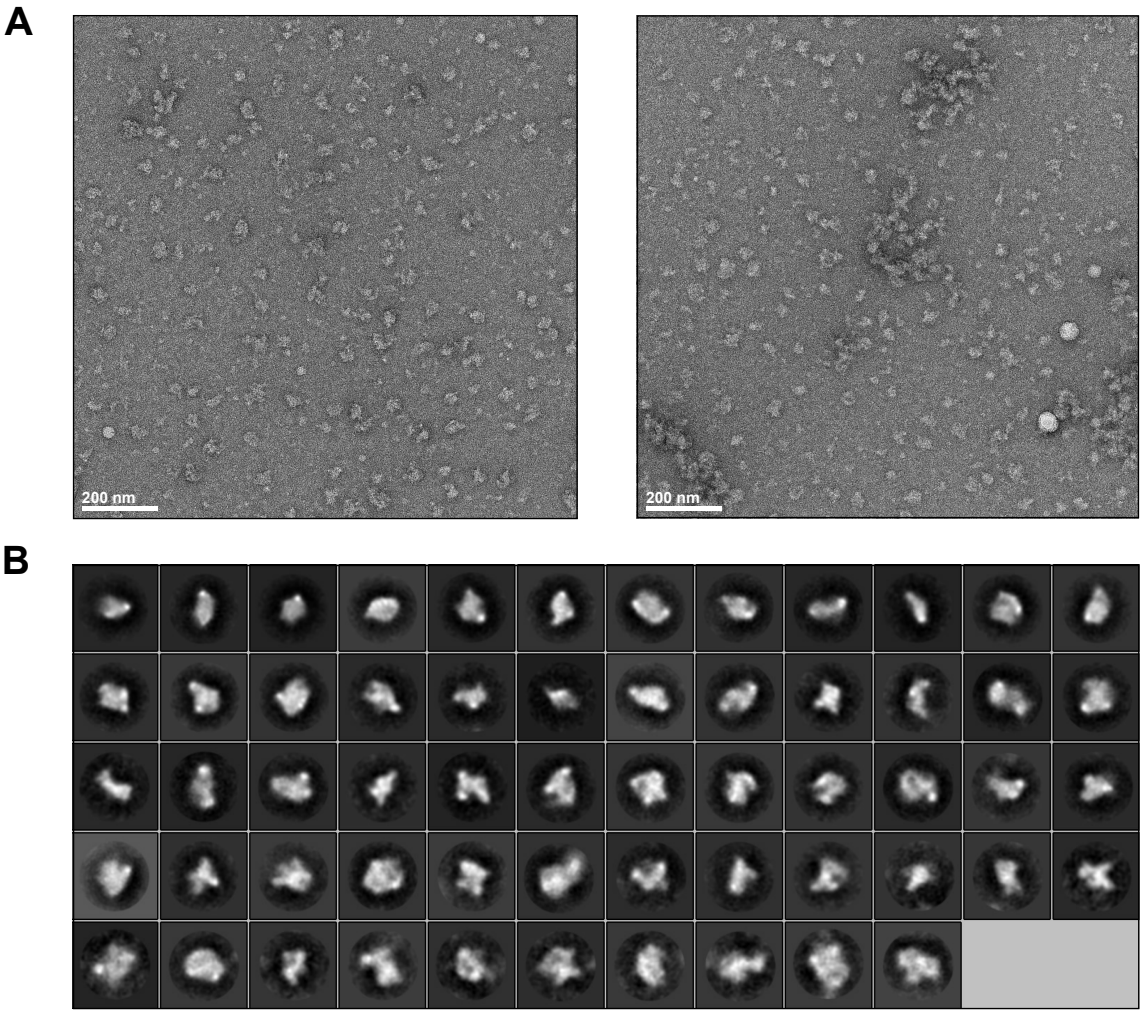


Figure 19: Negative-stain EM of the uncross-linked endogenous TREX complex. *(A)* Two representative negative-stain micrographs of the uncross-linked *S. cerevisiae* TREX complex. Scale bar represents 200 nm. *(B)* A representative reference-free 2D class average of the TREX complex. The black circle indicates 500 Å, the box size is 630 Å. The single-particle analysis was performed by Dr. Christin Radon (University of Potsdam).

4.1.4. The TREX subunit Tho2

The Tho2 protein was initially identified as a suppressor of the Transcriptional defect of Hpr1 by Overexpression (Piruat & Aguilera 1998). Tho2 represents the largest subunit within the TREX complex, forming the foundational platform of the entire complex in collaboration with Hpr1 (Schuller et al. 2020).

4.1.4.1. The unstructured C-terminal domain of Tho2

XL-MS analysis of the endogenous yeast TREX complex showed that most cross-links were located in flexible regions. In particular, the C-terminal end of Tho2 had numerous intra- and intermolecular cross-links, raising the question of whether deleting this flexible domain affects TREX assembly or cell growth.

To analyze TREX complex formation, the Hpr1 subunit was endogenously tagged with a C-terminal FLAG-TEV-Protein A tag, and the C-terminus of Tho2 was endogenously deleted (Fig. 20B). The TREX complex with either full-length Tho2 or C-terminally truncated Tho2 (tho2-ΔC) was then isolated by tandem affinity purification of tagged Hpr1. Both TEV eluates showed that THO subunits (Tho2, Mft1, Thp2, Tex1) and TREX components (Sub2, Yra1, Gbp2, Hrb1) co-purified with Hpr1 (Fig. 20A).

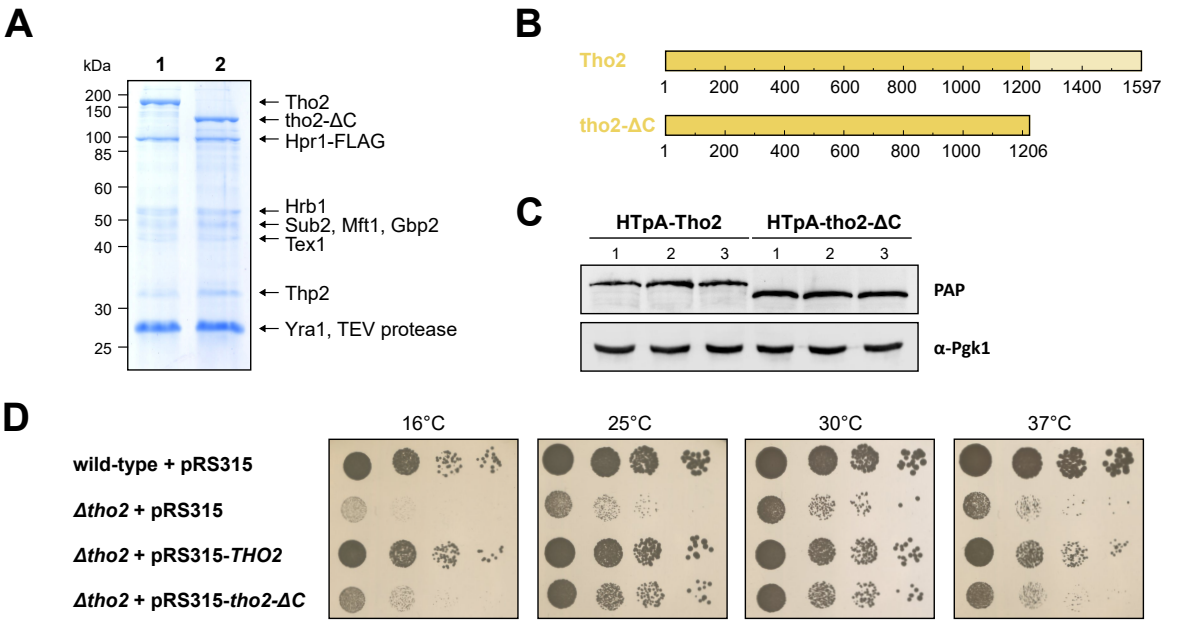


Figure 20: The unstructured C-terminal domain of Tho2 is not necessary for TREX complex formation but for growth at low and high temperatures. *(A)* SDS-PAGE of the purified endogenous *S. cerevisiae* TREX complex containing full-length Tho2 (lane 1) or C-terminally truncated Tho2, referred to as tho2-ΔC (lane 2). Molecular weight marker bands are indicated in kDa. *(B)* Schematic representation of full-length Tho2 and tho2-ΔC. *(C)* Protein levels of HTpA-Tho2 and HTpA-tho2-ΔC in whole cell extracts were determined by Western blot for three biological replicates. Pgk1 served as a loading control. *(D)* Dot spots of wild-type and the indicated THO2 mutant (W303 background). Shown are 10-fold serial dilutions that were incubated for two days at 25°C, 30°C and 37°C or six days at 16°C. Figure taken from Kern et al. 2023, licensed under a Creative Commons License (Attribution-NonCommercial 4.0 International).

To confirm that removing the C-terminal region does not affect tho2-ΔC expression compared to Tho2, Western blot analysis was conducted on whole-cell extracts

across three separate biological replicates, revealing similar protein levels for both (Fig. 20C). A dot spot assay showed that deleting the C-terminus of Tho2 caused severe growth defects at both low (16°C) and high (37°C) temperatures, comparable to those seen in Δtho2 (Fig. 20D).

4.1.4.2. Role of Tho2 in mRNA stability

To investigate the effect of Tho2 overexpression on transcript stability, the MCP-MS2 system was used. Tho2 was fused to MCP and expressed via a plasmid, while RPS1A RNA was genomically labeled with 24x MS2v6 repeats in its 3' UTR. As a control, unlabeled RPL4A RNA was used (Fig. 21). Since MCP-Tho2 was introduced via a plasmid, cells also produced untagged wild-type Tho2, essentially

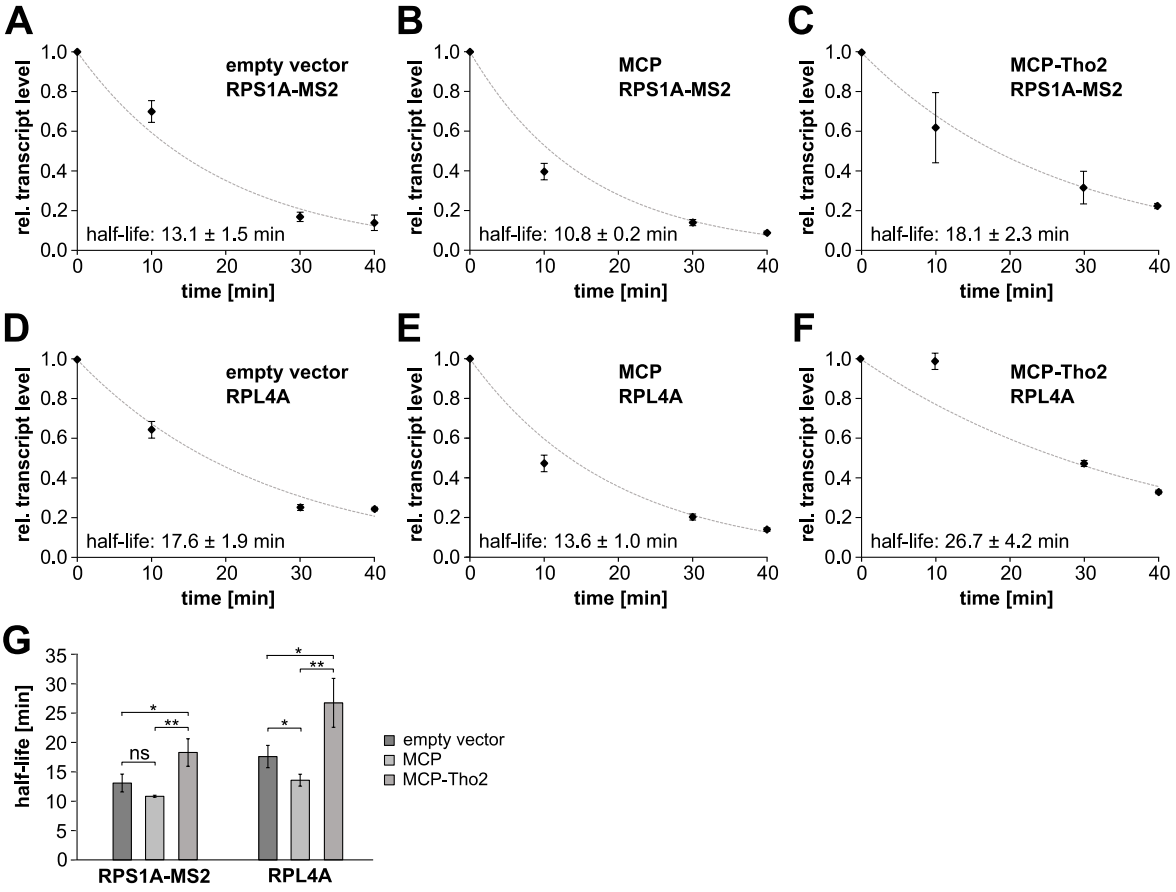


Figure 21: Overexpression of Tho2 leads to an increased half-life of transcripts. Relative transcript level of RSP1A, which is genomically labeled with 24x MS2v6 repeats in its 3' UTR, in *(A)* cells containing an empty vector, *(B)* cells containing a plasmid expressing MCP, and *(C)* cells containing a plasmid expressing MCP-Tho2. Relative transcript level of unlabeled RPL4A in *(D)* cells containing an empty vector, *(E)* cells containing a plasmid expressing MCP, and *(F)* cells containing a plasmid expressing MCP-Tho2. The calculated medium half-lives of three biological replicates are indicated in each graph. *(G)* Comparison of the medium half-lives of RPS1A-MS2, and RPL4A. The half-life of both transcripts is significantly increased in cells expressing MCP-Tho2 when compared to the empty vector ($p \leq 0.05$) and to the MCP-only control ($p \leq 0.01$).

resulting in Tho2 overexpression, which was confirmed by Western blot analysis (Supplementary Fig. S2). Yeast cells were treated with thiolutin to inhibit transcription, and RNA was extracted at various time points for qPCR analysis. This experiment was also performed with MCP-Npl3 and MCP-Nab2 (Supplementary Fig. S2 and S3).

The half-life of RPS1A-MS2 in cells with MCP-Tho2 was 18.1 ± 2.3 min, which was significantly longer than the empty vector control (13.1 ± 1.5 min, $p \leq 0.05$) and the MCP-only control (10.8 ± 0.2 min, $p \leq 0.01$) (Fig. 21A-C, G). A similar effect was observed for RPL4A, with its half-life significantly increasing from 17.6 ± 1.9 min in the empty vector control ($p \leq 0.05$) and 13.6 ± 1.0 min in the MCP-only control ($p \leq 0.01$) to 26.7 ± 4.7 min in cells expressing MCP-Tho2 (Fig. 21D-F, G).

4.2. The Sub2-Yra1-Tho1 complex from *C. thermophilum*

Building on the analysis of the complete TREX complex from yeast, the Sub2-Yra1-Tho1 subcomplex from *C. thermophilum* was studied to better understand the interaction between TREX and Tho1. *C. thermophilum* was selected as a model organism because its thermophilic nature improves protein stability, making it especially suitable for structural analyses.

4.2.1. The TREX subunit Sub2 in *C. thermophilum* (*ct*Sub2)

The ATP-binding DEAD-box RNA helicase Sub2 is a key component of the TREX complex involved in transcription elongation, mRNA export, pre-mRNA splicing, and spliceosome assembly (Jensen et al., 2001; Libri et al., 2001; Sträßer & Hurt, 2001; Zhang & Green, 2001; Sträßer et al., 2002; West & Milgrom, 2002). To highlight the similarity of Sub2 among humans, *S. cerevisiae*, and *C. thermophilum*, homolog sequences were compared using *Clustal Omega* (McWilliam et al., 2013) and *Jalview* (Waterhouse et al., 2009). The alignment shows high similarity, particularly in the Q-motif and the two RecA-like domains, while the more variable N-terminal region exhibits lower similarity (Fig. 22). These conserved regions underline Sub2's critical functions, although differences in regulatory mechanisms or domain organization may exist among organisms.



Figure 22: Sequence alignment of *H. sapiens* UAP56/DDX39B, *S. cerevisiae* Sub2 and *C. thermophilum* Sub2. The indicated Q-motif, DECD-box as well as the two RecA-like domains are highly conserved between the three species, whereas the first N-terminal amino acids are less conserved. The sequence alignment was done using *Clustal Omega* (McWilliam et al. 2013) and *Jalview* (Waterhouse et al. 2009). The *PROSITE* database (Sigrist et al. 2013) was used for assignment of domains and motifs.

4.2.1.1. Purification of *ct*Sub2

For high-resolution structural determination, *ct*Sub2 was expressed with an N-terminal His-tag in Rosetta DE3 *E. coli* cells and purified using affinity purification on a HiTrap TALON column, followed by anion exchange chromatography on a HiTrap Q column (Fig. 23A). Size exclusion chromatography on a Superdex 75 column yielded a single peak at the expected elution volume (~13 ml), and SDS-PAGE confirmed the protein's high purity and integrity, with no visible degradation bands (Fig. 23A, B).

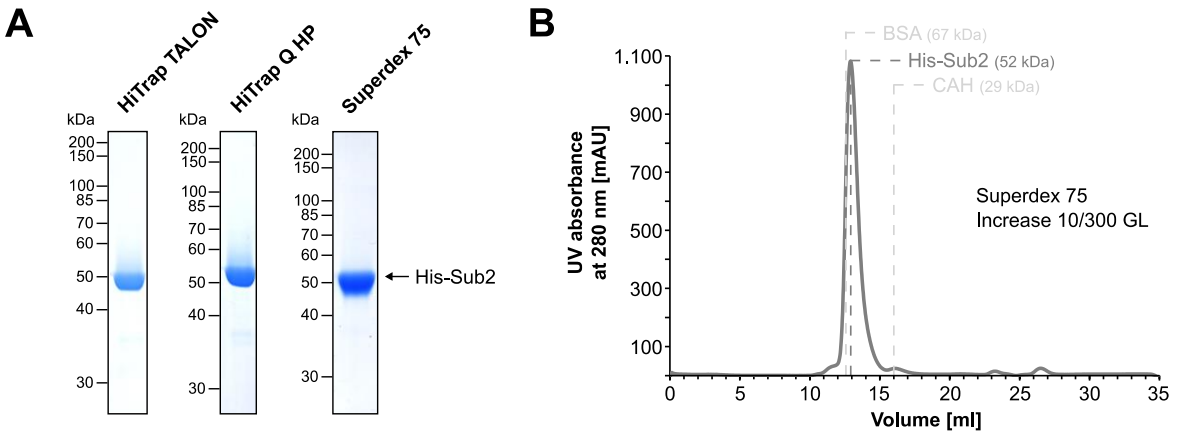


Figure 23: The purification of *ct*Sub2 yielded high amounts of pure protein. *(A)* The His-tagged *ct*Sub2 was expressed in Rosetta DE3 *E. coli* cells and purified in three steps, using an affinity

column (HiTrap TALON), an anion exchange column (HiTrap Q) and a size exclusion column (Superdex 75). The eluates were separated on a 12% SDS-PAGE gel that was stained with Coomassie Brilliant Blue. *(B)* The size exclusion chromatogram shows a single peak of His-*ct*Sub2 in its anticipated elution volume (approximately fraction 13 ml). Indicated are two marker proteins: BSA (67 kDa) that elutes approximately after the same volume as His-*ct*Sub2 and CAH (29 kDa) that elutes after approximately 16 ml.

4.2.1.2. Crystal structure of *ct*Sub2

To determine the structure of Sub2 from *C. thermophilum*, freshly purified *ct*Sub2 was crystallized at 10 mg/ml in a solution containing 1.4 M ammonium sulfate, 200 mM ammonium acetate, and 100 mM trisodium citrate (performed by Michaela Stumpf, Justus-Liebig-University Gießen). The resulting rod-shaped crystals were analyzed at the Max Planck Institute for Medical Research in Heidelberg and were solved at a resolution of 2.7 Å (done by Karin Fritz-Wolf, Justus-Liebig-University Gießen).

The first 49 residues at the N-terminus were not resolved, and the protein, free of nucleic acids, was bound only to SO_4 ions, showing *ct*Sub2 in the open conformation (Fig. 24C). The structure of *ct*Sub2 comprises two RecA-like domains characterized by central beta sheets surrounded by alpha helices. Alignment of these domains with their homologs in published yeast (Fig. 24A) and human (Fig. 24B) Sub2 structures revealed high structural similarity at the individual domain level. However, notable conformational differences were observed when comparing the overall arrangements of the RecA-like domains. In the human homolog's open conformation, the angle between RecA-like domains 1 and 2 is approximately 110°, whereas in *ct*Sub2, this angle increases to 220° (Fig. 24B-C). This suggests a more extensive conformational change is required for *ct*Sub2 to transition from open to closed states during interaction with RNA and ATP.

The transition of *ct*Sub2 from open to closed conformation was modeled using *PyMol* (Schrödinger & DeLano 2020), with the yeast structure in its closed conformation serving as a reference (Fig. 24D). This model highlights the degree of structural reorganization required for *ct*Sub2 to adopt its active state, and illustrates how this conformational change repositions the RecA-like domains to form the RNA-binding groove.

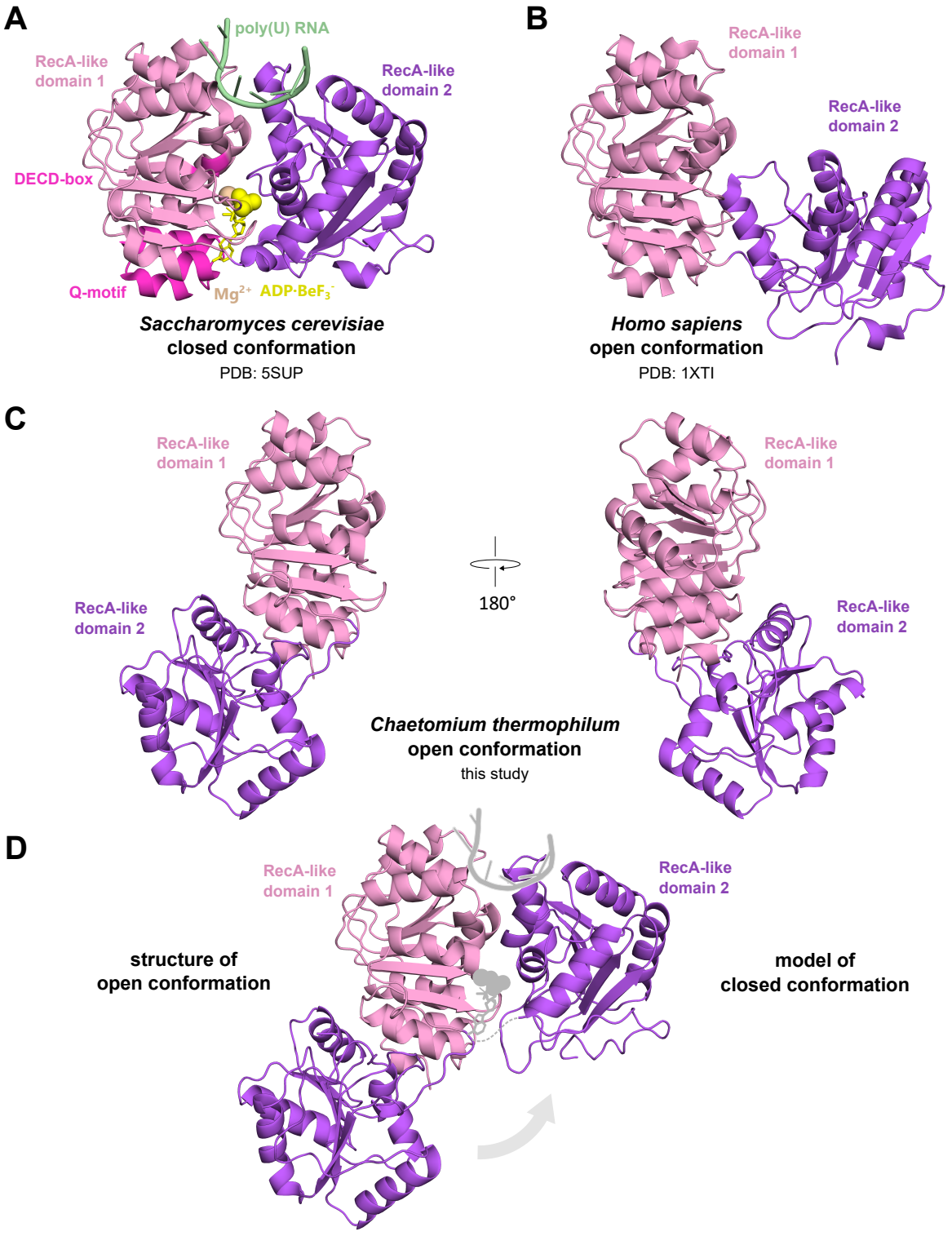


Figure 24: The crystal structure of *ct*Sub2 displays the protein in the open conformation. *(A)* The published *S. cerevisiae* structure displays Sub2 in the closed conformation, where the protein is bound to a poly(U) RNA and the ATP analog ADP·BeF_3^- (Ren et al. 2017; PDB: 5SUP). *(B)* The published human Sub2 structure shows the protein in the open conformation, where no substrates are bound (Shi et al. 2004; PDB: 1XTI). *(C)* The structure of *ct*Sub2 has a resolution of 2.7 Å and shows the protein in the open conformation. *(D)* The conformational change from open to closed state was modelled for *ct*Sub2 with *PyMol* (Schrödinger & DeLano 2020), using the yeast structure as reference.

4.2.2. The Tho1 protein in *C. thermophilum* (*ct*Tho1)

Tho1 is a nuclear protein capable of interacting with double-stranded DNA, RNA, and chromatin and has functions in transcription, mRNA export, and the assembly of ribonucleoprotein complexes. Through sequence alignment, it was determined that the human protein SARNP shares homology with the yeast Tho1 protein (Jimeno et al. 2006).

Sequence comparisons between Tho1 homologs in *H. sapiens*, *S. cerevisiae*, and *C. thermophilum* using *Clustal Omega* (McWilliam et al. 2013) and *Jalview* (Waterhouse et al. 2009) showed that the conserved N-terminal SAP domain, which binds double-stranded DNA, and the C-terminal domain, involved in RNA binding and protein interactions, are also present in *C. thermophilum*. Although these regions are conserved, functional differences may still exist between species. Overall, Tho1 has lower sequence conservation than Sub2, with less conserved linker and C-terminal regions (Fig. 25).

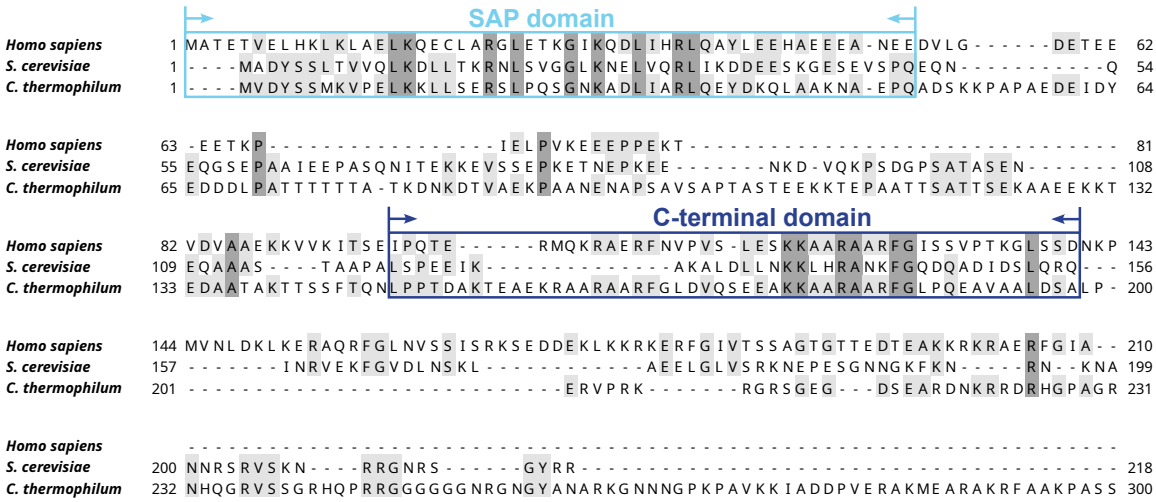


Figure 25: Sequence alignment of *H. sapiens* SARNP, *S. cerevisiae* Tho1 and *C. thermophilum* Tho1. The indicated SAP and C-terminal domains are more conserved regions between the three species, whereas the region between the two domains and at the C-terminal end are less conserved. The sequence alignment was done using *Clustal Omega* (McWilliam et al. 2013) and *Jalview* (Waterhouse et al. 2009). The *PROSITE* database (Sigrist et al. 2013) and structure predictions by *AlphaFold* (Jumper et al. 2021; Varadi et al. 2022) were used for assignment of domains and motifs.

4.2.2.1. Purification of *ct*Tho1

To obtain a high-resolution structure of *ct*Tho1, the protein was purified with the intention of crystallization, similar to *ct*Sub2. A His-tag was added to the N-terminus of *ct*Tho1, and it was expressed in Rosetta DE3 *E. coli* cells. Initial purification was

performed using a HiTrap TALON column, which showed *ct*Tho1 at ~45 kDa on an SDS-PAGE gel, slightly higher than its expected 35 kDa size. The protein was then further purified using a cation exchange HiTrap SP HP column, improving its purity (Fig. 26A). Finally, a Superdex 75 size exclusion column was used, with the protein eluting as a single peak, although earlier than expected compared to marker proteins (Fig. 26B). The final purified protein was free from visible contaminants (Fig. 26A), but attempts at crystallization were unsuccessful.

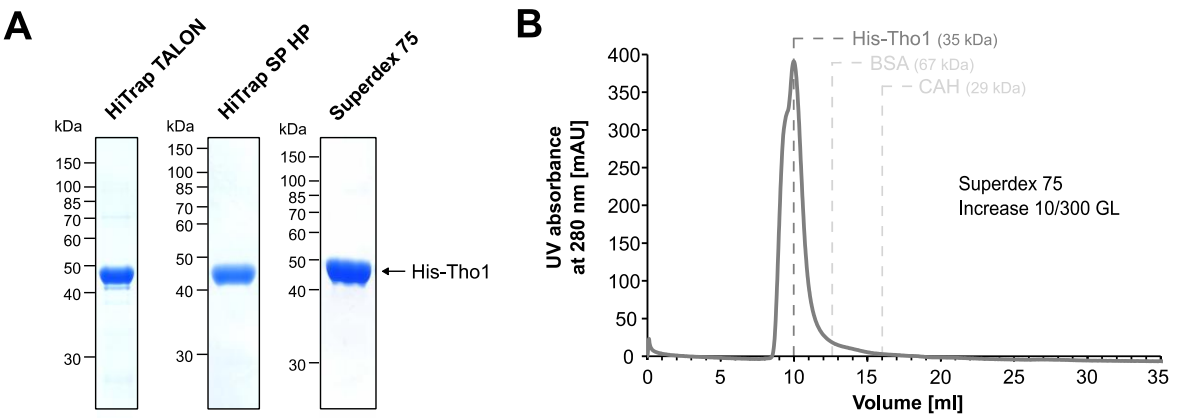


Figure 26: The purification of *ct*Tho1 yielded high amounts of pure protein. *(A)* The His-tagged *ct*Tho1 was expressed in Rosetta DE3 *E. coli* cells and purified in three steps, using an affinity column (HiTrap TALON), a cation exchange column (HiTrap SP HP) and a size exclusion column (Superdex 75). The eluates were separated on a 12% SDS-PAGE gel that was stained with Coomassie Brilliant Blue. *(B)* The size exclusion chromatogram shows a single peak of the protein. The protein eluted earlier than anticipated. Indicated are two marker proteins: BSA (67 kDa) that elutes approximately after 13 ml and CAH (29 kDa) that elutes after approximately 16 ml.

4.2.2.2. Structural analysis of *ct*Tho1

As no full-length Tho1 structures are available for *H. sapiens*, *S. cerevisiae*, or *C. thermophilum*, *AlphaFold* predictions (Jumper et al. 2021; Varadi et al. 2022) for these homologs were compared (Fig. 27).

The yeast Tho1 prediction aligns with published NMR structures, showing alpha-helical domains (Fig. 27A). The human SAP domain is similar to yeast, but its C-terminal domain differs in the 3D orientation of alpha helices. The human C-terminal end (CTE) is not as unstructured as the yeast CTE, but instead contains three alpha helices (Fig. 27B). The *C. thermophilum* SAP domain aligns reasonably well with the yeast and human SAP domain. However, the three-dimensional orientation of the alpha helices within the C-terminal domain differ when comparing it to the human and yeast domains. Like human Tho1, the *C. thermophilum* CTE is more structured

than the yeast CTE. Instead of the three additional alpha helices found in the human CTE, the *C. thermophilum* CTE contains a single alpha helix (Fig. 27C).

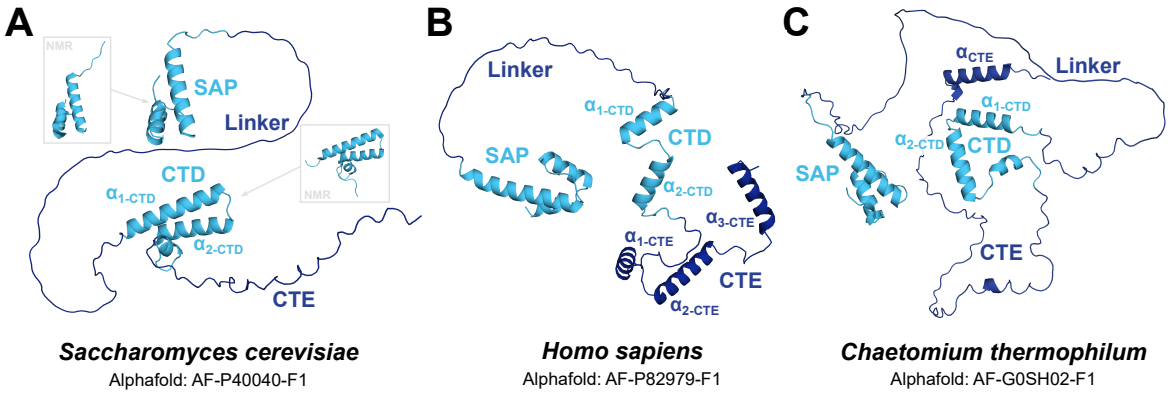


Figure 27: Structural comparison of the Tho1 proteins from *S. cerevisiae*, *H. sapiens* and *C. thermophilum*. *(A) AlphaFold* prediction of yeast (AF-P40040-F1), *(B)* human (AF-P82979-F1) and *(C) C. thermophilum* Tho1 (AF-G0SH02-F1). For all proteins, the SAP and C-terminal domains (CTD) are indicated. Additionally, the alpha helices of the CTD as well as potential alpha helices in the C-terminal end (CTE) are indicated. The NMR structures of the yeast SAP and CTD (Jacobsen et al. 2016) are shown in *(A)*. *AlphaFold* (Jumper et al. 2021; Varadi et al. 2022). Structures were visualized and processed with *PyMOL* (Schrödinger & DeLano 2020).

4.2.2.3. Generation and purification of *ct*Tho1 deletion mutants

Based on the *AlphaFold* predictions, structural differences in the CTD and CTE suggest functional variations across species. To investigate the functions of the different domains in *C. thermophilum*, twelve *ct*Tho1 deletion mutants were generated (Fig. 28A-C). In general, either entire protein domains, regions, or individual alpha helices were deleted (Fig. 28C).

For purification, all *ct*Tho1 mutants were N-terminally tagged with a Strep-tag II and expressed in Rosetta DE3 *E. coli* cells. Proteins were initially purified via affinity purification, followed by a cation exchange column to remove contaminants like negatively charged nucleic acids (RNA and DNA) (Fig. 29).

Reasonably high and pure quantities of the twelve *ct*Tho1 mutants were obtained, though *ct*Tho1(ΔSAP-ΔLinker), *ct*Tho1(ΔCTD-ΔCTE), *ct*Tho1-CTE, and *ct*Tho1-Linker mutants showed degradation bands at lower molecular weights on SDS-PAGE. Mutants containing the unstructured Linker region [(*ct*Tho1(ΔCTE), *ct*Tho1(ΔCTD-ΔCTE), *ct*Tho1-Linker, *ct*Tho1(ΔSAP-ΔCTE), *ct*Tho1(ΔCTD), *ct*Tho1(Δα$_{CTE}$), *ct*Tho1(Δα$_{1\text{-}CTD}$), and *ct*Tho1(ΔCTD-Δα$_{CTE}$)] migrated higher than expected on SDS-PAGE, relative to their calculated molecular weights (Fig. 29).

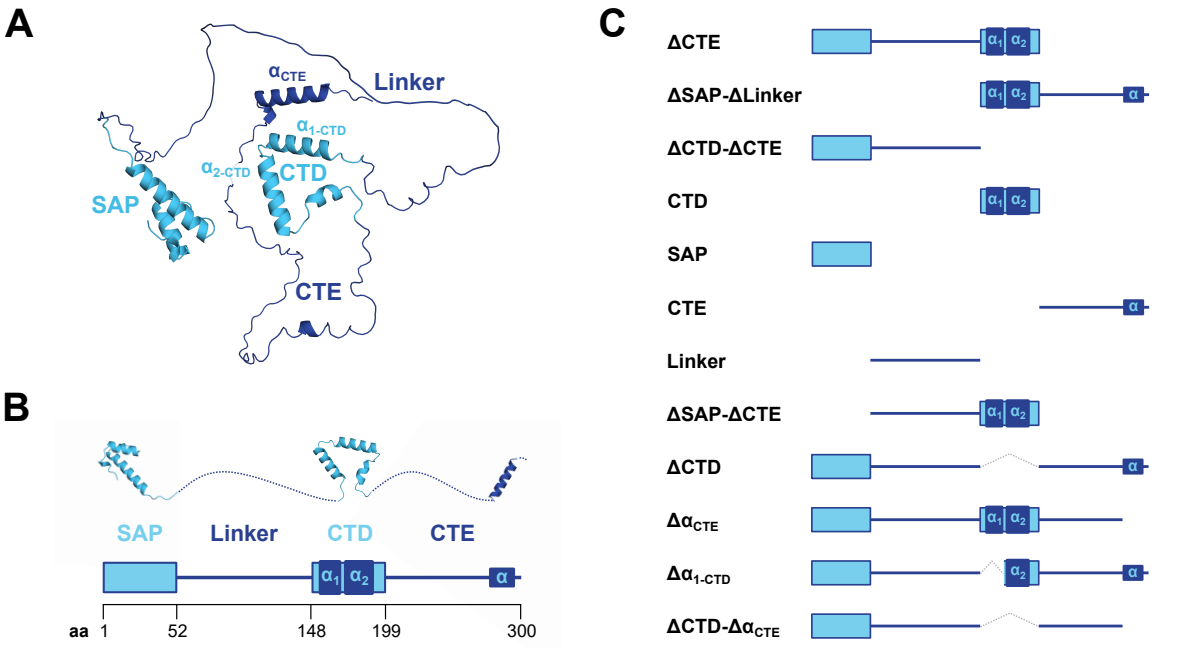


Figure 28: Various *ct*Tho1 mutants were generated based on the structure prediction. *(A) AlphaFold* prediction (Jumper et al. 2021; Varadi et al. 2022) of *ct*Tho1, showing the SAP, CTD, linker, and CTE domains, including two alpha helices in the CTD and one in the CTE. *(B)* Domain organization of *ct*Tho1 with amino acids (aa) marked for each domain. *(C)* Schematic of the twelve *ct*Tho1 mutants and their nomenclature.

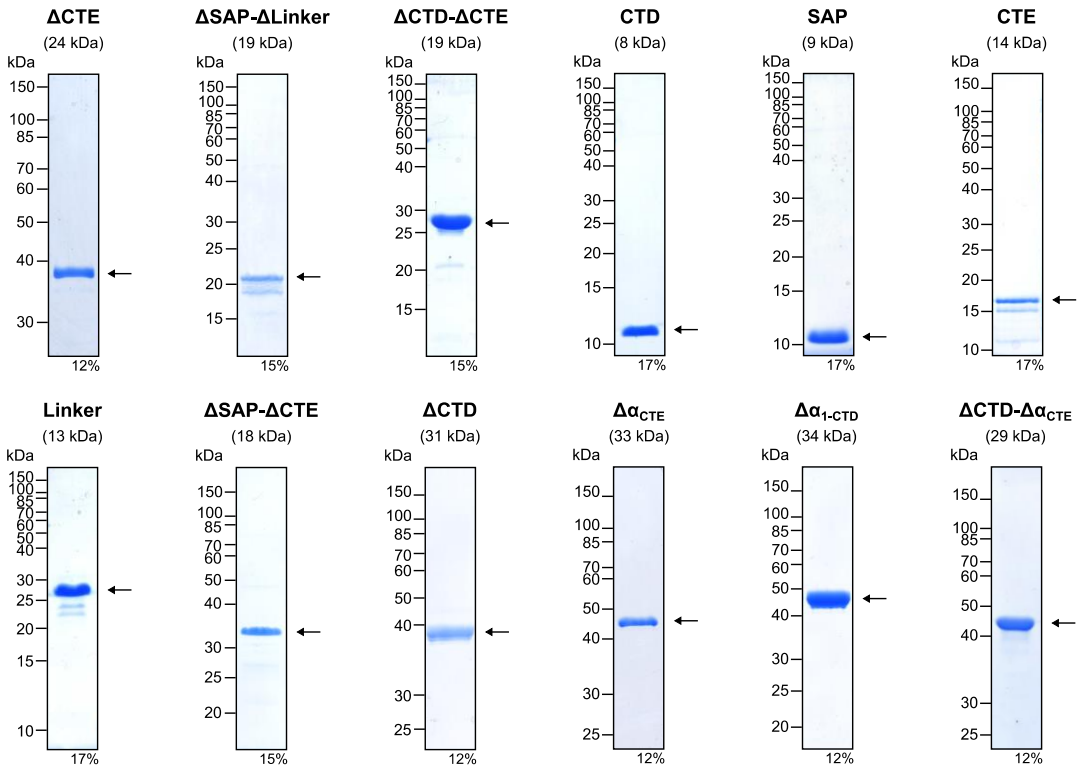


Figure 29: The *ct*Tho1 mutants were purified in reasonably high and pure quantities. All mutants were N-terminally Strep-tagged, expressed in *E. coli* Rosetta DE3, and purified using affinity and cation exchange columns (StrepTrap HP, HiTrap SP HP). Proteins were separated on either 12%, 15%, or 17% SDS-PAGE gels and stained with Coomassie Blue, with the target protein indicated by the arrow.

4.2.2.4. Secondary structure analysis of *ct*Tho1 mutants

After purifying all *ct*Tho1 mutants, CD spectroscopy was used to analyze their secondary structure content, including α-helices, β-sheets, and random coils. Distinct CD spectra enabled the structural characterization of each mutant (Fig. 30A). Proteins rich in α-helices exhibit a CD spectrum with negative peaks at 222 and 208 nm and a positive peak at 190-195 nm. β-sheet structures show a positive peak at 195-200 nm and a negative peak at 217-220 nm. Random coils lack distinct peaks beyond 210 nm, displaying weak, featureless signals in the UV region, making their identification challenging (Greenfield 2006, Fig. 30A).

The *AlphaFold* prediction of *ct*Tho1 suggests a mix of α-helices and random coils, reflected in its CD spectrum with a negative minimum around 222 nm (α-helices) and a negative peak at 200–205 nm (random coils) (Fig. 30B). Comparing the CD spectra of *ct*Tho1 and its deletion mutants shows that most mutants, except *ct*Tho1-SAP, *ct*Tho1-CTE, and *ct*Tho1-Linker, have similar spectra (Fig. 30C-F, 27J-N). The *ct*Tho1-SAP mutant displays features characteristic of α-helices, with negative peaks at 222 and 208 nm and a positive peak at 190-195 nm (Fig. 30G). The *ct*Tho1-CTE mutant lacks features indicating organized secondary structures (Fig. 30H), while the *ct*Tho1-Linker mutant shows a spectrum typical of random coils (Fig. 30I). Overall, the CD spectra align with the predicted secondary structures of each *ct*Tho1 mutant.

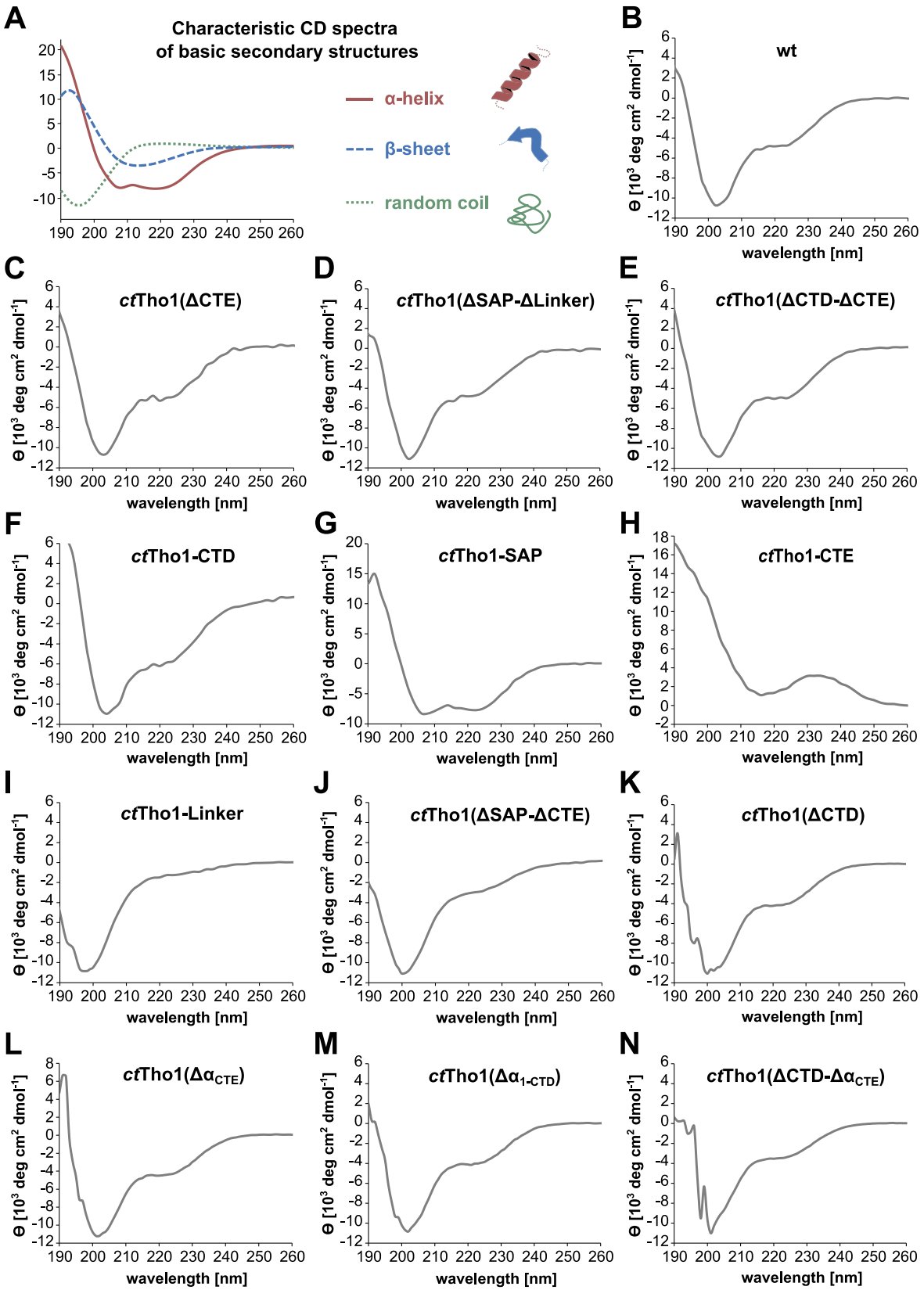


Figure 30: The secondary structure content of all *ct*Tho1 mutants aligns with the secondary structure prediction. *(A)* CD spectra of α-helices, β-sheets, and random coils. Far-UV CD spectra of *(B)* full-length *ct*Tho1 (wt), *(C)* *ct*Tho1(ΔCTE), *(D)* *ct*Tho1(ΔSAP-ΔLinker), *(E)* *ct*Tho1(ΔCTD-ΔCTE), *(F)* *ct*Tho1-CTD, *(G)* *ct*Tho1-SAP, *(H)* *ct*Tho1-CTE, *(I)* *ct*Tho1-Linker, *(J)* *ct*Tho1(ΔSAP-ΔCTE), *(K)* *ct*Tho1(ΔCTD), *(L)* *ct*Tho1(Δα$_{CTE}$), *(M)* *ct*Tho1(Δα$_{1-CTD}$), and *(N)* *ct*Tho1(ΔCTD-Δα$_{CTE}$).

4.2.2.5. DNA- and RNA-binding of *ct*Tho1 mutants

To identify regions of *ct*Tho1 necessary for DNA and RNA binding, Electrophoretic Mobility Shift Assays (EMSAs) were performed using two RNAs and two DNAs of varying lengths (Fig. 31).

For the longer RNA (37 nt), all mutants except *ct*Tho1(ΔCTE), *ct*Tho1(ΔCTD-ΔCTE), *ct*Tho1-SAP, and *ct*Tho1-Linker showed binding. The mutants *ct*Tho1-CTD, *ct*Tho1-CTE, and *ct*Tho1(ΔSAP-ΔCTE) had lower RNA-binding affinity, while *ct*Tho1(ΔSAP-ΔLinker), *ct*Tho1(ΔCTD), *ct*Tho1($\Delta\alpha_{CTE}$), *ct*Tho1($\Delta\alpha_{1\text{-}CTD}$), and *ct*Tho1(ΔCTD-$\Delta\alpha_{CTE}$) had shift patterns similar to full-length *ct*Tho1, with three to four distinct bands (Fig. 31A). The shorter RNA (16 nt) was also bound by the same mutants as the longer RNA, except *ct*Tho1-CTD, which no longer showed binding. Mutants binding the shorter RNA displayed two to three distinct bands, mirroring the full-length protein (Fig. 31B).

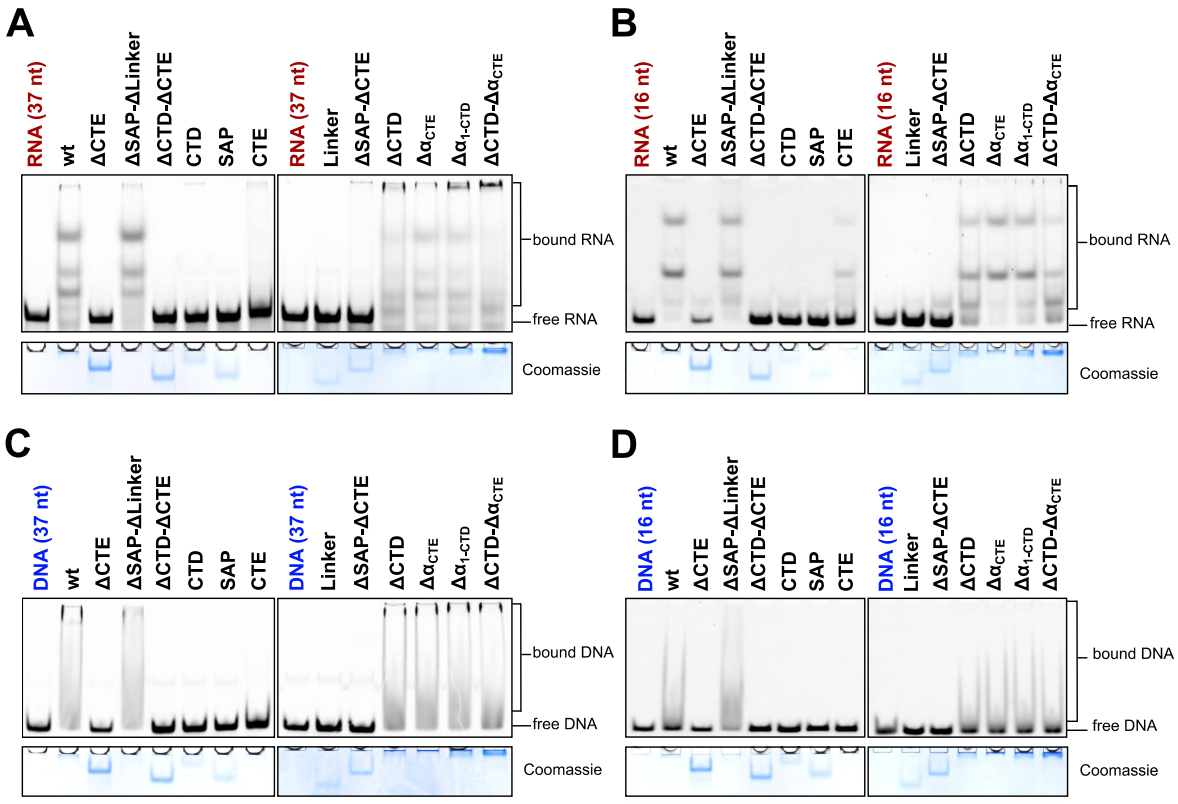


Figure 31: The CTE is the most essential region of *ct*Tho1 for RNA- and DNA-binding. The twelve *ct*Tho1 mutants as well as the full-length *ct*Tho1 (wt) were analyzed in an Electrophoretic Mobility Shift Assay (EMSA). For this, the mutants were pre-incubated with *(A)* 37 nt long 3'-Cy5-RNA, *(B)* 16 nt long 5'-Cy3-RNA, *(C)* 37 nt long 3'-Cy5-DNA and *(D)* 16 nt long 5'-Cy3-DNA and subsequently separated on a 10% TBE gel. Additionally, the used nucleic acid without protein was loaded onto the gel. The fluorescently labeled nucleic acids were detected with the laser scanner Typhoon FLA 9500.

For DNA, the longer DNA (37 nt) was bound by *ct*Tho1(ΔSAP-ΔLinker), *ct*Tho1-CTE, *ct*Tho1(ΔCTD), *ct*Tho1(Δα$_{CTE}$), *ct*Tho1(Δα$_{1\text{-}CTD}$), and *ct*Tho1(ΔCTD-Δα$_{CTE}$) mutants, though *ct*Tho1-CTE binding was weak, appearing as a shifted smear (Fig. 31C). The same mutants bound the shorter DNA (16 nt), except *ct*Tho1-CTE, which showed no binding (Fig. 31D).

In summary, *ct*Tho1 mutants lacking the CTE showed significantly reduced or no binding to both RNA and DNA.

4.2.2.6. Annealing activity of *ct*Tho1 mutants

To identify regions of *ct*Tho1 required for annealing DNA/DNA, DNA/RNA, and RNA/RNA, a FRET-based assay was performed using complementary single-stranded nucleic acids labeled with Cy3 (donor) and Cy5 (acceptor) dyes. Annealing by Tho1 brings the dyes into proximity, generating a FRET signal with a Cy5 emission peak at 670-680 nm (Fig. 32A-B).

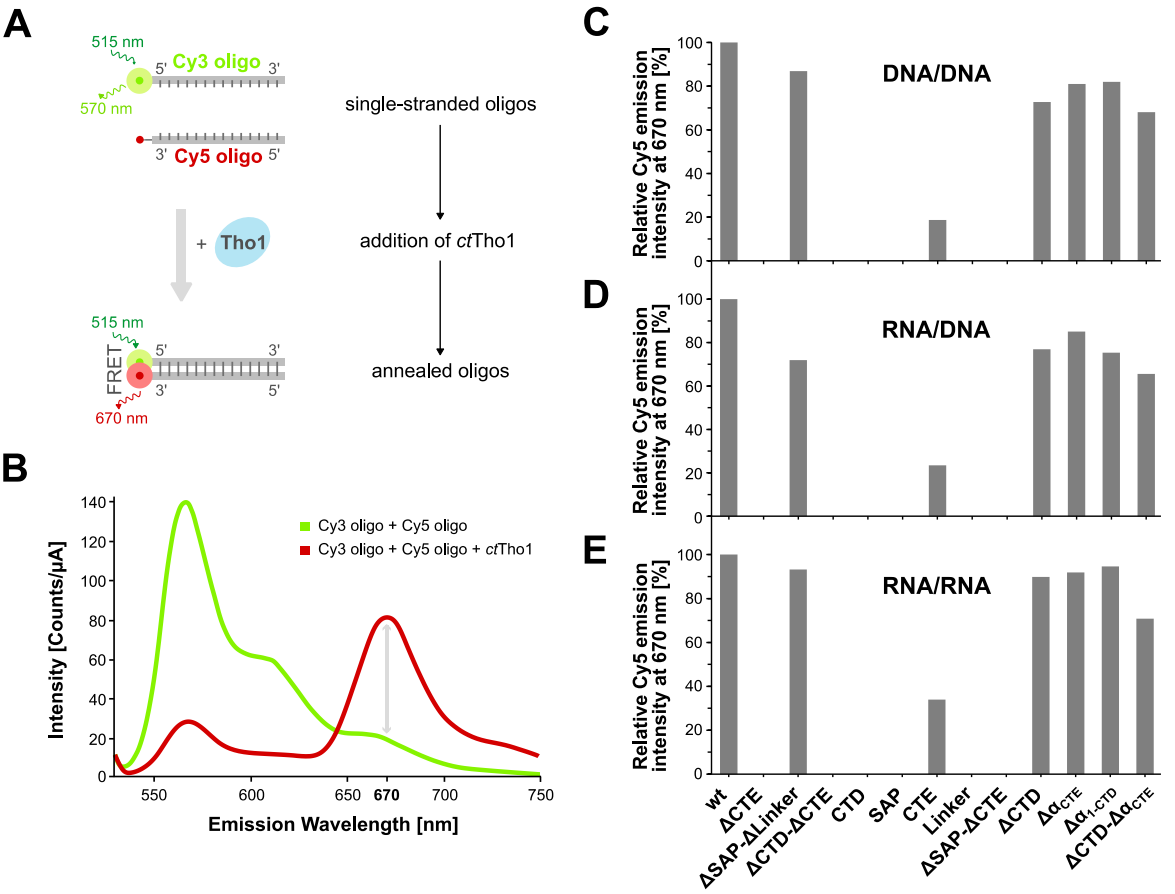


Figure 32: The CTE of *ct*Tho1 is necessary for annealing nucleic acids. *(A)* Schematic of the FRET assay using Cy3- and Cy5-labeled complementary nucleic acids. *(B)* Emission spectra before (green) and after (red) adding ctTho1, with Cy5 emission at 670 nm used to quantify annealing efficiency. *(C)* DNA/DNA, *(D)* RNA/DNA, and *(E)* RNA/RNA annealing activities of *ct*Tho1 and mutants, with *ct*Tho1 activity normalized to 100%.

Comparison of relative Cy5 emission intensities (full-length *ct*Tho1 set to 100%) revealed that the mutants *ct*Tho1-ΔCTE, *ct*Tho1(ΔCTD-ΔCTE), *ct*Tho1-CTD, *ct*Tho1-SAP, *ct*Tho1-Linker, and *ct*Tho1(ΔSAP-ΔCTE) completely lost annealing activity. Other mutants retained annealing capability but with reduced efficiency (Fig. 32C-E). DNA/DNA-annealing activity decreased by ~15% in *ct*Tho1(ΔSAP-ΔLinker), ~80% in *ct*Tho1-CTE, ~30% in *ct*Tho1(ΔCTD), ~20% in *ct*Tho1-$\Delta\alpha_{CTE}$) and *ct*Tho1($\Delta\alpha_{1\text{-}CTD}$), and ~30% in *ct*Tho1(ΔCTD-$\Delta\alpha_{CTE}$) (Fig. 32C). RNA/DNA-annealing activity dropped by ~30% in *ct*Tho1(ΔSAP-ΔLinker), ~80% in *ct*Tho1-CTE, ~25% in *ct*Tho1(ΔCTD), ~15% in *ct*Tho1($\Delta\alpha_{CTE}$), ~25% in *ct*Tho1($\Delta\alpha_{1\text{-}CTD}$), and ~35% in *ct*Tho1(ΔCTD-$\Delta\alpha_{CTE}$) (Fig. 32D). RNA/RNA-annealing activity showed a smaller reduction: ~10% in *ct*Tho1(ΔSAP-ΔLinker) and *ct*Tho1(ΔCTD), ~70% in *ct*Tho1-CTE, ~10% in *ct*Tho1($\Delta\alpha_{CTE}$), ~5% in *ct*Tho1($\Delta\alpha_{1\text{-}CTD}$), and ~30% in *ct*Tho1(ΔCTD-$\Delta\alpha_{CTE}$) (Fig. 32E).

In summary, deleting the CTE region abolishes nucleic acid annealing, while the presence of CTE alone preserves DNA/DNA-, RNA/DNA-, and RNA/RNA-annealing functions, though with significantly reduced efficiency compared to *ct*Tho1 and other functional mutants.

4.2.2.7. Interaction of *ct*Tho1 mutants with *ct*Sub2

Based on the known Tho1-Sub2 interaction in other species (Dufu et al. 2010; Jacobsen et al. 2016), the next aim was to identify the specific regions of *ct*Tho1 essential for this interaction by examining complex formation with *ct*Sub2 through *in vitro* reconstitution. Purified His-*ct*Sub2 was mixed with Strep-*ct*Tho1 mutants, followed by Strep-tag II pull-downs to analyze complex formation via SDS-PAGE (Fig. 33). Control experiments confirmed no non-specific His-*ct*Sub2 binding to the resin (Fig. 33B).

With full-length *ct*Tho1, nearly stoichiometric complex formation with His-*ct*Sub2 was observed (Fig. 33C). The *in vitro* reconstitution of His-*ct*Sub2 and the Strep-tagged mutants *ct*Tho1(ΔCTD-ΔCTE), *ct*Tho1(ΔCTE), *ct*Tho1(ΔCTD), *ct*Tho1($\Delta\alpha_{1\text{-}CTD}$), and *ct*Tho1(ΔCTD-$\Delta\alpha_{CTE}$) are shown in Figure 33D-H, while the *in vitro* reconstitutions of the remaining *ct*Tho1 mutants are shown in Supplementary Fig. S4. The *ct*Tho1(ΔCTD-ΔCTE) mutant completely lost *ct*Sub2 interaction (Fig. 33D). While the *ct*Tho1(ΔCTE) mutant retained interaction comparable to full-length

*ct*Tho1 (Fig. 33E), *ct*Tho1(ΔCTD) significantly reduced binding (Fig. 33F). Deleting the first alpha helix in the CTD [*ct*Tho1(Δα$_{1\text{-CTD}}$)] further impaired interaction (Fig. 33G), and the *ct*Tho1(ΔCTD-Δα$_{CTE}$) mutant abolished interaction entirely (Fig. 33H). In summary, the CTD domain of *ct*Tho1 is essential for *ct*Sub2 interaction, while the alpha helix in the CTE likely enhances, but is not essential for, this interaction.

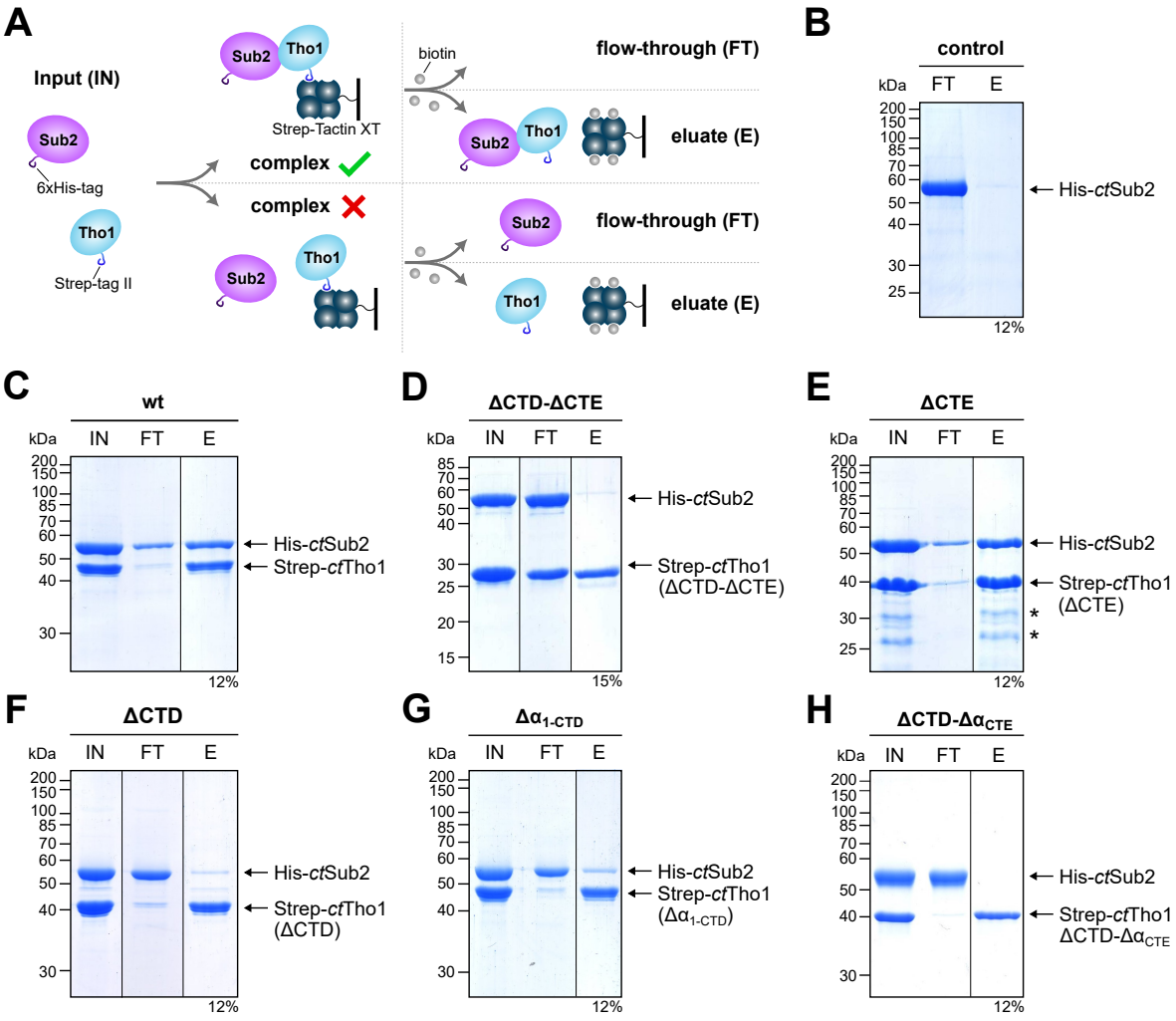


Figure 33: The CTD of *ct*Tho1 is primarily required for the interaction with *ct*Sub2. *(A)* Schematic over-view of the *in vitro* reconstitution and Strep-tag II pull-down experiment. When there is a strong complex formation between *ct*Tho1 and *ct*Sub2, the flow-through is expected to be empty, while the eluate should contain both proteins. When there is no complex formation, *ct*Sub2 is expected to be found in the flow-through, while *ct*Tho1 is present in the eluate. *(B)* As a control experiment, His-*ct*Sub2 was incubated without any Strep-*ct*Tho1 mutant. The Strep-tag II pull-down showed that there is no non-specific binding of His-*ct*Sub2 to the Strep-Tactin XT resin. *(C-H) In vitro* reconstitutions and Strep-tag II pull-downs of *(C)* full-length His-*ct*Tho1 (wt), *(D) ct*Tho1(ΔCTD-ΔCTE), *(E) ct*Tho1(ΔCTE), *(F) ct*Tho1(ΔCTD), *(G) ct*Tho1(Δα$_{1\text{-CTD}}$), and *(H) ct*Tho1(ΔCTD-Δα$_{CTE}$). Input (IN), flow-through (FT) and eluate (E) were analyzed on a 12% or 15% SDS-PAGE gel that was stained with Coomassie Brilliant Blue.

4.2.2.8. Overview of functional analyses in *ct*Tho1 mutants

Tho1 is composed of several domains, including the structured SAP and CTD domains, and the unstructured Linker and CTE regions. Functional analyses revealed that the CTE is primarily responsible for RNA and DNA binding, as well as *ct*Tho1's nucleic acid annealing activity. The CTD domain is essential for the interaction of *ct*Tho1 with the TREX complex component *ct*Sub2, while the alpha helix in the CTE likely enhances this interaction. Table 22 summarizes the functional characteristics of all *ct*Tho1 mutants.

Table 22: Summary of the functional characteristics of *ct*Tho1 mutants. Nucleic acid binding and interaction with *ct*Sub2 were rated relative to full-length *ct*Tho1 (++ means comparable to full-length *ct*Tho1, + means reduced binding or interaction, and - means no binding or interaction). The annealing activity for full-length *ct*Tho1 was set to 100%, with percentages for each mutant provided for comparison.

	DNA binding		RNA binding		DNA/DNA annealing	RNA/DNA annealing	RNA/RNA annealing	Interaction with *ct*Sub2
	37 nt	16 nt	37 nt	16 nt				
Full-length	++	++	++	++	100%	100%	100%	++
ΔCTE	-	-	-	-	-	-	-	++
ΔSAP-ΔLinker	++	++	++	++	87%	72%	93%	++
ΔCTD-ΔCTE	-	-	-	-	-	-	-	-
CTD	-	-	+	-	-	-	-	+
SAP	-	-	-	-	-	-	-	-
CTE	+	-	+	+	19%	23%	34%	+
Linker	-	-	-	-	-	-	-	-
ΔSAP-ΔCTE	-	-	+	+	-	-	-	++
ΔCTD	++	++	++	++	72%	77%	90%	+
$\Delta\alpha_{CTE}$	++	++	++	++	81%	85%	92%	++
$\Delta\alpha_{1\text{-}CTD}$	++	++	++	++	82%	75%	95%	+
ΔCTD-$\Delta\alpha_{CTE}$	++	++	++	++	68%	66%	71%	-

4.2.3. The reconstituted *ct*Sub2-*ct*Tho1 complex

To investigate the interaction between *ct*Sub2 and *ct*Tho1, the aim was to obtain a high-resolution structure of their complex. The proteins were purified using affinity, ion exchange, and size exclusion chromatography. After purifying His-*ct*Sub2 and Strep-*ct*Tho1 separately, they were incubated together, followed by Strep-tag II pull-down and further purification using Superdex 75 size exclusion chromatography. The size exclusion chromatography aimed to separate individual *ct*Sub2 and *ct*Tho1 proteins from the *ct*Sub2-*ct*Tho1 complex. Although the Superdex 75 column has a resolution limit for proteins over 70 kDa, it was chosen as the best option within the

available columns. Due to the reconstituted complex's molecular weight exceeding this limit, it was expected to elute within the exclusion volume (8-9 ml).

Experiments with full-length proteins revealed that after size exclusion chromatography *ct*Sub2 was approximately twice as abundant as *ct*Tho1 in the reconstituted complex, as shown by SDS-PAGE analysis (Fig. 34A). The size exclusion chromatogram confirmed the expected elution behavior of the complex, with no secondary peaks for the individual proteins (Fig. 34B).

As ATP is essential for Sub2's RNA helicase activity (La Cruz et al. 1999), it was analyzed whether binding of *ct*Sub2 to the non-hydrolysable ATP analog ADP·AlF_3 affects the stability of the *ct*Sub2-*ct*Tho1 complex. SDS-PAGE analysis after size exclusion chromatography showed a similar stoichiometry of *ct*Sub2 and *ct*Tho1 as observed without the ATP analog, though degradation bands appeared below ctTho1 (Fig. 34C). The size exclusion chromatogram revealed the complex eluting within the expected exclusion volume, with ADP·AlF_3 eluting between 17 and 18 ml (Fig. 34D).

Given that unstructured regions within proteins can reduce the probability of protein crystallization (Musselman et al. 2021), a truncated *ct*Sub2-*ct*Tho1 complex was subjected to *in vitro* reconstitution and size exclusion chromatography. The truncated complex also displayed an imbalance in stoichiometry, with ctSub2 being approximately two to three times more abundant than ctTho1 (Fig. 34E). The size exclusion chromatogram showed an additional peak corresponding to *ct*Tho1 not bound to *ct*Sub2 (Fig. 34F).

The different reconstituted *ct*Sub2-*ct*Tho1 complexes were tested in various conditions that promote crystallization, using an automated robotic system to screen for optimal conditions (performed by M. Stumpf). Although no crystals were formed under the tested conditions, the *in vitro* reconstitutions and size exclusion chromatography in general yielded stable complexes, making them suitable for further structural analysis.

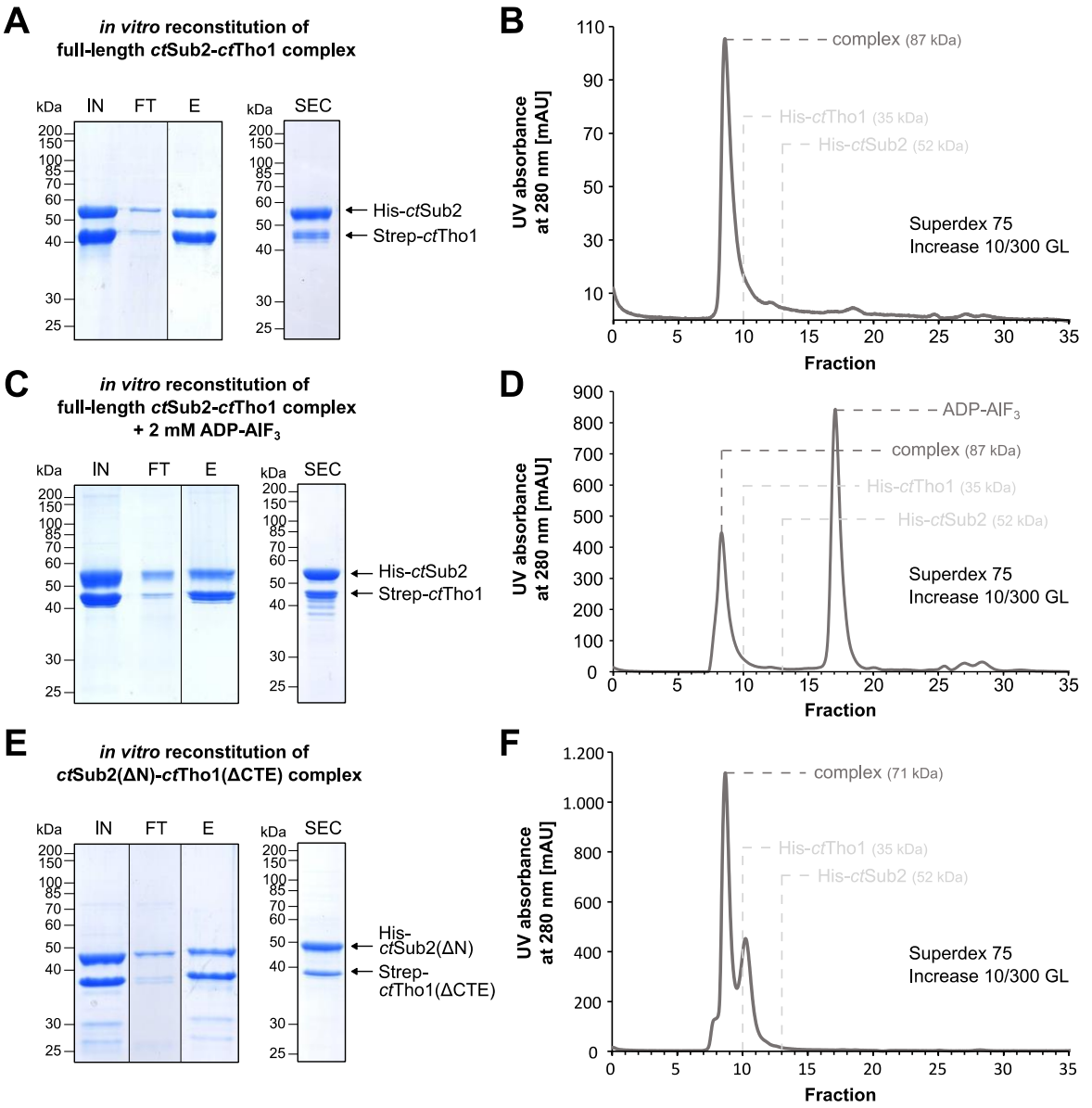


Figure 34: The S-Y complex displays a stoichiometry where *ct*Sub2 is approximately twice as abundant as *ct*Tho1. *(A, C, E)* The *(A)* full-length *ct*Sub2-*ct*Tho1 without and *(C)* with the addition of ADP·AlF$_3$ as well as the *(E)* truncated *ct*Sub2(ΔN)-*ct*Tho1(ΔCTE) complex were reconstituted *in vitro* and further purified by size exclusion chromatography. The eluates were separated on a 12% SDS-PAGE gel that was stained with Coomassie Brilliant Blue. *(B, D)* The size exclusion chromatogram shows a single peak of the *(B)* full-length *ct*Sub2-*ct*Tho1 complex and of the *(D)* full-length *ct*Sub2-*ct*Tho1 complex with ADP·AlF$_3$ in the anticipated exclusion volume. *(F)* The size exclusion chromatogram shows a peak of the *ct*Sub2(ΔN)-*ct*Tho1(ΔCTE) complex in the anticipated exclusion volume. A second peak appears at the volume where individual *ct*Tho1 protein is expected. *(B, D, F)* The elution volumes of individual His-*ct*Tho1 and His-*ct*Sub2 are indicated in every size exclusion chromatogram.

4.2.4. The RNA annealing protein Yra1 in *C. thermophilum* (*ct*Yra1)

Yra1 (ALYREF in humans) is part of the TREX complex and serves as an mRNA export adapter, recruiting and stabilizing the export receptor Mex67 (NXF1 in

humans) on mRNA (Sträßer & Hurt 2000). Yra1 interacts with Sub2, which in turn binds to Tho1, forming a trimeric Sub2-Yra1-Tho1 complex (Dufu et al. 2010; Xie et al. 2023).

Sequence comparison of Yra1 from *H. sapiens*, *S. cerevisiae*, and *C. thermophilum* using *Clustal Omega* (McWilliam et al. 2013) and *Jalview* (Waterhouse et al. 2009) show the two conserved motifs, the N-box and C-box (Fig. 35), which interact with Sub2 (Hautbergue et al. 2009; Ren et al. 2017). Unlike human and yeast proteins, *ct*Yra1 has an additional 30 amino acids following the C-box. All Yra1 homologs feature a conserved RNA recognition motif (RRM) (Fig. 35) that lacks typical RNA-binding abilities, relying instead on intrinsically disordered regions (IDRs) linking the N-box, RRM, and C-box for RNA interaction (Zenklusen et al. 2001).

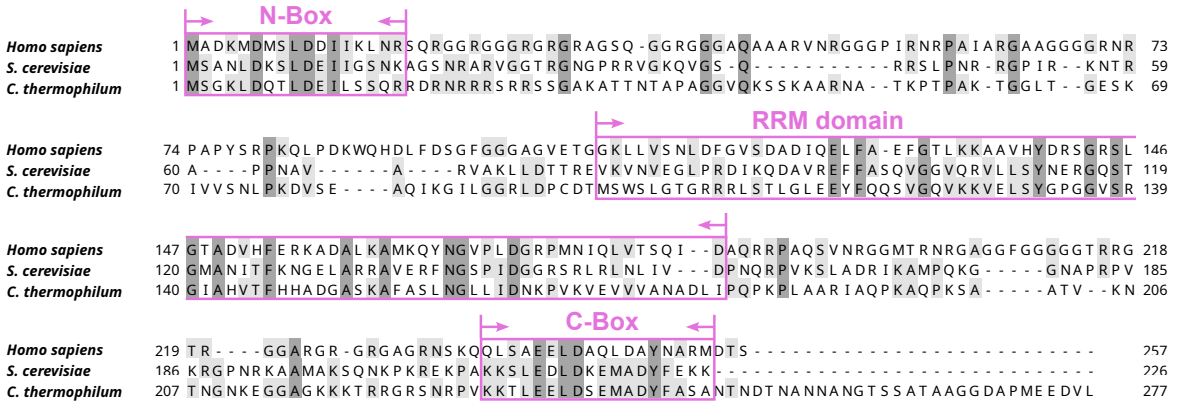


Figure 35: Sequence alignment of *H. sapiens* ALYREF, *S. cerevisiae* Yra1 and *C. thermophilum* Yra1. The indicated N-Box, RRM and C-Box are more conserved regions between the three species, whereas the region between the N-Box and the RRM, and the RRM and the C-Box are less conserved. The sequence alignment was done using *Clustal Omega* (McWilliam et al. 2013) and *Jalview* (Waterhouse et al. 2009). The *PROSITE* database (Sigrist et al. 2013) and structure predictions by *AlphaFold* (Jumper et al. 2021; Varadi et al. 2022) were used for assignment of domains and motifs.

4.2.4.1. Purification of *ct*Yra1

For further analysis of *ct*Yra1, the protein was purified by fusing a His- or Strep-tag to its C-terminus and expressing it in Rosetta DE3 *E. coli* cells. The tag enabled affinity purification, followed by cation exchange chromatography to eliminate contaminants and residual nucleic acids. SDS-PAGE analysis showed that *ct*Yra1 migrated above the 30 kDa marker, exceeding its calculated molecular weight of 29 kDa, with both His- and Strep-tagged versions showing reasonable purity (Fig. 36A-B).

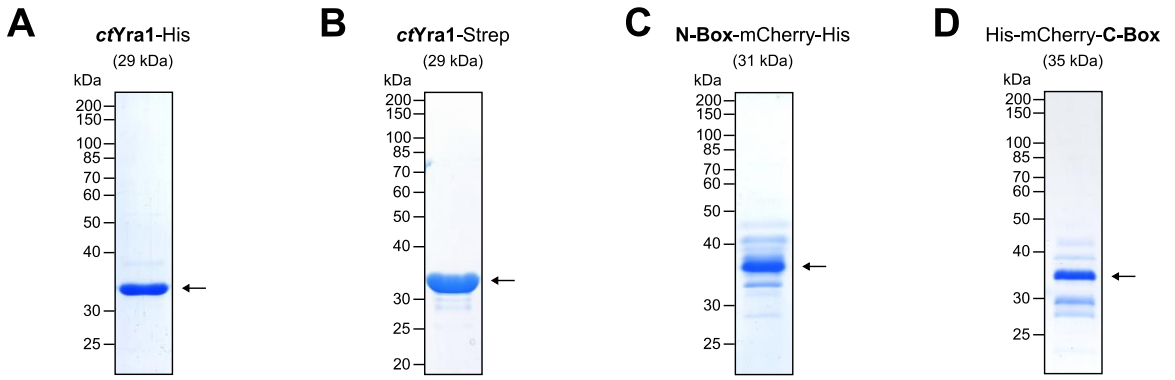


Figure 36: The *ct*Yra1 proteins and mutants purified in reasonably high and pure quantities. In general, the proteins *(A) ct*Yra1-His, *(B) ct*Yra1-Strep, *(C)* N-Box-mCherry-His, and *(D)* His-mCherry-C-Box were expressed in Rosetta DE3 *E. coli* cells and subsequently purified in two steps, using an affinity column and an ion exchange column. The *ct*Yra1 proteins were separated on a 12% SDS-PAGE gel that was stained with Coomassie Brilliant Blue. The arrow indicates the target protein. Other bands visible on the gel are remaining contaminations or degradation products.

The N-Box and C-Box of *ct*Yra1 were individually expressed as fusion proteins: the N-Box with C-terminal mCherry and His-tag, and the C-Box with N-terminal His-tag and mCherry. Both were purified via affinity chromatography followed by ion exchange chromatography. SDS-PAGE analysis showed that N-Box-mCherry-His migrated above the 35 kDa marker, higher than its calculated 31 kDa (Fig. 36C), while His-mCherry-C-Box migrated at its expected 35 kDa (Fig. 36D). Despite some contaminants and degradation products, both proteins were sufficiently pure for further testing (Fig. 36C-D).

4.2.4.2. *In vitro* reconstitutions with *ct*Yra1

The purified *ct*Yra1 proteins were tested for their ability to form complexes with *ct*Sub2 and *ct*Sub2-*ct*Tho1 *in vitro*.

For the *ct*Sub2-*ct*Yra1 complex, GST-*ct*Sub2 was mixed with *ct*Yra1-His, and a His pull-down followed by SDS-PAGE showed minimal *ct*Sub2 binding (Fig. 37A). Adding RNA did not improve complex formation (Fig. 37B). To assess the role of Sub2-binding motifs, N-Box-mCherry-His and His-mCherry-C-Box were tested with GST-*ct*Sub2. GST pull-downs revealed no complex formation with the N-Box (Fig. 37C), while the C-Box showed minimal binding (Fig. 37D). For the *ct*Sub2-*ct*Yra1-*ct*Tho1 (S-Y-T) complex, His-*ct*Sub2, His-*ct*Tho1, and *ct*Yra1-Strep were incubated, and a Strep pull-down showed only minimal complex formation (Fig. 37E).

In summary, stoichiometric reconstitution of *ct*Sub2-*ct*Yra1 or S-Y-T complexes proved challenging. Neither RNA addition nor using individual Sub2-binding motifs significantly enhanced complex formation *in vitro*.

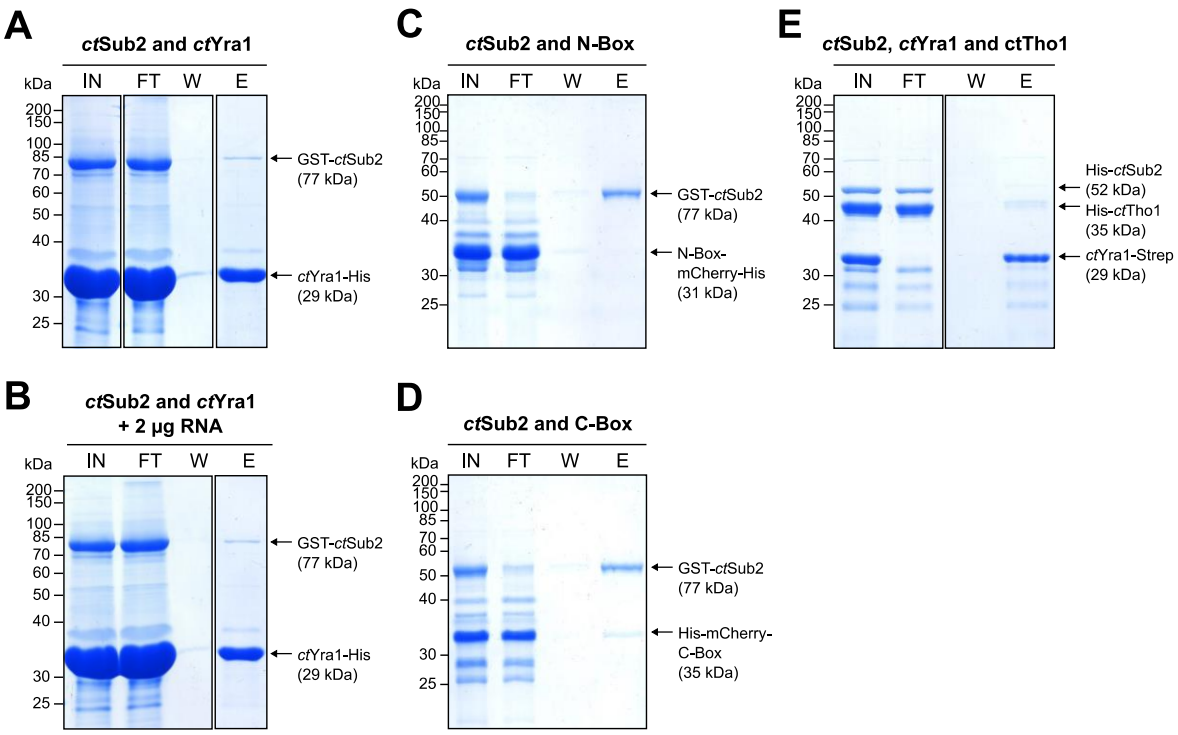


Figure 37: The *in vitro* reconstitution of the *ct*Sub2-*ct*Yra1-(*ct*Tho1) complex is not efficient. *(A-B)* His pull-downs of GST-*ct*Sub2 *in vitro* reconstituted with *(A)* full-length *ct*Yra1-His, and *(B)* full-length *ct*Yra1-His with additional RNA. GST pull-downs of GST-*ct*Sub2 with *(C)* N-Box-mCherry-His, and *(D)* His-mCherry-C-Box. *(E)* Strep pull-down of in vitro reconstituted His-*ct*Sub2, His-*ct*Tho1, and *ct*Yra1-Strep. Input (IN), flow-through (FT), last wash step (W) and eluate (E) were analyzed on a 12% SDS-PAGE gel that was stained with Coomassie Brilliant Blue.

4.2.5. Structure of the *ct*Sub2-*ct*Tho1-*ct*Yra1 (S-Y-T) complex

Due to the insufficient *in vitro* reconstitution of the S-Y-T complex in equimolar amounts, co-expression of the complex was tested as an alternative to potentially obtain more suitable S-Y-T complexes for further structural analysis. To potentially gain a structure of the complete TREX complex in association to *ct*Tho1, the pentameric THO complex from *C. thermophilum* was also co-expressed, purified and analyzed by negative stain EM (Supplementary Fig. S5).

4.2.5.1. Co-expression and purification of different S-Y-T complexes

A polycistronic plasmid encoding His-*ct*Sub2, His-*ct*Yra1, and Strep-*ct*Tho1 (S-Y-T) was created and expressed in Rosetta DE3 *E. coli*. Co-expression of variants

S-Y-T(ΔCTE) and S-Y-T(ΔSAP-ΔLinker), containing the *ct*Tho1 mutants Strep-*ct*Tho1(ΔCTE) or Strep-*ct*Tho1(ΔSAP-ΔLinker), was also tested. The complexes were isolated via Strep-tag affinity chromatography and analyzed by SDS-PAGE (Fig. 38A-C). The nucleic acid content was assessed by treating the eluates with RNase A, RNase H, or DNase I, followed by TBE gel electrophoresis and EtBr staining (Fig. 38D-F). RNase A specifically degrades single-stranded RNA, RNase H targets RNA in RNA-DNA hybrids, and DNase I digests DNA, enabling the identification of RNA, DNA, or RNA-DNA hybrids in the purified complex.

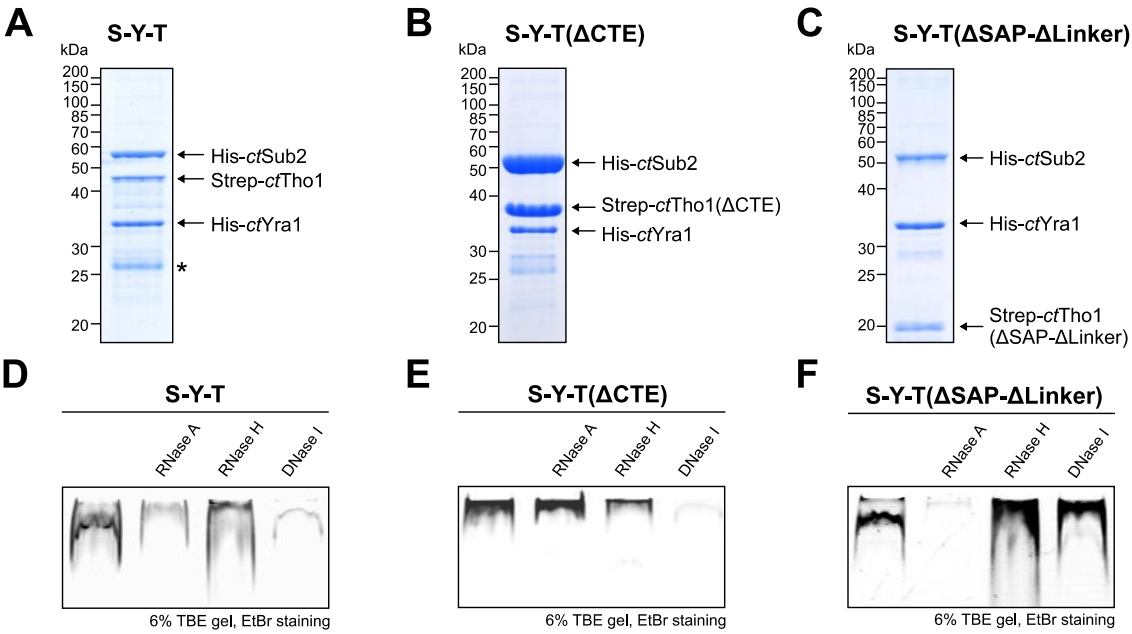


Figure 38: Deletions in *ct*Tho1 result in alterations in S-Y-T stoichiometry and nucleic acid content during co-expression. *(A-C)* Purifications utilizing the Strep tag II for affinity chromatography of *(A)* the full-length S-Y-T complex, *(B)* the S-Y-T(ΔCTE) complex, and *(C)* the S-Y-T(ΔSAP-ΔLinker) complex. The eluates were separated on a 12% SDS-PAGE gel stained with Coomassie Brilliant Blue. *(D-F)* The nucleic acid content of *(D)* the S-Y-T complex, *(E)* the S-Y-T(ΔCTE) complex, and *(F)* the S-Y-T(ΔSAP-ΔLinker) complex was analyzed by native PAGE using a 6% TBE gel stained with ethidium bromide (EtBr). The purified complex was either treated with RNase A (lane 2), RNase H (lane 3), or DNase I (lane 4).

Co-expression of the full-length S-Y-T complex results in an almost stoichiometric complex (Fig. 38A), which is strongly associated with nucleic acids, primarily DNA, as DNase I treatment nearly eliminates the signal (Fig. 38D, lane 4). The signal also decreases with RNase A treatment, indicating the presence of single-stranded RNA (Fig. 38D, lane 2). The co-expressed and purified S-Y-T(ΔCTE) complex has an unequal stoichiometry, with *ct*Sub2 predominating, followed by *ct*Tho1(ΔCTE) and

*ct*Yra1 (Fig. 38B). Native PAGE analysis shows strong nucleic acid association, mainly DNA, but no significant amounts of RNA (Fig. 38E, lanes 2-4). The S-Y-T(ΔSAP-ΔLinker) complex is nearly stoichiometric (Fig. 38C) and is strongly assocated to nucleic acids, primarily RNA, as RNase A treatment eliminates the signal (Fig. 38F, lane 2). Unlike the other complexes, it lacks DNA, as DNase I and RNase H treatments do not affect the signal (Fig. 38F, lanes 3-4).

In summary, removing the CTE from *ct*Tho1 alters complex stoichiometry and nucleic acid content, with the resulting complex mainly containing DNA. Deletion of the SAP domain and linker results in a complex that contains only RNA.

4.2.5.2. Nuclease treatment of the S-Y-T complex

The next step was to test whether nuclease treatment during purification could remove nucleic acids from the S-Y-T complex while maintaining its stoichiometry and integrity. Full-length S-Y-T complexes were affinity-purified and subjected to size exclusion chromatography, either untreated or treated with DNase I and a mix of RNases (A, T1, III, H, and I). Nuclease treatment significantly increased protein yield without altering stoichiometry (Fig. 39A, D). Native PAGE showed effective removal of nucleic acids, although a smear of 100-200 base pairs appeared post-treatment (Fig. 39B, E). SDS-PAGE analysis of the peak fraction obtained from size exclusion chromatography confirmed the complexes remained structurally intact with consistent stoichiometry (Fig. 36A, D). The size exclusion chromatogram indicated that both complexes eluted above 600 kDa, exceeding their calculated 116 kDa molecular weight (Fig. 39C, E).

In summary, nuclease treatment during purification increases protein yield without compromising the complex's stoichiometry or integrity, although nucleic acids of 100-200 base pairs remain.

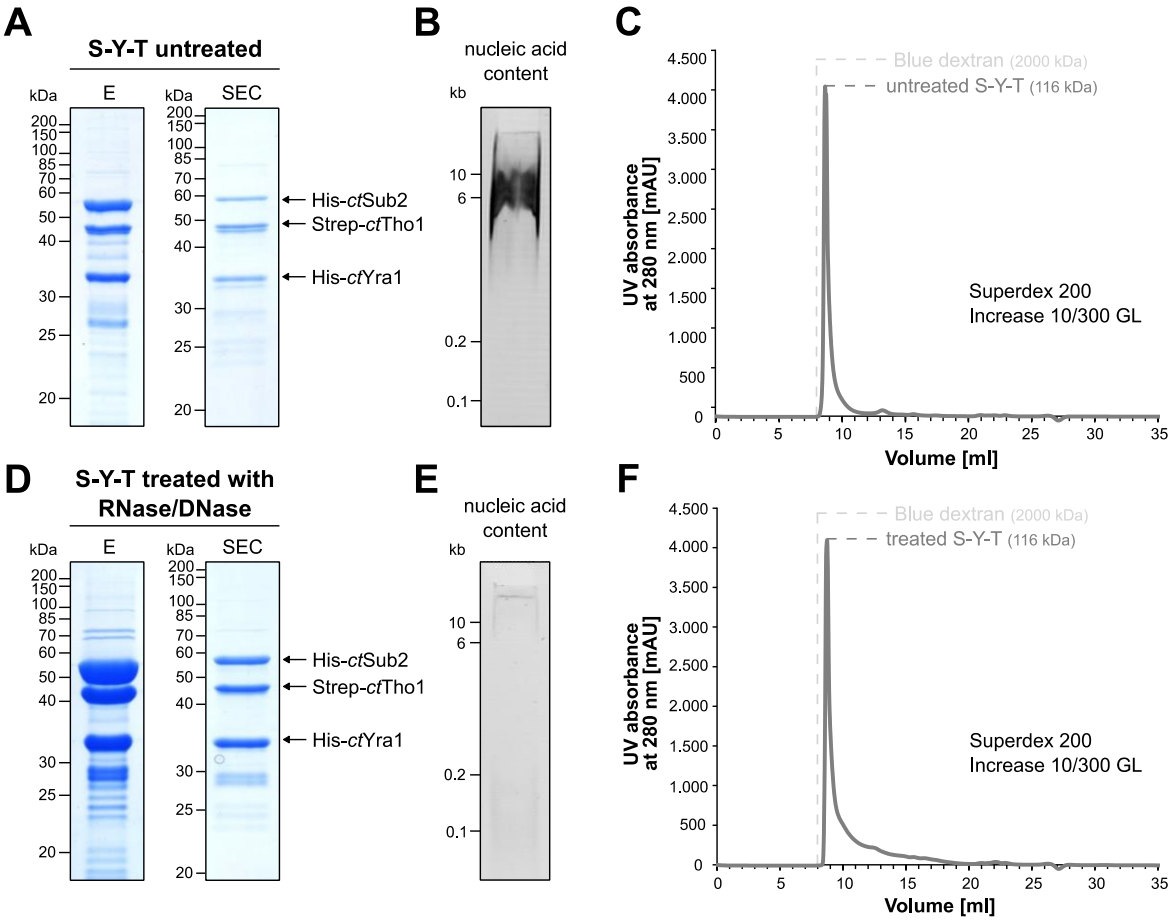


Figure 39: Nuclease treatment of S-Y-T improves protein yield without compromising the stoichiometry or integrity of the complex. *(A, D)* Analysis of the affinity purification (left) and of the peak fraction of size exclusion chromatography (right) on a 12% SDS-PAGE gel stained with Coomassie Brilliant Blue of *(A)* the untreated, and *(D)* the RNase mix and DNase I treated S-Y-T complex. *(B, E)* The nucleic acid content of *(B)* the untreated, and *(E)* the treated S-Y-T complex was analyzed by native PAGE using a 6% TBE gel stained with ethidium bromide (EtBr). *(C, F)* The size exclusion chromatogram of *(C)* the untreated, and *(F)* the treated S-Y-T complex shows that both complexes eluate as a single peak in the exclusion volume of the selected Superdex 200 column. Indicated is the elution volume of Blue dextran, which is commonly used as a marker protein for the exclusion volume of the column.

4.2.5.3. Binding of the S-Y-T complex to a specific RNA

Despite nuclease treatment, smaller nucleic acids remain in the S-Y-T complex, raising the question of whether it continues to bind the *E. coli*-derived RNA even when a specific RNA is added in excess. To investigate, the nuclease-treated complex was tested for binding to a 37 nt RNA in an EMSA and compared to the untreated complex (Fig. 40A). The untreated complex showed no binding, while the nuclease-treated complex exhibited binding, indicated by two shifted bands (Fig. 40A).

The EMSA suggested that nuclease-treated S-Y-T could bind to specific RNA, but

it could not determine whether the entire complex or just one component was responsible for the shift. To clarify this, an RNA pull-down was performed using the same 37 nt RNA used in the EMSA, but with a biotin attached to its 5' end instead of a Cy5 fluorophore. The S-Y-T complex, tagged with HA for detection, was incubated with the RNA. Western blot analysis of the pull-down revealed significant enrichment of the S-Y-T complex in the RNA eluate, compared to the negative control (Fig. 40B).

In summary, both EMSA and RNA pull-down experiments show that the nuclease-treated S-Y-T complex can bind to a specific RNA, while the untreated complex cannot.

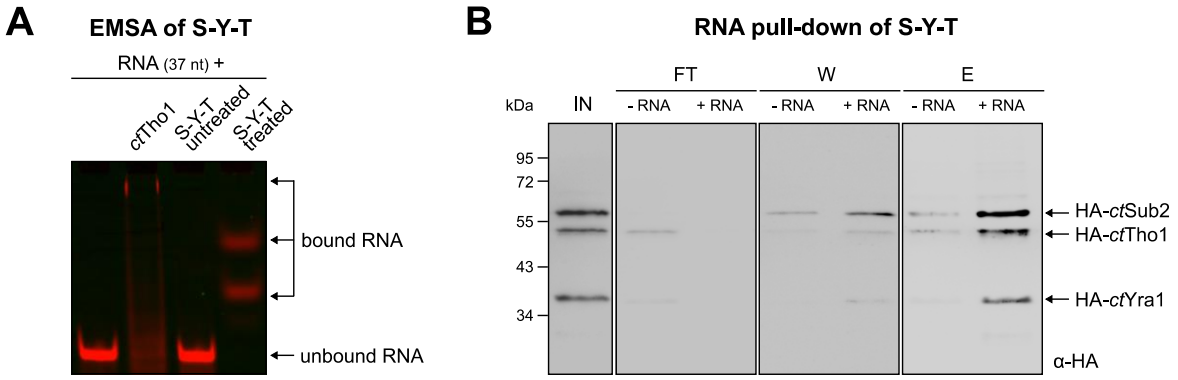


Figure 40: The nuclease treated S-Y-T complex can bind to a specific RNA introduced to it. *(A)* The untreated (lane 3) and nuclease treated S-Y-T complex (lane 4) were analyzed in an Electrophoretic Mobility Shift Assay (EMSA). For this, the complexes were pre-incubated with 37 nt long Cy5-labelled RNA and subsequently separated on a 10% TBE gel. Additionally, the used RNA without protein was loaded onto the gel (lane 1). The RNA + *ct*Tho1 was used as a positive control (lane 2). The fluorescently labeled nucleic acids were detected with the laser scanner Typhoon FLA 9500. *(B)* The nuclease treated S-Y-T, in which all of its components were N-terminally tagged with HA, was additionally analyzed in an RNA pull-down. The identical 37 nt long RNA utilized in the EMSA was employed, but in this case, it featured a biotin attached to its 5' end instead of a Cy5. The flow-through (F), final wash step (W), and the eluate (E) were analyzed by Western blot and compared to a negative control where no RNA was added. The S-Y-T complex is significantly enriched in the eluate when RNA was introduced.

4.2.5.4. Stability of the S-Y-T complex

The nuclease-treated S-Y-T complex was used for further studies, as it exhibited improved yield without strong association to non-specific nucleic acids, yet had the ability to bind specific RNA. The complex stability was evaluated under two storage conditions (4°C and -80°C) and varying salt concentrations (100-1000 mM NaCl).

For temperature stability, the size exclusion chromatography peak fraction of the freshly purified complex (Fig. 41A) was split into two. One sample was stored at 4°C for three days, while the other was flash frozen and kept at -80°C. After three days

at 4°C, significant degradation or precipitation was observed, as shown by barely visible SDS-PAGE bands (Fig. 41B, lane IN), compared to the fresh sample (Fig. 41A). Purification using the Strep tag II of *ct*Tho1 confirmed stoichiometric recovery of all three components in the eluate (Fig. 41B, lane E), with minimal amounts of *ct*Sub2 in the flow-through (Fig. 41B, lane FT). In contrast, the flash-frozen sample stored at -80°C showed no significant protein loss (Fig. 41C, lane IN). Affinity purification revealed stoichiometric amounts of all three components in the eluate (Fig. 41C, lane E), with minor amounts of *ct*Sub2 and *ct*Yra1 in the flow-through (Fig. 41C, lane FT).

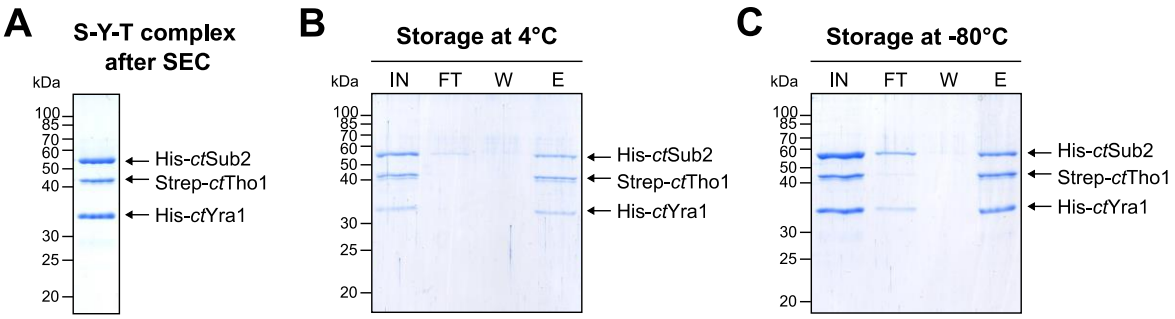


Figure 41: The S-Y-T complex remains stable during short-term storage at 4°C and long-term storage at -80°C. *(A)* The peak fraction from the size exclusion chromatography of a freshly purified S-Y-T complex on a 12% SDS-PAGE gel stained with Coomassie Brilliant Blue. All components are present in equal stoichiometric amounts. The S-Y-T complex was then stored at *(B)* 4°C for three days, or *(C)* flash frozen in liquid nitrogen and kept at -80°C. *(B, C)* The complex was subsequently purified by utilizing the Strep tag II of *ct*Tho1. The Input (IN, sample before purification), flow-through (FT), final wash step (W), and eluate (E) were analyzed on a 12% SDS-PAGE gel and Coomassie Brilliant Blue staining.

The stability of the S-Y-T complex was tested under increasing salt concentrations, which can either stabilize complexes by mitigating electrostatic repulsions or disrupt complexes by competing with binding interactions (Zhou & Pang 2018). Affinity purification was performed using a buffer with 100 mM NaCl, gradually increasing salt concentrations during washing steps. SDS-PAGE analysis shows that each wash step with increasing salt concentrations removes a portion of *ct*Sub2 from the complex. At 400 mM NaCl, *ct*Yra1 began dissociating from the complex, and by 1 M NaCl, only *ct*Sub2 and *ct*Tho1 remained in stoichiometric amounts (Fig. 42A).

Size exclusion chromatography at 150, 250, and 300 mM NaCl revealed that at 150 mM, the complex eluted as a single peak containing all three components (Fig. 42B, C). At 250 mM, *ct*Yra1 predominated in the exclusion volume peak, while *ct*Sub2 and *ct*Tho1 formed a separate peak (Fig. 42B, D). The elution of the 29 kDa

*ct*Yra1 in the exclusion volume suggests transient interactions or aggregation affecting its behavior in size exclusion chromatography. At 300 mM, *ct*Yra1 spread across later fractions, with *ct*Sub2 and *ct*Tho1 remaining in the second peak (Fig. 42B, E). This suggests that higher salt concentrations may resolve *ct*Yra1's transient interactions or aggregation, resulting in its reduced elution within the exclusion volume.

In summary, the S-Y-T complex is stable for short-term storage at 4°C but requires flash freezing for long-term storage. It remains intact at salt concentrations up to 150-200 mM. Above 250 mM, *ct*Yra1 dissociates, leaving a stable *ct*Sub2-*ct*Tho1 complex resistant to salt concentrations up to 1 M.

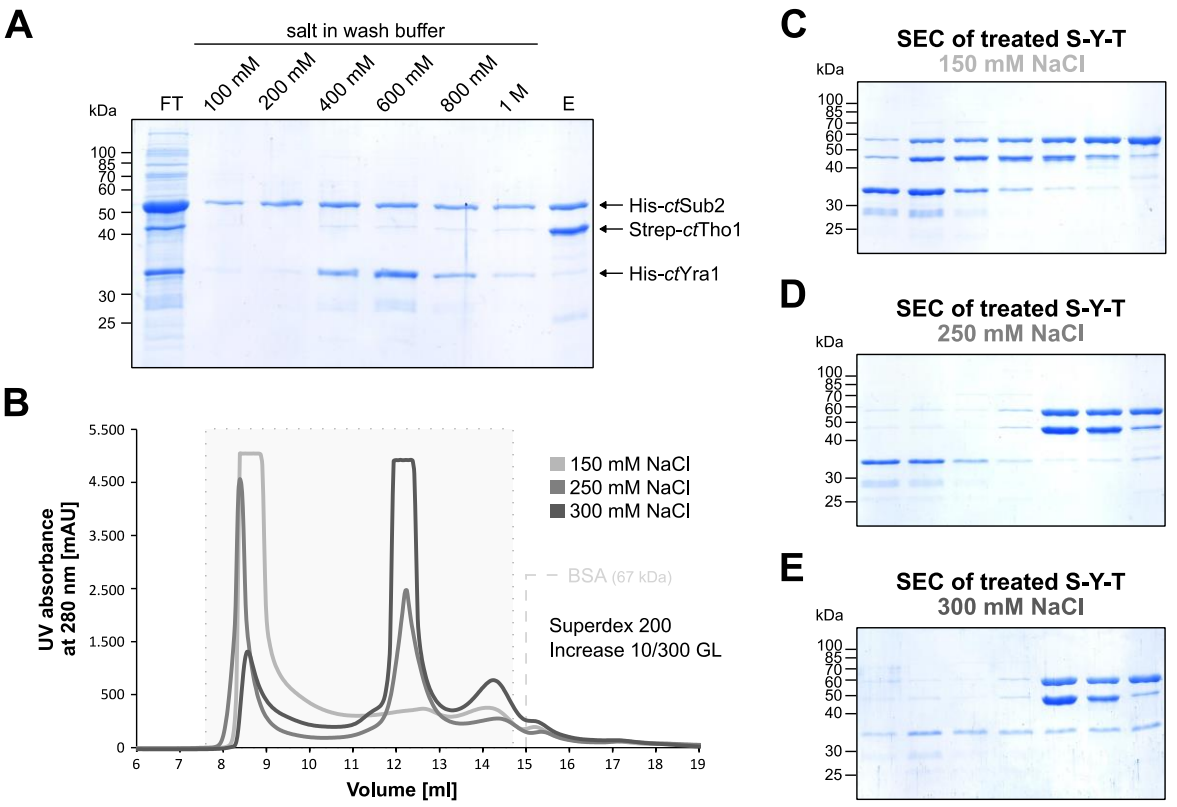


Figure 42: The S-Y-T complex retains its stability until *ct*Yra1 dissociates at a salt concentration of 250 mM. *(A)* The S-Y-T complex was affinity purified by the Strep tag II if ctTho1 in buffer containing 100 mM NaCl. The salt concentration was incrementally raised in each subsequent wash step. Flow-through (FT), wash steps (100, 200, 400, 600, 800, and 1000 mM), and eluate (E) were analyzed on a 12% SDS-PAGE gel stained with Coomassie Brilliant Blue. *(B)* The size exclusion chromatograms of S-Y-T in buffer with 150 mM, 250 mM, and 300 mM NaCl. When salt concentration is increased, the peak in the exclusion volume decreases, while a new peak at approximately 12 ml emerges. Indicated is the elution volume of BSA (67 kDa). The light grey background indicates fractions analyzed by SDS-PAGE. *(C-E)* Fractions obtained from the size exclusion chromatography at *(C)* 150 mM, *(D)* 250 mM, and *(E)* 300 mM NaCl, within elution volumes spanning from 7.7 to 14.7 ml, were examined on a 12% SDS-PAGE gel stained with Coomassie Brilliant Blue.

4.2.5.5. Limited proteolysis of the S-Y-T complex

The S-Y-T complex was analyzed using limited proteolysis with trypsin to identify flexible or exposed protein regions and assess the complex's stability. The rate and extent of proteolysis reflect complex stability, with stable regions resisting cleavage and flexible regions being more susceptible (Fontana et al. 1997; Tsai et al. 2002). Proteolysis was performed with 1 µg and 5 µg of trypsin, with 1 µg providing better resolution of early time points (Fig. 43A) and 5 µg better resolving later time points (Fig. 43B). At 1 µg, *ct*Sub2 remained stable for up to 30 minutes, while *ct*Tho1 and *ct*Yra1 were cleaved as early as 2 minutes. With 5 µg trypsin, three stable bands emerged after 30 minutes and persisted for over 1.5 hours. Intact Protein LC-MS analysis of these three bands (performed by Amin Fahim, Institute of Biochemistry, Goethe University Frankfurt) identified two bands as *ct*Sub2 fragments (45.088 kDa and 21.865 kDa), while a third protein (24.598 kDa) likely corresponds to the middle band but could not be attributed to an S-Y-T fragment.

In summary, *ct*Sub2 is the most stable and structured component of the S-Y-T complex, while the more flexible *ct*Tho1 and *ct*Yra1 are rapidly cleaved, indicating that they remain flexible or less structured even within the complex.

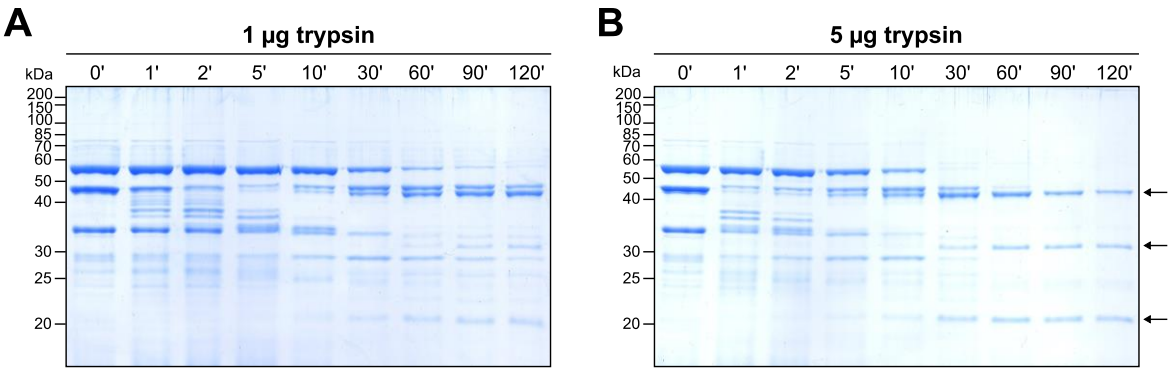


Figure 43: Limited proteolysis of S-Y-T shows that *ct*Sub2 is the most stable and structured component of the complex. The purified S-Y-T complex was treated with *(A)* 1 µg trypsin, or *(B)* 5 µg trypsin. The reaction was stopped with protease inhibitor after 1, 2, 5, 10, 30, 60, 90, and 120 minutes. Every time point was analyzed on a 12% SDS-PAGE gel stained with Coomassie Brilliant Blue. The arrows indicate three bands that were analyzed by LC-MS.

4.2.5.6. Structural analysis of the S-Y-T complex by cryo-EM

To determine the structure of the S-Y-T complex from *C. thermophilum*, both nuclease-treated and untreated complexes were analyzed using cryo-EM on TEM grids of gold or copper (Tatjana Schusser & Prof. Dr. Robert Tampé, Goethe

University Frankfurt). Best results were obtained on QUANTIFOIL R 1.2/1.3 plus C2 grids, featuring a 12 nm carbon film with 1.2 µm circular holes and an additional 2 nm carbon layer on a 300 copper mesh (Fig. 44).

The untreated S-Y-T complex displayed smaller, more heterogeneous complexes (Fig. 44A-B), while the treated complex was more homogeneous, with an average size of 20 nm and clear contrast, indicating possible RNA association (Fig. 44C-D).

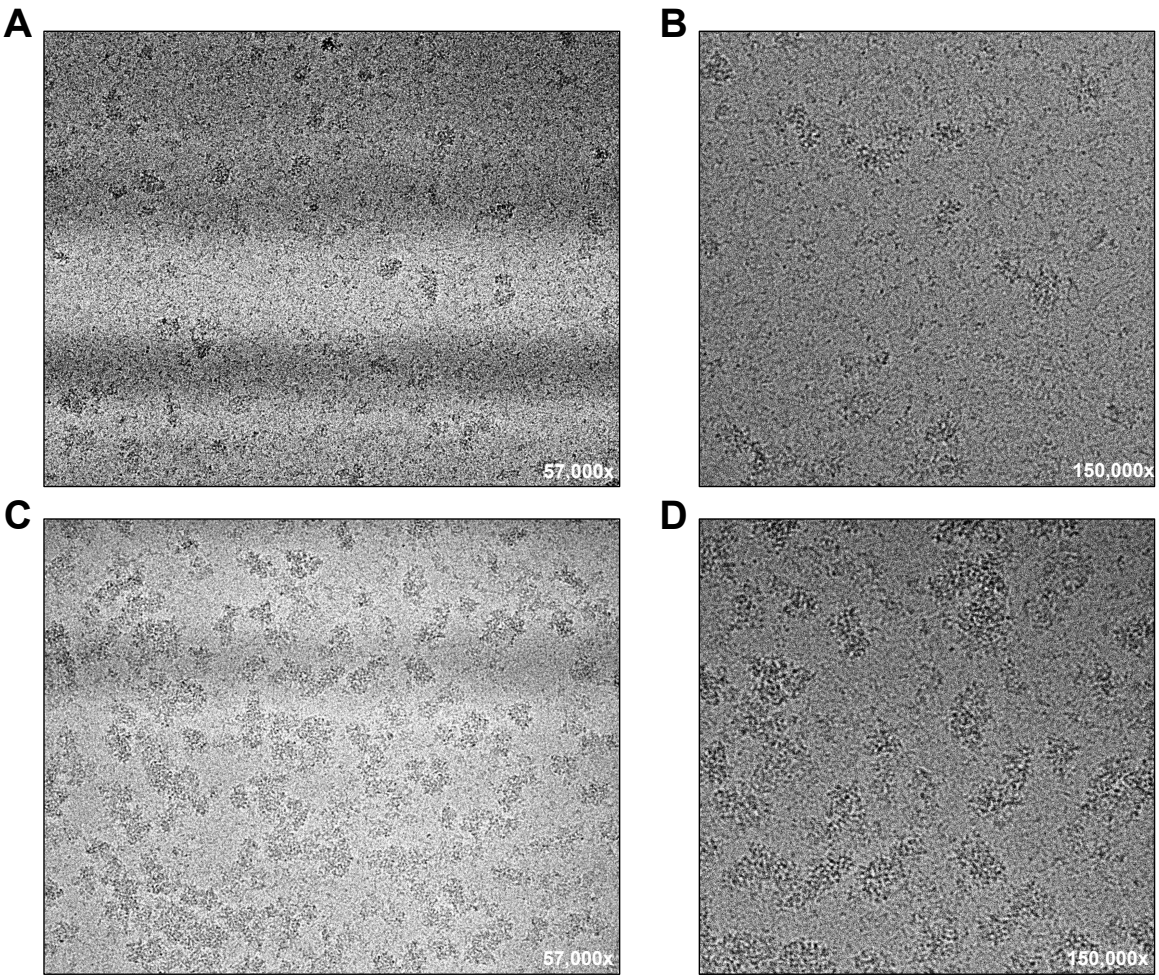


Figure 44: The cryo-EM analysis of S-Y-T reveals RNA bound complexes with a size of 20 nm. The samples were applied to the TEM grids QUANTIFOIL R 1.2/1.3 plus C2 on a 300 copper mesh. *(A-B)* The untreated S-Y-T shows smaller, heterogenous complexes at *(A)* a magnification of 57,000x, and *(B)* 150,000x. *(C-D)* The treated S-Y-T complex displays homogenous complexes measuring approximately 20 nm at a magnification of *(C)* 57,000x and *(D)* 150,000x.

5. Discussion

The TREX (transcription/export) complex is critical for nuclear mRNP assembly and linking transcription to mRNA export (Jimeno et al. 2002; Strässer et al. 2002; Zenklusen et al. 2002). This study aimed to characterize its structural arrangement in *S. cerevisiae*, with a more detailed focus on the C-terminal region of the TREX component Tho2. Additionally, the interaction between the TREX complex and Tho1 was investigated through analysis of the trimeric complex of Sub2, Yra1, and Tho1. For this, recombinant proteins from *C. thermophilum* were used to improve stability and enable detailed structural and functional analysis of the complex and its individual components.

5.1. The endogenous yeast TREX complex

The TREX complex is vital for gene expression and mRNA export, yet its complete structure was not known at the beginning of this study. Solving the structure of this dynamic and multifunctional complex is key to understanding its diverse roles in cellular processes. Structural insights can clarify how TREX components interact with each another, and how they engage with RNA and other cellular machinery during transcription and mRNA export.

5.1.1. Successful purification of the yeast TREX complex

The complete TREX complex was isolated from *S. cerevisiae* using tandem affinity purification (TAP). SDS-PAGE revealed an overabundance of Yra1 (Fig. 13A), consistent with findings by Bonneau et al. (2023). Their study showed that RNase treatment during purification restored Yra1 to equimolar levels, suggesting that, without RNase, longer RNA transcripts with multiple TREX assemblies are co-purified. Yra1's ability to bind RNA independently may explain its overrepresentation during RNase-free purification.

5.1.2. Structural insights into the yeast TREX complex by XL-MS

To investigate TREX's structure, the purified complex was cross-linked and analyzed by XL-MS. BS^3 formed medium-range links (11.4 Å), while ADH-DMTMM created both zero-length and medium-range links (11.4 Å; Leitner et al. 2014).

In total, over 550 cross-links were identified, with more than 445 non-redundant site pairs, particularly between THO components and Sub2. The most common cross-links between different subunits were observed between Tho2 and Hpr1, and between Mft1 and Thp2, which aligns with their known tight interactions in the stable THO core complex. The majority of cross-links were found in the C-terminal regions of Tho2 and Hpr1, as well as within the first 250 residues of Mft1 and Thp2. The observed cross-linking pattern is typical for such flexible regions, yet the pronounced cross-linking preference for these regions within the complex is unusual. To obtain a more balanced distribution of cross-links across the entire complex, variations in cross-linker concentrations or alternative cross-linking reagents could be tested. For instance, Xie et al. (2021) used EDC and DSS cross-linking for the recombinant THO-Gbp2 complex, successfully identifying the involvement of the C-terminal domain of Tho2 in interaction with Gbp2.

Despite the presence of Yra1 in the purified TREX sample, no inter-molecular cross-links between Yra1 and other components were observed (Fig. 14), suggesting that Yra1 might not be a stable component of the complex under the purification conditions used. The lack of RNase treatment during purification might have led to RNA co-purification, where Yra1 was not tightly associated with the complex. The lack of inter-molecular cross-links for Yra1 may also be due to insufficient cross-linking efficiency or structural constraints limiting Yra1's accessibility for cross-linking. The absence of cross-links could have also resulted from more technical limitations, such as constraints in the mass spectrometry instrumentation that can result in the non-detection of specific cross-links.

Intra-molecular cross-links detected in Yra1 with ADH-DMTMM (Fig. 14B) suggest multiple Yra1 copies may have been cross-linked. Bonneau et al. (2023) also analyzed cross-linked TREX complexes by XL-MS and observed intra-molecular cross-links in Yra1. They proposed a model suggesting the participation of multiple copies of Yra1 in mRNP compaction (Bonneau et al. 2023).

In general, XL-MS can provide supplementary structural insights into regions that are not accessible by traditional techniques. The distance restraints from XL-MS are useful for integrative/hybrid modeling approaches (Braitbard et al. 2019; Rout and Sali 2019; Koukos and Bonvin 2020). However, most intra-molecular cross-links in this study connect regions that are close in the primary sequence and have few

interactions with more ordered regions of the complex, preventing a reliable structural model of TREX based solely on the XL-MS data.

5.1.3. Compatibility of cross-links with recombinant yeast TREX complexes

When comparing the cross-linking data to the recombinant yeast TREX structure that was published by Schuller et al. (2020; PDB ID: 7apx), approximately 40% of the cross-links were compatible with the structure. Of these cross-links, roughly 80% of the interactions were in agreement with the monomer, with a maximum distance of 35 Å. This distance is based on the fact that chemical cross-linking typically forms covalent bonds between side chains less than 30-35 Å apart (Merkley et al. 2014). Only three cross-links exclusively aligned with the dimer.

The remaining 20% of cross-links were primarily found within Thp2 and Mft1 or between these two proteins. These cross-links did not align with the recombinant structure, suggesting either cross-linking-induced artifacts or alternative conformations in the native complex. Overall, it seems unlikely that the incompatible cross-links are due to cross-linking-induced artifacts, as these cross-links were consistently observed across multiple independent experiments and with different cross-linking chemistries. Additionally, most identifications had high scores, reducing the likelihood of false positives. The cross-linking conditions were optimized to prevent excessive cross-linking, and recent observations showed that using low reagent concentrations for the ADH-DMTMM reaction resulted in fewer ADH-incorporating cross-links, increasing zero-length cross-links by DMTMM (Mohammadi et al. 2021). Moreover, Xie et al. (2021) used the same DSS concentration as BS^3 (0.5 mM) and found a higher correlation with the structure, suggesting that the cross-linking conditions had a limited effect.

Rather than being due to cross-linking-induced artifacts, the results likely reflect a conformational difference in some of the purified complexes. This suggests that the endogenous TREX complex may exist in multiple structural states, which are crucial for its *in vivo* roles in mRNA biogenesis, in contrast to the more homogenous recombinant complex. These findings align with the dynamic nature of TREX and highlight the importance of using various cross-linking approaches to capture its full structural diversity.

5.1.4. Electron microscopy of the yeast TREX complex

The XL-MS analysis showed that more than half of the non-redundant cross-links are located in areas with limited resolution in the recently published THO–Sub2 structures (such as Schuller et al. 2020). Consequently, the question arose as to whether the arrangement of the cross-linked endogenous TREX complex differs from that of the recombinant THO–Sub2 complex. In addressing this question, the freshly prepared complex was analyzed by negative-stain electron microscopy using uranyl acetate (UANS EM). Overall, three distinct complex preparations were assessed using negative-stain EM. However, only the sample with the BS^3 cross-linked TREX complex produced a sufficient number of usable particles for single-particle analysis, leading to the generation of a 3D reconstruction with an overall resolution of 26 Å (Fig. 17). The overall appearance of the reconstructed BS^3 cross-linked TREX complex closely aligns with the published cryo-EM structures of the recombinant THO–Sub2 complex (Schuller et al. 2020). Notably, unlike the recombinant TREX complex, the endogenous complex includes Hrb1, Gbp2, Yra1, and the C-terminal end of Tex1, which was truncated in the recombinantly purified complexes. Nevertheless, these extra proteins do not appear to alter the overall structure of TREX. However, minor variations in domain orientation and additional densities between the native and recombinant structures were visible, likely due to the presence of these additional proteins. Unfortunately, their precise location could not be definitively determined at the resolution that was obtained for the cross-linked complex. The utilization of 0.5 mM BS^3 for cross-linking resulted in predominantly capturing the TREX complex in its dimeric state, which was also the preferred oligomeric state that was observed for the recombinant yeast TREX complex (Schuller et al. 2020).

Negative-stain images of the BS^3 cross-linked complex showed less heterogeneity compared to the uncross-linked sample, where particles varied greatly in size and shape (Fig. 19). This prevented the generation of suitable 2D classes for 3D reconstruction of the uncross-linked complex. However, images revealed particles resembling dimeric and monomeric states, with larger particles suggesting tetrameric forms or aggregation (Fig. 19B). Similar heterogeneity was observed in an uncross-linked endogenous TREX cryo-EM sample by Bonneau et al. (2023), who found that treatment with benzonase, a nuclease using RNA and DNA as a

substrate, produced more homogeneous particles, suggesting RNA or DNA binding contributes to the observed heterogeneity.

Negative-stain EM of the ADH-DMTMM cross-linked complex revealed particles resembling the uncross-linked form (Fig. 18), suggesting milder cross-linking than BS^3. While the heterogeneous particles were unsuitable for 3D reconstruction, three 2D classes could be identified: 29,660 particles resembling the dimeric state (Fig. 18C), 14,444 likely representing the monomeric form (Fig. 18D), and 5,016 potentially depicting a tetrameric assembly (Fig. 18B).

In summary, negative-stain EM analysis of BS^3, ADH-DMTMM cross-linked, and uncross-linked TREX complexes suggests that the endogenous yeast complex primarily assembles into a dimer. This study shows that the TREX dimer resembles the recombinant THO-Sub2 complex in shape and size, supported by findings of Bonneau et al. (2023). Similarly, Pacheco-Fiallos et al. (2023) confirmed structural consistency between the endogenous human TREX and the recombinant human THO–UAP56 complex as reported by Pühringer et al. (2020). The uncross-linked and ADH-DMTMM cross-linked samples captured the endogenous yeast TREX in distinct assemblies, suggesting its dynamic *in vivo* functions and structures.

5.1.5. Functional analysis of flexible regions in TREX components

The majority of identified cross-links were located in the flexible regions of TREX components, especially the long C-terminal end of Tho2. This study showed that the TREX complex forms even in the absence of the Tho2 C-terminus (Fig. 20A), consistent with findings by Xie et al. (2021), who reconstituted the THO-Sub2 complex using truncated subunits. However, dot spot assays revealed that cells lacking the Tho2 C-terminus exhibit growth defects at 16°C and 37°C (Fig. 20D), as also reported by Pena et al. (2012). Additionally, Pena et al. (2012) demonstrated that this region is necessary for THO complex interaction with nucleic acids, and its removal reduces THO recruitment to active chromatin, causing defects in mRNA biogenesis. Recent work by Beauvais et al. (2023) shows that the Tho2 C-terminus is crucial for recruiting Rrp6, a nuclear RNA exosome component (Mitchell et al. 1997; Briggs et al. 1998), implicating its role in recognizing defective mRNPs for degradation. Furthermore, studies on the human Tho2 homolog (THOC2) link C-terminal deletions to neurodevelopmental disorders (Kumar et al. 2018). These

findings highlight the critical role of the flexible Tho2 region in mRNP biogenesis, particularly under stress conditions.

In addition to numerous cross-links in the flexible region of Tho2, the XL-MS analysis in this study also detected many cross-links in the flexible C-terminal region of Hpr1. Findings from the literature show that, similar to the C-terminus of Tho2, the flexible region of Hpr1 is dispensable for the formation of the THO/TREX complex (Gwizdek et al. 2005; Xie et al. 2021). However, the Hpr1 C-terminus plays key roles in mRNA biogenesis. For instance, it interacts with the ubiquitin-associated domain of Mex67 (Gwizdek et al. 2006), and deletion of its last 88 amino acids reduces Hpr1 degradation rates (Gwizdek et al. 2005). Additionally, the Hpr1 C-terminus is a target for sumoylation (Bretes et al. 2014), with sumoylation and desumoylation proposed to regulate the THO complex's association with mRNPs. It has also been suggested that Hpr1 sumoylation may inhibit interactions between the Tho2 C-terminus and nucleic acids (Bretes et al. 2014).

In summary, the observations from this study and the findings from the literature highlight the critical roles of TREX's flexible regions. Even though these regions may not be resolved in structures, they play crucial roles in mRNA biogenesis, particularly under stress conditions.

5.1.6. Role of Tho2 in mRNA stability

The stability of mRNA can be influenced by various factors. For instance, certain sequences in the 3' UTR of mRNAs, known as AU-rich elements (AREs), are often associated with rapid mRNA degradation (Caput et al. 1986; Helfer et al. 2012). Furthermore, the length of the poly(A) tail is essential for stability, as reduction in the poly(A) tail length is commonly linked to the decay of mRNA (Passmore & Coller 2022). Not only the poly(A) tail protects the mRNA from degradation, but also the 5' cap increases the transcript stability (Topisirovic et al. 2011; Hui et al. 2014; Mauer et al. 2017). Other factors that impact mRNA stability can be the RNA secondary structure (Emory et al. 1992; Mauger et al. 2019) and post-transcriptional modifications, such as methylation (Zhao et al. 2017) or pseudouridylation (Sun et al. 2023). Environmental factors or cellular stress conditions, such as heat shock or oxidative stress, can also influence mRNA stability (Nilsson et al. 1984; Bond 2006). The rate of mRNA synthesis as well as transcription factors and regulatory elements

in the promoter region can affect both transcription and stability (Haimovich et al. 2013; Eser et al. 2014; Das et al. 2017; Espinar et al. 2018). MicroRNAs (miRNAs) have been shown to bind to specific sequences in mRNA, leading to degradation or translational repression (Wilczynska et al. 2015; O'Brien et al. 2018). Finally, proteins that bind to mRNA can either enhance or reduce its stability. For example, proteins like HuR can stabilize mRNA (Xu et al. 2005), while decay-promoting proteins can lead to mRNA degradation (Gherzi et al. 2004; Chou et al. 2006).

Given that Tho2 directly interacts with RNA and serves as the primary platform for TREX complex formation (Pena et al. 2006), an increased recruitment of Tho2 possibly affects transcript stability. Using the MCP-MS2 system, it was observed that overexpression of MCP-Tho2 significantly increased the stability of both tagged (RPS1A-MS2) and untagged (RPL4A) transcripts (Fig. 21). This indicates that Tho2 stabilizes mRNAs independently of the MS2 tag, likely due to its intrinsic RNA-binding capability. The effect was similar for other RNA-binding proteins, such as Npl3 and Nab2 (Supplemental Fig. S2), suggesting a broader role in mRNA stabilization. These findings imply that Tho2 may universally bind and stabilize transcripts when overexpressed, but further studies are needed to confirm this mechanism.

5.2. The Sub2-Yra1-Tho1 (S-Y-T) complex from *C. thermophilum*

The investigation of the Sub2-Yra1-Tho1 subcomplex from *C. thermophilum* offers insights into the molecular interactions between the TREX complex and the RNA-binding protein Tho1. The thermophilic properties of the model organism *C. thermophilum* were beneficial for preserving protein stability, enabling structural analysis that is more challenging with less stable organisms.

5.2.1. The structure of Sub2 from *C. thermophilum* (*ct*Sub2)

Sub2 is a member of the DEAD-box RNA helicase superfamily (La Cruz et al. 1999), and has functions in pre-mRNA splicing (Kistler & Guthrie 2001; Libri et al. 2001; Zhang & Green 2001) as well as mRNA export (Sträßer & Hurt 2001). High-resolution structures of the Sub2 protein from *S. cerevisiae* have been extensively documented. Nevertheless, these structures illustrate Sub2 exclusively in the closed conformation, where the protein is associated with RNA and an ATP analog (Ren et

al. 2017; Schuller et al. 2020; Chen et al. 2021; Xie et al. 2021). Structures of the human Sub2 homolog UAP56/DDX39B reveal the protein in the open conformation (Shi et al. 2004) as well as in the closed conformation (Xie et al. 2023).

In this study, the crystal structure of Sub2 from *C. thermophilum* was successfully resolved, revealing *ct*Sub2 in the open conformation (Fig. 24C). Comparing *ct*Sub2 with yeast and human structures, the RecA-like domains showed strong similarity, but a significant difference was observed in the angle between these domains. In human UAP56/DDX39B, the angle is around 110°, while in *ct*Sub2, it is approximately 220° (Fig. 24). This variation may be due to the thermophilic nature of *C. thermophilum*, as many investigations have demonstrated that thermophilic organisms have evolved their proteins to ensure stability at elevated temperatures (Szilágyi & Závodszky 2000; Taylor & Vaisman 2010; Vijayabaskar & Vishveshwara 2010; Meruelo et al. 2012). The open conformation of *ct*Sub2 may enhance its stability under such conditions. Further studies on *ct*Sub2 in the closed conformation would help clarify its dynamic functions in thermophilic organisms.

5.2.2. The role of *ct*Tho1 in mRNA biogenesis

Tho1 is a nuclear protein involved in transcription, mRNA export, and ribonucleoprotein complex formation, interacting with DNA, RNA, and chromatin (Jimeno et al. 2006).

In this study, the cellular functions of *ct*Tho1 were investigated using specific deletion mutants. DNA- and RNA-binding was assessed in EMSAs (Fig. 31), annealing activity was analyzed in FRET-based assays (Fig. 32), and interactions with *ct*Sub2 were investigated in pull-down experiments (Fig. 33). These experiments demonstrated the crucial role of the flexible C-terminal end (CTE) of *ct*Tho1 in binding nucleic acids and facilitating the annealing of DNA/DNA, DNA/RNA, and RNA/RNA hybrids (Fig. 31-32), in line with findings from homologous yeast Tho1 mutants (Miosga 2022). The 27 positively charged amino acids in the CTE, including 18 arginines (R) and 9 lysines (K), likely facilitate interactions with negatively charged nucleic acids (Fig. 45). In contrast, the SAP domain, proposed to bind DNA (Aravind & Koonin 2000), appears dispensable for nucleic acid binding and annealing (Fig. 31-32), possibly requiring longer or more specific DNA sequences for interaction.



	aa(+)	aa(-)
SAP	8	8
Linker	11	20
CTD	9	8
CTE	27	10

Figure 45: The positive charge of the *ct*Tho1 CTE most likely facilitates the interaction with nucleic acids. The amino acid (aa) sequence of wild-type *ct*Tho1is shown, with the SAP domain highlighted in light blue and the CTD in dark blue. The positive charge of arginine (R) and lysine (K), along with the negative charge of aspartic acid (D) and glutamic acid (E), is indicated. Within the SAP domain, there are eight positively charged aa and eight negatively charged aa. The Linker consists of 11 positive and 20 negative aa, the CTD has nine positive and eight negative aa, and the CTE comprises 27 positive and ten negative aa.

In conclusion, the *in vitro* experiments suggest important roles for the flexible region in *ct*Tho1, but it remains unclear which of these functions are carried out *in vivo*. Previous studies showed that overexpressing Tho1 alleviates mRNA export defects and hyperrecombination in *Δhpr1* and other THO complex null mutants (Jimeno et al. 2006), suggesting a potential *in vivo* role for the CTE in resolving R-loops. R-loops frequently form during transcription, where the recently transcribed RNA strand binds with the corresponding DNA strand, creating a loop with the non-template DNA strand. Although R-loops are involved in regular cellular functions like transcription and DNA replication, they have also been linked to genomic instability, leading to issues such as DNA damage, replication stress, and chromosomal rearrangements (Belotserkovskii et al. 2018). Tho1 may aid in R-loop resolution by facilitating DNA reannealing and displacing RNA.

The interaction between *ct*Tho1 and *ct*Sub2, mediated by the C-terminal domain (CTD), could assist in this process. Pull-down experiments in this study confirm this interaction (Fig. 33), supported by a recent structure showing the binding of yeast Tho1 CTD peptides to two human UAP56/DDX39B molecules (Fig. 46). The yeast Tho1 CTD contains a DIM (DDX39B interacting motif) in each of its two α-helices, which is essential for interacting with UAP56/DDX39B (Xie et al. 2023; Fig. 46). Similarly, the *C. thermophilum* Tho1 homolog also has two DIM motifs in its CTD, along with an additional third DIM motif in the CTE's alpha helix. The CTE of human

Tho1 homolog SARNP even has three alpha helices, each with an extra DIM motif. This suggests that while yeast Tho1 might bind two Sub2 molecules, human SARNP could interact with five UAP56/DDX39B molecules, and *C. thermophilum* Tho1 might bind three *ct*Sub2 molecules simultaneously. Supporting this hypothesis, pull-down experiments in this study showed that the *ct*Tho1(ΔCTD) mutant retains some interaction with *ct*Sub2, though with lower affinity, and the interaction is completely lost only when both the CTD and CTE alpha helix are deleted (Fig. 33).

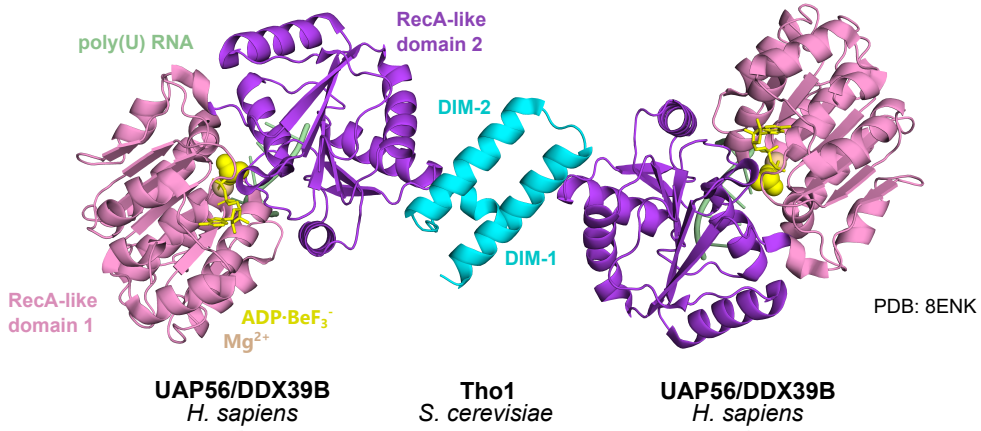


Figure 46: The yeast Tho1 CTD forms a complex with two UAP56/DDX39B molecules. The structural analysis (PDB: 8ENK) reveals the interaction of the *S. cerevisiae* Tho1, engaging through two DIM motifs in each α-helix with the RecA-like domain 2 of a total of two UAP56/DDX39B molecules. The structure shows the binding of the UAP56/DDX39B molecules in closed conformation to poly(U) RNA and ADP-BeF_3^- (Xie et al. 2023). Structure was visualized and processed with *PyMOL* (Schrödinger & DeLano 2020).

Taken together, these findings support a model where Sub2, as part of the co-transcriptionally recruited TREX complex, facilitates the interaction between Tho1 and newly synthesized RNA, particularly in resolving transcription-induced R-loops. Tho1 potentially enhances the helicase activity of Sub2 by reannealing the two DNA strands, ultimately resulting in the resolution of R-loops.

5.2.3. *In vitro* interactions of *ct*Sub2 with *ct*Tho1 and *ct*Yra1

In this study the *ct*Sub2-*ct*Tho1 complex was reconstituted *in vitro* to gain insights into its structural and functional properties, revealing information about its stoichiometry and dynamics (Fig. 34). Size exclusion chromatography of the complex showed that *ct*Sub2 is present in a stoichiometry about two to three times higher than *ct*Tho1. This ratio remained consistent even after deleting specific regions of the proteins, such as the C-terminal end (CTE) of *ct*Tho1 and the flexible

N-terminus of *ct*Sub2, or the addition of ATP analogs (Fig. 34). This finding suggests a stable association between *ct*Tho1 and *ct*Sub2, with one *ct*Tho1 protein likely interacting with up to three *ct*Sub2 molecules, consistent with the recently published structure by Xie et al. (2023; Fig. 47) and the presence of three DIM motifs in *ct*Tho1. Additionally, the reconstituted *ct*Sub2-*ct*Tho1 complex eluted earlier than expected in size exclusion chromatography, indicating that the complex may be larger than a simple *ct*Sub2-*ct*Tho1 dimer (Fig. 34). This supports the hypothesis that the complex forms a higher-order structure. Notably, *ct*Tho1 alone also eluted earlier than expected, which is likely due to the protein's unstructured regions.

In addition to the *ct*Sub2-*ct*Tho1 complex, the ability of *ct*Yra1 to form a complex with both *ct*Sub2 and *ct*Tho1 was tested. Previous studies have shown that recombinant human homologs of Sub2, Tho1, and Yra1 can form a trimeric complex, with UAP56/DDX39B interacting directly with SARNP and ALY/REF, while SARNP and ALY/REF do not interact with each other (Dufu et al. 2010). Although the *ct*Sub2-*ct*Tho1 complex was successfully reconstituted, the *ct*Sub2-*ct*Tho1-*ct*Yra1 trimer could not be formed *in vitro* (Fig. 37E). This possibly suggests that forming a trimeric complex in *C. thermophilum* may require additional factors, such as RNA or other cofactors, not present in the tested *in vitro* conditions. Reconstitution with *ct*Yra1 may be improved by adding varying RNA lengths, adding ATP or analogs, or adjusting protein tag positions.

In conclusion, these findings provide insights into the formation and stoichiometry of the *ct*Sub2-*ct*Tho1 complex, its potential interactions with *ct*Yra1, and suggest that optimizing conditions could promote the formation of a stable trimeric complex.

5.2.4. Structural analysis of the *ct*Sub2-*ct*Yra1-*ct*Tho1 (S-Y-T) complex

Sufficient amounts of *ct*Sub2-*ct*Yra1-*ct*Tho1 complexes were successfully obtained for structural analysis by co-expressing the complex from a polycistronic plasmid in *E. coli*. Co-expression offers several advantages over *in vitro* reconstitution, such as providing a physiological environment for complex formation and achieving correct stoichiometry of components (Romier et al. 2006; Stefan et al. 2015).

All tested co-expressed complexes strongly co-purified with nucleic acids. Analysis revealed that the *ct*Sub2-*ct*Yra1-*ct*Tho1 complex binds both DNA and RNA. Deleting the CTE of *ct*Tho1 yielded a DNA-only complex, while deleting the SAP domain and

linker eliminated DNA binding, leaving an RNA-only complex (Fig. 38). The results show that the SAP domain of *ct*Tho1 is needed for DNA binding of the complex, as it interacted with *E. coli* DNA. However, the SAP domain failed to bind shorter DNA fragments in EMSA and FRET assays (Fig. 31-32), supporting the hypothesis that the SAP domain requires a minimum DNA length for stable interaction. Shorter fragments likely lack the necessary binding interface for effective interaction. While the SAP domain of *ct*Tho1 seems primarily responsible for DNA binding in the trimeric complex, the results indicate that the CTE of *ct*Tho1, along with *ct*Yra1 and *ct*Sub2, facilitates RNA binding. In general, the complex likely has a global affinity for nucleic acids rather than sequence specificity, as shown by its ability to bind prokaryotic *E. coli* RNA and DNA.

For structural studies, contaminations of the complex with nonspecific nucleic acids should be reduced, as such heterogeneity can hinder obtaining high-resolution structures. The co-purified *E. coli* nucleic acids potentially obscure specific interactions or introduce variability into the complex. Nuclease treatment during purification addressed this, while also improving protein yield and preserving the stability of the trimeric complex (Fig. 39). The observation that fragments of 100-200 base pairs remained bound suggests that these nucleic acids are tightly integrated within the complex, likely contributing to its structural integrity or formation (Fig. 39). EMSAs and RNA pull-downs confirmed that the nuclease-treated complex could still bind specific RNA, indicating that RNA-binding functionality was preserved (Fig. 40). However, it remains unclear whether the added RNA replaces the tightly bound *E. coli* nucleic acids or binds alongside them. Further studies are needed to resolve this and ensure the complex reflects its functional state for structural analysis.

The nuclease-treated *ct*Sub2-*ct*Yra1-*ct*Tho1 complex remained stable during short-term storage at 4°C and long-term storage at -80°C (Fig. 41), which is essential for experimental reproducibility, particularly in structural studies. Complex stability at physiological salt concentrations (up to 150–200 mM) indicates that the complex can function effectively under conditions similar to those *in vivo*. The dissociation of *ct*Yra1 at higher salt concentrations (>250 mM) suggests its dynamic association with the stable *ct*Sub2-*ct*Tho1 core. The interaction of *ct*Yra1 might be regulated by ionic strength, possibly reflecting its transient role in binding or releasing nucleic acids during the complex's functional cycle. Limited proteolysis and LC-MS showed

that *ct*Sub2, as the most structured protein, is highly resistant to trypsin digestion, while *ct*Tho1 and *ct*Yra1 are digested faster, confirming their greater flexibility and less structured nature. This flexibility may be functionally important, allowing these proteins to adapt their conformations during nucleic acid binding or release. However, while essential for function, it introduces challenges in maintaining an intact and homogenous complex, which is important for structural analysis. Overall, these findings suggest a modular organization of the trimeric complex with a stable *ct*Sub2-*ct*Tho1 core that provides structural integrity and a more dynamic *ct*Yra1 that likely facilitates regulatory or interaction-specific roles.

Cryo-EM analysis revealed that nuclease treatment enhanced the homogeneity of *ct*Sub2-*ct*Yra1-*ct*Tho1 complexes, making them better suited for high-quality 2D class generation than untreated complexes (Fig. 44). The *ct*Sub2-*ct*Yra1-*ct*Tho1 complex formed larger assemblies exceeding the expected molecular weight of a single trimeric complex and was shown to contain RNA, confirming results of earlier experiments (Fig. 38-40). It is possible that multiple complexes associate with a single RNA molecule, resulting in larger assemblies. Additionally, the three DIM motifs of *ct*Tho1 could bind up to three *ct*Sub2 molecules, contributing to the formation of larger complexes. To achieve high-resolution reconstruction, improved complex preparation and particle labeling – such as immunogold labeling or the use of ~50 kDa Fabs (monoclonal fragments antigen binding) for domain-specific labeling – are options for precise identification of individual components within the complex (Goldberg & Fišerová 2016; Wu et al. 2012).

These findings highlight the dynamic nature of the *ct*Sub2-*ct*Yra1-*ct*Tho1 complex, with distinct roles for each component in nucleic acid binding and structural integrity. Cryo-EM images revealed improved homogeneity of the complex after nuclease treatment and provided insights into the larger assemblies formed by the complex. Further optimization of complex preparation will be crucial for achieving high-resolution structures of this multifunctional complex.

5.3. Conclusions and outlook

This study provided structural and functional insights into the TREX complex. Cross-linking mass spectrometry and negative-stain EM confirmed structural alignment of the endogenous TREX complex with published models of recombinant complexes

and revealed that the native yeast TREX complex predominantly forms dimers (Kern et al. 2023). Moreover, this study highlighted the importance of flexible protein regions, showing that the C-terminus of yeast Tho2 is essential for cellular stress response, while the C-terminal end of *ct*Tho1 facilitates nucleic acid binding. Co-expression experiments demonstrated the stable assembly of the *ct*Sub2-*ct*Tho1-*ct*Yra1 complex in *E. coli*, with a highly stable *ct*Sub2-*ct*Tho1 core. Future experiments could optimize *ct*Yra1 binding to the *ct*Sub2-*ct*Tho1 complex by testing different RNAs, ATP analogs, or protein tags. These efforts, alongside techniques like immunogold or Fab labeling, may enhance 3D structure reconstruction from cryo-EM images.

This study showed that the *ct*Sub2-*ct*Tho1-*ct*Yra1 complex forms larger, RNA-bound assemblies. Based on these findings and on studies by Xie et al. (2023), it is likely that in *C. thermophilum*, *ct*Tho1 binds up to three *ct*Sub2 molecules, which then might be compacted by multiple *ct*Yra1 proteins via its N- and C-box (Fig. 47). One possible model to explain the function of the complex composed of *ct*Sub2, *ct*Tho1, and *ct*Yra1 involves their coordinated role in mRNA processing, packaging, and transport. The ATP-dependent helicase Sub2 is known for its role in splicing and mRNA export (Kistler & Guthrie 2001; Libri et al. 2001; Sträßer & Hurt 2001; Zhang & Green 2001), while Tho1, for instance, has been linked to R-loop resolution (Piruat & Aguilera 1998; Jimeno et al. 2006). It is possible that, during transcription, Tho1 collaborates with Sub2 to unwind R-loops by displacing the RNA from the DNA and reannealing the two DNA strands. After R-loop resolution, Yra1 possibly joins the complex and subsequently facilitates the export of the mRNA that can now dissociate from the DNA (Fig. 48). In general, these three proteins – and additional proteins and protein complexes like TREX – may work together to ensure the efficient and accurate processing, transport, and export of mRNA within the cell. Experimental validation would be crucial to confirm and refine this proposed model.

The pentameric THO complex from *C. thermophilum* was successfully co-expressed and analyzed by EM (Supplemental Fig. S4), revealing predominantly monomeric particles. In future experiments, reconstituting *ct*THO with the *ct*Sub2-*ct*Tho1-*ct*Yra1 complex could provide additional insights on TREX-Tho1 interactions and their role in mRNA assembly.

In conclusion, this study advances understanding of the TREX complex, revealing key interactions between Sub2, Tho1, and Yra1, and providing a foundation for future research into its role in mRNA processing, packaging, and export.

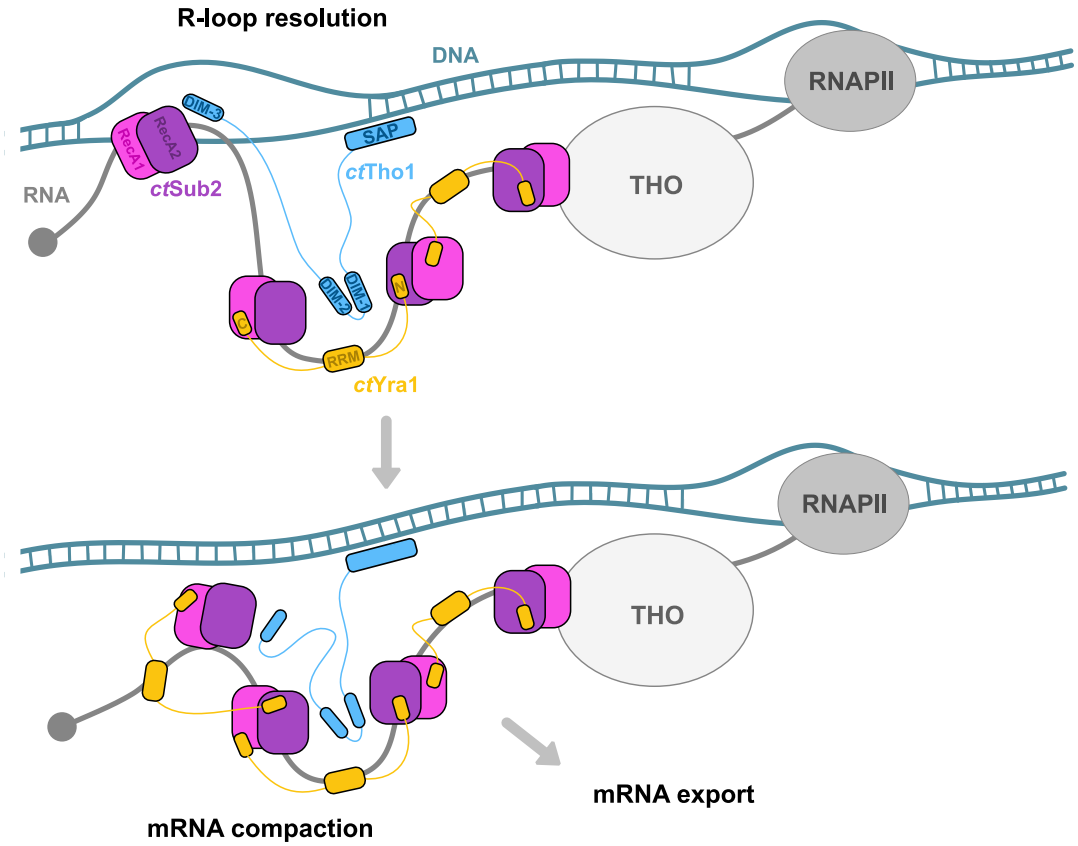


Figure 47: Model illustrating the potential role of the *ct*Sub2-*ct*Yra1-*ct*Tho1 complex in mRNA assembly. While transcription is ongoing, the formation of R-loops is possible. It is conceivable that the ATP-dependent helicase Sub2 works in conjunction with Tho1 to unravel these R-loop structures, displacing the RNA from the DNA and reannealing the two DNA strands. Following the resolution of R-loops, Yra1 potentially becomes part of the complex, aiding in mRNA compaction, and subsequently facilitating the export of mRNA, which can now dissociate from the DNA. In *C. thermophilum*, *ct*Tho1 possesses three DIM motifs, enabling the interaction of one ctTho1 molecule with up to three *ct*Sub2 molecules. In conjunction with *ct*Yra1, this results in the formation of larger *ct*Sub2-*ct*Tho1-*ct*Yra1 complexes. The model was modified from Xie et al. (2023) and expanded to encompass potential functions and structural organization in *C. thermophilum*.

References

Abovich, N.; **Liao**, X. C.; **Rosbash**, M. (1994): The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. In: *Genes & development* 8 (7), S. 843–854. DOI: 10.1101/gad.8.7.843.

Adams, R. L.; **Wente**, S. R. (2020): Dbp5 associates with RNA-bound Mex67 and Nab2 and its localization at the nuclear pore complex is sufficient for mRNP export and cell viability. In: *PLoS genetics* 16 (10), e1009033. DOI: 10.1371/journal.pgen.1009033.

Aebi, U.; **Heggeler**, B.; **Onorato**, L.; **Kistler**, J.; **Showe**, M. K. (1977): New method for localizing proteins in periodic structures: Fab fragment labeling combined with image processing of electron micrographs. In: *Proceedings of the National Academy of Sciences of the United States of America* 74 (12), S. 5514–5518. DOI: 10.1073/pnas.74.12.5514.

Aguilera, A.; **García-Muse**, T. (2012): R loops: from transcription byproducts to threats to genome stability. In: *Molecular cell* 46 (2), S. 115–124. DOI: 10.1016/j.molcel.2012.04.009.

Alberts, B.; **Johnson**, A.; **Lewis**, J.; **Walter**, P.; **Raff**, M.; **Roberts**, K. (2002): Molecular Biology of the Cell 4th Edition: International Student Edition. 4, überarbeitet: Routledge.

Alcázar-Román, A. R.; **Tran**, E. J.; **Guo**, S.; **Wente**, S. R. (2006): Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. In: *Nature cell biology* 8 (7), S. 711–716. DOI: 10.1038/ncb1427.

Amlacher, S.; **Sarges**, P.; **Flemming**, D.; **van Noort**, V.; **Kunze**, R.; **Devos**, D. P. et al. (2011): Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. In: *Cell* 146 (2), S. 277–289. DOI: 10.1016/j.cell.2011.06.039.

Aravind, L.; **Koonin**, E. V. (2000): SAP - a putative DNA-binding motif involved in chromosomal organization. In: *Trends in biochemical sciences* 25 (3), S. 112–114. DOI: 10.1016/S0968-0004(99)01537-6.

Ausubel, F. M.; **Brent**, R.; **Kingston**, R. E.; **Moore**, D. D.; **Seidman**, J. G.; **Smith**, J. A.; **Struhl**, K. (Hg.) (1994): Current Protocols in Molecular Biology. New York: Wiley.

Batisse, J.; **Batisse**, C.; **Budd**, A.; **Böttcher**, B.; **Hurt**, E. (2009): Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. In: *The Journal of biological chemistry* 284 (50), S. 34911–34917. DOI: 10.1074/jbc.M109.062034.

Beauvais, V.; **Moreau**, K.; **Zunar**, B.; **Hervouet-Coste**, N.; **Novacic**, A.; **Le Dantec**, A. et al. (2023): Tho2 is critical for the recruitment of Rrp6 to chromatin in response to perturbed mRNP biogenesis. In: *RNA (New York, N.Y.)*. DOI: 10.1261/rna.079707.123.

Belotserkovskii, B. P.; **Tornaletti**, S.; **D'Souza**, A. D.; **Hanawalt**, P. C. (2018): R-loop generation during transcription: Formation, processing and cellular outcomes. In: *DNA repair* 71, S. 69–81. DOI: 10.1016/j.dnarep.2018.08.009.

Bock, T.; **Chen**, W.-H.; **Ori**, A.; **Malik**, N.; **Silva-Martin**, N.; **Huerta-Cepas**, J. et al. (2014): An integrated approach for genome annotation of the eukaryotic thermophile Chaetomium thermophilum. In: *Nucleic acids research* 42 (22), S. 13525–13533. DOI: 10.1093/nar/gku1147.

Bolger, T. A.; **Folkmann**, A. W.; **Tran**, E. J.; **Wente**, S. R. (2008): The mRNA export factor Gle1 and inositol hexakisphosphate regulate distinct stages of translation. In: *Cell* 134 (4), S. 624–633. DOI: 10.1016/j.cell.2008.06.027.

Bond, U. (2006): Stressed out! Effects of environmental stress on mRNA metabolism. In: *FEMS yeast research* 6 (2), S. 160–170. DOI: 10.1111/j.1567-1364.2006.00032.x.

Boreikaitė, V.; **Passmore**, L. A. (2023): 3'-End Processing of Eukaryotic mRNA: Machinery, Regulation, and Impact on Gene Expression. In: *Annual review of biochemistry* 92, S. 199–225. DOI: 10.1146/annurev-biochem-052521-012445.

Bousquet-Antonelli, C.; **Presutti**, C.; **Tollervey**, D. (2000): Identification of a regulated pathway for nuclear pre-mRNA turnover. In: *Cell* 102 (6), S. 765–775. DOI: 10.1016/s0092-8674(00)00065-9.

Braitbard, M.; **Schneidman-Duhovny**, D.; **Kalisman**, N. (2019): Integrative Structure Modeling: Overview and Assessment. In: *Annual review of biochemistry* 88, S. 113–135. DOI: 10.1146/annurev-biochem-013118-111429.

Brambilla, M.; **Martani**, F.; **Bertacchi**, S.; **Vitangeli**, I.; **Branduardi**, P. (2019): The Saccharomyces cerevisiae poly (A) binding protein (Pab1): Master regulator of mRNA metabolism and cell physiology. In: *Yeast* (Chichester, England) 36 (1), S. 23–34. DOI: 10.1002/yea.3347.

Bretes, H.; **Rouviere**, J. O.; **Leger**, T.; **Oeffinger**, M.; **Devaux**, F.; **Doye**, V.; **Palancade**, B. (2014): Sumoylation of the THO complex regulates the biogenesis of a subset of mRNPs. In: *Nucleic acids research* 42 (8), S. 5043–5058. DOI: 10.1093/nar/gku124.

Briggs, M. W.; **Burkard**, K. T.; **Butler**, J. S. (1998): Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. In: *The Journal of biological chemistry* 273 (21), S. 13255–13263. DOI: 10.1074/jbc.273.21.13255.

Bucheli, M. E.; **Buratowski**, S. (2005): Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. In: *The EMBO journal* 24 (12), S. 2150–2160. DOI: 10.1038/sj.emboj.7600687.

Caput, D.; **Beutler**, B.; **Hartog**, K.; **Thayer**, R.; **Brown-Shimer**, S.; **Cerami**, A. (1986): Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. In: *Proceedings of the National Academy of Sciences of the United States of America* 83 (6), S. 1670–1674. DOI: 10.1073/pnas.83.6.1670.

Carmody, S. R.; **Wente**, S. R. (2009): mRNA nuclear export at a glance. In: *Journal of cell science* 122 (Pt 12), S. 1933–1937. DOI: 10.1242/jcs.041236.

Casamassimi, A.; **Ciccodicola**, A. (2019): Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation. In: *International journal of molecular sciences* 20 (6). DOI: 10.3390/ijms20061281.

Casañal, A.; **Kumar**, A.; **Hill**, C. H.; **Easter**, A. D.; **Emsley**, P.; **Degliesposti**, G. et al. (2017): Architecture of eukaryotic mRNA 3'-end processing machinery. In: *Science (New York, N.Y.)* 358 (6366), S. 1056–1059. DOI: 10.1126/science.aao6535.

Chan, S.-P.; **Cheng**, S.-C. (2005): The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. In: *The Journal of biological chemistry* 280 (35), S. 31190–31199. DOI: 10.1074/jbc.M505060200.

Chanarat, S.; **Burkert-Kautzsch**, C.; **Meinel**, D. M.; **Sträßer**, K. (2012): Prp19C and TREX: interacting to promote transcription elongation and mRNA export. In: *Transcription* 3 (1), S. 8–12. DOI: 10.4161/trns.3.1.19078.

Chanarat, S.; **Seizl**, M.; **Strässer**, K. (2011): The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. In: *Genes & development* 25 (11), S. 1147–1158. DOI: 10.1101/gad.623411.

Chanarat, S.; **Sträßer**, K. (2013): Splicing and beyond: the many faces of the Prp19 complex. In: *Biochimica et biophysica acta* 1833 (10), S. 2126–2134. DOI: 10.1016/j.bbamcr.2013.05.023.

Chávez, S.; **Beilharz**, T.; **Rondón**, A. G.; **Erdjument-Bromage**, H.; **Tempst**, P.; **Svejstrup**, J. Q. et al. (2000): A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. In: *The EMBO journal* 19 (21), S. 5824–5834. DOI: 10.1093/emboj/19.21.5824.

Chen, C.; **Tan**, M.; **Wu**, Z.; **Zhang**, Y.; **He**, F.; **Lu**, Y. et al. (2021): Structural and functional insights into R-loop prevention and mRNA export by budding yeast THO-Sub2 complex. In: *Science bulletin* 66 (23), S. 2347–2352. DOI: 10.1016/j.scib.2021.08.004.

Chen, X.; **Xu**, Y. (2022): Structural insights into assembly of transcription preinitiation complex. In: *Current opinion in structural biology* 75, S. 102404. DOI: 10.1016/j.sbi.2022.102404.

Cheng, H.; **Dufu**, K.; **Lee**, C.-S.; **Hsu**, J. L.; **Dias**, A.; **Reed**, R. (2006): Human mRNA export machinery recruited to the 5' end of mRNA. In: *Cell* 127 (7), S. 1389–1400. DOI: 10.1016/j.cell.2006.10.044.

Chou, C.-F.; **Mulky**, A.; **Maitra**, S.; **Lin**, W.-J.; **Gherzi**, R.; **Kappes**, J.; **Chen**, C.-Y. (2006): Tethering KSRP, a decay-promoting AU-rich element-binding protein, to mRNAs elicits mRNA decay. In: *Molecular and cellular biology* 26 (10), S. 3695–3706. DOI: 10.1128/MCB.26.10.3695-3706.2006.

Clancy, S.; **Brown**, W. (2008) Translation: DNA to mRNA to Protein. In: *Nature Education* 1(1):101.

Cole, C. N.; **Hammell**, C. M. (1998): Nucleocytoplasmic transport: driving and directing transport. In: *Current biology* : CB 8 (11), R368-72. DOI: 10.1016/S0960-9822(98)70239-8.

Colot, H. V.; **Stutz**, F.; **Rosbash**, M. (1996): The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. In: *Genes & development* 10 (13), S. 1699–1708. DOI: 10.1101/gad.10.13.1699.

Connelly, S.; **Manley**, J. L. (1988): A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. In: *Genes & development* 2 (4), S. 440–452. DOI: 10.1101/gad.2.4.440.

Cordin, O.; **Tanner**, N. K.; **Doère**, M.; **Linder**, P.; **Banroques**, J. (2004): The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. In: *The EMBO journal* 23 (13), S. 2478–2487. DOI: 10.1038/sj.emboj.7600272.

Couvillion, M.; **Harlen**, K. M.; **Lachance**, K. C.; **Trotta**, K. L.; **Smith**, E.; **Brion**, C. et al. (2022): Transcription elongation is finely tuned by dozens of regulatory factors. In: *eLife* 11. DOI: 10.7554/eLife.78944.

Coyne, A. N.; **Rothstein**, J. D. (2022): Nuclear pore complexes - a doorway to neural injury in neurodegeneration. In: *Nature reviews. Neurology* 18 (6), S. 348–362. DOI: 10.1038/s41582-022-00653-6.

Crick, F. (1970): Central dogma of molecular biology. In: *Nature* 227 (5258), S. 561–563. DOI: 10.1038/227561a0.

Das, S.; **Sarkar**, D.; **Das**, B. (2017): The interplay between transcription and mRNA degradation in *Saccharomyces cerevisiae*. In: *Microbial cell (Graz, Austria)* 4 (7), S. 212–228. DOI: 10.15698/mic2017.07.580.

Decroly, E.; **Ferron**, F.; **Lescar**, J.; **Canard**, B. (2011): Conventional and unconventional mechanisms for capping viral mRNA. In: *Nature reviews. Microbiology* 10 (1), S. 51–65. DOI: 10.1038/nrmicro2675.

Dermody, J. L.; **Dreyfuss**, J. M.; **Villén**, J.; **Ogundipe**, B.; **Gygi**, S. P.; **Park**, P. J. et al. (2008): Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. In: *PloS one* 3 (9), e3273. DOI: 10.1371/journal.pone.0003273.

Dufu, K.; **Livingstone**, M. J.; **Seebacher**, J.; **Gygi**, S. P.; **Wilson**, S. A.; **Reed**, R. (2010): ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. In: *Genes & development* 24 (18), S. 2043–2053. DOI: 10.1101/gad.1898610.

Eaton, J. D.; **Francis**, L.; **Davidson**, L.; **West**, S. (2020): A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. In: *Genes & development* 34 (1-2), S. 132–145. DOI: 10.1101/gad.332833.119.

Emory, S. A.; **Bouvet**, P.; **Belasco**, J. G. (1992): A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. In: *Genes & development* 6 (1), S. 135–148. DOI: 10.1101/gad.6.1.135.

Eser, P.; **Demel**, C.; **Maier**, K. C.; **Schwalb**, B.; **Pirkl**, N.; **Martin**, D. E. et al. (2014): Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. In: *Molecular systems biology* 10 (1), S. 717. DOI: 10.1002/msb.134886.

Espinar, L.; **Schikora** T., **Miquel** À.; **Domingo**, J.; **Carey**, L. B. (2018): Promoter architecture determines cotranslational regulation of mRNA. In: *Genome research* 28 (4), S. 509–518. DOI: 10.1101/gr.230458.117.

Fabre, E.; **Hurt**, E. (1997): Yeast genetics to dissect the nuclear pore complex and nucleocytoplasmic trafficking. In: *Annual review of genetics* 31, S. 277–313. DOI: 10.1146/annurev.genet.31.1.277.

Fischer, T.; **Strässer**, K.; **Rácz**, A.; **Rodriguez-Navarro**, S.; **Oppizzi**, M.; **Ihrig**, P. et al. (2002): The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. In: *The EMBO journal* 21 (21), S. 5843–5852. DOI: 10.1093/emboj/cdf590.

Flaherty, S. M.; **Fortes**, P.; **Izaurralde**, E.; **Mattaj**, I. W.; **Gilmartin**, G. M. (1997): Participation of the nuclear cap binding complex in pre-mRNA 3' processing. In: *Proceedings of the National Academy of Sciences of the United States of America* 94 (22), S. 11893–11898. DOI: 10.1073/pnas.94.22.11893.

Fontana, A.; **Polverino de Laureto**, P.; **Filippis**, V. de; **Scaramella**, E.; **Zambonin**, M. (1997): Probing the partly folded states of proteins by limited proteolysis. In: *Folding & design* 2 (2), R17-26. DOI: 10.1016/S1359-0278(97)00010-2.

Fujinaga, K.; **Huang**, F.; **Peterlin**, B. M. (2023): P-TEFb: The master regulator of transcription elongation. In: *Molecular Cell* 83 (3), S. 393–403. DOI: 10.1016/j.molcel.2022.12.006.

Fukuda, S.; **Wu**, D. W.; **Stark**, K.; **Pelus**, L. M. (2002): Cloning and characterization of a proliferation-associated cytokine-inducible protein, CIP29. In: *Biochemical and biophysical research communications* 292 (3), S. 593–600. DOI: 10.1006/bbrc.2002.6680.

Gajos, M.; **Jasnovidova**, O.; **van Bömmel**, A.; **Freier**, S.; **Vingron**, M.; **Mayer**, A. (2021): Conserved DNA sequence features underlie pervasive RNA polymerase pausing. In: *Nucleic acids research* 49 (8), S. 4402–4420. DOI: 10.1093/nar/gkab208.

Gehring, N. H.; **Roignant**, J.-Y. (2021): Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. In: *Trends in genetics : TIG* 37 (4), S. 355–372. DOI: 10.1016/j.tig.2020.10.008.

Gehring, N. H.; **Wahle**, E.; **Fischer**, U. (2017): Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. In: *Trends in biochemical sciences* 42 (5), S. 369–382. DOI: 10.1016/j.tibs.2017.02.004.

Gherzi, R.; **Lee**, K.-Y.; **Briata**, P.; **Wegmüller**, D.; **Moroni**, C.; **Karin**, M.; **Chen**, C.-Yi (2004): A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. In: *Molecular cell* 14 (5), S. 571–583. DOI: 10.1016/j.molcel.2004.05.002.

Gilbert, W.; **Guthrie**, C. (2004): The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. In: *Molecular Cell* 13 (2), S. 201–212. DOI: 10.1016/s1097-2765(04)00030-9.

Goldberg, M. W.; **Fišerová**, J. (2016): Immunogold Labeling for Scanning Electron Microscopy. In: *Methods in molecular biology (Clifton, N.J.)* 1474, S. 309–325. DOI: 10.1007/978-1-4939-6352-2_20.

Gómez-González, B.; **García-Rubio**, M.; **Bermejo**, R.; **Gaillard**, H.; **Shirahige**, K.; **Marín**, A. et al. (2011): Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. In: *The EMBO journal* 30 (15), S. 3106–3119. DOI: 10.1038/emboj.2011.206.

Greenfield, N. J. (2006): Using circular dichroism spectra to estimate protein secondary structure. In: *Nature protocols* 1 (6), S. 2876–2890. DOI: 10.1038/nprot.2006.202.

Gromadzka, A. M.; **Steckelberg**, A.-L.; **Singh**, K. K.; **Hofmann**, K.; **Gehring**, N. H. (2016): A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. In: *Nucleic acids research* 44 (5), S. 2348–2361. DOI: 10.1093/nar/gkw009.

Gross, T.; **Siepmann**, A.; **Sturm**, D.; **Windgassen**, M.; **Scarcelli**, J. J.; **Seedorf**, M. et al. (2007): The DEAD-box RNA helicase Dbp5 functions in translation termination. In: *Science (New York, N.Y.)* 315 (5812), S. 646–649. DOI: 10.1126/science.1134641.

Güttler, T.; **Görlich**, D. (2011): Ran-dependent nuclear export mediators: a structural perspective. In: *The EMBO journal* 30 (17), S. 3457–3474. DOI: 10.1038/emboj.2011.287.

Gwizdek, C.; **Iglesias**, N.; **Rodriguez**, M. S.; **Ossareh-Nazari**, B.; **Hobeika**, M.; **Divita**, G. et al. (2006): Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. In: *Proceedings of the National Academy of Sciences of the United States of America* 103 (44), S. 16376–16381. DOI: 10.1073/pnas.0607941103.

Haberle, V.; **Stark**, A. (2018): Eukaryotic core promoters and the functional basis of transcription initiation. In: *Nature reviews. Molecular cell biology* 19 (10), S. 621–637. DOI: 10.1038/s41580-018-0028-8.

Hackmann, A.; **Wu**, H.; **Schneider**, U.-M.; **Meyer**, K.; **Jung**, K.; **Krebber**, H. (2014): Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. In: *Nature communications* 5, S. 3123. DOI: 10.1038/ncomms4123.

Haimovich, G.; **Choder**, M.; **Singer**, R. H.; **Trcek**, T. (2013): The fate of the messenger is pre-determined: a new model for regulation of gene expression. In: *Biochimica et biophysica acta* 1829 (6-7), S. 643–653. DOI: 10.1016/j.bbagrm.2013.01.004.

Hampsey, M. (1998): Molecular genetics of the RNA polymerase II general transcriptional machinery. In: *Microbiology and molecular biology reviews : MMBR* 62 (2), S. 465–503. DOI: 10.1128/mmbr.62.2.465-503.1998.

Harlen, K. M.; **Trotta**, K. L.; **Smith**, E. E.; **Mosaheb**, M. M.; **Fuchs**, S. M.; **Churchman**, L. S. (2016): Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. In: *Cell reports* 15 (10), S. 2147–2158. DOI: 10.1016/j.celrep.2016.05.010.

Hashii, Y.; **Kim**, J. Y.; **Sawada**, A.; **Tokimasa**, S.; **Hiroyuki**, F.; **Ohta**, H. et al. (2004): A novel partner gene CIP29 containing a SAP domain with MLL identified in infantile myelomonocytic leukemia. In: *Leukemia* 18 (9), S. 1546–1548. DOI: 10.1038/sj.leu.2403450.

Hautbergue, G. M.; **Hung**, M.-L.; **Walsh**, M. J.; **Snijders**, A. P. L.; **Chang**, C.-T.; **Jones**, R. et al. (2009): UIF, a New mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. In: *Current biology: CB* 19 (22), S. 1918–1924. DOI: 10.1016/j.cub.2009.09.041.

Heath, C. G.; **Viphakone**, N.; **Wilson**, S. A. (2016): The role of TREX in gene expression and disease. In: *The Biochemical journal* 473 (19), S. 2911–2935. DOI: 10.1042/BCJ20160010.

Helfer, S.; **Schott**, J.; **Stoecklin**, G.; **Förstemann**, K. (2012): AU-rich element-mediated mRNA decay can occur independently of the miRNA machinery in mouse embryonic fibroblasts and Drosophila S2-cells. In: *PloS one* 7 (1), e28907. DOI: 10.1371/journal.pone.0028907.

Henriques, J. A.; **Vicente**, E. J.; **Da Leandro Silva**, K. V.; Schenberg, A. C. (1989): PSO4: a novel gene involved in error-prone repair in Saccharomyces cerevisiae. In: *Mutation research* 218 (2), S. 111–124. DOI: 10.1016/0921-8777(89)90017-7.

Herzel, L.; **Ottoz**, D. S. M.; **Alpert**, T.; **Neugebauer**, K. M. (2017): Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. In: *Nature reviews. Molecular cell biology* 18 (10), S. 637–650. DOI: 10.1038/nrm.2017.63.

Hilleren, P. J.; **Parker**, R. (2003): Cytoplasmic Degradation of Splice-Defective Pre-mRNAs and Intermediates. In: *Molecular Cell* 12 (6), S. 1453–1465. DOI: 10.1016/S1097-2765(03)00488-X.

Hocine, S.; **Singer**, R. H.; **Grünwald**, D. (2010): RNA processing and export. In: *Cold Spring Harbor perspectives in biology* 2 (12), a000752. DOI: 10.1101/cshperspect.a000752.

Hogg, R.; **McGrail**, J. C.; **O'Keefe**, R. T. (2010): The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. In: *Biochemical Society transactions* 38 (4), S. 1110–1115. DOI: 10.1042/BST0381110.

Hosoda, N.; **Kim**, Y. K.; **Lejeune**, F.; **Maquat**, L. E. (2005): CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. In: *Nature structural & molecular biology* 12 (10), S. 893–901. DOI: 10.1038/nsmb995.

Huertas, P.; **Aguilera**, A. (2003): Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. In: *Molecular cell* 12 (3), S. 711–721. DOI: 10.1016/j.molcel. 2003.08.010.

Hui, M. P.; **Foley**, P. L.; **Belasco**, J. G. (2014): Messenger RNA degradation in bacterial cells. In: *Annual review of genetics* 48, S. 537–559. DOI: 10.1146/annurev-genet-120213-092340.

Hurt, E.; **Luo**, M.-J.; **Röther**, S.; **Reed**, R.; **Strässer**, K. (2004): Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. In: *Proceedings of the National Academy of Sciences of the United States of America* 101 (7), S. 1858–1862. DOI: 10.1073/ pnas.0308663100.

Iglesias, N.; **Tutucci**, E.; **Gwizdek**, C.; **Vinciguerra**, P.; **Dach**, E. von; **Corbett**, A. H. et al. (2010): Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. In: *Genes & development* 24 (17), S. 1927–1938. DOI: 10.1101/gad.583310.

Izaurralde, E.; **Lewis**, J.; **McGuigan**, C.; **Jankowska**, M.; **Darzynkiewicz**, E.; **Mattaj**, I. W. (1994): A nuclear cap binding protein complex involved in pre-mRNA splicing. In: *Cell* 78 (4), S. 657–668. DOI: 10.1016/0092-8674(94)90530-4.

Jacobsen, J. O. B.; **Allen**, M. D.; **Freund**, S. M. V.; **Bycroft**, M. (2016): High-resolution NMR structures of the domains of Saccharomyces cerevisiae Tho1. In: *Acta crystallographica. Section F, Structural biology communications* 72 (Pt 6), S. 500–506. DOI: 10.1107/S2053230X16007597.

Jaenicke, R. (1991): Protein stability and molecular adaptation to extreme conditions. In: *European journal of biochemistry* 202 (3), S. 715–728. DOI: 10.1111/j.1432-1033.1991.tb16426.x.

Jaenicke, R.; **Böhm**, G. (1998): The stability of proteins in extreme environments. In: *Current opinion in structural biology* 8 (6), S. 738–748. DOI: 10.1016/s0959-440x(98)80094-8.

Jensen, T. H.; **Boulay**, J.; **Rosbash**, M.; **Libri**, D. (2001): The DECD box putative ATPase Sub2p is an early mRNA export factor. In: *Current biology : CB* 11 (21), S. 1711–1715. DOI: 10.1016/s0960-9822(01)00529-2.

Jeong, S. (2017): SR Proteins: Binders, Regulators, and Connectors of RNA. In: *Molecules and cells* 40 (1), S. 1–9. DOI: 10.14348/molcells.2017.2319.

Jimeno, S.; **Rondón**, A. G.; **Luna**, R.; **Aguilera**, A. (2002): The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. In: *The EMBO journal* 21 (13), S. 3526–3535. DOI: 10.1093/emboj/cdf335.

Jimeno, S.; **Luna**, R.; **García-Rubio**, M.; **Aguilera**, A. (2006): Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. In: *Molecular and cellular biology* 26 (12), S. 4387–4398. DOI: 10.1128/MCB.00234-06.

Johnson, S. A.; **Kim**, H.; **Erickson**, B.; **Bentley**, D. L. (2011): The export factor Yra1 modulates mRNA 3' end processing. In: *Nature structural & molecular biology* 18 (10), S. 1164–1171. DOI: 10.1038/nsmb.2126.

Johnson, S. A.; **Cubberley**, G.; **Bentley**, D. L. (2009): Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. In: *Molecular cell* 33 (2), S. 215–226. DOI: 10.1016/j.molcel.2008.12.007.

Jovanovic-Talisman, T.; **Zilman**, A. (2017): Protein Transport by the Nuclear Pore Complex: Simple Biophysics of a Complex Biomachine. In: *Biophysical journal* 113 (1), S. 6–14. DOI: 10.1016/j.bpj.2017.05.024.

Jumper, J.; **Evans**, R.; **Pritzel**, A.; **Green**, T.; **Figurnov**, M.; **Ronneberger**, O. et al. (2021): Highly accurate protein structure prediction with AlphaFold. In: *Nature* 596 (7873), S. 583–589. DOI: 10.1038/s41586-021-03819-2.

Kashyap, A. K.; **Schieltz**, D.; **Yates**, J.; **Kellogg**, D. R. (2005): Biochemical and genetic characterization of Yra1p in budding yeast. In: *Yeast (Chichester, England)* 22 (1), S. 43–56. DOI: 10.1002/yea.1185.

Katahira, J. (2012): mRNA export and the TREX complex. In: *Biochimica et biophysica acta* 1819 (6), S. 507–513. DOI: 10.1016/j.bbagrm.2011.12.001.

Kellner, N.; **Schwarz**, J.; **Sturm**, M.; **Fernandez-Martinez**, J.; **Griesel**, S.; **Zhang**, W. et al. (2016): Developing genetic tools to exploit Chaetomium thermophilum for biochemical analyses of eukaryotic macromolecular assemblies. In: *Scientific reports* 6, S. 20937. DOI: 10.1038/srep20937.

Kendirgi, F.; **Rexer**, D. J.; **Alcázar-Román**, A. R.; **Onishko**, H. M.; **Wente**, S. R. (2005): Interaction between the shuttling mRNA export factor Gle1 and the nucleoporin hCG1: a conserved mechanism in the export of Hsp70 mRNA. In: *Molecular biology of the cell* 16 (9), S. 4304–4315. DOI: 10.1091/mbc.E04-11-0998.

Kern, C.; **Radon**, C.; **Wende**, W.; **Leitner**, A.; **Sträßer**, K. (2023): Cross-linking mass spectrometric analysis of the endogenous TREX complex from *S. cerevisiae*. In: *RNA (New York, N.Y.)*. DOI: 10.1261/rna.079758.123.

Kervestin, S.; **Jacobson**, A. (2012): NMD: a multifaceted response to premature translational termination. In: *Nature reviews. Molecular cell biology* 13 (11), S. 700–712. DOI: 10.1038/nrm3454.

Khong, A.; **Parker**, R. (2020): The landscape of eukaryotic mRNPs. In: *RNA (New York, N.Y.)* 26 (3), S. 229–239. DOI: 10.1261/rna.073601.119.

Kistler, A. L.; **Guthrie**, C. (2001): Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. In: *Genes & development* 15 (1), S. 42–49. DOI: 10.1101/gad.851301.

Köhler, A.; **Hurt**, E. (2007): Exporting RNA from the nucleus to the cytoplasm. In: *Nature reviews. Molecular cell biology* 8 (10), S. 761–773. DOI: 10.1038/nrm2255.

Köhler, A.; **Pascual-García**, P.; **Llopis**, A.; **Zapater**, M.; **Posas**, F.; **Hurt**, E.; **Rodríguez-Navarro**, S. (2006): The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. In: *Molecular biology of the cell* 17 (10), S. 4228–4236. DOI: 10.1091/mbc.E06-02-0098.

Konarska, M. M.; **Grabowski**, P. J.; **Padgett**, R. A.; **Sharp**, P. A. (1985): Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. In: *Nature* 313 (6003), S. 552–557. DOI: 10.1038/313552a0.

Koukos, P. I.; **Bonvin**, A. M. J. J. (2020): Integrative Modelling of Biomolecular Complexes. In: *Journal of molecular biology* 432 (9), S. 2861–2881. DOI: 10.1016/j.jmb.2019.11.009.

Kress, T. L.; **Krogan**, N. J.; **Guthrie**, C. (2008): A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. In: *Molecular Cell* 32 (5), S. 727–734. DOI: 10.1016/j.molcel.2008.11.013.

Kumar, A.; **Clerici**, M.; **Muckenfuss**, L. M.; **Passmore**, L. A.; **Jinek**, M. (2019): Mechanistic insights into mRNA 3'-end processing. In: *Current opinion in structural biology* 59, S. 143–150. DOI: 10.1016/j.sbi.2019.08.001.

Kumar, R.; **Gardner**, A.; **Homan**, C. C.; **Douglas**, E.; **Mefford**, H.; **Wieczorek**, D. et al. (2018): Severe neurocognitive and growth disorders due to variation in THOC2, an essential component of nuclear mRNA export machinery. In: *Human mutation* 39 (8), S. 1126–1138. DOI: 10.1002/humu.23557.

Kwak, H.; **Lis**, J. T. (2013): Control of transcriptional elongation. In: *Annual review of genetics* 47, S. 483–508. DOI: 10.1146/annurev-genet-110711-155440.

La Cruz, J. de; **Kressler**, D.; **Linder**, P. (1999): Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. In: *Trends in biochemical sciences* 24 (5), S. 192–198. DOI: 10.1016/S0968-0004(99)01376-6.

Leaw, C. L.; **Ren**, E. C.; **Choong**, M. L. (2004): Hcc-1 is a novel component of the nuclear matrix with growth inhibitory function. In: *Cellular and molecular life sciences: CMLS* 61 (17), S. 2264–2273. DOI: 10.1007/s00018-004-4205-x.

Lee, M. S.; **Henry**, M.; **Silver**, P. A. (1996): A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. In: *Genes & development* 10 (10), S. 1233–1246. DOI: 10.1101/gad.10.10.1233.

Le Hir, H.; **Saulière**, J.; **Wang**, Z. (2016): The exon junction complex as a node of post-transcriptional networks. In: *Nature reviews. Molecular cell biology* 17 (1), S. 41–54. DOI: 10.1038/nrm.2015.7.

Leitner, A.; **Joachimiak**, L. A.; **Unverdorben**, P.; **Walzthoeni**, T.; **Frydman**, J.; **Förster**, F.; **Aebersold**, R. (2014): Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (26), S. 9455–9460. DOI: 10.1073/pnas.1320298111.

Libri, D.; **Graziani**, N.; **Saguez**, C.; **Boulay**, J. (2001): Multiple roles for the yeast SUB2/yUAP56 gene in splicing. In: *Genes & development* 15 (1), S. 36–41. DOI: 10.1101/gad.852101.

Logan, J.; **Falck-Pedersen**, E.; **Darnell**, J. E.; **Shenk**, T. (1987): A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. In: *Proceedings of the National Academy of Sciences of the United States of America* 84 (23), S. 8306–8310. DOI: 10.1073/pnas.84.23.8306.

Lopez Martinez, D.; **Svejstrup**, J. Q. (2024): Mechanisms of RNA Polymerase II Termination at the 3'-End of Genes. In: *Journal of molecular biology*, S. 168735. DOI: 10.1016/j.jmb.2024.168735.

Löscher, M.; **Fortschegger**, K.; **Ritter**, G.; **Wostry**, M.; **Voglauer**, R.; **Schmid**, J. A. et al. (2005): Interaction of U-box E3 ligase SNEV with PSMB4, the beta7 subunit of the 20 S proteasome. In: *The Biochemical journal* 388 (Pt 2), S. 593–603. DOI: 10.1042/BJ20041517.

Lund, M. K.; **Guthrie**, C. (2005): The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. In: *Molecular cell* 20 (4), S. 645–651. DOI: 10.1016/j.molcel.2005.10.005.

Luo, M. L.; **Zhou**, Z.; **Magni**, K.; **Christoforides**, C.; **Rappsilber**, J.; **Mann**, M.; **Reed**, R. (2001): Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. In: *Nature* 413 (6856), S. 644–647. DOI: 10.1038/35098106.

Luo, W.; **Johnson**, A. W.; **Bentley**, D. L. (2006): The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. In: *Genes & development* 20 (8), S. 954–965. DOI: 10.1101/gad.1409106.

Lykke-Andersen, J.; **Bennett**, E. J. (2014): Protecting the proteome: Eukaryotic cotranslational quality control pathways. In: *The Journal of cell biology* 204 (4), S. 467–476. DOI: 10.1083/jcb.201311103.

Macara, I. G. (2001): Transport into and out of the nucleus. In: *Microbiology and molecular biology reviews: MMBR* 65 (4), 570-94, table of contents. DOI: 10.1128/MMBR.65.4.570-594.2001.

Magistris, P. de (2021): The Great Escape: mRNA Export through the Nuclear Pore Complex. In: *IJMS* 22 (21), S. 11767. DOI: 10.3390/ijms222111767.

Mahajan, K. N.; **Mitchell**, B. S. (2003): Role of human Pso4 in mammalian DNA repair and association with terminal deoxynucleotidyl transferase. In: *Proceedings of the National Academy of Sciences of the United States of America* 100 (19), S. 10746–10751. DOI: 10.1073/pnas.1631060100.

Mandel, C. R.; **Bai**, Y.; **Tong**, L. (2008): Protein factors in pre-mRNA 3'-end processing. In: *Cellular and molecular life sciences: CMLS* 65 (7-8), S. 1099–1122. DOI: 10.1007/s00018-007-7474-3.

Mars, J.-C.; **Ghram**, M.; **Culjkovic-Kraljacic**, B.; **Borden**, K. L. B. (2021): The Cap-Binding Complex CBC and the Eukaryotic Translation Factor eIF4E: Co-Conspirators in Cap-Dependent RNA Maturation and Translation. In: *Cancers* 13 (24). DOI: 10.3390/cancers13246185.

Martínez-Lumbreras, S.; **Taverniti**, V.; **Zorrilla**, S.; **Séraphin**, B.; **Pérez-Cañadillas**, J. M. (2016): Gbp2 interacts with THO/TREX through a novel type of RRM domain. In: *Nucleic acids research* 44 (1), S. 437–448. DOI: 10.1093/nar/gkv1303.

Masuda, S.; **Das**, R.; **Cheng**, H.; **Hurt**, E.; **Dorman**, N.; **Reed**, R. (2005): Recruitment of the human TREX complex to mRNA during splicing. In: *Genes & development* 19 (13), S. 1512–1517. DOI: 10.1101/gad.1302205.

Mauer, J.; **Luo**, X.; **Blanjoie**, A.; **Jiao**, X.; **Grozhik**, A. V.; **Patil**, D. P. et al. (2017): Reversible methylation of m6Am in the 5' cap controls mRNA stability. In: *Nature* 541 (7637), S. 371–375. DOI: 10.1038/nature21022.

Mauger, D. M.; **Cabral**, B. J.; **Presnyak**, V.; **Su**, S. V.; **Reid**, D. W.; **Goodman**, B. et al. (2019): mRNA structure regulates protein expression through changes in functional half-life. In: *Proceedings of the National Academy of Sciences of the United States of America* 116 (48), S. 24075–24083. DOI: 10.1073/pnas.1908052116.

McWilliam, H.; **Li**, W.; **Uludag**, M.; **Squizzato**, S.; **Park**, Y. M.; **Buso**, N. et al. (2013): Analysis Tool Web Services from the EMBL-EBI. In: *Nucleic acids research* 41 (Web Server issue), W597-600. DOI: 10.1093/nar/gkt376.

Meinel, D. M.; **Burkert-Kautzsch**, C.; **Kieser**, A.; O'**Duibhir**, E.; **Siebert**, M.; **Mayer**, A. et al. (2013): Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. In: *PLoS genetics* 9 (11), e1003914. DOI: 10.1371/journal.pgen.1003914.

Meinel, D. M.; **Sträßer**, K. (2015): Co-transcriptional mRNP formation is coordinated within a molecular mRNP packaging station in S. cerevisiae. In: *BioEssays: news and reviews in molecular, cellular and developmental biology* 37 (6), S. 666–677. DOI: 10.1002/bies.201400220.

Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. (2014): Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. In: *Protein science: a publication of the Protein Society* 23 (6), S. 747–759. DOI: 10.1002/pro.2458.

Meruelo, A. D.; **Han**, S. K.; **Kim**, S.; **Bowie**, J. U. (2012): Structural differences between thermophilic and mesophilic membrane proteins. In: *Protein science : a publication of the Protein Society* 21 (11), S. 1746–1753. DOI: 10.1002/pro.2157.

Minocha, R.; **Popova**, V.; **Kopytova**, D.; **Misiak**, D.; **Hüttelmaier**, S.; **Georgieva**, S.; **Sträßer**, K. (2018): Mud2 functions in transcription by recruiting the Prp19 and TREX complexes to transcribed genes. In: *Nucleic acids research* 46 (18), S. 9749–9763. DOI: 10.1093/nar/gky640.

Miosga, M. B. (2022): Function of the RNA helicase Sub2 from *Saccharomyces cerevisiae* and its regulation by Tho1. Dissertation, *Universitätsbibliothek Gießen*. DOI: http://dx.doi.org/10.22029/jlupub-2972.

Mitchell, S. F.; **Parker**, R. (2014): Principles and properties of eukaryotic mRNPs. In: *Molecular cell* 54 (4), S. 547–558. DOI: 10.1016/j.molcel.2014.04.033.

Mitchell, P.; **Petfalski**, E.; **Shevchenko**, A.; **Mann**, M.; **Tollervey**, D. (1997): The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'--5' exoribonucleases. In: *Cell* 91 (4), S. 457–466. DOI: 10.1016/S0092-8674(00)80432-8.

Moeller, H. von; **Basquin**, C.; **Conti**, E. (2009): The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. In: *Nature structural & molecular biology* 16 (3), S. 247–254. DOI: 10.1038/nsmb.1561.

Mohammadi, A.; **Tschanz**, A.; **Leitner**, A. (2021): Expanding the Cross-Link Coverage of a Carboxyl-Group Specific Chemical Cross-Linking Strategy for Structural Proteomics Applications. In: *Analytical chemistry* 93 (4), S. 1944–1950. DOI: 10.1021/acs.analchem.0c03926.

Müller-McNicoll, M.; **Neugebauer**, K. M. (2013): How cells get the message: dynamic assembly and function of mRNA-protein complexes. In: *Nature reviews. Genetics* 14 (4), S. 275–287. DOI: 10.1038/nrg3434.

Müller-McNicoll, M.; **Neugebauer**, K. M. (2014): Good cap/bad cap: how the cap-binding complex determines RNA fate. In: *Nature structural & molecular biology* 21 (1), S. 9–12. DOI: 10.1038/nsmb.2751.

Musselman, C. A.; **Kutateladze**, T. G. (2021): Characterization of functional disordered regions within chromatin-associated proteins. In: *iScience* 24 (2), S. 102070. DOI: 10.1016/j.isci.2021.102070.

Napetschnig, J.; **Blobel**, G.; **Hoelz**, A. (2007): Crystal structure of the N-terminal domain of the human protooncogene Nup214/CAN. In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (6), S. 1783–1788. DOI: 10.1073/pnas.0610828104.

Nilsson, G.; **Belasco**, J. G.; **Cohen**, S. N.; **Gabain**, A. von (1984): Growth-rate dependent regulation of mRNA stability in Escherichia coli. In: *Nature* 312 (5989), S. 75–77. DOI: 10.1038/312075a0.

O'Brien, J.; **Hayder**, H.; **Zayed**, Y.; **Peng**, C. (2018): Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. In: *Frontiers in endocrinology* 9, S. 402. DOI: 10.3389/fendo.2018.00402.

O'Donnell, M.; **Langston**, L.; **Stillman**, B. (2013): Principles and concepts of DNA replication in bacteria, archaea, and eukarya. In: *Cold Spring Harbor perspectives in biology* 5 (7). DOI: 10.1101/cshperspect.a010108.

Pacheco-Fiallos, B.; **Vorländer**, M. K.; **Riabov-Bassat**, D.; **Fin**, L.; **O'Reilly**, F. J.; **Ayala**, F. I. et al. (2023): mRNA recognition and packaging by the human transcription-export complex. In: *Nature* 616 (7958), S. 828–835. DOI: 10.1038/s41586-023-05904-0.

Passmore, L. A.; **Coller**, J. (2022): Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. In: *Nature reviews. Molecular cell biology* 23 (2), S. 93–106. DOI: 10.1038/s41580-021-00417-y.

Peña, A.; **Gewartowski**, K.; **Mroczek**, S.; **Cuéllar**, J.; **Szykowska**, A.; **Prokop**, A. et al. (2012): Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. In: *The EMBO journal* 31 (6), S. 1605–1616. DOI: 10.1038/emboj.2012.10.

Peyro, M.; **Soheilypour**, M.; **Nibber**, V. S.; **Dickson**, A. M.; **Mofrad**, M. R. K. (2021): FG nucleoporins feature unique patterns that distinguish them from other IDPs. In: *Biophysical journal* 120 (16), S. 3382–3391. DOI: 10.1016/j.bpj.2021.06.031.

Piruat, J. I.; **Aguilera**, A. (1998): A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. In: *The EMBO journal* 17 (16), S. 4859–4872. DOI: 10.1093/emboj/17.16.4859.

Portman, D. S.; **O'Connor**, J. P.; **Dreyfuss**, G. (1997): YRA1, an essential Saccharomyces cerevisiae gene, encodes a novel nuclear protein with RNA annealing activity. In: *RNA* (New York, N.Y.) 3 (5), S. 527–537.

Pühringer, T.; **Hohmann**, U.; **Fin**, L.; **Pacheco-Fiallos**, B.; **Schellhaas**, U.; **Brennecke**, J.; **Plaschka**, C. (2020): Structure of the human core transcription-export complex reveals a hub for multivalent interactions. In: *eLife* 9. DOI: 10.7554/eLife.61503.

Ramanathan, A.; Robb**,** G. B.; **Chan**, S.-H. (2016): mRNA capping: biological functions and applications. In: *Nucleic acids research* 44 (16), S. 7511–7526. DOI: 10.1093/nar/gkw551.

Rambout, X.; **Maquat**, L. E. (2020): The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. In: *Genes & development* 34 (17-18), S. 1113–1127. DOI: 10.1101/gad.339986.120.

Rayala, H. J.; **Kendirgi**, F.; **Barry**, D. M.; **Majerus**, P. W.; **Wente**, S. R. (2004): The mRNA export factor human Gle1 interacts with the nuclear pore complex protein Nup155. In: *Molecular & cellular proteomics: MCP* 3 (2), S. 145–155. DOI: 10.1074/mcp.M300106-MCP200.

Reed, R.; **Cheng**, H. (2005): TREX, SR proteins and export of mRNA. In: *Current opinion in cell biology* 17 (3), S. 269–273. DOI: 10.1016/j.ceb.2005.04.011.

Ren, Y.; **Schmiege**, P.; **Blobel**, G. (2017): Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. In: *eLife* 6. DOI: 10.7554/eLife.20070.

Rodríguez-Molina, J. B.; **West**, S.; **Passmore**, L. A. (2023): Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. In: *Molecular cell* 83 (3), S. 404–415. DOI: 10.1016/j.molcel.2022.12.021.

Rodríguez-Navarro, S.; **Fischer**, T.; **Luo**, M.-J.; **Antúnez**, O.; **Brettschneider**, S.; **Lechner**, J. et al. (2004): Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. In: *Cell* 116 (1), S. 75–86. DOI: 10.1016/s0092-8674(03)01025-0.

Romier, C.; **Ben Jelloul**, M.; **Albeck**, S.; **Buchwald**, G.; **Busso**, D.; **Celie**, P. H. N. et al. (2006): Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. In: *Acta crystallographica. Section D, Biological crystallography* 62 (Pt 10), S. 1232–1242. DOI: 10.1107/S0907444906031003.

Rondón, A. G.; **Jimeno**, S.; **Aguilera**, A. (2010): The interface between transcription and mRNP export: from THO to THSC/TREX-2. In: *Biochimica et biophysica acta* 1799 (8), S. 533–538. DOI: 10.1016/j.bbagrm.2010.06.002.

Rondón, A. G.; **Jimeno**, S.; **García-Rubio**, M.; **Aguilera**, A. (2003): Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. In: *The Journal of biological chemistry* 278 (40), S. 39037–39043. DOI: 10.1074/jbc.M305718200.

Rout, M. P.; **Aitchison**, J. D.; **Suprapto**, A.; **Hjertaas**, K.; **Zhao**, Y.; **Chait**, B. T. (2000): The yeast nuclear pore complex: composition, architecture, and transport mechanism. In: *The Journal of cell biology* 148 (4), S. 635–651. DOI: 10.1083/jcb.148.4.635.

Rout, M. P.; **Sali**, A. (2019): Principles for Integrative Structural Biology Studies. In: *Cell* 177 (6), S. 1384–1403. DOI: 10.1016/j.cell.2019.05.016.

Ruskin, B.; **Krainer**, A. R.; **Maniatis**, T.; **Green**, M. R. (1984): Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. In: *Cell* 38 (1), S. 317–331. DOI: 10.1016/0092-8674(84)90553-1.

Santos-Rosa, H.; **Moreno**, H.; **Simos**, G.; **Segref**, A.; **Fahrenkrog**, B.; **Panté**, N.; **Hurt**, E. (1998): Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores. In: *Molecular and cellular biology* 18 (11), S. 6826–6838. DOI: 10.1128/mcb.18.11.6826.

Saulière, J.; **Murigneux**, V.; **Wang**, Z.; **Marquenet**, E.; **Barbosa**, I.; **Le Tonquèze**, O. et al. (2012): CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. In: *Nature structural & molecular biology* 19 (11), S. 1124–1131. DOI: 10.1038/nsmb.2420.

Schlautmann, L. P.; **Gehring**, N. H. (2020): A Day in the Life of the Exon Junction Complex. In: *Biomolecules* 10 (6). DOI: 10.3390/biom10060866.

Schmitt, C.; **Kobbe**, C. von; **Bachi**, A.; **Panté**, N.; **Rodrigues**, J. P.; **Boscheron**, C. et al. (1999): Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. In: *The EMBO journal* 18 (15), S. 4332–4347. DOI: 10.1093/emboj/18.15.4332.

Schrödinger, L. & **DeLano**, W. (2020): PyMOL. Online verfügbar unter http://www.pymol.org/pymol.

Schuller, S. K.; **Schuller**, J. M.; **Prabu**, J. R.; **Baumgärtner**, M.; **Bonneau**, F.; **Basquin**, J.; **Conti**, E. (2020): Structural insights into the nucleic acid remodeling mechanisms of the yeast THO-Sub2 complex. In: *eLife* 9. DOI: 10.7554/eLife.61467.

Segref, A.; **Sharma**, K.; **Doye**, V.; **Hellwig**, A.; **Huber**, J.; **Lührmann**, R.; **Hurt**, E. (1997): Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. In: *The EMBO journal* 16 (11), S. 3256–3271. DOI: 10.1093/emboj/16.11.3256.

Shatkin, A. J.; **Manley**, J. L. (2000): The ends of the affair: capping and polyadenylation. In: *Nature structural biology* 7 (10), S. 838–842. DOI: 10.1038/79583.

Sherman, F. (1991): Getting started with yeast. In: *Methods in enzymology* 194, S. 3–21. DOI: 10.1016/0076-6879(91)94004-V.

Shi, H.; **Cordin**, O.; **Minder**, C. M.; **Linder**, P.; **Xu**, R.-M. (2004): Crystal structure of the human ATP-dependent splicing and export factor UAP56. In: *Proceedings of the National Academy of Sciences of the United States of America* 101 (51), S. 17628–17633. DOI: 10.1073/pnas.0408172101.

Shi, M.; **Zhang**, H.; **Wu**, X.; **He**, Z.; **Wang**, L.; **Yin**, S. et al. (2017): ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. In: *Nucleic acids research* 45 (16), S. 9640–9653. DOI: 10.1093/nar/gkx597.

Shoemaker, C. J.; **Green**, R. (2012): Translation drives mRNA quality control. In: *Nature structural & molecular biology* 19 (6), S. 594–601. DOI: 10.1038/nsmb.2301.

Sigrist, C. J. A.; **Castro**, E. de; **Cerutti**, L.; **Cuche**, B. A.; **Hulo**, N.; **Bridge**, A. et al. (2013): New and continuing developments at PROSITE. In: *Nucleic acids research* 41 (Database issue), D344-7. DOI: 10.1093/nar/gks1067.

Sihn, C.-R.; **Cho**, S. Y.; **Lee**, J. H.; **Lee**, T. R.; **Kim**, S. H. (2007): Mouse homologue of yeast Prp19 interacts with mouse SUG1, the regulatory subunit of 26S proteasome. In: *Biochemical and biophysical research communications* 356 (1), S. 175–180. DOI: 10.1016/j.bbrc.2007.02.134.

Singh, G.; **Kucukural**, A.; **Cenik**, C.; **Leszyk**, J. D.; **Shaffer**, S. A.; **Weng**, Z.; **Moore**, M. J. (2012): The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. In: *Cell* 151 (4), S. 750–764. DOI: 10.1016/j.cell.2012.10.007.

Sikorski, R. S.; **Hieter**, P. (1989): A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. In: *Genetics* 122 (1), S. 19–27. DOI: 10.1093/genetics/122.1.19.

Singh, G.; **Pratt**, G.; **Yeo**, G. W.; **Moore**, M. J. (2015): The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. In: *Annual review of biochemistry* 84, S. 325–354. DOI: 10.1146/annurev-biochem-080111-092106.

Singh, P.; **Saha**, U.; **Paira**, S.; **Das**, B. (2018): Nuclear mRNA Surveillance Mechanisms: Function and Links to Human Disease. In: *Journal of molecular biology* 430 (14), S. 1993–2013. DOI: 10.1016/j.jmb.2018.05.009.

Soucek, S.; **Corbett**, A. H.; **Fasken**, M. B. (2012): The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. In: *Biochimica et biophysica acta 1819* (6), S. 546–554. DOI: 10.1016/j.bbagrm.2012.03.006.

Stefan, A.; **Ceccarelli**, A.; **Conte**, E.; **Montón Silva**, A.; **Hochkoeppler**, A. (2015): The multifaceted benefits of protein co-expression in Escherichia coli. In: *Journal of visualized experiments : JoVE* (96). DOI: 10.3791/52431.

Stewart, M. (2019): Polyadenylation and nuclear export of mRNAs. In: *The Journal of biological chemistry* 294 (9), S. 2977–2987. DOI: 10.1074/jbc.REV118.005594.

Strahm, Y.; **Fahrenkrog**, B.; **Zenklusen**, D.; **Rychner**, E.; **Kantor**, J.; **Rosbach**, M.; **Stutz**, F. (1999): The RNA export factor Gle1p is located on the cytoplasmic fibrils of the NPC and physically interacts with the FG-nucleoporin Rip1p, the DEAD-box protein Rat8p/Dbp5p and a new protein Ymr 255p. In: *The EMBO journal* 18 (20), S. 5761–5777. DOI: 10.1093/emboj/18.20.5761.

Strässer, 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. In: *The Journal of cell biology* 150 (4), S. 695–706. DOI: 10.1083/jcb.150.4.695.

Strässer, K.; **Hurt**, E. (2000): Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. In: *The EMBO journal* 19 (3), S. 410–420. DOI: 10.1093/emboj/19.3.410.

Strässer, K.; **Hurt**, E. (2001): Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. In: *Nature* 413 (6856), S. 648–652. DOI: 10.1038/35098113.

Strässer, K.; **Masuda**, S.; **Mason**, P.; **Pfannstiel**, J.; **Oppizzi**, M.; **Rodriguez-Navarro**, S. et al. (2002): TREX is a conserved complex coupling transcription with messenger RNA export. In: *Nature* 417 (6886), S. 304–308. DOI: 10.1038/nature746.

Strobl, L. J.; **Eick**, D. (1992): Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. In: *The EMBO journal* 11 (9), S. 3307–3314. DOI: 10.1002/j.1460-2075.1992.tb05409.x.

Stutz, F.; **Bachi**, A.; **Doerks**, T.; **Braun**, I. C.; **Séraphin**, B.; **Wilm**, M. et al. (2000): REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. In: *RNA (New York, N.Y.)* 6 (4), S. 638–650. DOI: 10.1017/s1355838200000078.

Sugiura, T.; **Sakurai**, K.; **Nagano**, Y. (2007): Intracellular characterization of DDX39, a novel growth-associated RNA helicase. In: *Experimental cell research* 313 (4), S. 782–790. DOI: 10.1016/j.yexcr.2006.11.014.

Sun, H.; **Li**, K.; **Liu**, C.; **Yi**, C. (2023): Regulation and functions of non-m6A mRNA modifications. In: *Nature reviews. Molecular cell biology* 24 (10), S. 714–731. DOI: 10.1038/s41580-023-00622-x.

Szilágyi, A.; **Závodszky**, P. (2000): Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. In: *Structure (London, England : 1993)* 8 (5), S. 493–504. DOI: 10.1016/S0969-2126(00)00133-7.

Taschuk, F.; **Cherry**, S. (2020): DEAD-Box Helicases: Sensors, Regulators, and Effectors for Antiviral Defense. In: *Viruses* 12 (2). DOI: 10.3390/v12020181.

Taylor, T. J.; **Vaisman**, I. I. (2010): Discrimination of thermophilic and mesophilic proteins. In: *BMC structural biology* 10 Suppl 1 (Suppl 1), S5. DOI: 10.1186/1472-6807-10-S1-S5.

Terry, L. J.; **Wente**, S. R. (2007): Nuclear mRNA export requires specific FG nucleoporins for translocation through the nuclear pore complex. In: *The Journal of cell biology* 178 (7), S. 1121–1132. DOI: 10.1083/jcb.200704174.

Thomas, B. J.; **Rothstein**, R. (1989): Elevated recombination rates in transcriptionally active DNA. In: *Cell* 56 (4), S. 619–630. DOI: 10.1016/0092-8674(89)90584-9.

Topisirovic, I.; **Svitkin**, Y. V.; **Sonenberg**, N.; **Shatkin**, A. J. (2011): Cap and cap-binding proteins in the control of gene expression. In: *Wiley interdisciplinary reviews. RNA* 2 (2), S. 277–298. DOI: 10.1002/wrna.52.

Tran, E. J.; **Zhou**, Y.; **Corbett**, A. H.; **Wente**, S. R. (2007): The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. In: *Molecular cell* 28 (5), S. 850–859. DOI: 10.1016/j.molcel.2007.09.019.

Tsai, C.-J.; **Polverino de Laureto**, P.; **Fontana**, A.; **Nussinov**, R. (2002): Comparison of protein fragments identified by limited proteolysis and by computational cutting of proteins. In: *Protein science: a publication of the Protein Society* 11 (7), S. 1753–1770. DOI: 10.1110/ps.4100102.

Tseng, C.-K.; **Cheng**, S.-C. (2013): The spliceosome catalyzes debranching in competition with reverse of the first chemical reaction. In: *RNA (New York, N.Y.)* 19 (7), S. 971–981. DOI: 10.1261/rna.038638.113.

Varadi, M.; **Anyango**, S.; **Deshpande**, M.; **Nair**, S.; **Natassia**, C.; **Yordanova**, G. et al. (2022): AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. In: *Nucleic acids research* 50 (D1), D439-D444. DOI: 10.1093/nar/gkab1061.

Vijayabaskar, M. S.; **Vishveshwara**, S. (2010): Comparative analysis of thermophilic and mesophilic proteins using Protein Energy Networks. In: *BMC bioinformatics* 11 Suppl 1 (Suppl 1), S49. DOI: 10.1186/1471-2105-11-S1-S49.

Wade, J. T.; **Struhl**, K. (2008): The transition from transcriptional initiation to elongation. In: *Current opinion in genetics & development* 18 (2), S. 130–136. DOI: 10.1016/j.gde.2007.12.008.

Wahl, M. C.; **Will**, C. L.; **Lührmann**, R. (2009): The spliceosome: design principles of a dynamic RNP machine. In: *Cell* 136 (4), S. 701–718. DOI: 10.1016/j.cell.2009.02.009.

Waterhouse, A. M.; **Procter**, J. B.; **Martin**, D. M. A.; **Clamp**, M.; **Barton**, G. J. (2009): Jalview Version 2--a multiple sequence alignment editor and analysis workbench. In: *Bioinformatics (Oxford, England)* 25 (9), S. 1189–1191. DOI: 10.1093/bioinformatics/btp033.

Wegener, M.; **Müller-McNicoll**, M. (2019): View from an mRNP: The Roles of SR Proteins in Assembly, Maturation and Turnover. In: *Advances in experimental medicine and biology* 1203, S. 83–112. DOI: 10.1007/978-3-030-31434-7_3.

Wende, W.; **Friedhoff**, P.; **Sträßer**, K. (2019): Mechanism and Regulation of Co-transcriptional mRNP Assembly and Nuclear mRNA Export. In: *Advances in experimental medicine and biology* 1203, S. 1–31. DOI: 10.1007/978-3-030-31434-7_1.

Wente, S. R.; **Rout**, M. P. (2010): The nuclear pore complex and nuclear transport. In: *Cold Spring Harbor perspectives in biology* 2 (10), a000562. DOI: 10.1101/cshperspect.a000562.

Werner-Washburne, M.; **Braun**, E. L.; **Crawford**, M. E.; **Peck**, V. M. (1996): Stationary phase in Saccharomyces cerevisiae. In: *Molecular microbiology* 19 (6), S. 1159–1166. DOI: 10.1111/j.1365-2958.1996.tb02461.x.

West, R. W.; **Milgrom**, E. (2002): DEAD-box RNA helicase Sub2 is required for expression of lacZ fusions in Saccharomyces cerevisiae and is a dosage-dependent suppressor of RLR1 (THO2). In: *Gene* 288 (1-2), S. 19–27. DOI: 10.1016/S0378-1119(02)00482-1.

Whelan, M.; **Pelchat**, M. (2022): Role of RNA Polymerase II Promoter-Proximal Pausing in Viral Transcription. In: *Viruses* 14 (9). DOI: 10.3390/v14092029.

Wilczynska, A.; **Bushell**, M. (2015): The complexity of miRNA-mediated repression. In: *Cell death and differentiation* 22 (1), S. 22–33. DOI: 10.1038/cdd.2014.112.

Will, C. L.; **Lührmann**, R. (2011): Spliceosome structure and function. In: *Cold Spring Harbor perspectives in biology* 3 (7). DOI: 10.1101/cshperspect.a003707.

Wu, S.; **Avila-Sakar**, A.; **Kim**, J.; **Booth**, D. S.; **Greenberg**, C. H.; **Rossi**, A. et al. (2012): Fabs enable single particle cryoEM studies of small proteins. In: *Structure (London, England: 1993)* 20 (4), S. 582–592. DOI: 10.1016/j.str.2012.02.017.

Xie, Y.; **Clarke**, B. P.; **Kim**, Y. J.; **Ivey**, A. L.; **Hill**, P. S.; **Shi**, Y.; **Ren**, Y. (2021): Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2. In: *eLife* 10. DOI: 10.7554/elife.65699.

Xie, Y.; **Gao**, S.; **Zhang**, K.; **Bhat**, P.; **Clarke**, B. P.; **Batten**, K. et al. (2023): Structural basis for high-order complex of SARNP and DDX39B to facilitate mRNP assembly. In: *Cell reports* 42 (8), S. 112988. DOI: 10.1016/j.celrep.2023.112988.

Xu, Y. Z.; **Di Marco**, S.; **Gallouzi**, I.; **Rola-Pleszczynski**, M.; **Radzioch**, D. (2005): RNA-binding protein HuR is required for stabilization of SLC11A1 mRNA and SLC11A1 protein expression. In: *Molecular and cellular biology* 25 (18), S. 8139–8149. DOI: 10.1128/MCB.25.18.8139-8149.2005.

Yamaguchi, Y.; **Inukai**, N.; **Narita**, T.; **Wada**, T.; **Handa**, H. (2002): Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. In: *Molecular and cellular biology* 22 (9), S. 2918–2927. DOI: 10.1128/MCB.22.9.2918-2927.2002.

Yu, M.; **Heidari**, M.; **Mikhaleva**, S.; **Tan**, P. S.; **Mingu**, S.; **Ruan**, H. et al. (2023): Visualizing the disordered nuclear transport machinery in situ. In: *Nature* 617 (7959), S. 162–169. DOI: 10.1038/s41586-023-05990-0.

Zarnack, K.; **Balasubramanian**, S.; **Gantier**, M. P.; **Kunetsky**, V.; **Kracht**, M.; **Schmitz**, M. L.; **Sträßer**, K. (2020): Dynamic mRNP Remodeling in Response to Internal and External Stimuli. In: *Biomolecules* 10 (9). DOI: 10.3390/biom10091310.

Zenklusen, D.; **Vinciguerra**, P.; **Strahm**, Y.; **Stutz**, F. (2001): The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. In: *Molecular and cellular biology* 21 (13), S. 4219–4232. DOI: 10.1128/MCB.21.13.4219-4232.2001.

Zenklusen, D.; **Vinciguerra**, P.; **Wyss**, J.-C.; **Stutz**, F. (2002): Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. In: *Molecular and cellular biology* 22 (23), S. 8241–8253. DOI: 10.1128/MCB.22.23.8241-8253.2002.

Zhang, M.; **Green**, M. R. (2001): Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. In: *Genes & development* 15 (1), S. 30–35. DOI: 10.1101/gad.851701.

Zhao, B. S.; **Roundtree**, I. A.; **He**, C. (2017): Post-transcriptional gene regulation by mRNA modifications. In: *Nature reviews. Molecular cell biology* 18 (1), S. 31–42. DOI: 10.1038/nrm.2016.132.

Zhou, H.-X.; **Pang**, X. (2018): Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. In: *Chemical reviews* 118 (4), S. 1691–1741. DOI: 10.1021/acs.chemrev.7b00305.

Publications

Cross-linking mass spectrometric analysis of the endogenous TREX complex from *S. cerevisiae*.

Kern, Carina; Radon, Christin; Wende, Wolfgang; Leitner, Alexander; Sträßer, Katja

Published 2023 in *RNA (New York, N.Y.)* (DOI: 10.1261/rna.079758.123)

Acknowledgments

Zunächst möchte ich meiner Doktormutter Prof. Dr. Katja Sträßer dafür danken, dass sie mir die Möglichkeit gegeben hat, meine Dissertation am Institut für Biochemie zu schreiben. Ich bin ihr sehr dankbar für ihren Rat, ihre Unterstützung und ihre Anleitung während dieser Studie. Ihre aufschlussreichen Kommentare und Vorschläge waren grundlegend für die Durchführung der Experimente und die Fertigstellung dieser Arbeit.

Mein Dank gilt natürlich auch Apl. Prof Dr. Elena Evguenieva-Hackenberg, die das Zweitgutachten dieser Dissertation übernommen hat. Auch bei den anderen Mitgliedern des Prüfungskomitees möchte ich mich bedanken.

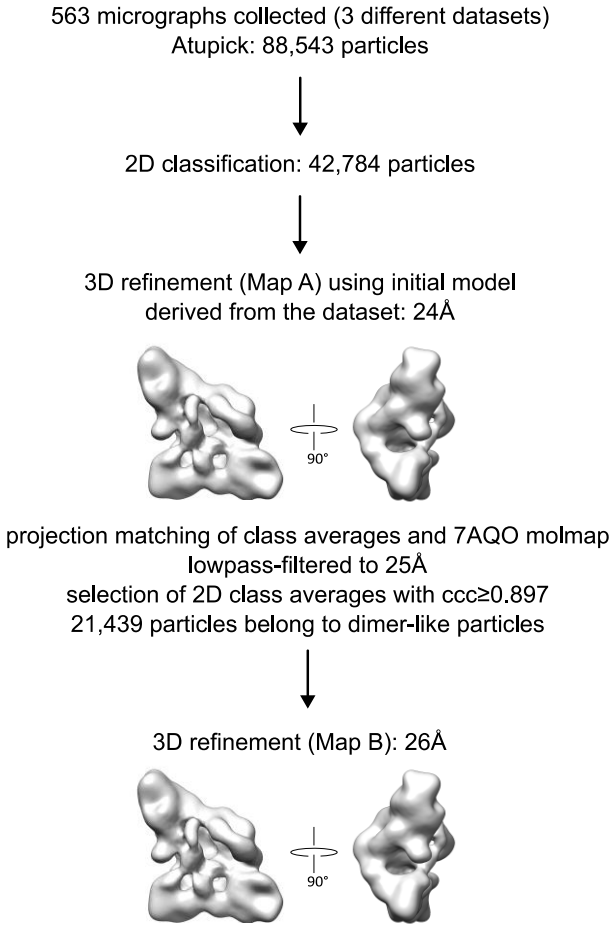
Des Weiteren möchte ich mich bei meinen Kollaborationspartnern für deren Beitrag zu dieser Arbeit bedanken: Alexander Leitner für das Durchführen und Auswerten der XL-MS Versuche, Christin Radon für das Durchführen und Auswerten der Elektronenmikroskopie, Petra Wendler für die hilfreichen Diskussionen über die EM-Bilder, Michaela Stumpf für das Durchführen der Kristallisationsversuche, Karin Fritz-Wolf für das Lösen und Auswerten der *ct*Sub2 Kristallstruktur, Ahim Fahid für das Durchführen der LC-MS Analysen, Tatjana Schusser für das Durchführen der Kryo-EM und Robert Tampé für das konstruktive und sehr hilfreiche Feedback zu den Experimenten mit dem S-Y-T Komplex.

Ganz besonders möchte ich mich auch bei Dr. Wolfgang Wende bedanken. Ich bin sehr dankbar für seine großzügige Unterstützung und Anleitung bei der Durchführung aller Experimente. Auch von seinen konstruktiven Kommentaren habe ich beim Verfassen dieser Arbeit sehr profitiert.

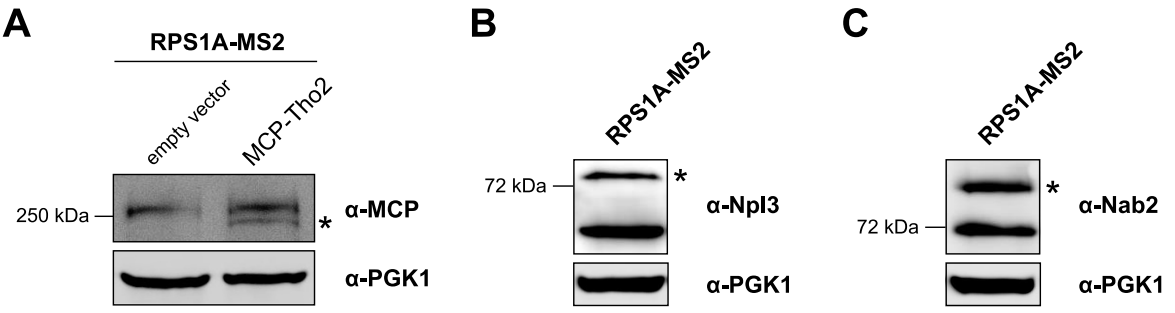
Mein besonderer Dank geht auch an alle übrigen Labormitglieder und studentischen Hilfskräfte. Vielen Dank für eure Hilfe und Unterstützung!

Nicht zuletzt möchte ich mich bei meiner Familie und meinen Freunden bedanken, die mich während der Erstellung dieser Arbeit unterstützt haben.

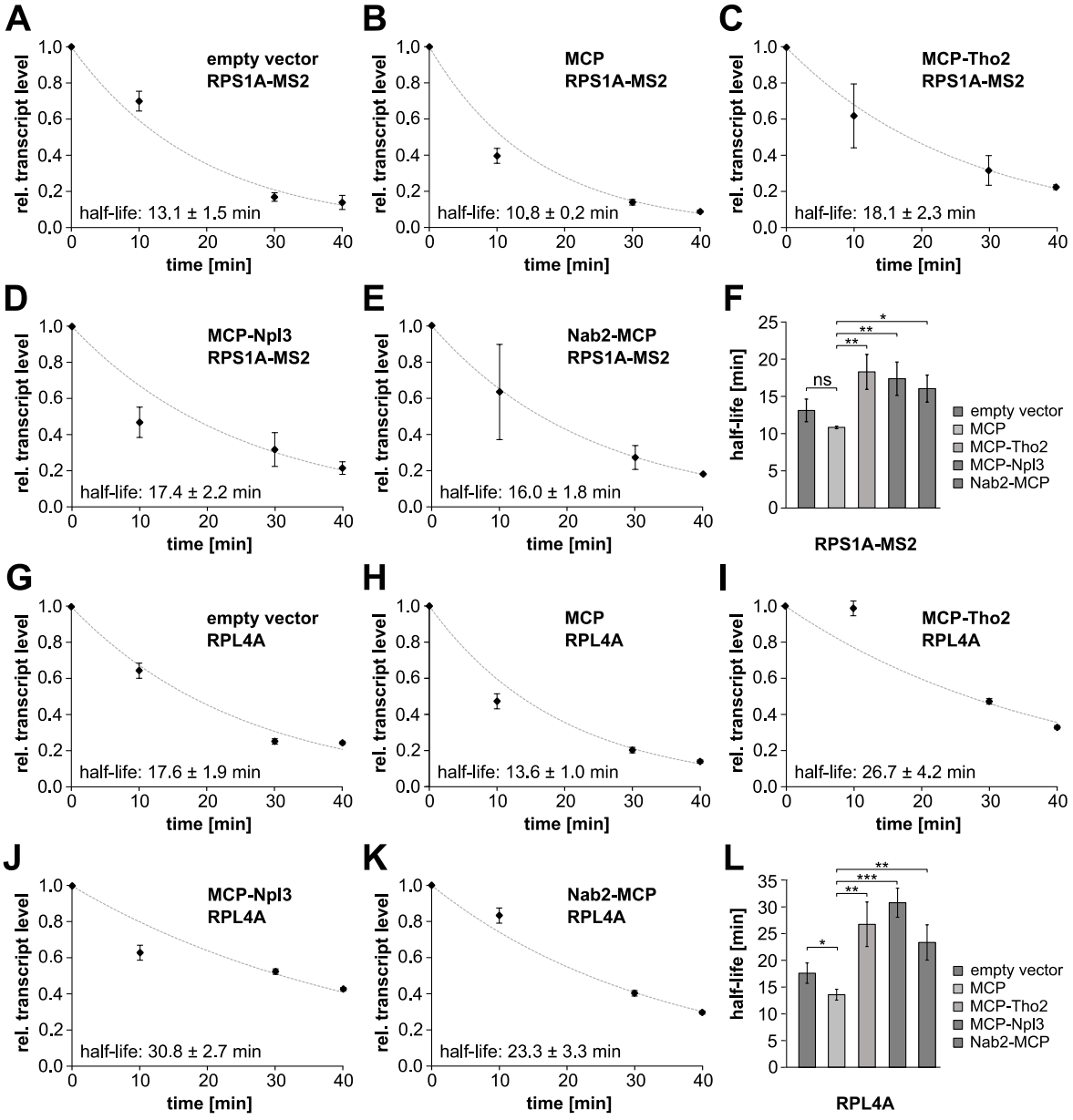
Appendix



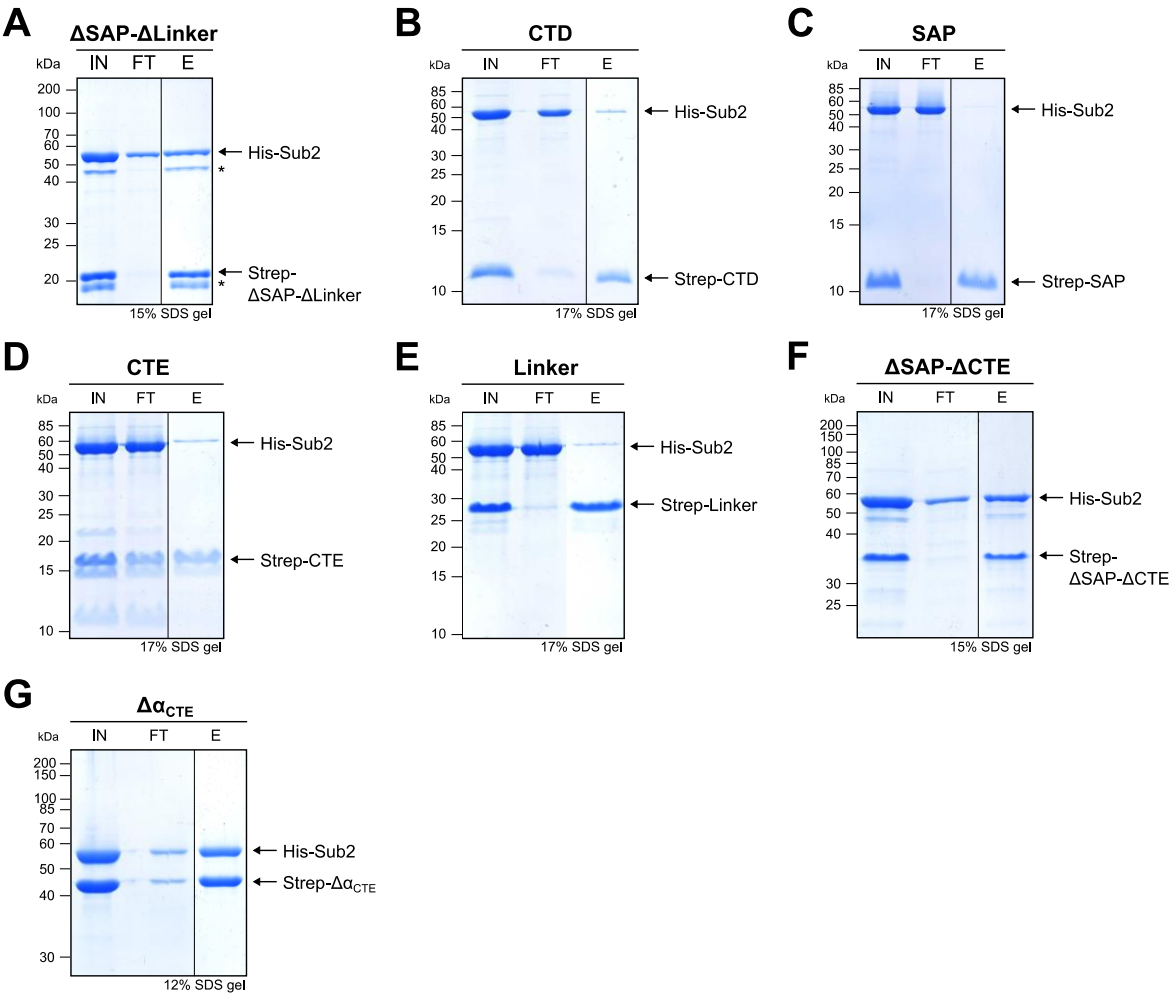
Supplementary Figure S1. Data processing workflow showing the image processing procedure with the number of particles and the resolution of the 3D reconstruction. The data was processed by Christin Radon (University of Potsdam).



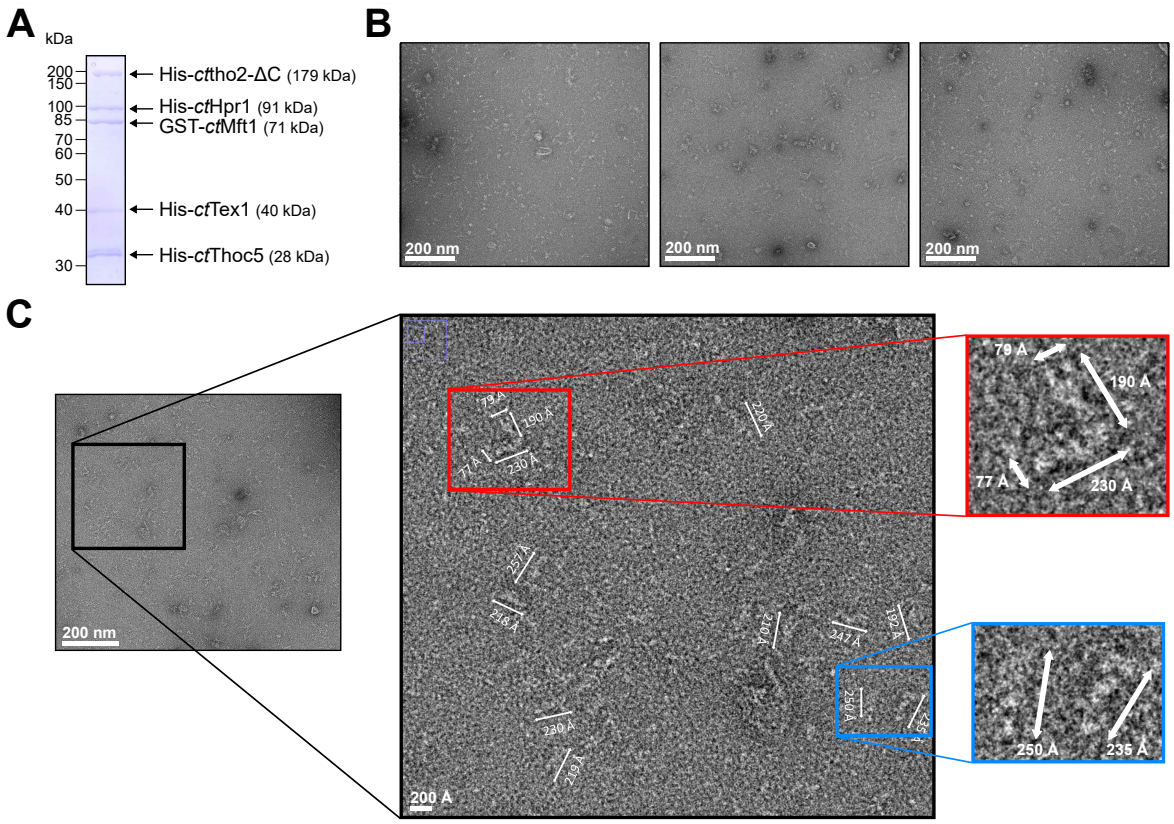
Supplementary Figure S2. Verification of protein expression by Western blot. The asterisk indicates the expressed *(A)* MCP-Tho1, *(B)* MCP-Npl3, and *(C)* Nab2-MCP proteins. Pgk1 served as a loading control.



Supplementary Figure S3. Overexpression of Tho2, Npl3 and Nab2 leads to an increased half-life of transcripts. Relative transcript level of RSP1A, which is genomically labeled with 24x MS2v6 repeats in its 3' UTR, in *(A)* cells containing an empty vector, *(B)* cells containing a plasmid expressing MCP, *(C)* cells containing a plasmid expressing MCP-Tho2, (D) cells containing a plasmid expressing MCP-Npl3, and (E) cells containing a plasmid expressing Nab2-MCP. Relative transcript level of unlabeled RPL4A in *(G)* cells containing an empty vector, *(H)* cells containing a plasmid expressing MCP, *(I)* cells containing a plasmid expressing MCP-Tho2, (J) cells containing a plasmid expressing MCP-Npl3, and (K) cells containing a plasmid expressing Nab2-MCP. The calculated medium half-lives of three biological replicates are indicated in each graph. Comparison of the medium half-lives of *(F)* RPS1A-MS2, and *(L)* RPL4A. The half-life of both transcripts is significantly increased in cells expressing MCP-Tho2, MCP-Npl3, or Nab2-MCP when comparing to cells expressing only MCP. * = p value < 0.05, ** = p value ≤ 0.01, *** = p value ≤ 0.001, students t-test.



Supplementary Figure S4. *In vitro* reconstitutions and pull-down experiments of the remaining *ct*Tho1 mutants. If there is a strong complex formation between *ct*Tho1 and *ct*Sub2, the flow-through is expected to be empty, while the eluate should contain both proteins. If there is no complex formation, *ct*Sub2 is expected to be found in the flow-through, while *ct*Tho1 is present in the eluate. *(A-G) In vitro* reconstitutions and Strep-tag II pull-downs of *(A)* ΔSAP-ΔLinker, *(B)* CTD, *(C)* SAP, *(D)* CTE, *(E)* Linker, *(F)* ΔSAP-ΔCTE, and *(G)* $\Delta\alpha_{CTE}$. Input (IN), flow-through (FT) and eluate (E) were analyzed on a 12%, 15%, or 17% SDS-PAGE gel that was stained with Coomassie Brilliant Blue.



Supplementary Figure S5. Negative-stain EM of the recombinant *ct*THO complex. *(A)* Representative purification of the recombinant *ct*THO complex by GST affinity purification. The eluate was analyzed on a 10% SDS-PAGE gel stained with Coomassie Brilliant Blue. Subunits corresponding to His-*ct*tho2-ΔC, His-*ct*Hpr1, GST-*ct*Mft1, His-*ct*Tex1 and His-*ct*Thoc5 were identified by mass spectrometry and are indicated accordingly. *(B)* Three representative negative-stain micrographs of the recombinant *ct*THO complex. On the grid prepared for negative-stain microscopy, the sample displayed higher levels of aggregation and a variety of heterogeneous particles. Scale bar represents 200 nm. *(C)* A representative close-up micrograph featuring measured particles. Most of the particles possess dimensions approximately in the range of 200 to 250 Å in length and 75 to 80 Å in width. In general, the particles closely resemble the size and shape of a monomer. Scale bar represents 200 Å. The red box highlights two exemplary particles, while the blue box shows another two.

Eigenständigkeitserklärung

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 Antiplagiat-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.

_____ _____

Ort, Datum **Unterschrift**