Insights into rRNA binding and lncRNAs in *Brassicaceae*

Dissertation with the aim of achieving a doctoral degree

at the faculty of Mathematics, Informatics, and Natural Sciences

Department of Biology

University of Hamburg

Submitted by

Rita João Pereira Fernandes

Hamburg, 2024

Date of disputation:

27/03/2024

Reviewers:

Prof. Dr. Julia Kehr

Prof. Dr. Stefan Hoth

Eidesstattliche Versicherung - Declaration on oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Rita João Pereira Fernandes

Hamburg, den 21.01.2024 Unterschrift

Abstract

Plants must quickly adapt to their surroundings due to their sessile nature. Their ability to adjust and thrive in diverse environmental conditions directly impacts their growth and development. This is especially critical for crop plants, where quantity and quality are vital to secure proper global nutrition. Plants have several ways to adapt or respond to external stimuli. One way is through the regulation of their proteome by maintaining the balance between protein synthesis and degradation. Another way is by emitting signaling molecules over short or long distances from affected areas to unaffected areas.

The synthesis of proteins is controlled by ribosomes and their biogenesis is a multistep process that involves the synthesis and maturation of rRNAs with the aid of hundreds of ribosome biogenesis factors (RBFs). Although many RBFs have been identified in plants, a substantial amount remains to be thoroughly studied compared to other organisms. Thus, one of the aims of this work was to characterize the structure and function of the Nucleolar RNA Chaperone-like 1 (NURC1) RBF in *A. thaliana*. SEC-SAXS results showed that NURC1 is highly flexible and adopts an elongated shape in solution. SEC-MALLS confirmed that NURC1 was a monomer with a molecular weight of ~28 kDa and a R_h of 2.4 nm. The homogeneity of the sample was compromised by the increase of protein concentration as observed through DLS, native electrophoresis and batch SAXS measurements. RNA binding was determined by MST. NURC1 bound the ITS2 RNA of *A. thaliana* with a K_d of 228 ± 83 nM. Furthermore, NURC1 displayed an RNA chaperone-like activity by FRET. The flexible termini of NURC1 seemed to be essential for structural stability and RNA recognition and/or biding since their truncation led to either increase of the sample polydispersity, decreased the binding affinity towards the ITS2 or reduced the RNA chaperone-like activity.

Long-distance communication is achieved through the phloem. Several classes of RNA have been identified in the phloem sap, including long non-coding RNAs (lncRNAs), which can potentially work as signals. The crop plant *B. napus* allows to readily collect fair amounts of phloem sap by exudation. Thus far, the population of lncRNAs of *B. napus* phloem sap remains to be explored. Hence, the second aim of this thesis involved the identification and characterization of lncRNAs in the phloem sap of *B. napus* by Illumina sequencing. For comparison, leaf samples were also analyzed. Over 3000 putative lncRNAs were identified in phloem and leaf samples. No structural differences were found between leaf and phloem lncRNAs. Differential expression and functional analysis (GO terms) were also performed. Parent genes associated with various molecular functions or pathways were identified, likely due to the diversity of processes that lncRNAs are involved.

Zusammenfassung

Pflanzen müssen sich aufgrund ihrer sessilen Natur schnell an ihre Umgebung anpassen. Ihre Fähigkeit, sich an unterschiedliche Umweltbedingungen anzupassen und dort zu gedeihen, wirkt sich direkt auf ihr Wachstum und ihre Entwicklung aus. Dies ist vor allem für Kulturpflanzen von entscheidender Bedeutung, da Quantität und Qualität für die Sicherstellung einer angemessenen globalen Ernährung entscheidend sind. Pflanzen haben mehrere Möglichkeiten, sich anzupassen oder auf äußere Reize zu reagieren. Eine Möglichkeit besteht in der Regulierung ihres Proteoms, indem sie das Gleichgewicht zwischen Proteinsynthese und -abbau aufrechterhalten. Ein anderer Weg ist die Aussendung von Signalmolekülen über kurze oder lange Strecken von betroffenen zu nicht betroffenen Teilen der Pflanze.

Die Synthese von Proteinen wird von Ribosomen gesteuert, und ihre Biogenese ist ein mehrstufiger Prozess, der die Synthese und Reifung von rRNAs mit Hilfe von Hunderten von Ribosomen-Biogenese-Faktoren (RBFs) umfasst. Obwohl viele RBFs in Pflanzen identifiziert wurden, ist eine beträchtliche Anzahl von ihnen im Vergleich zu anderen Organismen noch nicht gründlich untersucht worden. Eines der Ziele dieser Arbeit war es daher, die Struktur und Funktion des RBF Nucleolar RNA Chaperone-like 1 (NURC1) in *A. thaliana* zu charakterisieren. Die SEC-SAXS-Ergebnisse zeigten, dass NURC1 sehr flexibel ist und in Lösung eine längliche Form annimmt. SEC-MALLS bestätigte, dass NURC1 ein Monomer mit einem Molekulargewicht von ~28 kDa und einem R_h-Wert von 2,4 nm ist. Die Homogenität der Probe wurde durch den Anstieg der Proteinkonzentration beeinträchtigt, wie durch DLS, native Elektrophorese und SAXS-Batch-Messungen festgestellt wurde. Die RNA-Bindung wurde durch MST bestimmt. NURC1 band die ITS2-RNA von *A. thaliana* mit einer K_d von 228 ± 83 nM. Darüber hinaus zeigte NURC1 durch FRET eine RNA-Chaperon-ähnliche Aktivität. Die flexiblen Termini von NURC1 scheinen für die strukturelle Stabilität und die RNA-Erkennung bzw. -Bindung wesentlich zu sein, da ihre Verkürzung entweder zu einer Erhöhung der Polydispersität der Probe, einer Verringerung der Bindungsaffinität zur ITS2 oder einer Verringerung der RNA-Chaperon-ähnlichen Aktivität führte.

Die Kommunikation über große Entfernungen erfolgt über das Phloem. Im Phloemsaft wurden mehrere RNA-Klassen identifiziert, darunter lange nicht kodierende RNAs (lncRNAs), die möglicherweise als Signale dienen können. Die Nutzpflanze *B. napus* erlaubt das Sammeln größerer Mengen an Phloemsaft durch die Exsudation. Bislang ist die Population der lncRNAs im Phloemsaft von *B. napus* noch nicht erforscht. Daher bestand das zweite Ziel dieser Arbeit in der Identifizierung und Charakterisierung von lncRNAs im Phloem-Saft von *B. napus* durch Illumina-

Sequenzierung. Zum Vergleich wurden auch Blattproben analysiert. Über 3000 putative lncRNAs wurden in Phloem- und Blattproben identifiziert. Es wurden keine strukturellen Unterschiede zwischen Blatt- und Phloem-lncRNAs gefunden. Es wurden auch eine differenzierte Expressions- und Funktionsanalyse (GO-Terme) durchgeführt, bei der Parentale Gene identifiziert wurden. Diese stehen mit verschiedenen molekularen Funktionen oder Pfaden in Verbindung, was wahrscheinlich auf die Vielfalt der Prozesse zurückzuführen ist, an denen lncRNAs beteiligt sind.

Table of Contents

List of Abbreviations

ANOVA	Analysis of variance
ATP	Adenosine triphosphate
CC	Companion Cell
Cryo-EM	Cryogenic electron microscopy
CTAB	Centrimonium bromide
Cy3	Cyanine3
Cy5	Cyanine5
DLS	Dynamic Light Scattering
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
ETS	External Transcribed Spacer
FAO	Food and Agriculture Organization
FPKM	Fragments Per Kilobase of transcript per Million read pairs
FRET	Fluorescence Resonance Energy Transfer
gDNA	Genomic DNA
GO	Gene Ontology
IRP	Involved in rRNA Processing
ITS	Internal Transcribed Spacer
$K_{annealing}$	Annealing constant
K_d	Dissociation constant
lincRNA	Long intergenic non-coding RNA
lncRNA	Long non-coding RNA
LSU	Large Subunit
MALDI	Matrix Assisted Laser Desorption Ionization
MALLS	Multi-Angle Laser Light Scattering
MS	Mass Spectrometry
MSA	Multiple Sequence Alignment
NTPs	Nucleoside Triphosphate
NURC1	NUCLEOLAR RNA CHAPERONE-LIKE 1
ORF	Open Reading Frame

PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PD	Plasmodesmata
PPCs	Phloem Parenchyma Cells
P-proteins	Phloem proteins
PPUs	Pore-plasmodesma units
QC	Quality Control
QELS	Quasi-Elastic Light Scattering
RBFs	RNA Biogenesis Factors
rDNA	Ribosomal DNA
RFOs	Raffinose Family Oligosaccharides
R_g	Gyration Radius
R_h	Hydrodynamic Radius
r_{max}	Maximum radius
RNP	Ribonucleoprotein
RPs	Ribosomal Proteins
RRM	RNA Recognition Motif
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
SAXS	Small Angle X-ray Scattering
SDM	Site Directed Mutagenesis
SDS	Sodium dodecyl sulfate
SE	Sieve Element
SEC	Size Exclusion Chromatography
SEL	Size Exclusion Limit
snoRNA	Small Nucleolar RNA
SRP	Signal recognition particle
SSU	Small Subunit
STRS	STRESS RESPONSE SUPPRESSORS
TEV	Tobacco Etch Virus
TOF	Time of Flight

List of Figures

List of Tables

List of Tables

1. Introduction

1.1. Plants: scientific and economic relevance

Plants are vital for human survival. They supply not only food, but also clothing, shelter, medicine, fuel, and most importantly, they recycle CO_2 into O_2, the air we breathe[1]. The angiosperms, or flowering plants, are a diverse group of plants that produce seeds enclosed in a vessel. With over 250 000 species known, this clade dominates the terrestrial biota[2,3]. They provide practically all plant-based food, such as fruits and cereals. Thus, they are of great economic and agricultural importance. The *Brassicaceae* family belongs to this clade. *Arabidopsis thaliana* is a very well-known member of this family. Despite not being a crop plant, traits like fast and possible indoor growth, small size, self-compatibility, and small genome made this species an attractive model organism for plant science, including fields like genetics, molecular biology, and physiology[4,5]. Another popular member of this family is *Brassica napus*, commonly known as rapeseed or oilseed rape. During the 1970s, the appearance of "double low" rapeseed (low levels of toxic compounds such as erucic acid and glucosinolates), aroused special interest as an edible vegetable oil source worldwide, and was commercially named **C**an**ola** (Canadian Oil Low Acid) [6,7]. Canola's oil nutritional balance — i.e., lipid profile low in saturated fats, high in monosaturated fats, and rich in omega-3 fatty acids— led nutritionists to prefer its use over other types of oils[7]. Cultivars with high levels of toxic compounds are still exploited for production of lubricants, waxes, water repellents, biofuels, textiles, rubber, etc[8,9]. According to the Food and Agriculture Organization (FAO), in the past 20 years, canola oil production has been rising, with approximately 25 million tons produced globally in the year 2020, from which nearly 10 million were produced in Europe (**Figure 1.1A**). China was the largest producer of canola oil, followed by Germany and Canada (**Figure 1.1B**).

Genetically, *B. napus* is an allotetraploid (AACC, 2n=4x=38) that originated by spontaneous and inter-specific hybridizations between the diploid *Brassicas rapa* (AA, 2n=2x=20) and *oleracea* (CC, 2n=2x=18). Due to extensive breeding of *Brassicas* over time, and the fact that no truly wild *B. napus* populations are known, it is still under scrutiny which genotypes are the precise ancestors and the original hybrid[10–12]. However, there is evidence that *B. napus* A subgenome is closest to European turnip than other *B. rapa* subspecies[11,13]. The C subgenome origin remains elusive, but it might have evolved from a common ancestor of four *B. oleracea* subspecies (kohlrabi, broccoli, cauliflower, and Chinese kale)[11]. *Brassicas* are annual or biennial crops, and their life cycle can vary between 6 to 12 months depending on species or ecotype, which is considerably longer when compared to the life cycle of *A. thaliana* of six weeks[5,14]. Not all *Brassicas* are self-compatible, like for example, *B. rapa* and *B. oleracea*[15]. They also require considerable space to grow due to their larger size. These features can be inconvenient for laboratory research, where time and space demands are crucial. However, one great advantage of utilizing crops like *B. napus* is that its phloem sap can readily be collected in reasonable amounts[16–18].

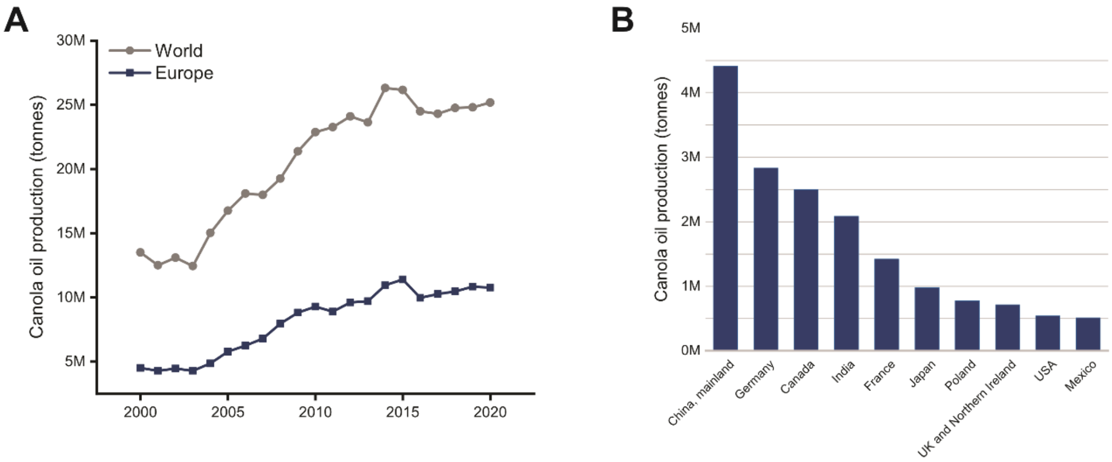


Figure 1.1− Canola oil production has increased over the 21st century. (**A**) Canola oil production in the World (grey, spheres) and Europe (blue, squares). (**B**) Amount of canola oil produced by different countries. Both graphs show averaged production values for the years 2000 until 2021 (FAOSTAT, 2023).

1.2. Ribosome biogenesis in plants

Plants can control their development or respond to external stimuli in several ways, including the adjustment of their proteome by regulating the balance between protein synthesis and degradation. Ribosomes are responsible for protein synthesis and are produced constitutively in eukaryotic cells. Their biogenesis is one of the most energy-demanding and tightly regulated cellular processes, that includes the synthesis, processing, and assembly of four ribosomal RNAs (rRNAs) with ~80 ribosomal proteins (RPs)[19]. Plants with impaired ribosome biogenesis are less efficient in stress response and show severe phenotypes, such as embryo lethality or delayed embryogenesis[20–22].

Figure 1.2 shows an overview of ribosome biogenesis. The protein synthesis machinery is highly conserved, and the cytosolic mature ribosome (80S) consists of a small subunit (SSU or 40S) that comprises the 18S rRNA associated with at least 33 RPs, and a large subunit (LSU or 60S) that contains the 25S/28S (plants/mammals), 5.8S and 5S rRNAs assembled to at least 47 RPs[19,23]. The 5S rRNA is transcribed by polymerase (Pol) III and processed in the nucleoplasm, whereas the polycistronic rRNA, which contains the other three rRNAs, is transcribed by Pol I and processed in the nucleolus[24,25].

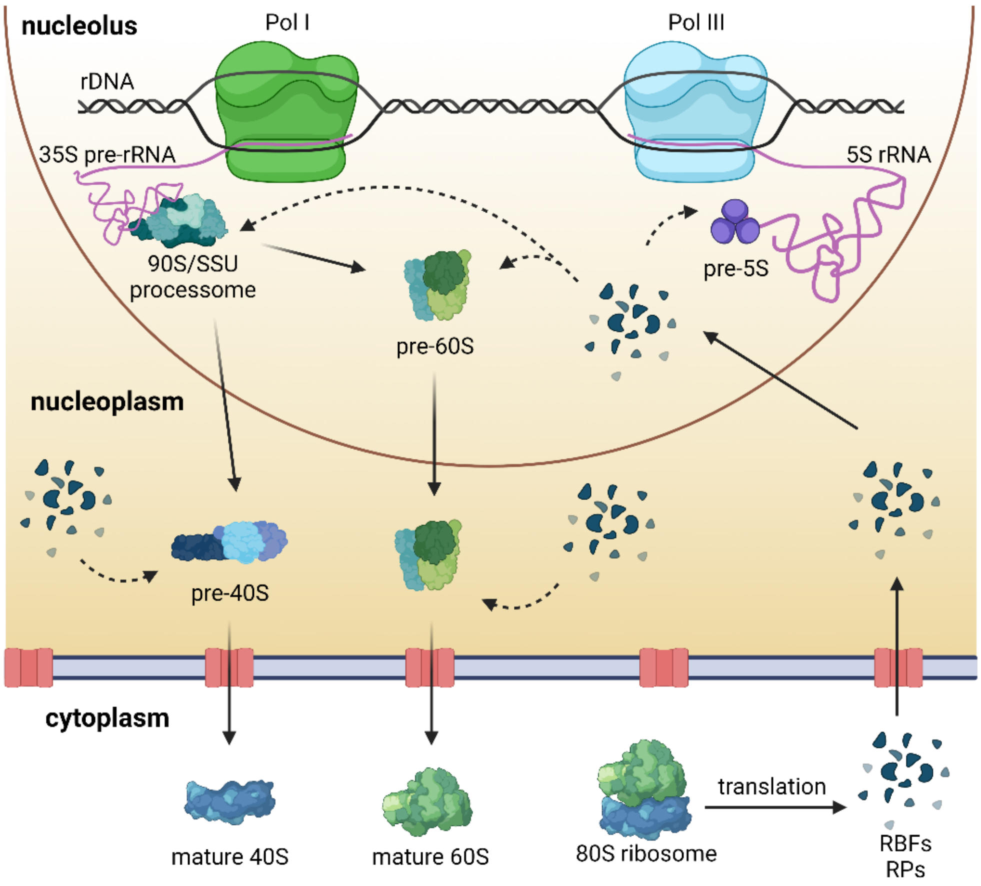


Figure 1.2− Schematic representation of ribosome biogenesis. The ribosome biogenesis process starts in the nucleolus, where the rDNA is transcribed into 35S and 5S pre-rRNAs, by Pol I and Pol III, respectively. Processing of the 35S pre-rRNA is carried out by the 90S/SSU processome. During this process, SSU and LSU rRNAs are split and follow different processing paths. Maturation steps in the nucleolus, nucleoplasm and cytoplasm were omitted for simplification. Lastly, cytosolic maturation occurs, and the processed subunits are assembled into the final mature 80S ribosome. In its turn, translation of processing factors, such as ribosome biogenesis factor (RBFs) and ribosomal proteins (RPs) occurs in the cytoplasm. These factors are then imported into the nucleus and integrated into the processome complex when required. 5S rRNA processing steps are not included. This image was created with Biorender and is based on ref. [26].

The polycistronic pre-rRNA, or 45S pre-rRNA in plants, includes the 18S, 5.8S and 25S sequences separated by internal transcribed spacers (ITS1 and ITS2) and flanked by external transcribed spacers (5'ETS and 3'ETS) (**Figure 1.3**). Processing of the pre-rRNAs occurs co-transcriptionally and relies on the catalytic activity of exo- and endonucleases for cleavage and removal of the spacer regions. The repertoire of nucleases plus the timing and order of processing event can vary between organisms[27]. Based on the first cleavage event, three different pathways are known to occur in plants. The first pathway or 5'ETS-first starts with the removal of the 5'-ETS by successive cleavage at the P'/B_2, P_2 and A_2 sites followed by complete cleavage of ITS1 (**Figure 1.3**). Like in yeast and mammals, it is dependent on the early binding of the SSU processome to the rDNA[28,29]. The plant SSU processome has been characterized in *B. oleracea*[30]. The second pathway, known as ITS1-first, begins with cleavage at the A_3 site of ITS1, resulting in the separation of SSU and LSU rRNAs (**Figure 1.3**). The third pathway or ITS2-first involves the cleavage of the ITS2 at the C_2 site, followed by the cleavage at the A_3 site of the ITS1 (**Figure 1.3**). This pathway is unique to plants, and it was described by Palm *et al.* 2019, who identified nine 'involved in rRNA processing' (IRP) factors required for successful rRNA maturation, one of which is involved in this pathway[31]. However, the nuclease responsible for the ITS2 cleavage has not been identified thus far[27].

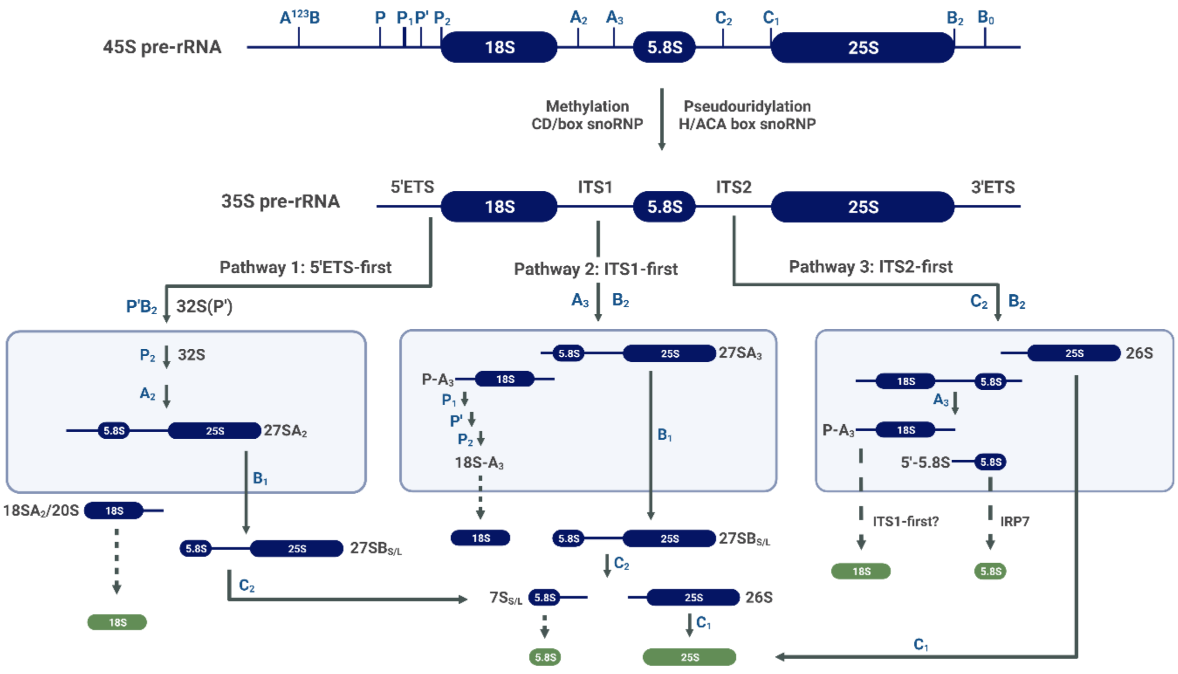


Figure 1.3− rRNA processing pathways. Processing of the 45S rRNA involves the removal of both ETS and ITS sequences. The different cleavage sites are shown. The early pre-rRNA is methylated and

pseudouridylated by C/D box and H/ACA box small-nucleolar RNAs (snoRNAs) ribonucleoprotein (RNP) complexes, respectively. These modifications result in the stability of the secondary and tertiary structure of the rRNA scaffold.[32] Further steps can be followed by three different pathways: 5'ETS-, ITS1- or ITS2-first. Pathway specific steps are inside boxes and common steps outside. The final mature rRNAs are shown in green. For simplification, several cleavage steps and RNA intermediaries were not included. This figure is based on ref. [25] and was created with Biorender.

Regardless of the pathway taken, cleavage of both ITS sequences always leads to the separation of SSU and LSU precursor rRNAs and their biogenesis follows separate routes that can be mediated by the 90S/SSU processome and by pre-60S particles/LSU processome, respectively (**Figure 1.2 and Figure 1.3**)[23,25]. These large ribonucleoprotein particles (RNPs) comprise numerous supporting elements, such as RPs and ribosome biogenesis factors (RBFs). RPs are thought to act as chaperones during structural assembly and are essential in pre-rRNA processing and nuclear export[33]. In *A. thaliana* and *B. napus*, many RPs are encoded by multiple paralogs and the composition of ribosomes is affected by the developmental stage, type of tissue and environmental stimuli[34,35]. Thus, cytosolic ribosomal heterogeneity is particularly high in plants. Additionally, the plethora of pre-ribosomal particles, which quickly change spatial- and temporally, makes the study of ribosome biogenesis a complex and challenging task.

1.3. Ribosome biogenesis factors

RBFs include a multitude of proteinaceous and RNA molecules that transiently associate with pre-ribosomal particles but are not part of the mature ribosome. They can act as chaperones as well as modification, processing, assembly, and remodeling factors, throughout a variety of biogenesis steps ranging from rRNA processing to nuclear export of pre-ribosomal particles [36,37].

Several proteomic and bioinformatic analyses of the plant nucleolus allowed the identification of approximately 300 putative RBFs based on orthology to yeast and human factors[37–40]. The distribution of RBFs in *A. thaliana* has been proposed, supporting the idea that all three − plants, mammals, and fungi− share a common assembly mechanism[37].

Many of the identified RBFs are associated with the pre-90S complex, like for example Has1. The yeast Has1 is a DEAD-box RNA helicase (DDX-RH) which is recruited either by Rlp7 or Nop15 into the pre-66S ribosome to regulate conformational changes within the 5.8S/25S rRNA[41]. More than 50 DDX-RHs are encoded in *Arabidopsis,* and many are implied not only in ribosome biogenesis, but also in other physiological functions[42]. For instance, the AtRH36 is involved, not only in the 18S rRNA processing, but also in female gametogenesis[43], and STRESS RESPONSE SUPPRESSORS (STRS1 and STRS2) function as negative regulators to abiotic stresses, such as, salt and heat[44]. Several proteins were associated with the pre-60S, like for example Nop15.

In vivo studies showed that Nop15 depletion led to substantial decrease of ribosomal subunits, and consequently, to abrupt cell growth arrest. Accumulation of the 27SA pre-RNA was observed, while the 27SB and 7S pre-RNAs were lost[45]. Co-precipitation with these pre-rRNAs supported the conclusion that Nop15 is essential in the synthesis of the 25S and 5.8S[45,46]. An additional role as trigger of cytokinesis in mitotic division was suggested, although the mechanism behind remains elusive[45]. The human ortholog (NIFK), has also been reported to play an important role in cell cycle progression, besides its role in the maturation of LSU rRNAs[37,47]. Moreover, overexpression of NIFK promotes lung cancer progression[48].

Nop15 has also been used as bait for rapid affinity-purification, and its interactome included 136 proteins, 89 of which were RBFs associated with 90S and pre-60S particles[49]. Examples of co-purified RBFs are Ebp2 and Nug1. In *Arabidopsis*, NSN1 (ortholog to Nug1) interacts with EBP2 (ortholog to Ebp2), and its depletion caused delayed 25S rRNA maturation/60S ribosome biogenesis and resulted in growth retardation and plant premature senescence[50]. The ortholog of yeast Nop15 protein and its interactome has yet to be characterized in plants.

1.3.1 Nop15 role in the ITS2 processing

Nop15 is a nucleolar RBF that is associated with the LSU processome. The structure of several states of yeast pre-60S has been shown by Cryogenic Electron Microscopy (Cryo-EM) and over 30 assembly factors, including unprocessed 25S, 5.8S and 5S rRNA intermediates, have been identified[51–54]. The GTPases Nog1 and Nog2 serve as hub proteins, allowing the recruitment and interaction of distant assembly factors[53]. The assembly factors are clustered around the pre-rRNA from the 5’-end of the 25S until the 3’-end of the 5.8S, with the unprocessed ITS2 in between[26,53]. In **Figure 1.4A** the structural overview of a pre-60S particle is depicted. The “foot-like” structure contains Nop15, which participates in a subcomplex with Cic1, Rlp7, Nop53 and Nop7 (**Figure 1.4B**)[55]. The high protein content at the ITS2 reflects the likely function of the RBFs to chaperone and protect from improper processing[53]. Chemical cross-linking coupled with mass spectrometry revealed that Nop15 crosslinks with the RPs L25 and Rlp7, and in turn Rlp7 crosslinks Cic1 and Nop53[55]. Moreover, Nop15 and Cic1 were found to bind directly to the ITS2[46,55].

The ITS2 can adopt two distinct secondary structures termed the ‘ring’ and the ‘hairpin’ (**Figure 1.4C**). Both forms have features in common, such as stems II, IV, V and VI. However, the two differ at stem III. In the hairpin model this segment forms long range interactions, whereas in the ring form is split into two independent stems (III.A and III.B). The ring form seems to be essential for recognition of the processing machinery, as a quality-control checkpoint, so that premature processing is prevented[56]. Moreover, mutations altering the structure of stem III.A results in failed processing, and the transition from ring to hairpin form seems to be necessary for efficient processing, but not required[56]. Nop15 binding site was mapped to stem III.A, and together with Cic1 secures the flexible and open structure of the ring form[46].

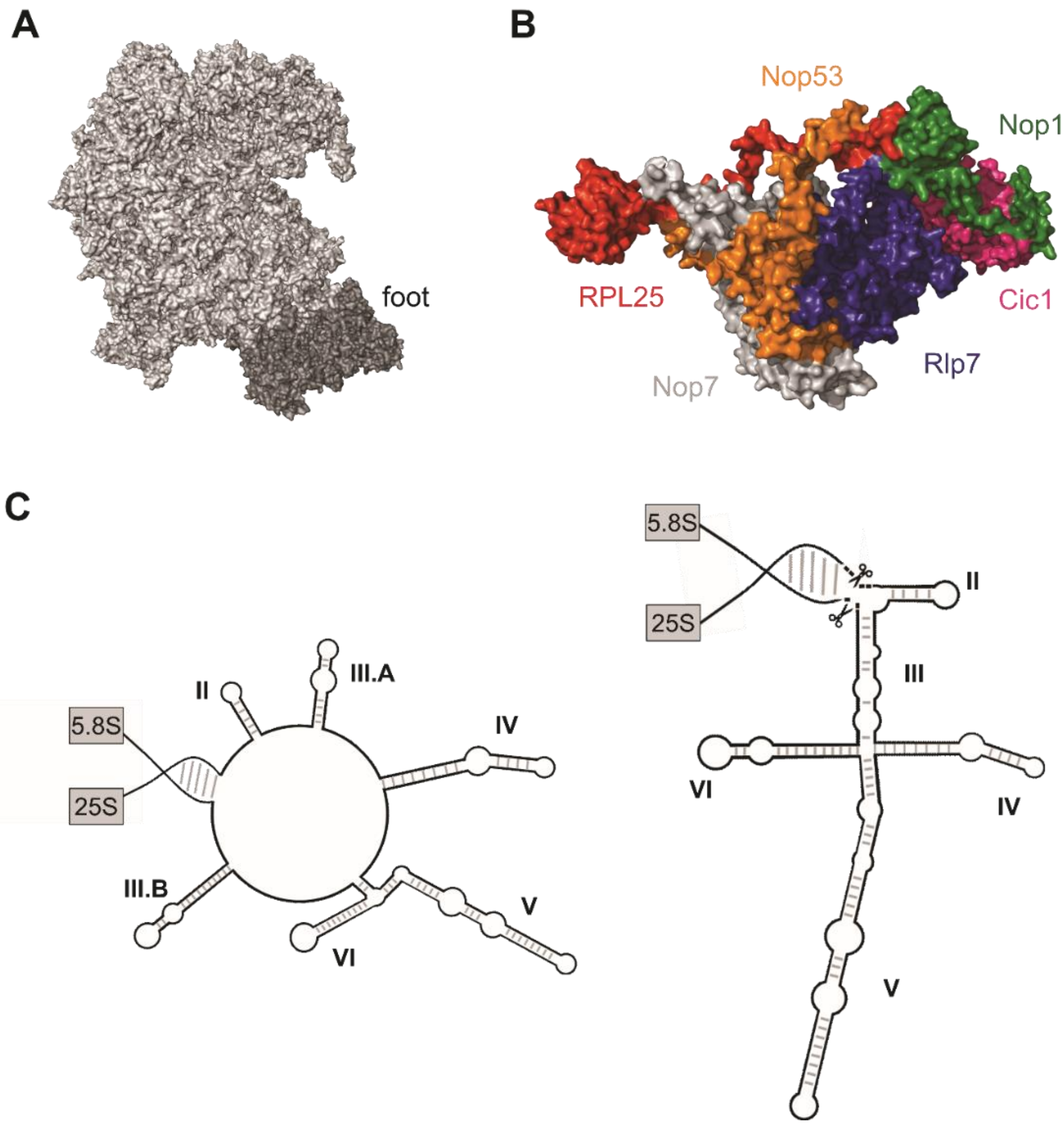


Figure 1.4− Nop15 is part of the pre-60S ribosomal particle and binds to the ITS2. (**A**) Surface illustration of the pre-60S particle. The 'foot-like' region is shown in dark grey. (**B**) Zoom-in of the 'foot-like' region. Nop15 (green) is localized in this region and interacts directly with Cic1 (hot pink), Rlp7 (deep blue) and RPL25 (red). Nop7 (light grey) and Nop53 (orange) are also present in this sub-complex. (**C**) The ITS2 RNA can adopt two secondary structures. The ring form (left) is recognized by the processing machinery and Nop15 binds to stem III.A, preventing incorrect processing of the ITS2. The 'hairpin' form (right) is the result of the pairing of stems III.A and III.B from the ring form, resulting in a structure thermodynamically more stable. Transition from the ring form into the hairpin form is essential for processing efficiency. Figures A and B were designed in Pymol and were based on the structures from PDB 3JCT ref.[55]. Figure C was based on ref. [56].

1.3.2 Nop15 structure and ITS2 recognition

The atomic structure of Nop15 has been resolved by techniques such as Cryo-EM[55] and X-ray crystallography[57]. Nop15 harbors a single RNA recognition motif (RRM) embedded between N- and C-terminal flexible regions (**Figure 1.5A**). The RRM motif is one of the most ubiquitous RNA binding domains and consists of approximately 90 amino acids with a characteristic $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology.[58,59] Nop15$_{81-191}$ crystal revealed an additional β strand (β_4') located between α_2 and β_4, and one short α-helix (α_3) appended to the core RRM (**Figure 1.5B**). β_1 and β_3 contain the RNA-binding motifs RNP2 and RNP1, respectively. Aliphatic residues from the C-terminal were shown to interact with aromatic residues of both RNPs, that typically form stacking interactions with RNA bases (**Figure 1.5**). Removal of the C-terminal aliphatic residues resulted in protein aggregation, suggesting that these interactions promote Nop15 solubility[57]. Moreover, Small-Angle X-ray Scattering (SAXS) and Nuclear Magnetic Resonance (NMR) experiments showed that the C-terminal residues are highly flexible in solution[57].

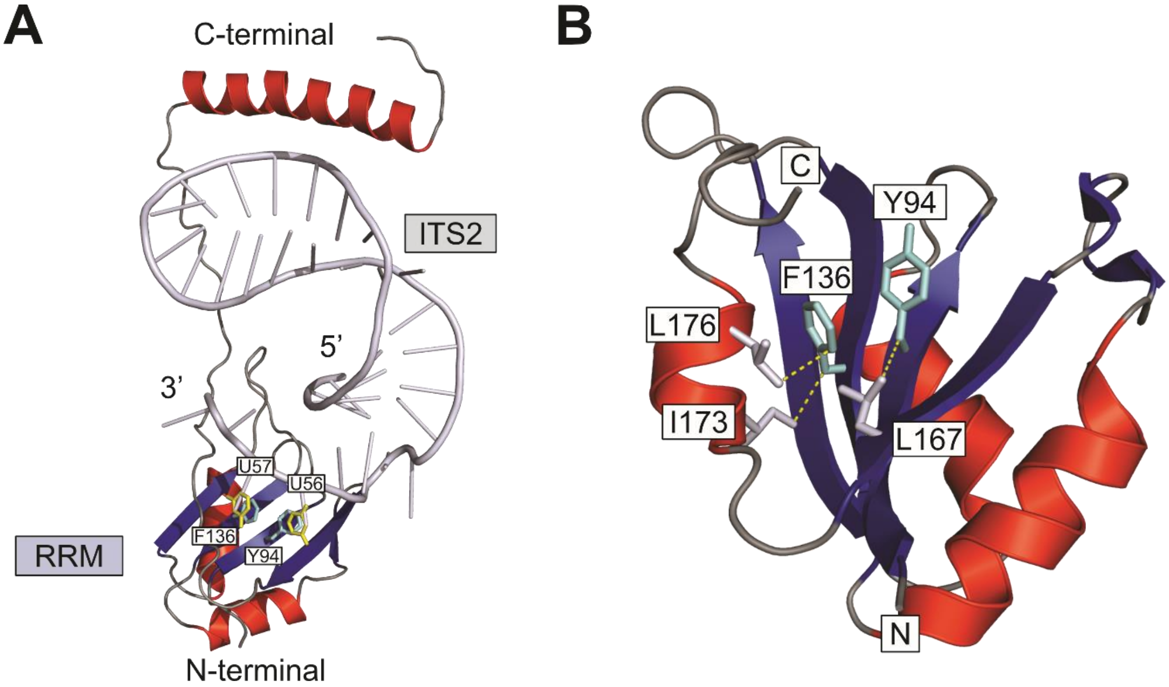


Figure 1.5− C-terminal of Nop15 refolds to interact with ITS2 RNA and is essential for structural stability in absence of RNA. (**A**) Cryo-EM model of Nop15-ITS2 complex (PDB ID: 3JCT). Nop15 harbors a canonical RMM. The aromatic residues Y94 and F136 (cyan sticks) of the RRM interact with nucleotides U56 and U57 (yellow sticks) from the 3'-end of ITS2 RNA. (**B**) Crystal structure of Nop15 (2.0 Å, PDB ID: 5T9P) ref. [57]. RNP2 (Y94) and RNP1 (F136) aromatic residues (cyan sticks) are masked by the aliphatic residues L167, I173 and L176 (white sticks). Atomic interaction distances are represented by dashed yellow lines. Distances are between 3.8-4.0 Å. The α_3-helix containing the aliphatic residues L173 and L176 is not

present in the Cryo-EM structure, thus refolding of the C-terminal is induced by RNA binding. Both structures were designed in Pymol.

The binding affinity of Nop15 to its pre-rRNA target (ITS2 stem III.A) has been determined. The K_d value was in the nanomolar range for the full-length Nop15, and the binding was improved when truncating the N-terminus[57]. This could be explained by the removal of an acid patch that competes with the RNA for the binding sites. However, removal of the C-terminal aliphatic residues resulted in an increase of the K_d value to the micromolar range[57]. When looking at the Cryo-EM structure, these residues interact with the ITS2 distal end, being essential for the Nop15-ITS2 complex stability (**Figure 1.5A**). Comparing both structures, the flexible C-terminal region structural rearrangement is essential for tight RNA binding[57].

1.4. Intercellular communication in plants

Besides adjusting their proteome, plants also developed a rapid way of communicating development- or environmental changes throughout the entire body. Information can be carried from impacted areas of the plant to unaffected areas in the form of signals, which can be mobile molecules, such as hormones and ribonucleic acids[60]. Furthermore, land plants had to adapt to both aerial and soil environments, which resulted in the specialization of auto- and auxotrophic organs. Consequently, translocation of photoassimilates from synthetic tissues to non-photosynthetic tissues is most crucial for robust growth. Plants communicate intercellularly over short- and long-distances, i.e., to neighbor cells or to distant plant parts, respectively. Short-distance or cell-to-cell communication can be achieved by three transport pathways: apoplastic, symplastic or transcellular (**Figure 1.6**). As the name suggests, the apoplastic pathway takes place in the apoplast, where molecules diffuse through the cell wall and finally enter the target cell by the plasma membrane via transporters[60]. The symplastic pathway occurs across the cytoplasm and the transport is mediated by the plasmodesmata (PD), a plasma membrane-lined cytoplasmatic channels that confer symplastic continuity between neighboring cells[61]. Lastly, the transcellular pathway involves the transport of molecules across the plasma membrane by import-export mechanisms, such as diffusion, secretion, and transporter- or receptor-mediated systems[62]. Long-distance communication is accomplished by the vascular system.

1.5. The vascular system and long-distance communication

Two conducting tissues form the plant vascular system, the xylem and phloem. The vascular architecture is characteristic of each species and of each part of the plant within that species[63]. In dicotyledons like *A. thaliana* and *B. napus*, the stem vasculature consists of bundles that have a radial pattern of abaxial phloem and adaxial xylem with an active procambium in between[63,64]. The xylem is responsible for the unidirectional transport of water and minerals from the roots to above-ground organs, whereas the phloem is responsible for transporting photoassimilates from photosynthetic tissues (sources) to auxotrophic tissues (sinks)[61].

1.5.1 The xylem

The xylem consists mainly of water-conducting tracheary elements fibres that provide mechanical support and stability due to the presence of thick, lignified secondary cell walls, and cells that can store carbohydrates, water, and mineral nutrients, called parenchyma[63,65]. Tracheary elements and fibres undergo programmed cell death (PCD), while parenchyma cells retain their living components at maturity[63–65]. The transport of water through the xylem is ruled by four forces: transpiration, water cohesion, capillary action, and root pressure[65]. Transpiration and cohesion are the driving forces, being the first caused by the evaporation of water from leaf stomata, which results in a negative water potential, and, consequently, the water molecules will move together, due to their cohesive nature, from the roots to the leaves[65]. As a result, the movement in the xylem is unidirectional.

1.5.2 The phloem

The main conductive tissue of the phloem is comprised of sieve elements (SE). Mature SE lack organelles, such as the nucleus, vacuoles, Golgi and rough endoplasmic reticulum (ER)[64]. This offers a low resistance, increasing the transport efficiency[64,66]. Yet, the SE are still living cells, since they retain their plasma membrane and some organelles, like the smooth ER, plastids and mitochondria, and contain phloem specific proteins (P-proteins)[64,67]. They have an elongated shape and are connected to each other by sieve plates, forming a continuous tube known as sieve tube. Both the sieve plates and the lateral cell walls of SE are equipped with PD-derived pores[64]. SE survival depends on their neighbor cells, the companion cells (CC), and it is common to refer to them as the SE-CC complex, since they originated from the same mother cell[67]. The

SE-CC complex is connected by pore-plasmodesma units (PPUs). PPUs differ from other PDs structurally, in that they are branched on the CC side and have only one pore on the SE side[64,67,68]. They also have a larger size exclusion limit (SEL) that allows the passage of molecules bigger than 60 kDa. In contrast, mesophyll and epidermal PD have a SEL of around 1 kDa[68].

One of the first steps of long-distance transport involves phloem loading. Sugars from mesophyll cells can reach the SE-CC complex either by the apoplastic or the symplastic pathway[64]. Both pathways were described above (**chapter 1.4.**). The loading of sugar molecules into the CCs can be divided into active or passive. Active loading implies that energy (ATP) is required for the movement to happen. This is true for molecules entering via apoplast since they move against the concentration gradient[64,69]. Passive loading is achieved simply by the diffusion of sugars down the concentration gradient that exists between the mesophyll (high) and phloem (low) cells[69]. Additionally, symplastic transported monosaccharides can be converted into raffinose in specialized CCs, known as intermediary cells[64,69]. This process is termed polymer trapping and requires energy, thus is considered an active loading mechanism[69]. The loading mechanism is dependent on the plant species, yet different mechanisms can be used by one species[70].

Currently, the mass flow hypothesis proposed by Münch[71] is widely accepted to explain how the movement in the phloem is accomplished. Briefly, the transport of molecules is driven by the difference in hydrostatic pressure between source (higher) and sink (lower) cells. This difference is prompted by the higher concentration of solute (e.g., sucrose) in the sieve tube of source tissues, as a result, water is drawn from the xylem by osmosis, increasing the cell turgor and causing mass/bulk flow from source to sink[69,72]. Recent experimental evidence supports that long-distance transport is indeed driven by pressure[73].

Lastly, phloem-unloading is likely accomplished by bulk flow from the SE-CC complex into adjacent phloem parenchyma cells (PPCs)[64]. Once there, the nutrients can take either the symplastic or apoplastic pathway to reach the final destination: the sink cells.

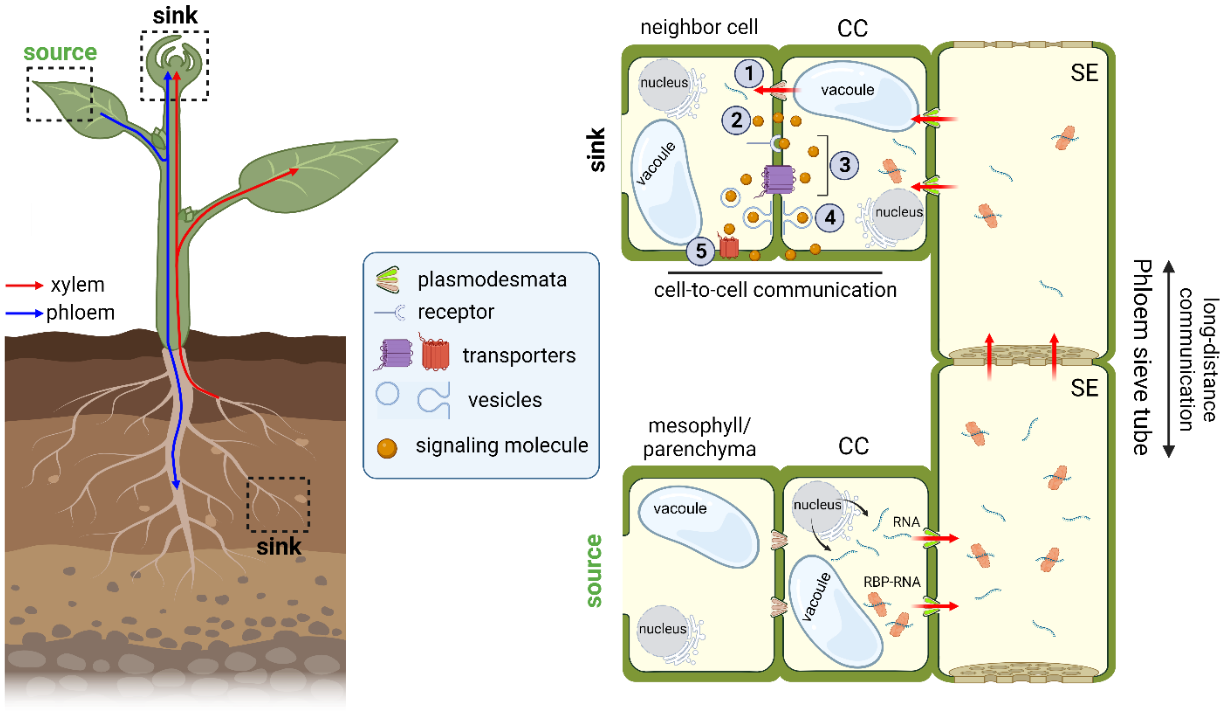


Figure 1.6− Transport pathways of signaling molecules in plants. (**Left**) Xylem unidirectional transport from the roots to the leaves and shoot apical meristematic tissues, and phloem bidirectional transport between source and sink tissues. (**Right**) Short- and long-distance transport pathways. Cell-to-cell transport can be symplastic via plasmodesmata (1), diffusion (2), transporters/receptors (3), secretion (4), or apoplastic (5). Long-distance transport occurs in the phloem. Signaling molecules (e.g., RNA) from source tissues are transported from the companion cell (CC) via plasmodesmata into the sieve element (SE). The RNA can travel through the phloem sieve tube freely or complexed with an RNA binding protein (RBP-RNA). Once the signal reaches the sink cells it can act as a positive or negative regulator, for example. Image created with BioRender.

1.5.2.1. The phloem as a highway for mobile RNAs

Phloem sap has a high sugar content, generally in the form of sucrose. Nonetheless, polyols and oligosaccharides of the raffinose family (RFOs) are also present in the sap[74]. Besides sugars, a myriad of molecules can be found in phloem, such as, inorganic molecules[75,76], phytohormones[77], amino acids[78,79], lipids[80], proteins[81–83] and RNAs[81,84,85].

All classes of RNAs have been found in the phloem sap, ranging from protein coding RNAs (mRNAs) to non-coding RNAs (ncRNAs), such as micro RNAs (miRNAs), small interfering RNAs (siRNAs) − so-called small ncRNAs (sncRNAs)−, and housekeeping RNAs, like transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs)[84]. Another class that gathered more attention in recent years are the long ncRNAs (lncRNAs)[86–88].

Protein coding viral RNAs in virus-infected plants provided the first evidence of intercellular RNA transport via PD, as well as its spreading through the phloem from source to sink tissues[84]. Therefore, non-autonomous endogenous RNAs can also reach the CC-SE complex via PD. The loading/unloading and transport of RNAs may be facilitated by binding proteins (RBPs), forming ribonucleoprotein (RNPs) complexes **(Figure 1.6)**[89,90]. When endogenous RNA was first discovered in the phloem sap, it was initially thought to be a contamination from CCs and PPCs[89]. Nevertheless, quantitative analysis indicates that many transcripts in the phloem sieve tube are highly enriched relatively to the surrounding vascular tissues, and comparative analysis shows that the vascular bundle transcriptomes are almost identical, whereas phloem sap transcriptomes only overlap about 50% of the transcripts[89].

Several thousand of potential mobile RNAs have been identified[91–95], but only some have been studied in detail regarding their signaling function[84]. Examples of phloem-mobile RNAs that act as signals are the *AtTCTP1* mRNA, known to stimulate the growth of lateral roots in primary roots, and miR399, that regulates phosphate homeostasis during phosphate starvation[96,97].

Subsequently, two questions arise: what is the origin of these RNAs and are all potential mobile RNAs signals? Mature SE are enucleated (**chapter 1.5.2**), hence present RNAs cannot be a product of transcription. Interestingly, even though RNAs and proteins involved in translation are present, there is no evidence of protein synthesis occurring in the sieve tubes[16,84]. Thus, RNAs and proteins must originate from neighboring CCs. Even though it cannot be ruled out that they are remnants of former intact SEs, several observations raise doubts that all mobile RNAs are leftovers[90]. The answer to the second question is not simple. As previously mentioned, the number of potential mobile RNAs is high. Kehr *et al*. 2022, proposed several criteria for qualifying a mobile RNA as a signaling molecule, such as: displaying population changes over time and in response to stimuli; being produced in source tissues, present in phloem and unloaded in target tissues; and exhibit functionality after transport (must decode some kind of process, e.g., negative/positive regulation)[90].

1.6. Long non-coding RNAs in plants

About 90% of the eukaryotic genome is transcribed, yet only approximately 2% of these transcripts are translated into proteins[98,99]. Once regarded as transcriptional noise, today ncRNAs are known to be key players in a plethora of cellular regulatory networks. As the name suggests, these transcripts have little to no coding potential. Their sizes can range from 20-30 nucleotides (sncRNAs) to >200 nucleotides (lncRNAs)[100]. Like mRNAs, lncRNAs can have 5'-end capping and 3' poly-A tails, can go through alternative splicing and they may have ORFs with potential to encode oligopeptides[101]. A vast majority of the knowledge about lncRNAs comes from studies in the animal kingdom, nevertheless, in recent years, plant lncRNAs have been receiving more attention.

Regarding their biogenesis, most plant lncRNAs are transcribed by RNA Pol II[102]. Plants have two additional polymerases, Pol IV and Pol V, that can also transcribe lncRNAs. Pol IV- and Pol V-dependent lncRNAs exhibit structural differences, such as a lack of poly-A tails, and they participate in the RNA-mediated DNA methylation (RdDM) pathway[103]. Pol IV transcripts are siRNAs precursors, whilst Pol V and some Pol II transcripts act as sRNA targets[100,103,104].

Three major classes of lncRNAs are broadly accepted with respect to their genomic location and/or transcription direction: intergenic, intronic and natural antisense transcripts (NATs) (**Figure 1.7A**)[100,102]. Further classes of lncRNAs have been identified in other organisms, such as promoter- and enhancer-related lncRNAs[102]. LncRNAs can regulate neighbor (cis) or distant (trans) genes, and their interactome comprises nucleic acids and proteins[98]. Generally, they can act as decoys or sponges, scaffolds, precursors, guides or signals to modulate transcription, translation or epigenetic modifications of their target genes (**Figure 1.7B)**[98,103,105].

Plant lncRNAs are known to modulate a variety of processes, such as root nodulation, stress responses, flowering, etc[101]. For example, the major repressor of flowering, FLOWERING LOCUS C (FLC), is repressed after a long period of low temperature known as vernalization[106]. This is accomplished by epigenetic modifications at the FLC chromatin, such as the repressive histone mark H3K27me3, mediated by the Polycomb Repression Complex 2 (PRC2)[99]. In Arabidopsis, three lncRNAs are involved in this repressing process. COLDAIR and COLDWRAP are sense lncRNAs transcribed from the first intron and promoter regions of the FLC locus, respectively[107,108]. Both have been shown to interact with PRC2 and their knockdown resulted in reduced vernalization response[107,108]. COOLAIR is a set of NATs implicated in early vernalization[104]. In contrast to COLDAIR and COLDWRAP, COOLAIR knockdown did not affect the vernalization response[109]. Moreover, it does not interact with PRC2, and seems to play a role in a PRC2-independent repression mechanism involving the reduction of H3K36me3, an antagonist of H3K27me3[110]. Conversely, a recent study showed that COOLAIR is not strictly necessary for successful vernalization[111]. In *B. napus*, lncRNAs have been shown to be involved in seed oil biosynthesis, and also to be differentially expressed in response to stresses, such as cold and drought[112–114]. Besides acting cell-autonomously, lncRNAs can also participate in long-distance communication. In cucumber, lncRNAs have been reported to be enriched and move through the phloem in response to phosphate deficiency[87]. In apple, genome-wide studies allowed tissue profiling of lncRNAs, including the phloem[86,115]. The long non-coding profile of phloem sap in other species, such as in *B. napus*, remains to be explored.

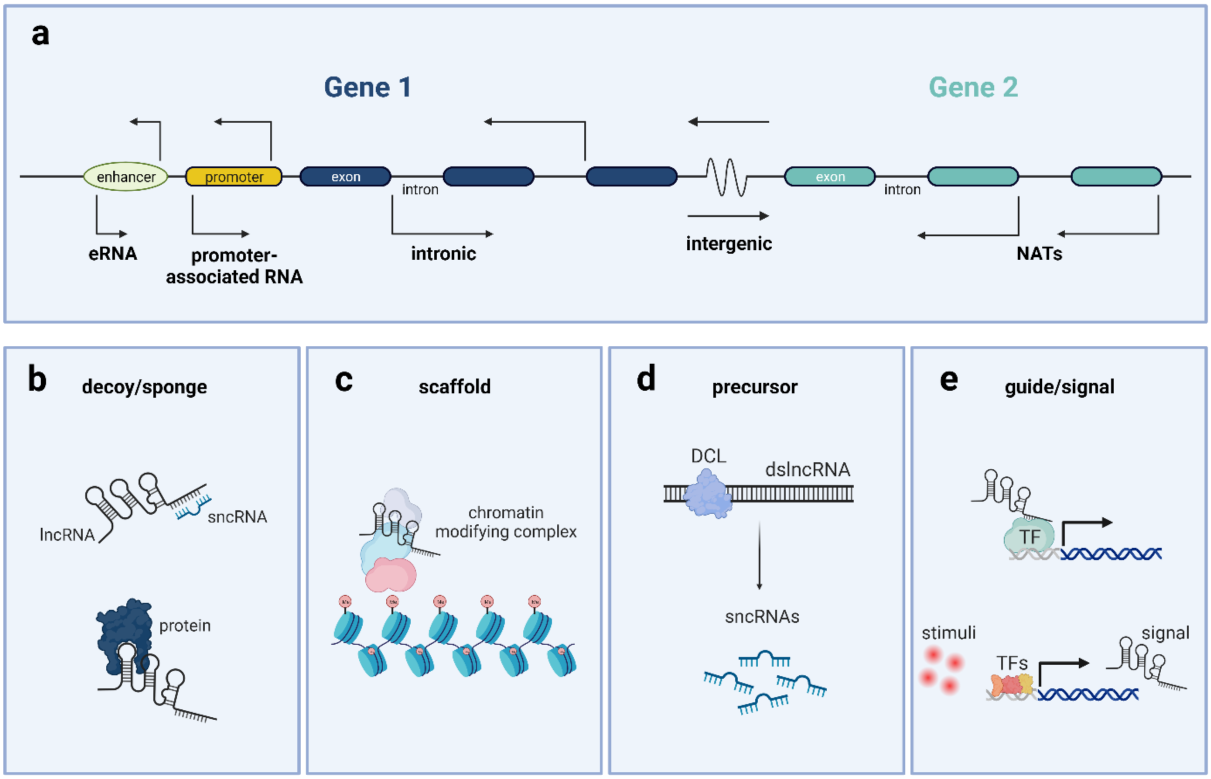


Figure 1.7- Classification of lncRNAs based on their genomic location and their functions.
(**a**) LncRNAs classes according to their genomic location. These include intronic or intergenic, promoter associated and enhancer RNAs (eRNA). Furthermore, they can be divided giving their transcription direction into sense or anti-sense, including natural anti-sense (NATs) lncRNAs. They can act as (**b**) decoys or sponges, (**c**) scaffolds, (**d**) precursors and/or (**e**) guides or signals. Created with Biorender.

1.7. Aims

As plants are sessile, they had to develop different mechanisms to ensure a high level of plasticity in order to control development and to adapt to different environmental changes. As mentioned in **chapter 1.2**, one way to achieve this is by controlling their proteome, relying on the balance between protein synthesis and degradation. The former is governed by ribosomes. Ribosome biogenesis is a pivotal process in all eukaryotes that counts with a multitude of maturation steps aided by numerous factors, such as RBFs. Many yeast RBF orthologs have been identified in *Arabidopsis*[37]. The yeast Nop15 protein is an RBF involved in the processing of polycistronic pre-rRNA. The absence of Nop15 led to impaired ribosome synthesis[45]. Moreover, the Nop15 interactome has been characterized and its structure has been resolved by Cryo-EM and X-ray crystallography[49,55,57]. However, the plant ortholog remains to be characterized. Unlike yeast, most plant RBFs lack structural and/or functional characterization. This characterization is vital to better understand the molecular mechanisms behind ribosome biogenesis in plants. Therefore, the first aim of this work was to characterize structural- and functionally the Nop15 ortholog in Arabidopsis, hereafter named NUCLEOLAR RNA CHAPERONE-LIKE 1 (NURC1). The structural characterization was done by the aid of techniques such as Dynamic Light Scattering (DLS), Small-Angle X-ray (SAXS) and Multi-Angle Light Scattering (MALLS). The functionality as chaperone and the binding capacity of NURC1 were studied by emission spectroscopy and MicroScale Thermophoresis (MST), respectively. Furthermore, truncated versions of NURC1 were used to explore the importance of different secondary structures.

Another adaptation mechanism involves long-distance communication of signaling molecules, such as RNAs (**chapter 1.4**). All classes of RNAs were found in the phloem sap[84]. Even though many have shown mobility, only a few have been linked to signaling roles. Recently, lncRNAs have gained more attention because of their regulatory functions in various cellular networks. Many of the studied plant lncRNAs are tissue specific and act in a cell-autonomous manner. Fewer studies addressed the non-cell autonomous roles of lncRNAs. Nonetheless, lncRNAs have been reported to be enriched in phloem sap and to move through the vascular system under stress conditions[87]. The lncRNA profile of phloem sap was determined in apple trees[86,115]. *B. napus* phloem sap lncRNA content remains to be explored. Hence, the second objective of this work was to identify lncRNAs in phloem sap of *B. napus*. Its phloem can be easily collected by puncturing the stem, which results in spontaneous exudation of fair amounts of sap. Total RNA was extracted from phloem samples and the RNA integrity checked by high-resolution automated electrophoresis. The lncRNA composition was determined by Next Generation Illumina Sequencing.

Both projects provide valuable insights into the roles of ncRNAs and their interaction partners in diverse plant cellular processes of two important plants − the model plant *A. thaliana* and the economically significant crop plant *B. napus*− that could further be applied in the biotechnological improvement of crops.

2. Material and Methods

2.1. Materials

2.1.1 Plants

A. thaliana Col-0 ecotype (AG Hoth, Universität Hamburg) and *B. napus* cv. *Drakkar* (AG Kehr, Universität Hamburg) were used in this work. *A. thaliana* was grown at 22°C, 70% humidity and short-day conditions in a Percival SE-41 (CLF Plant Climatics, Wertingen). *B. napus* (cv. *Drakkar*) were sown in Ø18 cm pots containing a soil-sand mixture of 3:1 and were grown at 15°C, 70% humidity and long-day conditions in a greenhouse.

2.1.2 Bacteria

Escherichia coli XL10 Gold (Agilent, Waldbronn) strain was used for initial transformation of the cloned plasmids for long term storage. *E. coli* strain BL21-CodonPlus (DE3)-RIPL (Agilent, Waldbronn) was used for the protein expression. Glycerol stocks were prepared with 30% glycerol and stored at -80°C.

2.1.3 Plasmids

Three different plasmids were used in this work and are listed in **Table 2.1**. Their maps can be found in **Figure 7.1** and **Figure 7.2**.

Table 2.1- List of the used plasmids. Plasmid name and manufacturer are shown.

Plasmid	Manufacturer
pET28a(+)	Merck, Millipore, Darmstadt (DE)
pET28a(+) GoldenGate cloning	Prepared in house
pUC57	Genescript, Rijswijk, (NL)

2.1.4 Chemicals and Oligonucleotides

All chemicals were purchased from AppliChem (Darmstadt, DE), Merck (Darmstadt, DE), Roth (Karlsruhe, DE), Serva (Heidelberg, DE) and Sigma-Aldrich (Taufkirchen, DE) unless stated otherwise. All oligonucleotides were synthesized by Eurofins Genomics (Ebersberg, DE) and they are listed in **Table 7.1**.

2.1.5 Buffers

The composition of the buffers used in this work is listed in **Table 2.2**.

Table 2.2- List of commonly used buffers. Buffer name and its components are shown.

Buffer	Components
CTAB buffer	3% (w/v) Cetrimonium bromide (CTAB) 0.5% (w/v) Polyvinylpyrrolidone 100 mM Tris-HCl pH 8.0 1.4 M NaCl 20 mM EDTA 2% (v/v) DTT
Lysis buffer	50 mM HNa_2PO_4 pH 7.5 200 mM NaCl 30 mM imidazole 5% glycerol (v/v) 1 M DTT 1 mM PMSF
Washing buffer	50 mM HNa_2PO_4 pH 7.5 1 M NaCl 30 mM imidazole 5% glycerol (v/v) 1 M DTT 1 mM PMSF
Elution Buffer	50 mM HNa_2PO_4pH 7.5 200 mM NaCl 1 M imidazole 5% glycerol (v/v) 1 M DTT 1 mM PMSF
Dialysis buffer	50 mM HNa_2PO_4 pH 7.5 200 mM NaCl 5% glycerol (v/v) 1 mM DTT 1 mM EDTA
SEC running buffer	50 mM HNa_2PO_4 pH 7.5 300 mM NaCl 5% glycerol (v/v) 1 mM DTT

Transcription buffer	50 mM Tris–HCl pH 7.5 15 mM $MgCl_2$ 5 mM DTT 2 mM spermidine
MTS buffer	20 mM Tris–HCl pH 8.0 150 mM NaCl 0.1 mg/ml BSA 1 mM DTT 0.1% Tween-20 (v/v).
SD-mix	4% SDS 40 mM DTT
Chaperone buffer	50 mM HEPES pH 7.5 3 mM $MgCl_2$ 1 mM DTT

2.1.6 Consumables

Consumables, such as pipet tips and tubes, were acquired from Sarstedt (Sarstedt, DE) and Eppendorf (Hamburg, DE). Other consumables are listed in **Table 2.3**.

Table 2.3- List of consumables. All consumables used in this work are listed together with the manufacturer information.

Product	Manufacturer
45 µm filtropur S	Sarstedt (Sarstedt, DE)
96-well flat black plate	TECAN (Männedorf, CH)
Agilent RNA 6000 Nano Kit	Agilent (Santa Clara, USA)
Complete Protease Inhibitor	Roche (Mannheim, DE)
GelRed DNA Stain	Biotium (Hayward, USA)
GeneRuler 1 kb plus DNA ladder	Thermo Fisher Scientific (Darmstadt, DE)
Gibson Assembly® Master Mix	New England Biolabs (Frankfurt am Main, DE)
HiLoad™ 16/600, Superdex™ 200 pg	GE Healthcare (Uppsala, SE)
HisTrap HP	GE Healthcare (Uppsala, SE)
Kimtech Science Precision Wipes Tissue	Kimberly-Clark Professional (Koblenz-Rheinhafen, DE)
Monolith NT.115 Standard treated capillaries	NanoTemper (Munich, DE)
NucleoSpin Gel and PCR Clean up kit	Macherey&Nagel (Düren, DE)
NucleoSpin Plasmid easy pure kit	Macherey&Nagel (Düren, DE)
PageRuler prestained 10 to 180 kDa	Fermentas (St. Leon-Rot, DE)
Parafilm M	Brand (Wertheim, DE)

RevertAid First Strand cDNA synthesis kit	Thermo Fisher Scientific (Darmstadt, DE)
RNA Clean & Concentrator-25	Zymo research (Freiburg, DE)
Superdex® 75 Increase 10/300 GL	GE Healthcare (Uppsala, SE)
Terazaki 72-well plate	Greiner Bio-one (Kremsmünster, AT)
TRIzol®	Thermo Fisher Scientific (Darmstadt, DE)
TRIzol® LS	Thermo Fisher Scientific (Darmstadt, DE)
Vivaspin concentrator	Sartorius (Göttingen, DE)

2.1.7 Enzymes

Table 2.4 shows the enzymes used in the different protocols.

Table 2.4- List of the enzymes. Enzyme name and manufacturer are shown.

Enzyme	Manufacturer
BsaI	New England Biolabs (Frankfurt am Main, DE)
DNase I	Thermo Fisher Scientific (Darmstadt, DE)
DpnI	New England Biolabs (Frankfurt am Main, DE)
FastAP alkaline phosphatase	Thermo Fisher Scientific (Darmstadt, DE)
HindIII	New England Biolabs (Frankfurt am Main, DE)
Inorganic pyrophosphatase	New England Biolabs (Frankfurt am Main, DE)
Lysozyme	AppliChem (Darmstadt, DE)
NdeI	New England Biolabs (Frankfurt am Main, DE)
Phusion high fidelity polymerase	Thermo Fisher Scientific (Darmstadt, DE)
RiboLock	Thermo Fisher Scientific (Darmstadt, DE)
RNase A	Carl Roth (Karlsruhe, DE)
T4 DNA Ligase	Thermo Fisher Scientific (Darmstadt, DE)
T7 polymerase	In house
Thrombin	GE Healthcare (Uppsala, SE)
Trypsin	Promega (Madison, US)
XhoI	Thermo Fisher Scientific (Darmstadt, DE)

2.1.8 Equipment

The equipment that was used to perform the experiments is listed below (**Table 2.5**).

Table 2.5- Equipment list. All equipment used in this thesis is listed together with the manufacturer information.

Equipment	Manufacturer
-20 °C freezer	Liebherr (Biberach, DE)
4 °C fridge	Liebherr (Biberach, DE)
-80 °C freezer	Binder (Tuttlingen, DE)
2000c UV/vis spectrophotometer	Thermo Fisher Scientific (Darmstadt, DE)
ÄKTA prime plus	GE Healthcare Bio-Sciences AB (Uppsala, SE)
Balance	Kern & Sohn GmbH (Balingen, DE)
Bioanalyzer 2100	Agilent (Santa Clara, USA)
Branson Sonifier 250	Branson Ultrasonics (Eemnes, NL)
Centrifuge Beckman Coulter Avanti JXN-30	Beckman Coulter Diagnostics (Krefeld, DE)
Centrifuge Hettich ROTINA 380R	Hettich (Tuttlingen, DE)
Centrifuge MiniSpin	Eppendorf (Hamburg, DE)
Centrifuge Sigma 1-16 k	Sigma (Osterode am Harz, DE)
Chemidoc Touch Gel	BioRad (Munich, DE)
Heating block Thermomixer comfort	Eppendorf (Hamburg, DE)
Monolith NT.115™	NanoTemper (Munich, DE)
NanoDrop One C	Thermo Fisher Scientific (Darmstadt, DE)
Peristaltic pump 2232 Microperplex S	Pharmacia LKB (Uppsala, SE)
pH-meter	Mettler-Toledo (Gießen, DE)
Precision balance ABJ	Kern&Sohn GmbH (Balingen, DE)
Shaker HT CH-4103	Infors HT AG (Bottmingen, CH)
SpectroLight 600	Xtal concepts (Hamburg, DE)
TECAN SPARK®	TECAN (Männedorf, CH)
Thermocycler PCR cycler T3000	Biometra (Göttingen, DE)
ThermoShaker TS1	Biometra (Göttingen, DE)
Ultraflex III MALDI-TOF-TOF	Bruker (Bremen, DE)
Vacuum pump	Aeromat KNF (Freiburg, DE)
Vortexer VF2	IKA® Labortechnik (Staufen, DE)

2.1.9 Bioinformatic tools and databases

All the software and databases used for data analysis are listed in **Table 2.6**.

Table 2.6- List of the used software and databases. All programs and databases used for the data analysis are shown.

Software	Publisher
Agilent 2100 Expert Software	Agilent (Santa Clara, USA)
ASTRA 7 software package	Wyatt Technology Corporation (Santa Barbara, USA)
ATSAS 3.0 package tools	ref.[116]
Bioconductor/DESeq2	ref.[117]
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/, ref.[118]
CNCI	ref.[119]
CPC	ref.[120]
Cufflinks: Cuffcompare and Cuffmerge	ref.[121]
EMBOSS water	https://www.ebi.ac.uk/Tools/psa/emboss_water/
EnsemblPlants	http://plants.ensembl.org/index.html
Fastp	ref.[122]
FlexControl	Bruker Daltonik (Bremen, DE)
g:Profiler	https://biit.cs.ut.ee/gprofiler/gost, ref.[123]
HISAT2	ref.[124]
HTSeq	ref.[125]
IGV 2.16.2	ref.[126]
iTasser	http://www.zhanggroup.org/I-TASSER/
MASCOT server	http://www.matrixscience.com/
mMass 5.5.0	http://www.mMass.org/
MO.Affinity Analysis Software version	NanoTemper (Munich, DE)
OriginPro 2021b	OriginLab (Northampton, USA)
Pfam	ref.[127]
ProtParam	http://www.web.expasy.org/protparam/
PyMOL 2.5.4	ref.[128]
R/RStudio 4.3.1	ref.[129]
SASpy	ref.[130]
SnapGene 5.2.3	SnapGene (San Diego, USA)
StringTie	ref.[131]

2.2. Methods

2.2.1 Nucleic acid extraction

2.2.1.1. RNA extraction

Total RNA was extracted from *A. thaliana* leaves. The plant material was frozen in liquid nitrogen and ground in a pre-cooled mortar. Approximately 100 mg of ground tissue was mixed with TRIzol® according to the manufacturer's instructions. After the chloroform phase separation, the upper aqueous phase was collected, and RNA washing, and precipitation was performed by RNA Clean & Concentrator-25 kit following the manufacturer's protocol.

2.2.1.2. gDNA extraction

Genomic DNA (gDNA) extraction was performed by following a general CTAB protocol. *A. thaliana* leaves were frozen in liquid nitrogen and consequently ground in a pre-cooled mortar. Approximately 100 mg of ground leaf tissue was used and 500 µl of CTAB buffer was added to it. The mix was homogenized by vortex, followed by 30 min of incubation at 60°C, mixing it every 10 minutes. The tubes were centrifuged for 5 min at 14 000 x g. The clean supernatant was transferred to a new tube and 10 µl of RNase A solution (10 mg/ml) was added. RNA digestion was performed at 37°C for 30 min. An equal volume of chloroform/isoamyl alcohol (24:1) was added, vortexed for 5 seconds, and then centrifuged for 1 min at 14 000 x g, to separate the phases. The aqueous upper phase was transferred to a new tube and the process was repeated until the upper phase was clear. Consequently, the aqueous phase was transferred to a new tube and the DNA precipitated by adding 0.7 volume cold isopropanol and was left at -20°C for 15-30 min (overnight is also possible for higher yield). The sample was centrifuged at 14 000 x g for 10 min. The supernatant was discarded, and the pellet washed with 500 µl of ice-cold 70% ethanol. The ethanol was decanted and the pellet air dried under a fume hood. The pellet was resuspended in double distilled water and stored at -20°C until use.

2.2.2 NURC1 cloning

2.2.2.1. Traditional cloning of NURC1 full-length

To obtain NURC1 full-length protein, total RNA was extracted from *A. thaliana* leaves and reverse transcribed by Reverse Transcription-Polymerase Chain Reaction (RT-PCR), with the RevertAid First Strand cDNA synthesis kit. The cDNA of full-length *NURC1* (AT5g04600) was cloned into pET28a(+) using *NdeI* and *XhoI* restriction sites by Dr. Anna Ostendorp (Universität Hamburg, Germany). Further, ligation was performed using T4 DNA ligase, and the ligation mix was transformed into *E. coli* by heat shock. The primers can be found in **Table 7.1** and the construct map in **Figure 7.3**.

2.2.2.2. Site-directed mutagenesis of NURC1$_{1\text{-}160}$ and NURC1$_{1\text{-}140}$

NURC1$_{1\text{-}160}$ and NURC1$_{1\text{-}140}$ were obtained by adding a stop codon at positions 140 and 160 through site directed mutagenesis (SDM) with Phusion polymerase. The typical reaction mixture can be found in **Table 2.7**, and the PCR program steps in **Table 2.8**. *DpnI* digestion was performed at 37°C for 1h for the removal of the template plasmid. The mutated construct was then transformed into *E. coli* by heat shock. All primers can be found in **Table 7.1** and the construct map in **Figure 7.3**.

Table 2.7- SDM reaction mixture. The components are listed together with their final concentration.

Component	Final concentration
HF or GC buffer	1X
dNTPs	200 µM
Primer mix	0.5 µM each primer
Template DNA	5-50 ng
Phusion DNA Polymerase	0.04 U/µl
ddH$_2$O	-

Table 2.8- PCR program for SDM.

Cycle step	Temperature (°C)	Time	Cycles
Initial denaturation	98	5 min	1
Denaturation	98	10 sec	16
Annealing	60-65	30 sec	
Extension	72	30 sec/kb	
Final Extension	72	10 min	1

2.2.2.3. Golden Gate cloning of NURC1$_{53\text{-}222}$

NURC1$_{53\text{-}222}$ was amplified from pET28a(+)-NURC1$_{FL}$ with designed primers that included *BsaI* recognition sites. The amplicon was inserted into a modified version of pET28a(+) through Golden Gate cloning with *BsaI*. The Golden Gate compatible pET28a(+) was designed and modified by Dr. Kim Lühmann (Universität Hamburg, Germany), and its map can be found in **Figure 7.2B**.

The primer sequences can be found in **Table 7.1** and the construct map in **Figure 7.4.** Typical concentrations are shown in **Table 2.9** and the PCR program in **Table 2.10**. Finally, the plasmid was transformed into *E. coli* by heat shock.

Table 2.9- Golden Gate reaction mixture. The components are listed together with their final concentration.

Component	Final concentration
Plasmid	100-200 ng
Insert	3-5x molar excess to plasmid
10x restriction enzyme buffer or T4 ligase buffer (e.g., CutSmart)	1X
10 mM ATP	0.5 mM
T4 Ligase (5 U/µl)	0.25 U/ µl
BsaI (10 U/µl)	0.5 U/ µl
ddH_2O	-

Table 2.10- PCR program for Golden Gate cloning.

Temperature (°C)	Time	Cycles
37	2 min	50
16	5 min	
37	60 min	1
50	10 min	1
80	10 min	1

2.2.2.4. Gibson Assembly of 45S rDNA

To produce ITS2 RNA, gDNA was extracted from *A. thaliana* by following the CTAB-Phenol/Chloroform extraction protocol (**chapter 2.2.1.2**). 45S rDNA was amplified with *HindIII* overhangs, gel extracted and purified using the NucleoSpin Gel and PCR Clean-up kit. The PCR fragment was cloned into pUC57 using the Gibson Assembly cloning method (**Table 2.11**). pUC57 was previously linearized by digestion with *HindIII* at 37°C for 1h, followed by the inactivation step at 80°C for 20 min. To avoid plasmid re-ligation, an alkaline phosphatase treatment was conducted at 37°C for 1h, followed by 20 min at 70°C for enzyme inactivation. Insert and plasmid were incubated at 50°C for 1h. The presence of ITS2 was confirmed by Sanger sequencing using designed primers. All primers are listed in **Table 7.1** and the construct map in **Figure 7.5**. The primer design of 45S rDNA and ITS2 sequences were based on the nucleotide sequences reported by Unfried and Gruendler[132].

Table 2.11- Gibson assembly reaction mixture. The components are listed together with their final concentration.

Component	Final concentration
Plasmid	100-200 ng
Insert	3-5x molar excess to plasmid
Gibson Assembly Master Mix	1X
ddH_2O	-

2.2.3 Protein expression and purification

The expression of NURC1 full-length and truncated versions was carried out at 24°C overnight in the *E. coli* strain BL21-CodonPlus (DE3)-RIPL by autoinduction. Harvesting of the cells was achieved by centrifugation at 4°C and 5000 rpm for 30 min (JA-10 rotor, Avanti JXN-30 Beckman Coulter). The pellets were resuspended in lysis buffer (**Table 2.2**), and protease inhibitor (1 pill, Roche), RNase A (q.s.) and DNaseI (q.s., PanReac AppliChem) were added to the resuspension. The lysis step was performed by adding 1 mg/ml of lysozyme to the mixture and incubating at 4°C, for 40 min with stirring. For complete cell rupture, sonication was employed (8-10 times 30 s on + 30 s off, 45% duty cycle, output control 5, Branson sonifier 250). The cell debris was removed through centrifugation at 4°C and 14 000 rpm for 30 min (JA-25.50 rotor, Avanti JXN-30 Beckman Coulter). The supernatant was filtered with a 45 µm filtropur S. Subsequently, immobilized matrix assisted chromatography (IMAC) was conducted using a 5 ml HisTrap HP. After immobilization, the column was washed with 50 ml of washing buffer (**Table 2.2**). The elution was carried out with a linear gradient from 30 mM to 1 M of imidazole on an Äkta system (Äkta prime plus). The next step involved the removal of the excessive imidazole and the 6xHis-tag. For that, dialysis and tag cleavage were combined and performed overnight at 4°C. Dialysis buffer was used in this step (**Table 2.2**), and thrombin (1 U/ml) was used for the tag proteolysis. After dialysis, the solution was centrifuged to remove any larger particles and concentrated. Further purification was performed by size exclusion chromatography (SEC) using a HiLoad™ 16/600, Superdex™ 200 pg column at 4 °C pre-equilibrated with running buffer (**Table 2.2**). The fraction purity was checked by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), and the protein identity was confirmed by Matrix Assisted Laser Desorption Ionization/Time of Flight (MALDI-

TOF) mass spectrometry (MS). Finally, the purest fractions were combined, concentrated, and stored at -80°C until use.

2.2.4 *In vitro* transcription and RNA labeling

Using pUC57-45S rDNA as template, full length ITS2 was amplified by touchdown PCR with Phusion polymerase, using a forward primer that contained the T7 promotor sequence. The amplicon corresponding to the ITS2 RNA size (187 nucleotides) was gel extracted and purified by NucleoSpin® Gel and PCR Clean-up kit. Since the ITS2 fragments were smaller than 100 nucleotides, two complementary DNA oligonucleotides containing the T7 promoter sequences at the 5’ ends were used for each fragment instead. They were annealed at 95°C for 5 min, and were let to cool down at room temperature for 10 min. The primers are listed on **Table 7.1**.

In vitro transcription was carried out either for 2-3h at 37°C and covered from light because of the presence of cy5 labeled UTP. The reaction mixture can be found in **Table 2.12**. The remaining DNA template was removed by DNaseI digestion. The final transcripts were purified with the RNA Clean & Concentrator kit. RNA concentration and quality were established via nanodrop (Nanodrop one C) and bioanalyzer (Agilent 2100 bioanalyzer).

Table 2.12- *In vitro* transcription reaction mixture. The components are listed together with their final concentration.

Component	Final concentration
DNA template	10 pmol
Transcription buffer	1X
NTPs ATP GTP CTP UTP Cy5-UTP	 2 mM 2 mM 2 mM 1.87 mM 0.13 mM
Inorganic phosphatase	0.005 U/µl
Ribolock	1 U/µl
DMSO	10% (v/v)
T7 RNA polymerase	0.1 mg/ml
ddH_2O	-

2.2.5 Microscale Thermophoresis

The binding affinity between the NURC1 proteins and ITS2 RNAs were determined by MST. The concentration of proteins was in the nano- to micromolar range, while the concentration of fluorescently labelled RNAs was kept fixed at 10 nM. All protein samples were centrifuged at maximum speed (20,000 x g) for 10 min at 4°C. Protein and RNA concentration was determined by absorbance at 280 nm and 260 nm, respectively, using a Nanodrop one C. A series of dilutions was prepared in MST buffer (**Table 2.2**). The protein-RNA mix was spun down and kept at room temperature for 5 min. The measurements were performed in a Nanotemper Monolith NT.115 device (100% red LED and medium MST power), using Monolith NT.115 capillaries. The data analysis was performed with the MO.Affinity analysis software.

2.2.5.1. SDS denaturation test

The SDS denaturation test (SD-test) was used to distinguish if fluorescent changes are caused by an interaction (e.g., binding) or by non-specific effects (e.g., loss of protein to aggregation or adsorption). The protein-RNA mixture was centrifuged for 10 min at maximum speed, using low binding tubes. Subsequently, 15 µl were carefully removed and mixed with 15 µl of SD-mix (**Table 2.2**) into a new low binding tube. All samples were incubated for 5 min at 95°C. The tubes were spun down before loading the samples into the capillaries.

2.2.6 Chaperone Assay

The RNA chaperone activity assay was based on the protocol by Rajkowitsch and Schroeder[133]. The same fluorescently labelled 21-nucleotide oligoribonucleotides were used to test the chaperone capacity of NURC1 proteins. The test was conducted by mixing 10 nM of $21R^+$ strand with 50 nM of protein and incubating at room temperature for 1 min in 50 mM HEPES pH 7.5, 3 mM $MgCl_2$ and 1 mM DTT. Henceforth, 10 nM of $21R^-$ strand was added to the mixture and homogenized by pipetting up and down. The fluorescence measurements were recorded immediately after adding the second RNA strand at 37°C in a 96-well flat black plate with a Spark multimode microplate reader. The data analysis was done in OriginPro 2021b and the data was fitted to an exponential one-phase association model.

2.2.7 Dynamic Light Scattering

DLS measurements were done using a Terazaki 72-well plate covered with paraffin oil to expose the sample solution drops in a SpectroLight 600 Instrument. A red-light diode laser at 660 nm was used and the scattered light was detected by a photomultiplier tube at an angle of 142°. All protein samples were centrifuged at maximum speed for 10 min at 4°C. The protein concentration was determined by absorbance at 280 nm via nanodrop (2000c UV/vis spectrophotometer). All samples were kept on ice until measured. The scattering measurements were taken 10 to 20 times for 10 s at 20°C in 5 µl drops. Scattering curves were averaged and the hydrodynamic radii distribution was determined based on the diffusion constants using the Stokes-Einstein equation.

2.2.8 Small-Angle X-ray Scattering

SAXS measurements were performed in batch mode, as well as coupled to SEC (SEC-SAXS) at the synchrotron beamline P12 at the PETRA III (DESY, Hamburg, Germany). Batch measurements were recorded by Dr. Yunyun Gao (CFEL, DESY, Hamburg, Germany). SEC-SAXS measurements were recorded by Dr. Sven Falke (CFEL, DESY, Hamburg, Germany). The samples were centrifuged at maximum speed for 10 min at 4°C. The SEC run was done in a Superdex® 75 Increase 10/300 GL column with a flow rate of 0.5 ml/min. The data was normalized to the intensity of the transmitted beam, radially averaged, and the scattering of the solvent-blank subtracted. For the batch mode the background subtraction was accomplished by doing a separate measurement of merely the buffer and for the SEC mode by using averaged frames from before and after the elution peak. The measurements were carried out at room temperature. Batch measurements were collected by 20 successive 0.045 s frames that were averaged. SEC-SAXS measurements were collected continuously with 1 s frames. Data processing and modelling were done utilizing ATSAS 3.0 package tools[116]. The beamline configuration and sample parameters are listed below (**Table 2.13**).

Table 2.13- SAXS and sample parameters. Detailed information about the batch and the SEC-SAXS experiments is shown. Table from ref.[136].

Sample information	Organism	*A. thaliana*
	Source	In house purification
	UniProt ID	Q9LZ65
	Calculated Extinction Coefficient [A280, 0.1%(w/v)]	0.459
	"v" ¯from chemical composition (cm3.g-1)	0.743
	Particle contrast from sequence and solvent constituent	2.727 (12.267-9.540)
	M from chemical composition (Da)	25262
SEC parameters	SEC column	Superdex 10/300 G
	Loading concentration	18.0 mg/ml
	Injection volume	87µl
	Flow rate	0.5 ml/min
	Solvent	50 mM HNa_2PO_4 pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT
SAXS parameters	X-ray source	PETRA III (DESY, Hamburg, Germany), Beamline P12
	Detector	Photon counting Pilatus 6M (423.6 x 434.6 mm^2)
	Wavelength (Å)	1.24
	Detector distance (m)	3
	Focal spot (mm)	0.2 x 0.12
	Exposure time (sec)	Batch: 0.045 SEC-SAXS: 0.995
	Sample temperature (°C)	20

2.2.9 Multi-Angle Laser Light Scattering

Multi-Angle Laser Light Scattering (MALLS) was coupled with SEC to determine the homogeneity and size of NURC1 proteins. All samples were centrifuged at maximum speed for 10 min at 4°C before SEC. A Superdex® 75 Increase 10/300 GL column was used for the SEC run with a flow rate of 0.5 ml/min. SEC-MALLS measurements were performed at P12 beamline with an on-line UV-vis/MALLS/QELS/RI system. Sample scattering was recorded with a miniDAWN® TREOS® MALLS detector, with an in-built Quasi-Elastic Light Scattering (QELS) module and an Optilab T-rEX (RI) refractometer. The light scattering of toluene was used to calibrate the MALLS system. The molecular weight was determined by taking three-angle MALLS scattering intensities combined with the protein concentration given by RI using the ASTRA 7 software package. The QELS module was used for hydrodynamic radius (R_h) determination. The solvent viscosity was corrected to 5% (v/v) glycerol.

2.2.10 MALDI-TOF MS

MALDI-TOF MS was performed to confirm the identity of the proteins. First, the protein bands were excised out of the SDS-PAGE gel and transferred into low binding tubes. Gel pieces were washed by incubating it in 100 µl of NH_4CO_3 for 5 min. After, the solvent was discarded and a 1:1 mixture of NH_4CO_3 and acetonitrile was added for destaining. Destaining was achieved by incubating the gel pieces at 37°C with occasional mixing until no color was visible. The solvent mix was removed, 100 µl of acetonitrile were added and incubated for 5 min. After incubation the solvent was removed, and the gel pieces were dried under the hood with an open lid. Finally, trypsin digestion was performed by adding 20 µl of trypsin (0.001 µg/µl) and incubating at 37°C for 2h. After digestion, the sample was loaded onto an AnchorChip 600/384 target as follows: 1 µl of TA30 (30% acetonitrile and 0.1% TFA in water) and 1 µl of digested protein sample were pipetted onto the chip and dried, followed by addition of 0.5 µl of HCCA matrix (alpha-cyano-4-hydroxycinnamic acid in TA30), and dried again. Peptide calibration standard (#206195, Bruker Daltronik) was used for calibration. The MS was conducted in an Ultraflex III MALDI-TOF-TOF using 20% laser power in reflection mode. The peptide mass fingerprint was analyzed with the Mmass software. Peak peaking was performed manually. Lastly, the selected peaks were compared to the UP6548_A_thaliana database restricted to *A. thaliana* through the MASCOT server. Other search parameters were included, such as oxidation (M) variable modification and a peptide tolerance of 1 Da.

2.2.11 Sanger Sequencing

Sanger sequencing was performed to confirm the cloning results. The sequencing was carried out at Eurofins genomics (Ebersberg, DE) and Microsynth Seqlab (Göttingen, DE).

2.2.12 Next-Generation Sequencing

2.2.12.1. Phloem and Leaf RNAs

The samples used for Illumina sequencing were prepared by Dr. Kim Lühmann (Universität Hamburg, Germany) as described in her thesis[134]. Shortly, *B. napus* phloem sap was collected using the exudation method. Total RNA was extracted from phloem sap and leaves, and the purity of phloem sap was assessed by RT-PCR. To ascertain that there was no significant contamination from surrounding cells damaged during sampling, two mRNAs were monitored: Thioredoxin h mRNA, which is known to be found in the phloem stream, and Rubisco small subunit that should ideally be absent from phloem samples[16]. RNA integrity was determined by high-resolution automated electrophoresis. The purest samples were sent in triplicates to Novogene (Cambridge, UK), where long non-coding library preparation and Illumina sequencing were conducted. The leaf lncRNAs analyzed in this thesis correspond to the input samples of AtGRP7 CnBr column, and the phloem lncRNAs correspond to the ones used in the circular RNA analysis from Dr. Kim Lühmann (Universität Hamburg, Germany)[134]. Leaf and phloem RNAs were not sampled from the same plants nor the same timepoints.

2.2.12.2. Bioinformatic analysis of lncRNAs

Most of the bioinformatic analysis was conducted by Novogene (Cambridge, UK). The quality control of the reads was done with the Fastp[122] software. The clean reads were then mapped to the reference genome with the Hierarchical Indexing for Spliced Alignment of Transcripts 2 (HISAT2)[124] software. *B. napus* AST_PRJEB5043_v1[135] reference genome was used for the alignment, available in the Ensembl Plants database. StringTie[131] was used to assemble and quantify known and novel transcripts based on the currently available annotation. It was run with '--library-type' settings, while the other parameters were set as default. Cuffmerge[121] was used to merge the transcripts. Furthermore, Cuffcompare[121] was used to classify the assembled transcripts based on the reference annotation. High-Throughput Sequencing (HTSeq)[125] was used to assign the mapped reads to genomic features and they were counted using the union counting mode. To identify lncRNAs, transcripts with low expression level, a single exon, length less than 200 nucleotides, and that were annotated to known genes were filtered out. Furthermore, the coding potential was determined by three tools: Coding Non-Coding Index (CNCI)[119], Coding Potential Calculator (CPC)[120] and Protein Families Database (Pfam)[127]. Only transcripts classified as "non-coding" by the three programs were labeled as non-coding transcripts.

The structural features analysis of lncRNAs was achieved by using the files from Novogene analysis. The plots were created with R/RStudio[129] and the scripts can be found in **Figure 7.6-Figure 7.8**. The differential expression analysis was conducted in house by using the DESeq2 program available in R/Bioconductor. Read counts provided by Novogene were used as input. Transcripts with counts lower than 5 were filtered out. Only common gene IDs between leaf and phloem samples were kept and gene IDs starting with "XLOC" were removed. Isoforms and/or transcripts that shared the same parent gene were merged by summing up their counts. The script used for the analysis can be found in **Figure 7.9**. Finally, the GO term analysis was performed on the g:Profiler[123] webserver by using the differentially expressed gene IDs as input. The search was performed by using *B. napus* as organism and a significance threshold of 0.05 using the Benjamini-Hochberg FDR method. The other options were set as default.

3. Results

One of the aims of this thesis was to analyze the structure and function of the *A. thaliana* RBF NURC1. The first part of this chapter is focused on this topic and covers the purification and characterization of NURC1 structure, as well as its RNA binding capacity. Many of the results presented in this chapter were published in Fernandes *et al.* 2023[136]. The second part comprises the second aim of this thesis: the identification and characterization of lncRNAs in *B. napus* phloem sap.

3.1. Characterization of the NUCLEOLAR RNA CHAPERONE-LIKE 1

3.1.1 Sequence and Structure Comparison

The analysis of available information on NURC1 (UniProt, 14th July 2023) showed that NURC1 was annotated as a nucleolar putative RNA binding protein and had a predicted RRM domain comprised of 78 amino acids located in the middle of its sequence that is flanked by the N- and C-termini. In addition, NURC1 and other plant putative RBFs were experimentally shown to localize in the nucleolus of *N. benthamiana* leaf cells[136].

In a next step, the NURC1 sequence was compared to the sequences of the human (NIFK) and yeast (Nop15) orthologs by performing a multiple sequence alignment (MSA) with Clustal Omega[118] (**Figure 3.1**). The N- and C-terminal regions displayed low conservation between the three proteins, while the middle residues containing the RRM domain appeared to be more conserved. An acid patch rich in aspartic (D) and glutamic acid (E) at the N-terminus seemed to be limited to Nop15. However, a smaller but similar patch was observed in the NIFK C-terminus. Poly-S patches were also observed at the N-terminus of Nop15 and NURC1.

```
Nop15     MVKSTSKTSTKETVTKQPTEEKPIQEKEELALETSSSSSDEEDEKDEDEIEGLAASDDEQ
NIFK      -------------------------------------------------MATF------S
NURC1     -----------------------MGAKAKKALK--------------KNMKKVA---ASA
                                                           :  .

Nop15     SGTHKIKRLNPKK---------QANEKKSKDKKTLEEYSGIIYVSRLPHGFHEKELSKYF
NIFK      GPAGPILSLNPQEDVEFQKEVAQVRKRITQRKKQEQLTPGVVYVRHLPNLLDETQIFSYF
NURC1     SSSQLPLPQNPKPSADFLPLEGGPARKAPVTTPPLQNKATVLYIGRIPHGFYETEIEAFF
          . :      **:             .:    .   :    ::*: ::*: : *.::  :*

Nop15     AQFGDLKEVRLARNKKTGNSRHYGFLEFVNKEDAMIAQESMNNYLLMGHLLQVRVLPKGA
NIFK      SQFGTVTRFRLSRSKRTGNSKGYAFVEFESEDVAKIVAETMNNYLFGERLLECHFMPPEK
NURC1     SQFGTVKRVRVARNKKTGKSKHFGFIQFEDPEVAEIAAGAMNDYLLMEHMLKVHVIEPEN
          :*** :...*::*.*:**:*: :.*::* . : * *.  :**:**:  ::*: :.:

Nop15     KIEKLYKYKKRVL-------VE------KGITKPVKQLKDNMKQKHEERIKKLAKSGIEF
NIFK      VHKELFKDWNIPFKQPSYPSVK--RYNRNRTLTQKLRMEERFKKKERLLRKKLAKKGIDY
NURC1     VKPNLWRGFKCNFKPVDSVQIERRQLNKERTLEEHRKMLQKIVKKDQKRRKRIEAAGIEY
             :*::  :  :       ::      :       :: :.: :*..   *::   **::

Nop15     KW----------------------------------------------------------
NIFK      DFPSLILQKTESISKTNRQTSTKGQVLRKKKKKVSGTLDTPEKTVDSQGPTPVCTPTFLE
NURC1     ECPELVGNTQPVPKRIKFSEE-D-------------------------------------
          .

Nop15     --------------------------------------------------
NIFK      RRKSQVAELNDDDKDDEIVFKQPISCVKEEIQETQTPTHSRKKRRRSSNQ
NURC1     --------------------------------------------------
```

Figure 3.1− MSA of NURC1 and its orthologs, human NIFK and yeast Nop15. The MSA was performed using Clustal Omega. The asterisks (*) indicate fully conserved residues, the colons (:) conservation of residues with similar properties and the periods (.) conservations of residues with weak similar properties. The amino acids are represented in different colors according to the following properties: small and hydrophobic (incl. aromatic Y) in red, acidic residues (D, E) in blue, basic residues (R, K) in magenta, hydroxyl + sulfhydryl + amine + G in green and others in grey. Nop15 acidic patch is highlighted in yellow. Putative acidic cluster of NIFK is highlighted in green. Poly-S patch of Nop15 and NURC1 is underlined in dark green.

The AlphaFold model of NURC1 was available (alphafold.ebi.ac.uk)[137]. The predicted secondary structure harbored a high confidence (score > 90, ranging from 0 to 100) RRM domain between its two terminal regions. The C-terminal tail had two predicted α-helices with some degree of confidence (90 > score > 70). Furthermore, both termini included unstructured regions. The secondary structures of Nop15 and NIFK were determined experimentally by techniques such as

X-ray crystallography, Cryo-EM and NMR[55,57,138]. The available crystal and NMR structures were obtained by truncating the flexible termini, thus losing structural information about these regions. For that reason, Nop15 Cryo-EM and NIFK AlphaFold models were selected to test the structure similarity to the NURC1 AlphaFold model using the Alignment plugin in Pymol instead. The structure alignments and the RRM sequence containing the RNP2 and RNP1 motifs can be found in **Figure 3.2**. Nop15 was used as target and the resulting Root Mean Square Deviation (RMSD) values for NIFK and NURC1 were 0.776 and 0.595 Å, respectively. All three RRMs superimposed to a high degree, whereas the termini did not. Looking at the sequences, the level of conservation is higher in this region, particularly the aromatic residues F105 and F108 in RNP1 and Y63 in RNP2 seem to be conserved across the species.

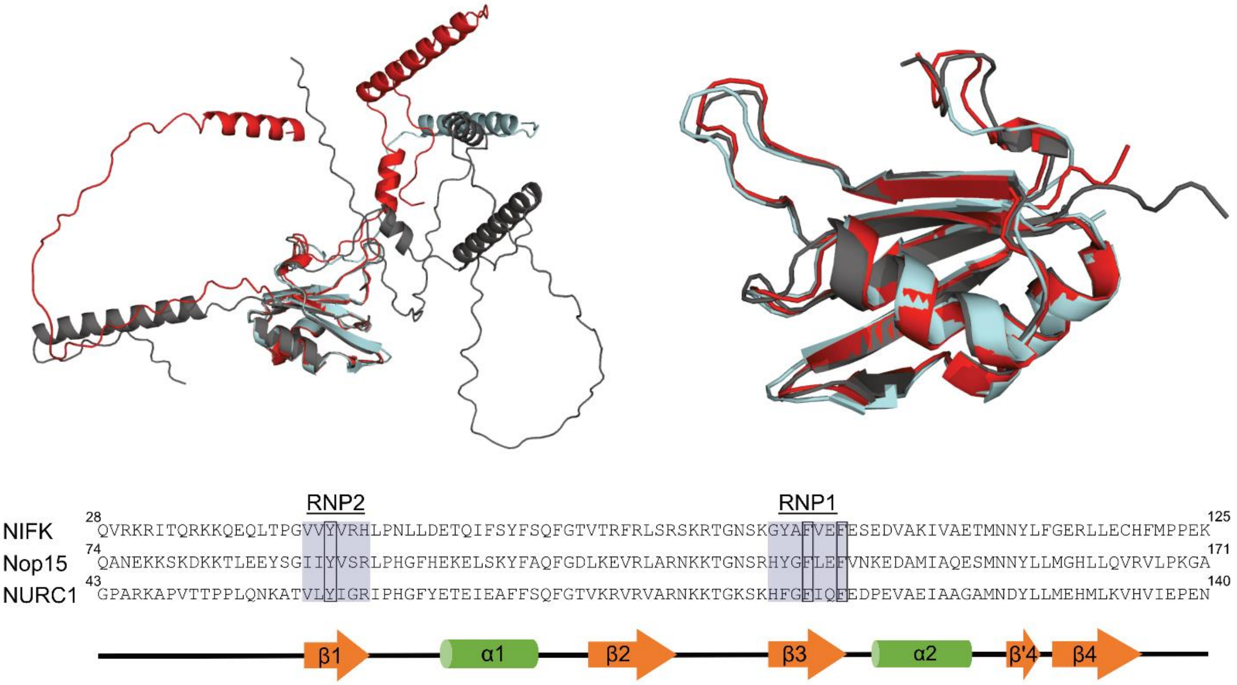


Figure 3.2− The RRM structure and RNPs aromatic residues are highly conserved across species. (**Top, left**) Structure alignment of NURC1 AlphaFold model (red, UniProt ID Q9LZ65), NIFK AlphaFold model (grey, UniProt ID Q9BYG3) and Nop15 CryoEM structure (light blue, PDB ID 3JCT) represented as cartoons. (**Top, right)** Zoom on the aligned RRMs. The superimposition was performed in Pymol using the alignment plugin (align, outlier rejection, 20 cycles and a cutoff of 2). (**Bottom**) Clustal Omega of all three protein sequences. RNPs are highlighted in sky blue and conserved amino acids are inside boxes. The RRM secondary structure is displayed on the bottom as a cartoon where β-sheets are represented as orange arrows and α-helices as green rods. Figure from ref.[136].

3.1.2 Purification of NURC1

To study the structure and function of the NURC1 full-length protein (NURC1$_{FL}$, ~25 kDa) *in vitro*, the protein had first to be purified. This was accomplished by heterologous expression in *E. coli*. Additionally, three truncated versions were generated. NURC1$_{1\text{-}160}$ (~18 kDa) that lacked the predicted α-helix at the C-terminus, NURC1$_{1\text{-}140}$ (~15 kDa) that had the C-tail completely removed, and lastly, NURC1$_{53\text{-}222}$ (~20 kDa), which missed the highly disordered N-terminus. All versions were cloned into the pet28a(+) expression vector with a 6xHis-tag located at the N-terminal. The expression was carried out overnight at 22°C by auto-induction. The purification was carried out as described in **chapter 2.2.3**.

The purity of NURC1$_{FL}$ was followed throughout the purification by SDS-PAGE (**Figure 3.3**). Large amounts of protein were observable after affinity chromatography with a size between 25-35 kDa, along with other bands at variable sizes (**Figure 3.3A)**. After size exclusion chromatography (SEC), the overall number of bands was significantly reduced and thick bands at around 25 kDa were visible (**Figure 3.3B**). NURC1$_{1\text{-}160}$, NURC1$_{1\text{-}140}$ and NURC1$_{53\text{-}222}$ SDS-PAGE results can be found in **Figure 7.10**. Generally, the samples seemed quite pure with some faint or thin bands at higher molecular weight in some of the fractions. Only the purest fractions were selected, combined, and used in further analysis. The IMAC chromatogram yielded two peaks at approximately 23 ml and 41 ml, whereas the SEC chromatogram of NURC1$_{FL}$ yielded a small and a large peak at around 50 ml and 62 ml, respectively (**Figure 3.3C and D**). NURC1$_{1\text{-}160}$, NURC1$_{1\text{-}140}$ and NURC1$_{53\text{-}222}$ chromatograms yielded two peaks, as well (**Figure 7.10**). The first and smaller peak had an elution volume at around 50 ml. NURC1$_{1\text{-}160}$ and NURC1$_{53\text{-}222}$ showed a noticeable larger peak at around 86 ml and 89 ml, respectively. NURC1$_{1\text{-}140}$ second peak had lower absorbance compared with the other versions and its elution peak was around 86 ml.

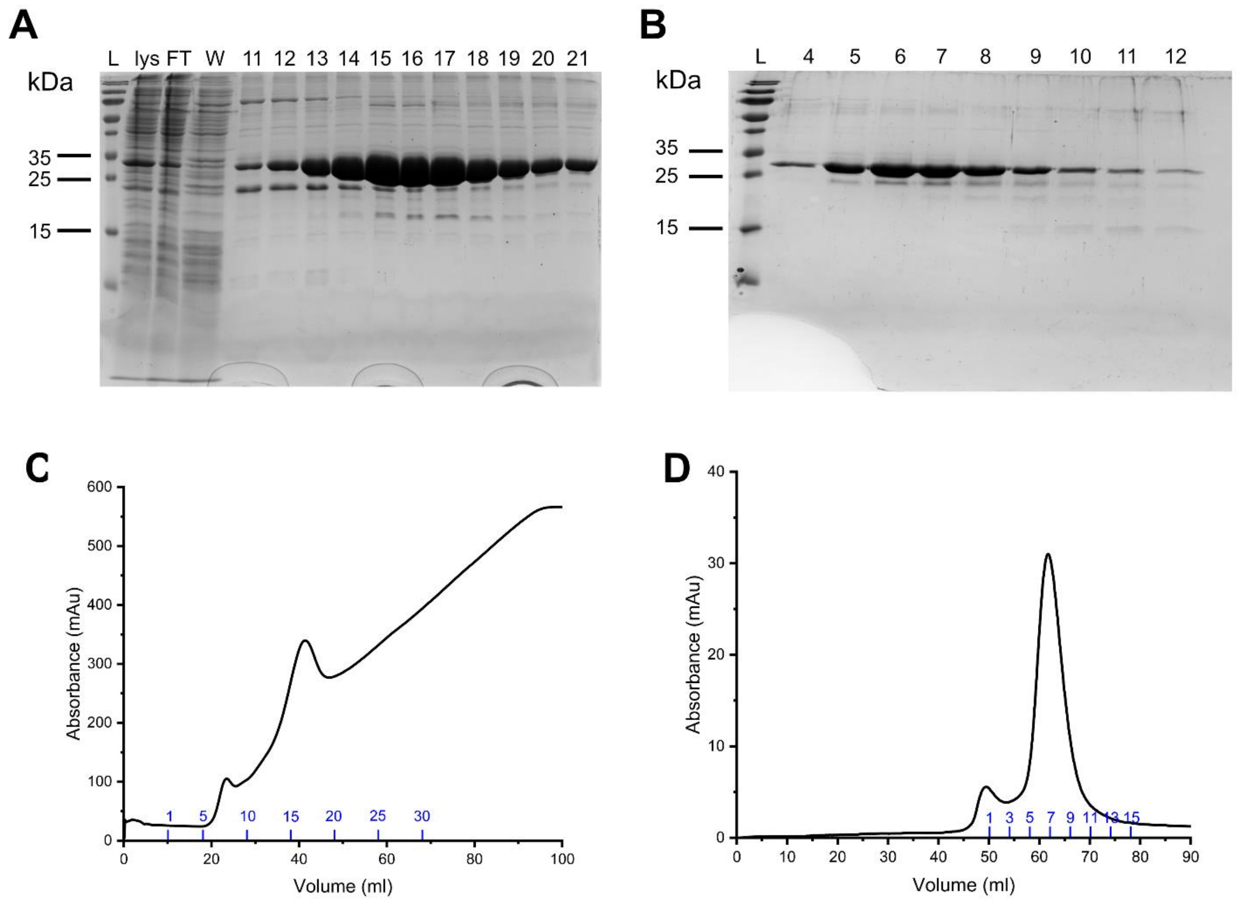
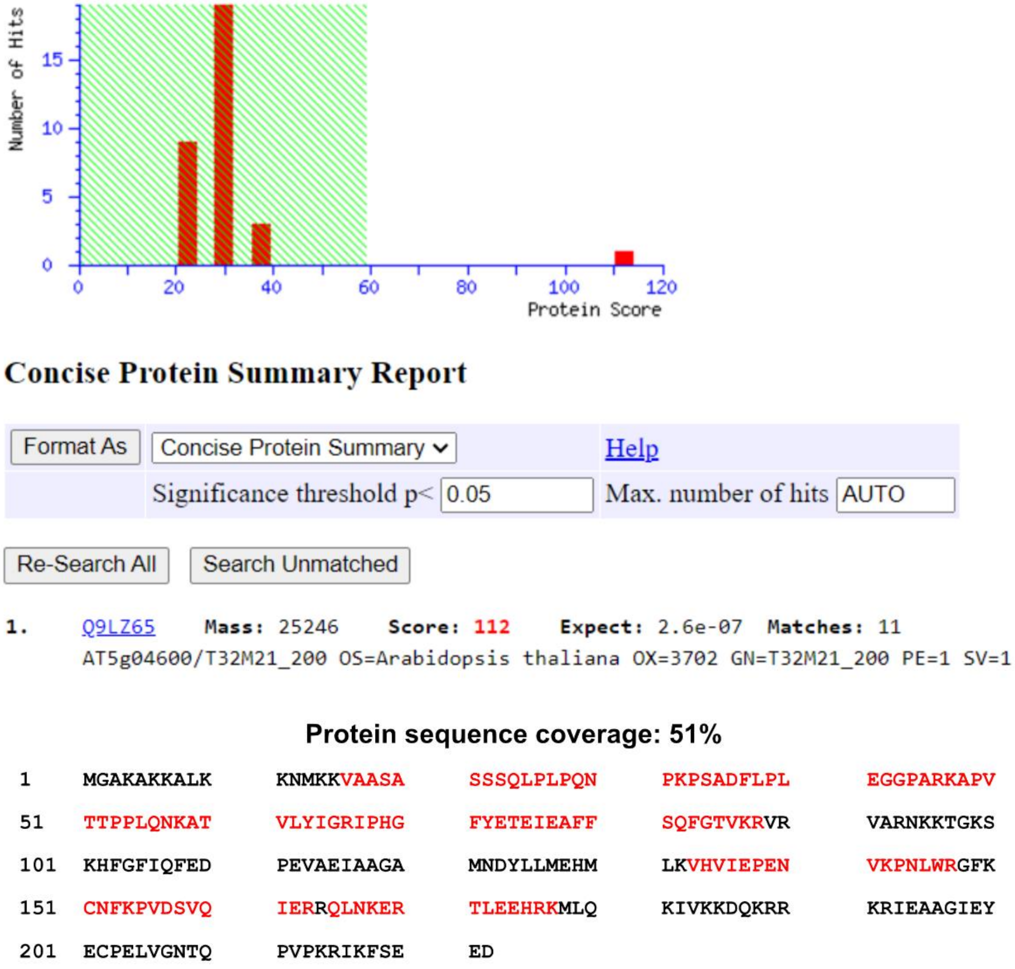


Figure 3.3− NURC1$_{FL}$ purification results. 12% SDS-PAGE after (**A**) IMAC and (**B**) SEC. L: PageRuler Prestained Protein Ladder (ThermoFisher Scientific); lys: lysate; FT: flowthrough; W: washing flowthrough; 11-21: IMAC elution fractions; 4-12: SEC elution fractions. (**C**) IMAC and (**D**) SEC chromatograms showing the absorbance (mAu) as a function of elution volume (ml). Äkta fractions are shown in blue and match those found in A and B.

MALDI-TOF MS was performed to confirm the protein identity (**Figure 3.4 and Figure 7.11**). MASCOT search against UP6548_A_thaliana database identified the AT5g04600 protein, i.e., NURC1, with a score of 112 and an expected value of 2.6×10^{-7}. Eleven peptides were matched which resulted in a sequence coverage of 51%.



Concise Protein Summary Report

Significance threshold p< 0.05 Max. number of hits AUTO

1. Q9LZ65 **Mass:** 25246 **Score:** 112 **Expect:** 2.6e-07 **Matches:** 11
AT5g04600/T32M21_200 OS=Arabidopsis thaliana OX=3702 GN=T32M21_200 PE=1 SV=1

Protein sequence coverage: 51%

```
1    MGAKAKKALK  KNMKKVAASA  SSSQLPLPQN  PKPSADFLPL  EGGPARKAPV
51   TTPPLQNKAT  VLYIGRIPHG  FYETEIEAFF  SQFGTVKRVR  VARNKKTGKS
101  KHFGFIQFED  PEVAEIAAGA  MNDYLLMEHM  LKVHVIEPEN  VKPNLWRGFK
151  CNFKPVDSVQ  IERRQLNKER  TLEEHRKMLQ  KIVKKDQKRR  KRIEAAGIEY
201  ECPELVGNTQ  PVPKRIKFSE  ED
```

Figure 3.4− MALDI-TOF MS established NURC1 identity. The score histogram shows the number of hits as a function of the protein score. Protein scores greater than 59 are statistically significant ($p<0.05$). UniProt ID (Q9LZ65), Mass, E-value (Expect), number of matched mass values (Matches), Protein name (AT5g04600/T32M21_200) and Organism (OS) are shown. NURC1 sequence has a coverage of 51% and the matched peptides are shown in red. The search was performed using the Mascot Search tool from mMass and the UP6548_A_thaliana database. Methionine oxidation was set as variable modification and 0.5 Da of peptide tolerance was allowed.

3.1.3 Impact of concentration and terminal truncation on sample homogeneity

It is usual practice to initially attempt to characterize the sample homogeneity before analyzing the structure of a protein. For that, techniques like native-PAGE and DLS were employed.

NURC1$_{FL}$ has a theoretical isoelectric point (pI) of 9.84. The separation of native basic proteins by cathodic discontinuous PAGE has been shown to yield better resolution and does not require binding of dyes when compared to other techniques[139]. Consequently, this method was used to check the composition of the NURC1_{FL} sample (**Figure 3.5**). Three protein stocks at different concentrations were used. Samples 1 and 3 corresponded to protein stocks with lower concentration, whereas sample 2 had a higher concentration. All samples showed smearing with darker bands at variable sizes that ended at 24 kDa. Samples 1 and 3 had four visible bands between 24 and 140 kDa, whilst sample 2 had additional bands above 140 kDa. These results suggested the presence of several oligomeric states, where higher molecular weight oligomers seemed to be concentration dependent.

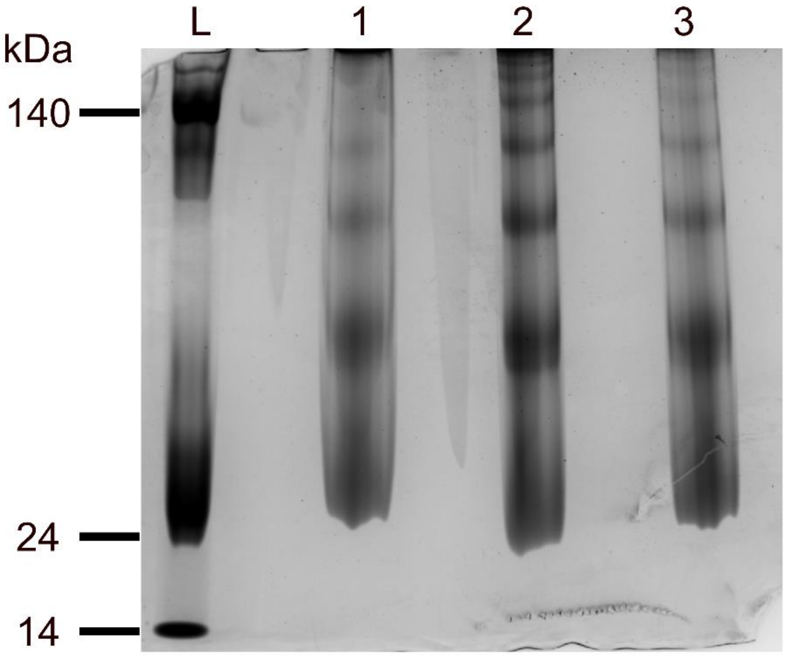


Figure 3.5− Native-PAGE of NURC1_{FL} at different concentrations. Cathodic discontinuous PAGE of NURC1_{FL} from stocks with concentration 0.7 mg/ml (**1**), 6.8 mg/ml (**2**) and 2 mg/ml (**3**). Protein stocks were centrifuged at maximum speed (20,000 × g) for 10 min at 4 °C, and 10 µg of protein were loaded into a 10% bis-acrylamide gel. The ladder (**L**) contained an equimolar mixture of lactate dehydrogenase (~140 kDa), TEV protease (~24 kDa), and lysozyme (~14 kDa).

DLS measurements revealed that the hydrodynamic radius (R_h) and the polydispersity level increased with protein concentration (**Figure 3.6A**), supporting the concentration dependency hypothesis. Moreover, DLS of $\text{NURC1}_{1\text{-}160}$, $\text{NURC1}_{1\text{-}140}$ and $\text{NURC1}_{53\text{-}222}$ showed that the truncations at the C-terminus induced polydispersity, and that the removal of the N-terminus induced substantial aggregation **(Figure 3.6B**).

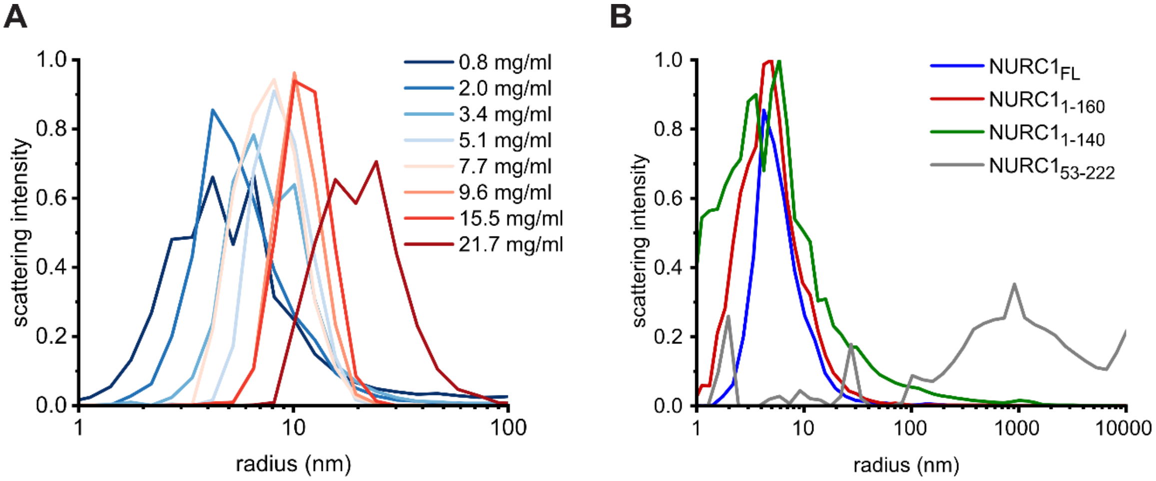


Figure 3.6− DLS scattering profiles of NURC1 and its truncated versions. (**A**) NURC1$_{FL}$ from lowest (dark blue) to highest (dark red) concentrations. (**B**) NURC1$_{FL}$ (2.0 mg/ml) in blue, NURC1$_{1-160}$ (2.4 mg/ml) in red, NURC1$_{1-140}$ (1.0 mg/ml) in green, and NURC1$_{53-222}$ (2.4 mg/ml) in grey. All samples were centrifuged at 20,000 x g for 10 min at 4°C. Figure from ref.[136].

3.1.4 Structural and Functional Analysis

3.1.4.1. Exploring the shape of NURC1 by SAXS

Crystallization presents a challenge when it comes to flexible macromolecules. There is currently no crystal structure of NURC1 available. The flexible regions are commonly truncated to increase the crystallization success. However, no crystals could be obtained from NURC1$_{FL}$ nor from the truncated proteins NURC1$_{1-160}$ and NURC$_{1-140}$ (data not shown). Consequently, SAXS was chosen to study NURC1$_{FL}$ structure in solution. Three scattering curves were recorded at different concentrations in batch mode (**Figure 3.7A**). The Guinier analysis of the scattering data showed that the radius of gyration (R_g) increased non-linearly (**Figure 3.7B and Figure 7.12**). Once more, suggesting a concentration dependency.

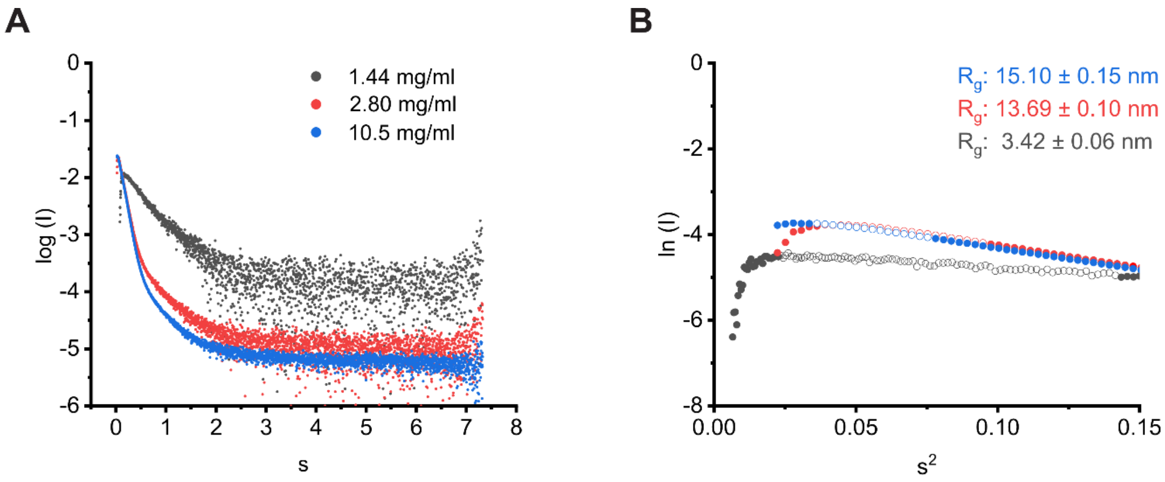


Figure 3.7− SAXS analysis of three NURC1$_{FL}$ concentrations in batch mode. (**A**) Scattering curves at the different concentrations: 1.44 mg/ml (black), 2.80 mg/ml (red) and 10.5 mg/ml (blue). (**B**) Guinier analysis. The curve fit is represented as open spheres and the resulting R_g values are shown as insets. Figure from ref.[136].

To overcome the concentration dependency of NURC1$_{FL}$, SEC-SAXS was performed. The SEC chromatogram yielded one peak (**Figure 3.8A**). The Guinier approximation resulted in a R_g of 3.52 ± 0.01 nm with sR_g limits between 0.47 and 1.29 (**Figure 3.8B**). The linear fit at low s values resulted in a Pearson’s R value of -0.998 and the double logarithmic scale scattering curve slope was null at low s values, thus sustaining that similar sized and monodispersed particles were present in solution. This was also supported by the SEC-MALLS results, where a single elution peak was observed that yielded a molecular weight of around 28 kDa **(Figure 7.13A and B**). From the in line QLES detection a R_h of 2.4 nm was obtained (**Figure 7.13C**). The R_g/R_h ratio can be used to obtain information about the shape of molecules. Globular proteins usually have a ratio of 0.77, whereas NURC1 showed a ratio of 1.5, thus suggesting an elongated shape. Another way to access conformational information is to look at the dimensionless Kratky plot. Globular proteins present a bell-shaped peak at $s{*}R_g = \sqrt{3} \approx 1.73$ with a height of 1.1[140]. The dimensionless Krakty plot of NURC1 is shown in **Figure 3.8C**. The bell-shape peak was visible, however, it was clearly right-shifted ($s{*}R_g = 3.26$) and slightly higher [$(s{*}R_g^2){*}I/I(0) = 1.62$], suggesting that NURC1 possesses an elongated shape and/or is partially unfolded. Additionally, the increase of values at $s{*}R_g > 7.5$ instead of converging with the axis provided evidence for flexibility. The pair distribution function P(r) revealed a non-Gaussian curve with a peak at approximately 2.5 nm (**Figure 3.8D**). A long tail that ends at a r_{max} of 12.4 nm and a small shoulder were observable, which were again characteristic of an elongated macromolecule. It is worth mentioning that the $R_g/I(0)$ ratios from the Guinier plot and P(r) function were in agreement, dismissing any presence of aggregates that primarily affect lower s values, which are used for the Guinier approximation. The molecular weight was determined by Bayesian Inference, resulting in an estimated molecular weight of around 33.8 kDa with a probability of 51.6%.

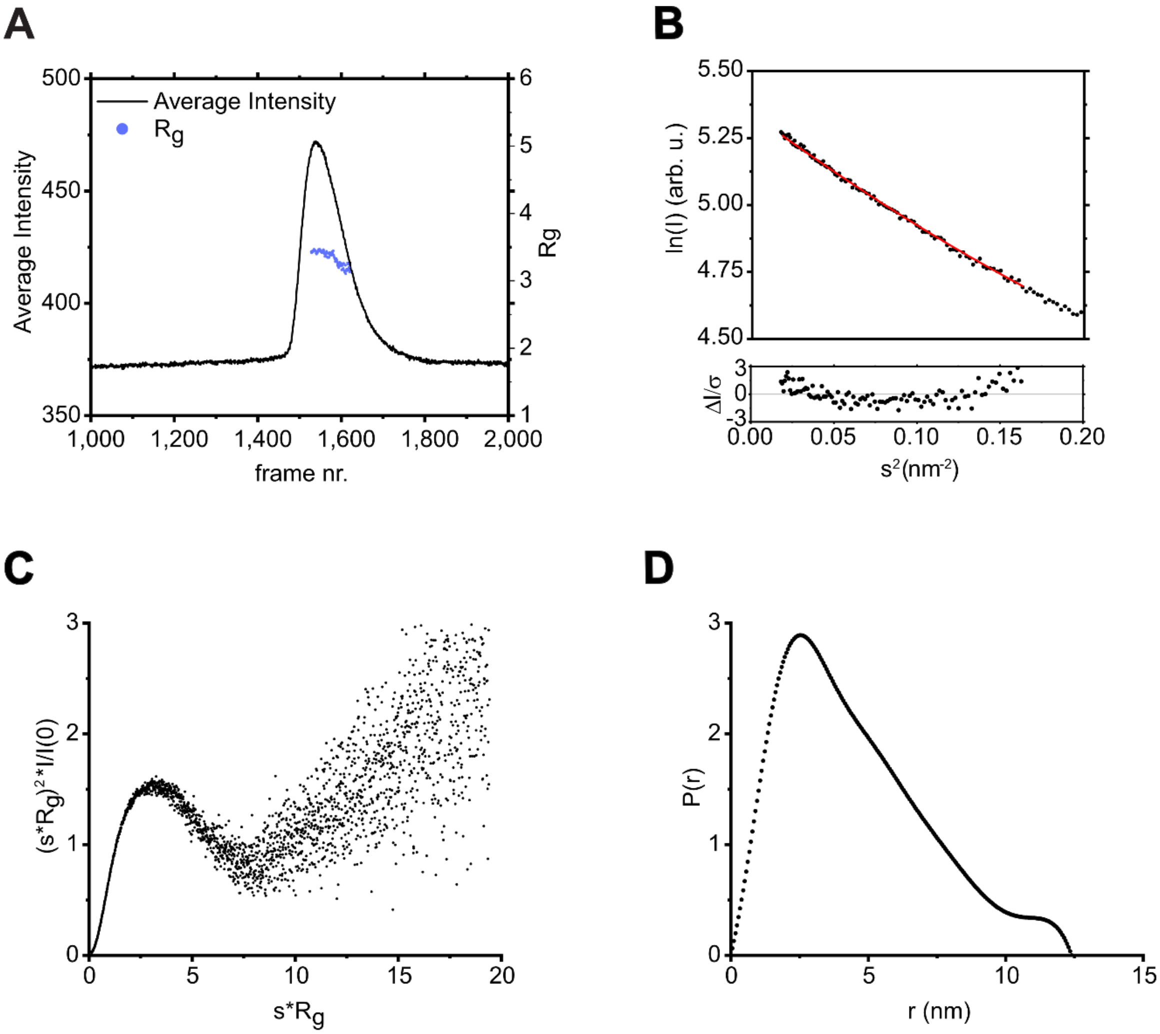


Figure 3.8− SEC-SAXS analysis of NURC1$_{FL}$. (**A**) SEC chromatogram with the average intensity (black line) and the R_g values for the selected range of frames (purple spheres). (**B**) The Guinier fit (red line) of the experimental data (black spheres) and the standardized residuals. (**C**) Dimensionless Krakty plot. (**D**) Pair distribution function. Figure adapted from ref.[136].

3.1.4.2. NURC1 structure modeling

The next step involved the evaluation of the scattering data by comparing it with known atomic structures. Four atomic models were used against NURC1$_{FL}$ experimental scattering data (**Table 7.2**). Two of them consisted of atomic resolution structures obtained by experimental data from X-ray crystallography and CryoEM measurements. The other two were created by bioinformatic tools, like i-Tasser and AlphaFold, that predict the tertiary structure from amino acid sequences. The program CRYSOL was used for modeling these structures into the envelopes determined by the SAXS data. The final chi-square (χ^2) values represent the discrepancy between the calculated scattering from the atomic model and the experimental data. Therefore, smaller χ^2 values indicate a better fit. **Table 7.2** contains the final χ^2 values for the tested homology models. The AlphaFold model showed the smallest value with a χ^2 of 11.97.

SREFLEX was used to estimate the flexibility of the AlphaFold model, offering a pool of conformers that might fit the data better. Ten models were returned and tested again against the experimental data with CRYSOL. The best fit resulted in a major improvement with a χ^2 of 3.04.

Ab initio models were generated with GASBOR and DAMMIN to access the shape of NURC1. The GASBOR model showed a χ^2 of 1.56, whereas the DAMMIN model had a χ^2 of 1.06. Both dummies support the elongated shape of NURC1. Furthermore, the *ab initio* models were superimposed with the best SREFLEX model (**Figure 3.9**). The superimposition revealed that both termini did not perfectly fit the model, and this was more prominent for the N-terminus that has flexibility and is highly disordered.

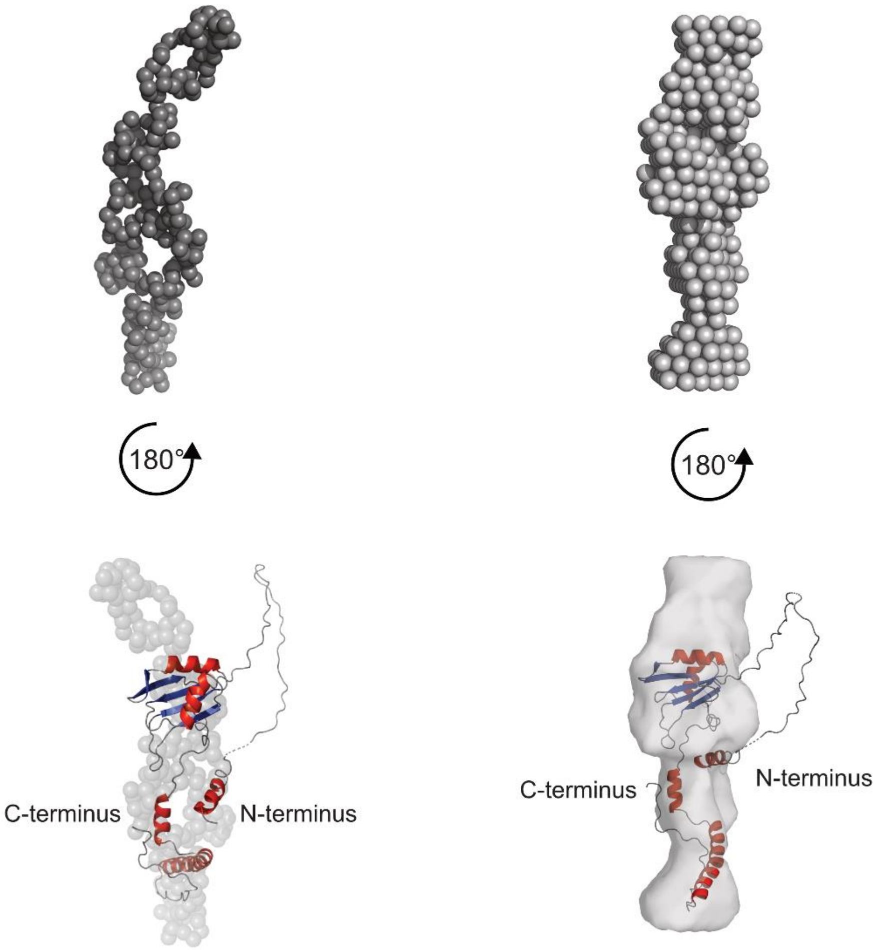


Figure 3.9− NURC1$_{FL}$ *ab initio* models and their superimposition with the SREFLEX model. (**A**) GASBOR P1 (top left, dark grey spheres) and DAMMIN (top right, light grey spheres) ab initio models based on NURC1$_{FL}$ SAXS. The superimposition of both dummies with the SREFLEX model is shown (bottom left and right). The superimposition was performed in Pymol using the supalm tool from SASpy ATSAS plugin. The superimposition with GASBOR P1 and DAMMIN P1 yielded normalized space discrepancy (NSD) values of 4.77 and 5.36, respectively. Figure from ref.[136].

3.1.4.3. Determining the RNA binding capacity of NURC1

Yeast Nop15 binding to the ITS2 RNA has been studied and is known to bind to the stem III.A of the ring form[46,57]. Likely, the plant ortholog NURC1 has a conserved target and RNA binding activity. Hence, the binding affinity of NURC1 to the plant ITS2 RNA was investigated by MST. For that, *A. thaliana* ITS2$_{FL}$ RNA was synthesized as described (**chapter 2.2.4**). In **Figure 3.10A** and **Figure 7.14**, the predicted structure of *A. thaliana* ITS2 RNA ring form is depicted. The *Arabidopsis* ITS2 RNA ring has four predicted stems, two less than yeast's ring structure (**Figure**

1.4C). A local alignment (EMBOSS Water) was performed between yeast ITS2 stem III.A and *Arabidopsis* $ITS2_{FL}$ to determine the putative binding site. The alignment resulted in a stretch of 20 bases that belong to stem III with similarity of 66.7% (**Figure 3.10B**). Accordingly, three additional RNAs were synthesized to screen for the real binding site, containing parts of ITS2, namely $ITS2_{1-79}$ (stem I and II), $ITS2_{82-162}$ (stem III), and $ITS2_{157-187}$ (stem IV). Their secondary structure predictions can be found in **Figure 7.14**.

The binding affinities between the different versions of NURC1 and ITS2 were determined (**Figure 7.15 and Table 7.3**). For instance, $NURC1_{FL}$ and $NURC1_{1-160}$ bound to $ITS2_{FL}$ with a K_d of 228 ± 83 nM and 116 ± 20 nM, respectively. In contrast, the removal of 20 additional amino acids in $NURC1_{1-140}$ led to an increase in the K_d value to 938 ± 167 nM, thus significantly reducing the binding affinity (**Figure 3.10C)**. This decrease was observable across all measurements with other ITS2 stem-loops (**Figure 7.15A**). Binding between $NURC1_{53-22}$ and $ITS2_{FL}$ was detectable (**Figure 7.16A**). However, it was not possible to determine the K_d value due to fluorescence quenching occurring at higher protein concentrations. The quenching could be induced either by protein aggregation or ligand binding. Therefore, the SDS denaturation test (SD-test) was conducted to determine the cause of quenching. The SD-test is used to distinguish fluorescence changes caused either by an interaction or by non-specific effects, such as aggregation or adsorption. Protein denaturation is achieved by using SDS as denaturation agent together with high temperature. This leads to the disruption of the interaction between protein and ligand. Consequently, if the fluorescence changes are still observable it means that the loss is due to non-specific interactions. After denaturation, $NURC1_{53-22}$ samples did not recover the fluorescence, indicating that quenching was most likely caused by protein aggregation (**Figure 7.16B**).

Overall, no significant differences in binding across the different ITS2 stems were observed, apart from $NURC1_{1-160}$ and $ITS2_{157-187}$, that bound significantly less (**Figure 7.15B**). $NURC1_{1-160}$ exhibited binding to $ITS2_{1-79}$ and $ITS2_{82-162}$ and resulted in K_d values of 353 ± 93 nM and 239 ± 124 nM, respectively (**Figure 3.10D**), whereas the binding to $ITS2_{157-187}$ was significantly reduced with a Kd of 673 ± 60 nM (**Figure 3.10D**). The MST traces and binding curves for all measurements can be found in the **Figure 7.17-Figure 7.20**.

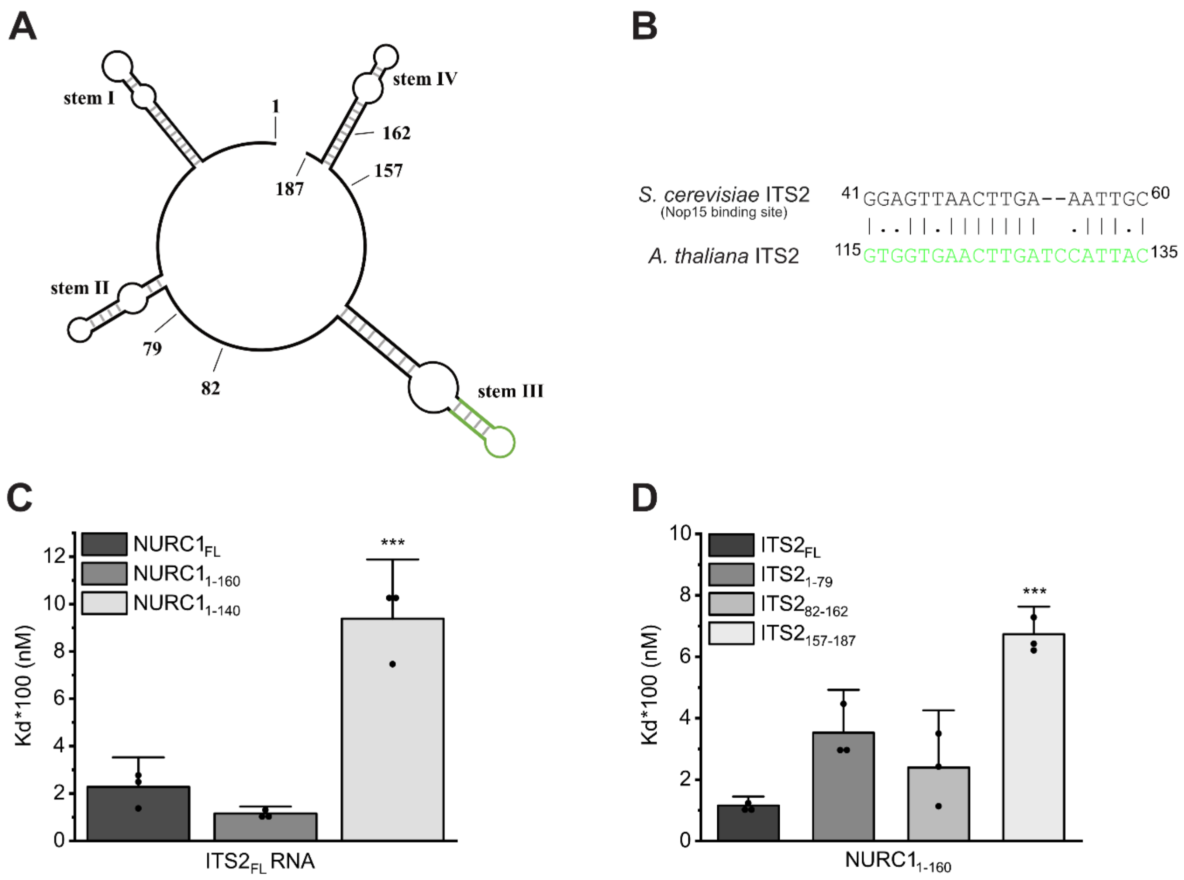


Figure 3.10− Analysis of plant ITS2 and its binding to NURC1 by MST. (**A**) *A. thaliana* ITS2 rRNA ring form has four stem-loops. The green region in stem III indicates the highest similarity with yeast's ITS2 stem III.A. (**B**) Local alignment of yeast's stem III.A and *Arabidopsis* ITS2 using EMBOSS Water. (**C**) Comparison of K_d values between different NURC1 versions to the $ITS2_{FL}$. (**D**) Comparison of K_d values between different ITS2 stems to $NURC1_{1\text{-}160}$. The statistical analysis was performed with one-way ANOVA (**p* < 0.05, ***p* < 0.01 and ****p* < 0.001) in OriginPro 2021b. Figure adapted from ref. [136].

3.1.4.4. Assessing the RNA chaperone activity of NURC1

Nop15 binding to the ITS2 ring form may prevent stem III.A mispairing and reduce misfolding of the ring form, that could otherwise lead to premature and/or improper processing of pre-rRNAs[57]. To determine if the stabilization of the ITS2 ring form in *Arabidopsis* may be mediated by NURC1, its putative chaperone activity was investigated by real-time FRET. Two complementary RNA strands labelled with different fluorophores (Cy3 and Cy5) were used to follow the chaperone activity of NURC1. The FRET index (i.e., ratio between the fluorophores fluorescence overtime) was used to determine the annealing constants ($K_{annealing}$). Overall, changes in the FRET index were observed in every sample (**Figure 3.11A**). The FRET index of NURC1-FL, $NURC1_{1\text{-}160}$, NURC1$_{1\text{-}140}$, and NURC1$_{53\text{-}222}$ increased much faster when compared to the RNA control, thus indicating faster annealing. Particularly, NURC1$_{FL}$ and NURC1$_{53\text{-}222}$ showed the fastest FRET increase between 0 and 50 seconds, followed by a plateau. TEV protease showed a similar performance to the RNA control. A one-phase exponential association fitting was applied to each curve to determine the $K_{annealing}$ values (**Figure 3.11B and Table 3.1**).

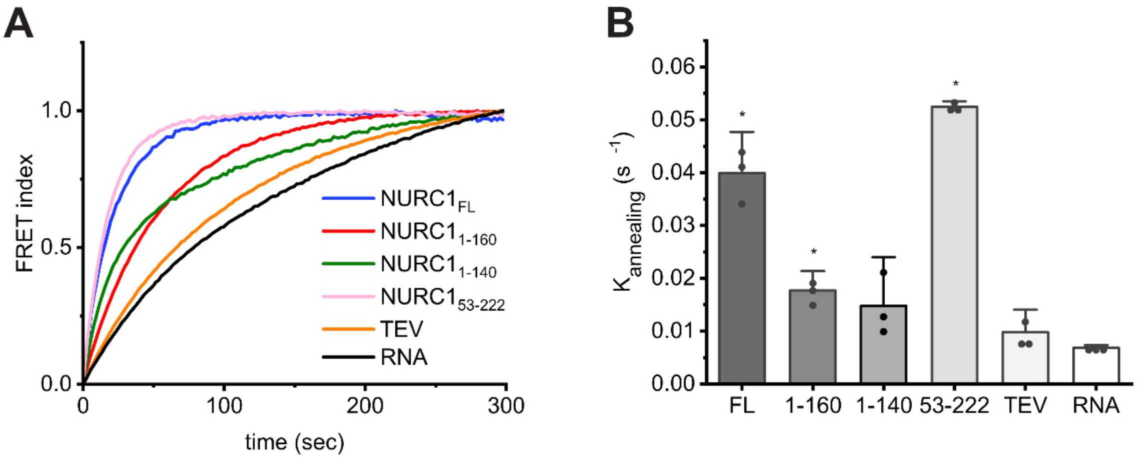


Figure 3.11− FRET analysis of NURC1 RNA chaperone activity. (**A**) Normalized FRET index changes overtime. The curves were obtained by averaging triplicates. The RNA sample consists only of the two complementary strands, and a 1:5 (RNA:protein) molar ratio was used in all samples. The fluorescence changes were recorded for 300 s at 37°C. (**B**) $K_{annealing}$ differences between samples shown in a bar plot. Triplicates were measured (black dots) and the results are plotted as mean + standard deviation. The statistical analysis was performed by comparing the samples to the RNA control with two-sample t-test ($^*p < 0.05$) in OriginPro 2021b. Figure from ref.[136].

The $K_{annealing}$ of NURC1$_{53\text{-}222}$ and NURC1$_{FL}$ showed significant differences compared to the RNA control with values of 0.0525 s^{-1} and 0.0399 s^{-1}, respectively. The lower $K_{annealing}$ values of 0.0177 s^{-1} for NURC1$_{1\text{-}160}$ and 0.0148 s^{-1} for NURC1$_{1\text{-}140}$ suggest that these truncations had a minor though significant effect on chaperone activity. As expected, TEV protease resulted in the lowest $K_{annealing}$ of 0.00978 s^{-1}.

Table 3.1- Chaperone Assay annealing constants. p-values were obtained by comparing each sample to the RNA control with a two-sample t-test. Table from ref.[136].

Protein	$K_{annealing}$ (s^{-1})	Standard deviation	p-value
NURC1$_{FL}$	0.0399	5.19×10^{-3}	3.86×10^{-4}
NURC1$_{1\text{-}160}$	0.0177	2.46×10^{-3}	1.63×10^{-3}
NURC1$_{1\text{-}140}$	0.0148	6.13×10^{-3}	8.81×10^{-2}
NURC1$_{53\text{-}222}$	0.0525	7.00×10^{-4}	5.59×10^{-8}
TEV	0.00978	2.85×10^{-3}	1.50×10^{-1}
RNA	0.00684	3.35×10^{-4}	-

3.2. Identification and characterization of lncRNAs in *B. napus* phloem sap

The next chapter is focused on the analysis of the lncRNA transcriptome from *B. napus* phloem. Furthermore, the leaf lncRNA content was also assessed for comparison reasons.

3.2.1 Analysis of lncRNAs in the sequencing data

RNA-seq consists of four main steps: RNA extraction, library preparation, sequencing, and data analysis. The latter is a multi-step process that involves a set of bioinformatics algorithms which need to be executed in a certain order. This set of programs is called a pipeline. The programs used for the analysis can differ from study to study, but the essential steps remain the same. **Figure 3.12** illustrates the common essential steps of a pipeline for lncRNAs identification. The results for each step of the pipeline will be presented in the following text and the programs used by Novogene or in-house will be specified accordingly.

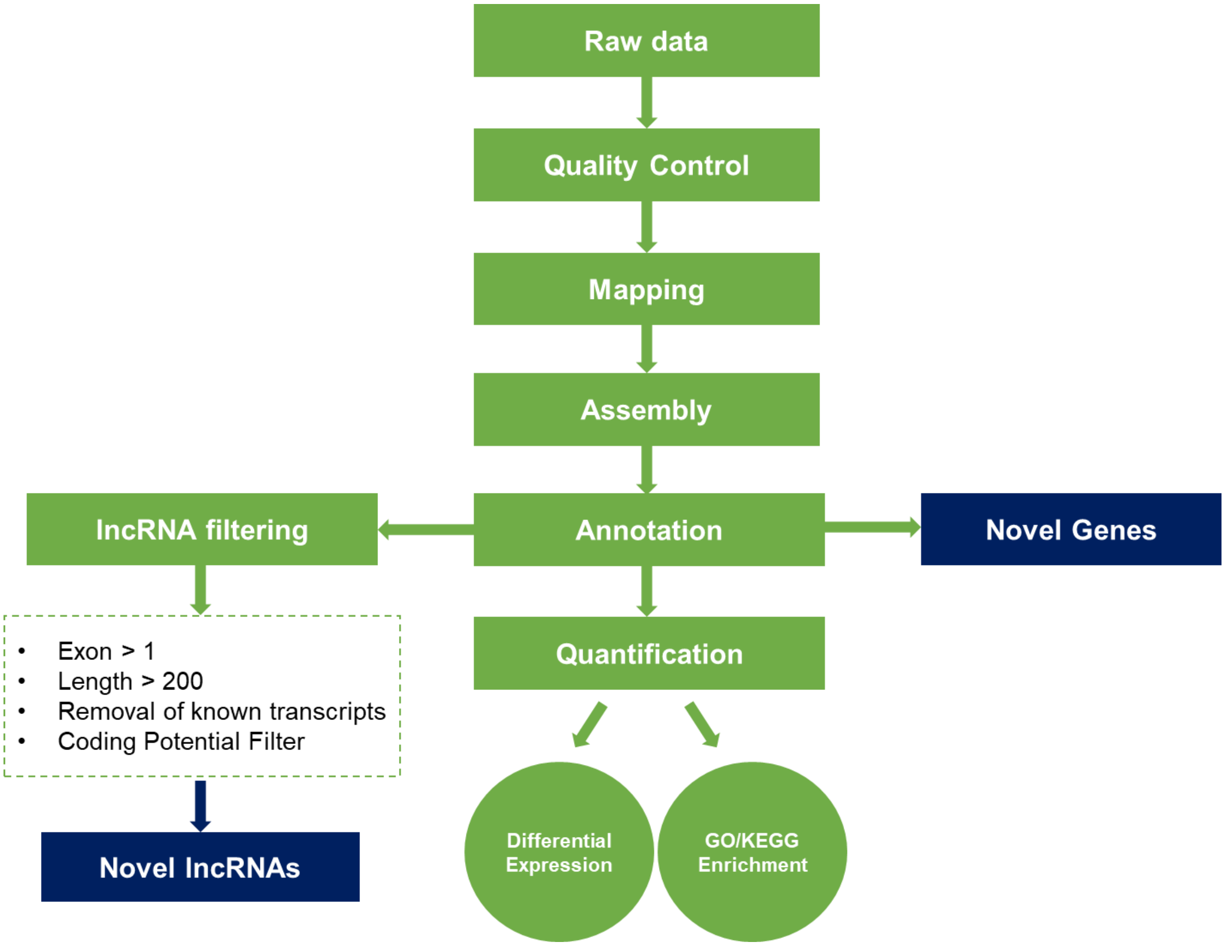


Figure 3.12− RNA-seq pipeline main steps for lncRNAs identification. Main steps include quality control, mapping, assembly, and annotation. Novel lncRNAs can be identified by applying a series of filters, and novel genes can also be identified. Quantification allows us to explore differentially expressed genes or transcripts, followed by Gene Ontology (GO)/Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis.

3.2.1.1. Quality control and mapping

The first step of the pipeline is quality control (QC) of the reads. Features such as the number of clean reads, GC content, base quality and error rates are determined. The QC summary can be found in **Table 3.2** and error rate plots in **Figure 7.21**-**Figure 7.22**. Cleaning of the raw reads resulted in over 30 million clean reads per sample with GC content below 50% and a Q30 above 90%.

Table 3.2- QC summary. Overview of phloem and leaf reads quality.

Sample	Raw reads	Clean reads	GC%	Q30 (%)
Phloem1	35447709	35219761	46.4	93.22
Phloem2	33352768	33185960	46.47	92.06
Phloem3	36172109	35869804	46.82	93.47
Leaf1	30673150	30328173	42.5	90.21
Leaf2	36325392	35934409	42.65	92.99
Leaf3	36482677	36095338	42.69	93.45

After obtaining clean reads, the next step was to map them to a reference genome or transcriptome, depending on the research goal. Given that lncRNAs can span diverse genomic locations, such as intronic, intergenic and promoter-associated regions (**Figure 1.7**), the genome was chosen instead of the transcriptome for the alignment. HISAT2[124] was used as the software aligner.

The phloem mapping summary is shown in **Table 3.3**. Over 90% of the reads mapped to the reference genome for all replicates and under 20% of them mapped to multiple sites. Of the 70% uniquely mapped reads, the strand distribution seemed fairly even and around 50% of the reads mapped entirely to a single exon (non-splice reads), while over 20% were segmented and mapped to two exons (splice reads).

The leaf mapping summary can be found in **Table 7.4**. Briefly, over 90% of the reads were mapped to the reference genome, and approximately 50% of the reads were mapped to multiple sites, more than double when compared to the multi-mapping of phloem samples.

Table 3.3- Phloem mapping summary. Overview of phloem mapped reads. Read-1 refers to left reads (5' end) and read-2 to right reads (3' end).

Sample name	Phloem1	Phloem2	Phloem3
Total reads	70439522	66371920	71739608
Total mapped	64720176 (91.88%)	61179344 (92.18%)	66122883 (92.17%)
Multiple mapped	13598890 (19.31%)	12058642 (18.17%)	12067223 (16.82%)
Uniquely mapped	51121286 (72.57%)	49120702 (74.01%)	54055660 (75.35%)
Read-1	25744528 (36.55%)	24800256 (37.37%)	27209751 (37.93%)
Read-2	25376758 (36.03%)	24320446 (36.64%)	26845909 (37.42%)
Reads map to '+'	25421423 (36.09%)	24431539 (36.81%)	26896819 (37.49%)
Reads map to '-'	25699863 (36.49%)	24689163 (37.20%)	27158841 (37.86%)
Non-splice reads	33589388 (47.69%)	31766896 (47.86%)	35110804 (48.94%)
Splice reads	17531898 (24.89%)	17353806 (26.15%)	18944856 (26.41%)

After mapping, Novogene used a python library named HTSeq[125] to determine which genomic features were more expressed. The genomic features were determined by aligning the reads to a reference annotation and were counted using the union counting mode. Leaf and phloem HTSeq results are shown in **Figure 3.13.** Unsurprisingly, the category "protein_coding" showed the higher expression, accounting for approximately 87% and 83% of all features expressed in leaf and phloem, respectively. In both tissues, around 11% of the reads were grouped into "Others". Other categories with lower expression levels included "sense_intronic", "snoRNA" (small-nucleolar RNA) and "SRP_RNA" (Signal Recognition Particle RNA). Surprisingly, SRP RNA was found to be more expressed in the phloem (5.25%) than in the leaf (1.29%).

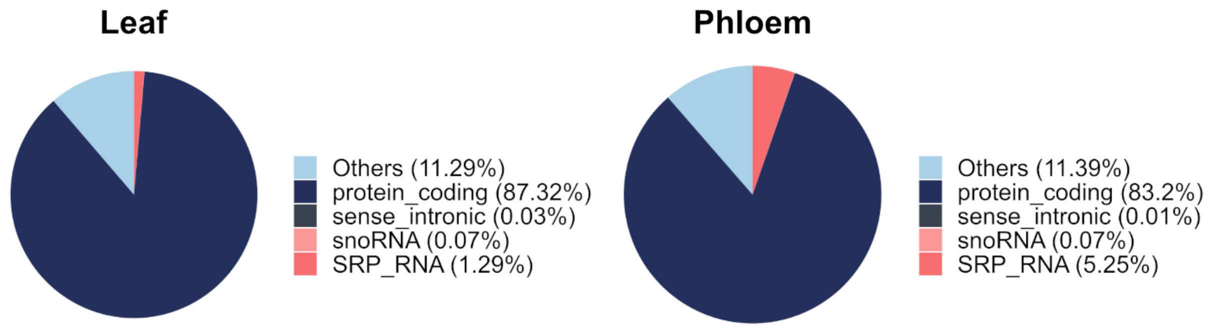


Figure 3.13− Mapping regions. Leaf (left) and phloem (right) mapped features expression levels in percentage determined by HTSeq. "Others" is represented in light blue, "protein_coding" in dark blue, "sense_intronic" in dark grey, "snoRNA" in light pink and "SRP_RNA" in salmon pink.

3.2.1.2. Annotation and quantification

Once the reads are mapped, they need to be assembled and quantified. During assembly the reads that were mapped to the same or nearby genomic location are pieced together to reconstruct a transcript model. Novogene employed the StringTie[131] program to assemble and quantify the mapped reads. The quantified reads were normalized to fragments per kilobase of transcript per million read pairs (FPKM), therefore sequencing depth and fragment size were taken into account. The FPKM distribution of leaf and phloem transcripts is shown in **Figure 3.14**. The FPKM distribution between replicates in each tissue seemed to be similar, as they exhibit a similar central tendency and spread of data.

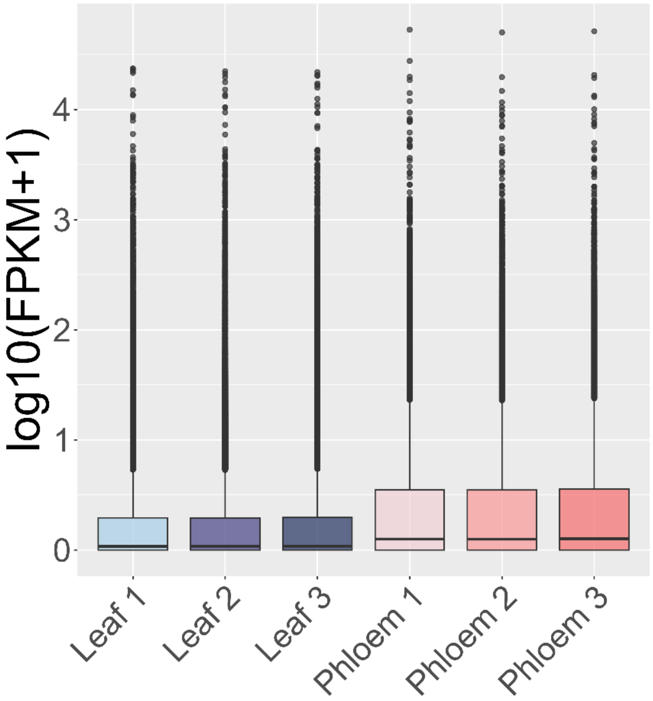


Figure 3.14− FPKM distribution of leaf and phloem transcripts. The FPKM distribution is represented as box plots. Leaf replicates are shown in shades of blue and phloem replicates in shades of pink.

Subsequently, Cuffmerge[121] was used to merge overlapping or similar transcripts into a single transcript to avoid redundancy. Moreover, Cuffcompare[121] was used together with the reference annotation to help identify the types of assembled transcripts, i.e., known or novel genes, transcripts or isoforms. Features like genomic location, strand, gene name and exon number are included in the output file. The algorithm also organizes the transcripts by classes (**Table 7.5**). Classes intron (i), unknown (u), generic exonic overlap (o), potentially novel isoforms (j), potential novel transcripts (e), and exonic overlap (x) are usually interesting classes for filtering lncRNAs.

3.2.1.3. Identification of lncRNAs

Identifying putative lncRNAs was one of the key objectives. To identify and predict lncRNAs a series of filters were applied to the assembled transcripts. Transcripts that had a low expression level, a single exon, a length inferior to 200 nucleotides, or that aligned to known genes were filtered out. Additionally, the coding potential was determined, which is a crucial step in the identification of lncRNAs. Three tools were used to analyze the transcripts coding potential: Coding Non-Coding Index (CNCI), Coding Potential Calculator (CPC) and Protein Families Database (Pfam). In leaves, a total of 1142 transcripts were classified as non-coding by the three programs (**Figure 3.15A**), while 2326 non-coding transcripts were identified in phloem samples (**Figure 3.15B**), more than double than in leaves. The lncRNAs class was also determined according to their genomic location. Long intergenic non-coding RNA (lincRNA) were the predominant class, representing approximately 47% and 46% of the lncRNAs in leaf and phloem, respectively (**Figure 3.15**). Sense overlapping, antisense and sense intronic lncRNAs were also detected with similar abundance in both tissues (**Figure 3.15**). The remaining transcripts were categorized into “others”, accounting for ~36% of the leaf, and ~39% of the phloem long non-coding content (**Figure 3.15**). Moreover, 2143 and 1173 potential novel genes were found in the phloem and in the leaf, respectively (data not shown).

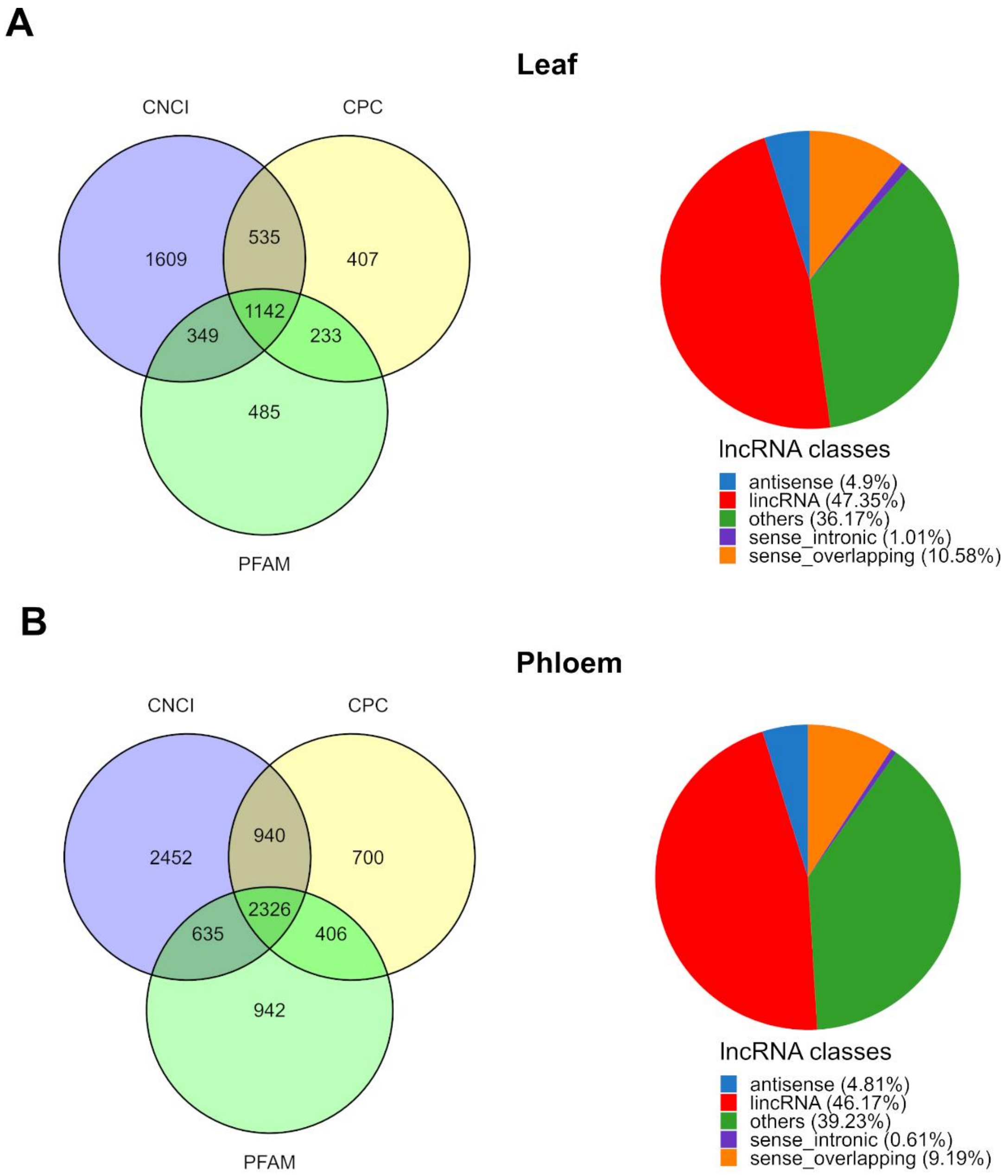


Figure 3.15− Coding potential analysis and classification of detected lncRNAs. The coding potential of lncRNAs was predicted with CNCI, CPC, and Pfam tools. (**A**) Coding potential of lncRNAs in the leaf (left) and their classes (right). (**B**) Coding potential of lncRNAs in the phloem (left) and their classes (right).

The expression level of the novel lncRNAs and their annotation can be found in the supplementary chapter (**Figure 7.23** and **Table 7.6-Table 7.7)**. The most expressed lncRNA in the phloem (TCONS_00053408) had a FPKM value of ~2,239.68 and was associated with gene ENSRNA049479067 that matches a SSU rRNA. Additionally, an antisense version of this transcript (TCONS_00053411) was also annotated to the same gene but it showed a lower FPKM value of ~55 (not shown). The lncRNAs have two exons: a larger one upstream and a smaller one downstream of the gene (**Figure 7.24**). XLOC_044198 and GSBRNA2T00091120001 gene related lncRNAs are also found to be highly expressed with FPKM values of ~1,039.21 and ~407.67. Whereas the most expressed lncRNA in leaf was linked to GSBRNA2T00002257001 gene with a FPKM of ~367.96, which is coding for a mitogen-activated protein kinase (BnaC04g03080D). The second most common lncRNA was the XLOC_012879 with a FPKM of ~166.79. Moreover, lncRNAs associated with SSU rRNA were also present in leaf (ENSRNA049449359 and ENSRNA049479067), however they were expressed at a lower level. Additionally, two Plant_SRP genes (ENSRNA049474326 and ENSRNA049474308) were annotated to several lncRNAs transcripts. In the phloem, two transcripts were found: TCONS_00007974 (antisense) and TCONS_00008124 (sense). The sense variant had a FPKM value of ~83.65 (**Table 7.6**), while the antisense variant had a lower value of ~10 (not shown). Only the sense version was identified in the leaf (TCONS_00004903) with a FPKM value of ~0.78 (not shown). This seemed to agree with the mapping results that the phloem has a higher percentage of reads mapped to SRP RNAs (**Figure 3.13**). The exon distribution of the phloem and leaf lncRNAs across the two SRP RNA genes is shown in **Figure 3.16**.

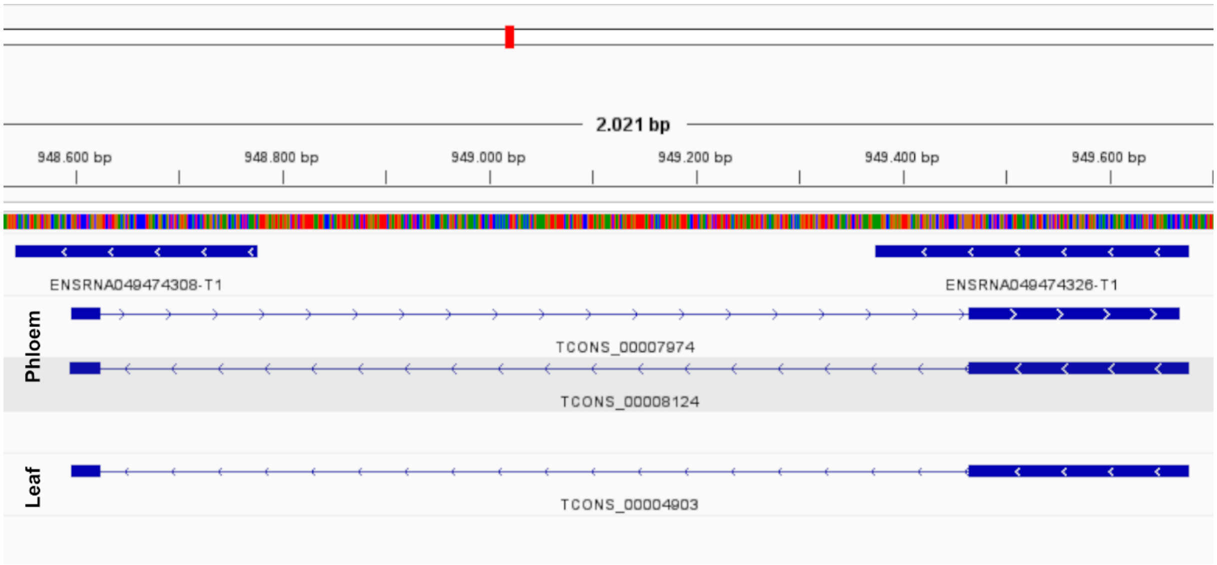
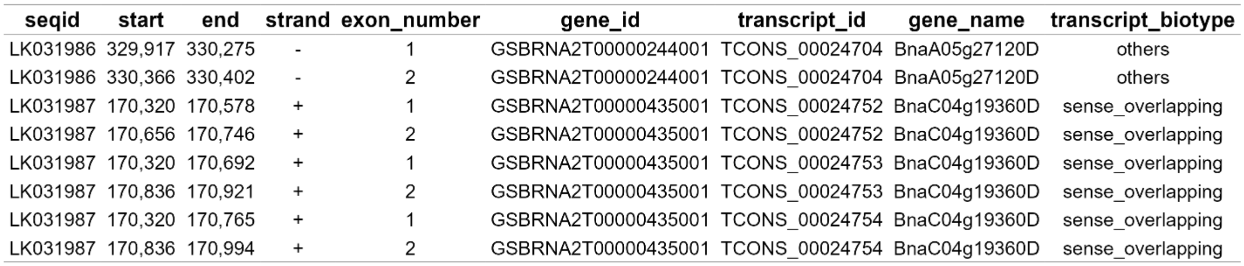


Figure 3.16− Leaf and phloem SRP RNA lncRNAs. The exon distribution of lncRNAs related to two SRP RNA genes is shown. The first exon spans a significant portion of the ENSRNA049474326 SRP RNA, and the second spans a small portion of the ENSRNA049474308 SRP RNA. The results were visualized with the IGV[126] software, and the image was obtained by print screen. Labels were added by using Affinity Publisher 2.

“XLOC” identifiers are produced during the Cuffcompare run, therefore they are run specific and represent potential novel loci, i.e., transcripts or genes that were not previously annotated in the reference genome. Further annotation details, such as location and strand, can be found in **Table 7.8** and **Table 7.9.**

Occasionally, multiple lncRNA transcripts may be annotated to the same parent gene. For example, transcripts TCONS_00024752, TCONS_00024753 and TCONS_00024754 were annotated to the GSBRNA2T00000435001 gene (**Table 3.4**). Each one of these transcripts has two exons, share the same directionality and starting position. However, the end position differs, which results in different transcript lengths. For instance, TCONS_00024752 is 426 nucleotides long, TCONS_00024753 is 601 nucleotides long, and TCONS_00024754 is 674 nucleotides long. These types of transcripts are usually referred to as isoforms, representing variants of the same lncRNA with distinct structural features, like exon length and/or exon composition.

Table 3.4- Example of lncRNAs isoforms annotation. GSBRNA2T00000435001 related lncRNAs isoforms in the phloem.



seqid	start	end	strand	exon_number	gene_id	transcript_id	gene_name	transcript_biotype
LK031986	329,917	330,275	-	1	GSBRNA2T00000244001	TCONS_00024704	BnaA05g27120D	others
LK031986	330,366	330,402	-	2	GSBRNA2T00000244001	TCONS_00024704	BnaA05g27120D	others
LK031987	170,320	170,578	+	1	GSBRNA2T00000435001	TCONS_00024752	BnaC04g19360D	sense_overlapping
LK031987	170,656	170,746	+	2	GSBRNA2T00000435001	TCONS_00024752	BnaC04g19360D	sense_overlapping
LK031987	170,320	170,692	+	1	GSBRNA2T00000435001	TCONS_00024753	BnaC04g19360D	sense_overlapping
LK031987	170,836	170,921	+	2	GSBRNA2T00000435001	TCONS_00024753	BnaC04g19360D	sense_overlapping
LK031987	170,320	170,765	+	1	GSBRNA2T00000435001	TCONS_00024754	BnaC04g19360D	sense_overlapping
LK031987	170,836	170,994	+	2	GSBRNA2T00000435001	TCONS_00024754	BnaC04g19360D	sense_overlapping

3.2.1.4. Characterization of lncRNAs

Further investigation was conducted to explore several structural features of lncRNAs across tissues. Attributes such as size, direction and number of exons were investigated (**Figure 3.17**). The size distribution of phloem and leaf lncRNA transcripts is shown in **Figure 3.17A** and **Figure 3.17B**, respectively. Most of the phloem and leaf transcripts fell within the size range of 400-800 nucleotides. Notably, phloem transcripts had the highest count at length ~600 nucleotides, whereas leaf transcripts exhibited a higher count at ~500 nucleotides. Regarding the directionality, it was evident that both antisense and sense transcripts were present at similar levels in both tissues (**Figure 3.17C**). However, phloem samples showed a slightly higher percentage of antisense transcripts compared to leaves, whereas leaf samples displayed a slightly higher percentage of sense transcripts compared to the phloem. Lastly, the exon distribution in both tissues revealed that nearly half of the transcripts comprise 2 exons, and approximately 25% contain 3 exons (**Figure 3.17D**). Further, a maximum exon number of 12 was observed in the phloem (0.086%). On the other hand, the maximum exon number found in leaf was 11 (0.088%).

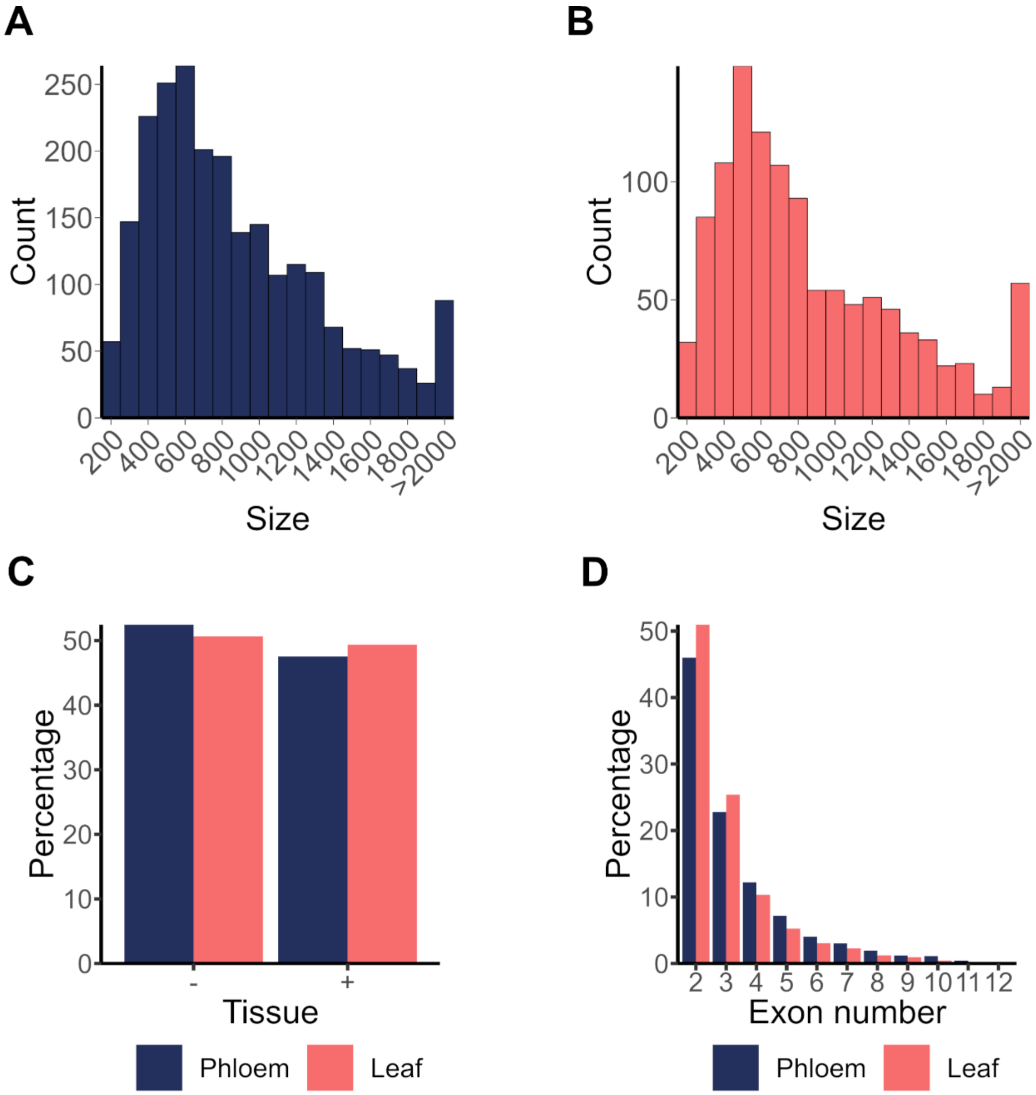


Figure 3.17− Exploring lncRNAs features. (**A**) Size distribution of lncRNAs from phloem sap. (**B**) Size distribution of lncRNAs from leaf samples. (**C**) Percentage of antisense (-) and sense (+) of lncRNAs in phloem (dark blue) and leaves (salmon pink). (**D**) Exon number distribution of lncRNAs in the phloem (dark blue) and in the leaf (salmon pink). Values are in percentage.

3.2.1.5. Differential expression analysis of common lncRNAs

The following step in the pipeline was the differential expression analysis of lncRNAs and was performed in-house. Differential expression analysis uses normalized count data and statistical tests to determine the differences in expression levels of genes or transcripts between conditions[141]. DESeq2[117] was chosen to perform differential expression analysis. This program takes count reads as input, then normalizes them according to library size and composition between samples and uses a negative binomial model to estimate the fold changes for each gene or transcript[117]. The count data of transcripts for each tissue was provided by Novogene. However, identification of common transcripts between phloem and leaf had to be done to provide a clean input file to DESeq2.

Data cleaning consisted of removing any transcript with counts lower than 5. In addition, lncRNA transcripts common to leaf and phloem had to be filtered based on the parent gene to which they were annotated, since the gene ID is universal, in contrast to the transcript ID, which is sample specific. Furthermore, Gene IDs starting with "XLOC" were removed since they are run-specific. Several transcripts matched the same gene ID. This was likely due to the presence of several isoforms. To avoid redundancy, the isoforms and/or transcripts that shared the same parent gene were merged by summing up their counts, resulting in a single and unique gene ID. Moreover, the directionality of the transcripts was not considered in this analysis. Finally, a list of 274 genes was obtained, which represented common parent genes between leaf and phloem (data not shown). DESeq2 was used in R/Bioconductor, and the script used for this analysis can be found in **Figure 7.9**. Across the 274 common parent genes, 67 showed significant differential expression (**Table 7.10**). The distribution between up- and downregulated parent genes was even, with a ratio close to 1. A volcano plot was generated to better visualize the results (**Figure 3.18**). For example, parent genes, such as, GSBRNA2T00105977001, ENSRNA049439209 and GSBRNA2T00003166001 were significantly downregulated in the phloem compared to the leaf. Whilst ENSRNA049479067, GSBRNA2T00157764001 and GSBRNA2T00015777001 were significantly upregulated.

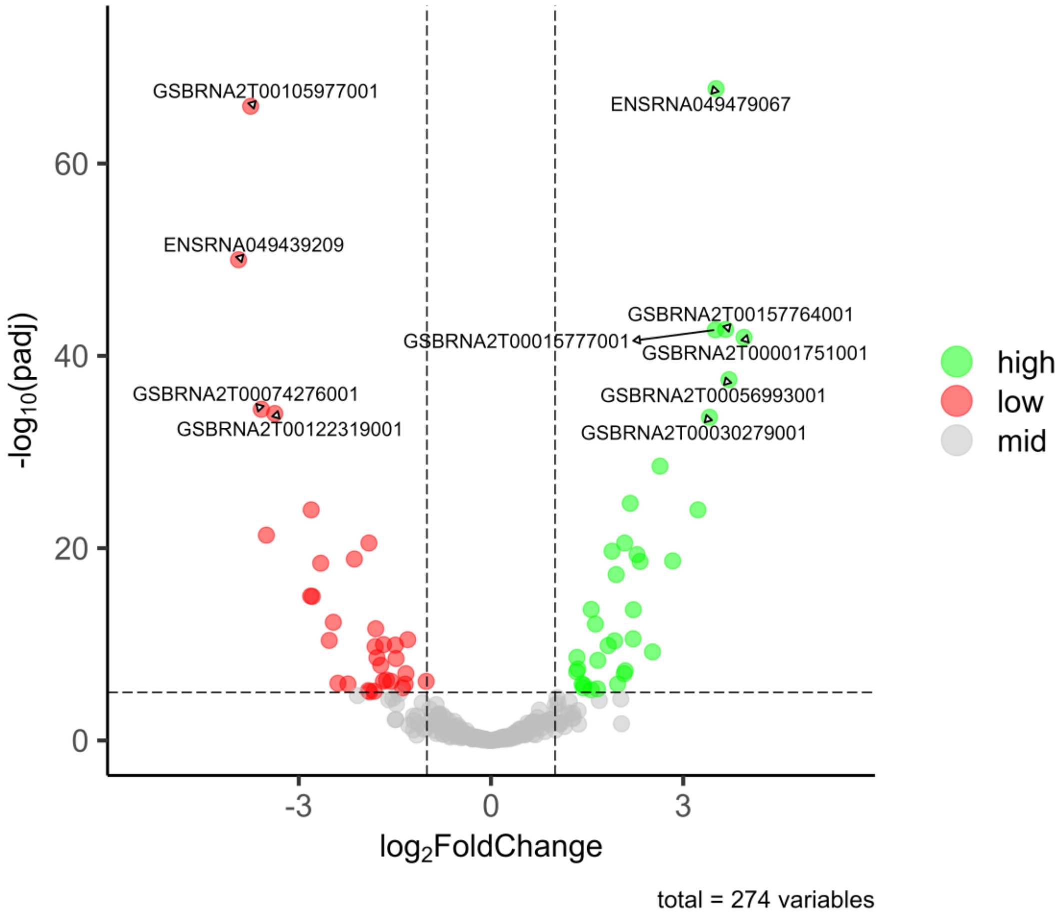


Figure 3.18− Volcano plot of differentially expressed parent genes between tissues. The volcano plot shows differentially expressed parent genes in phloem sap compared to the leaves. Differentially expressed genes are shown as upregulated (green/high), low or downregulated (red/low) and non-differentially expressed genes are shown grey/mid. Thresholds are shown as dashed lines. Vertical dashed lines between $\log_2(-1)$ and $\log_2(1)$ represent the minimum fold-change, and horizontal lines represent the minimum fold p-value [$-\log_{10}(1x10^{-5})$]. This figure was created in R/RStudio.

3.2.1.6. GO terms analysis

Gene ontology (GO) enrichment analysis was performed on the g:Profiler[123] webserver to gain insights into the expression profile of each tissue and to determine the predicted functions of the parent genes. The up- and downregulated parent genes in phloem samples were analyzed separately. The two gene lists were used as input and both searches were done by using the g:GOSt tool for *B. napus* organism. GO terms were divided according to the functional terms “molecular function”, “biological process”, and “cellular component”. In total, 171 GO terms were found in the downregulated gene set: 44 terms for molecular function, 96 terms for biological process, and 31 terms for cellular component. Whereas a total of 95 GO terms were found in the upregulated set of genes: 36 terms associated with molecular function, 48 terms with biological process, and 11 terms with cellular function. The GO terms in downregulated parent genes that were significantly enriched in the molecular function category were phospholipase A2 activity (GO:0004623) and glutamyl-tRNA reductase activity (GO:0008883) (**Figure 3.19A**). Upregulated genes did not have significantly enriched GO terms (**Figure 3.19B**). Furthermore, most of the GO terms were only counted once, with the exception for terms like binding (GO:0005488), protein binding (GO:0005515), RNA binding (GO:0003723), and ATP binding (GO:0005524), for example. However, these terms were not significantly enriched.

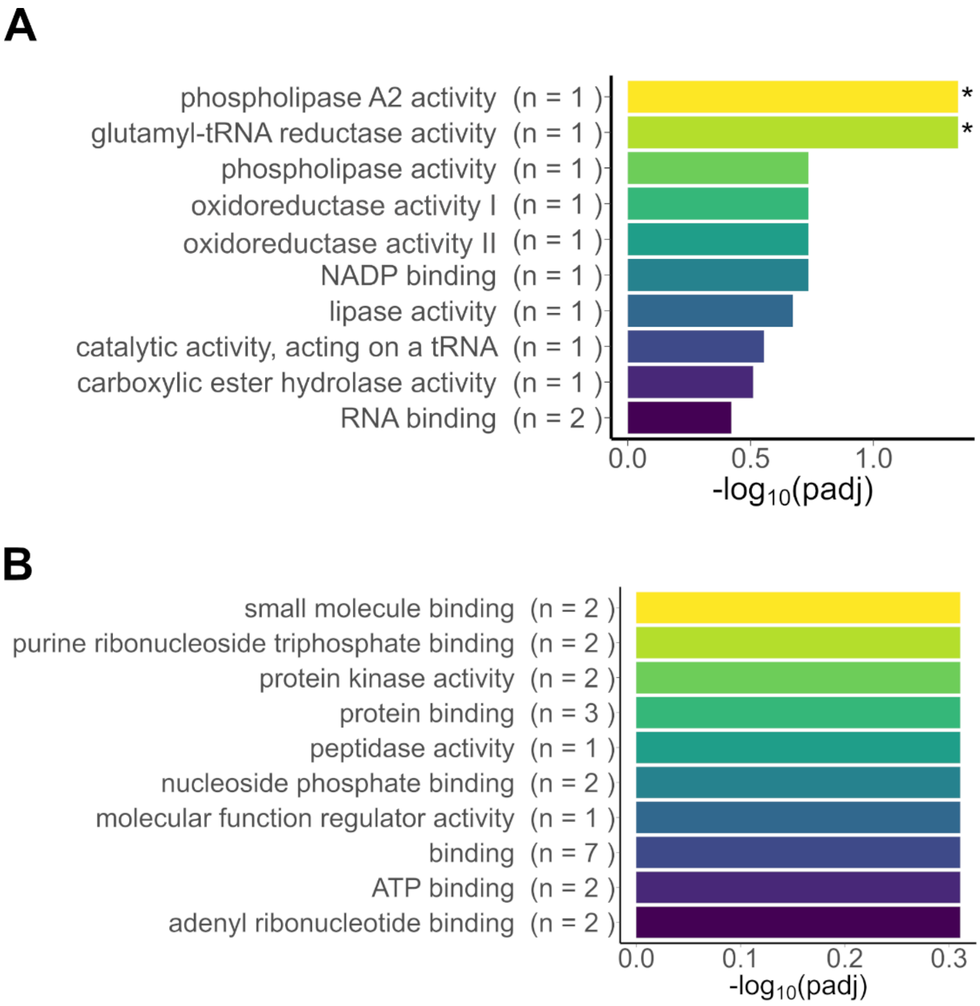


Figure 3.19− GO terms by molecular function of differentially expressed parent genes. (**A**) GO terms associated with downregulated genes. Oxidoreductase activity I: acting on the aldehyde or oxo groups of donors, NAD or NADP as acceptor. Oxidoreductase II: acting on the aldehyde or oxo group of donors. (**B**) GO terms associated with upregulated genes. Significant padj-values are shown with an *. The number of hits associated with a GO term is shown as n.

Not as many GO terms were found in the cellular component category, with only one term associated with the nuclear outer membrane (GO:0005640) being significantly enriched in the downregulated set (**Figure 3.20**). The upregulated set only had hits with padj values of 1, thus it is not presented here.

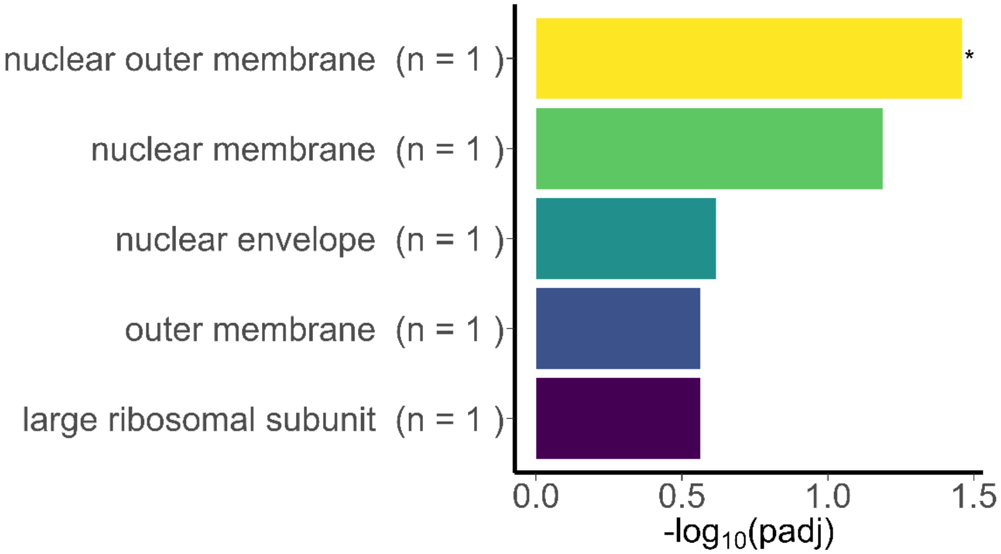


Figure 3.20− GO terms by cellular component of differentially expressed parent genes. GO terms associated with downregulated genes. Significant padj-values are shown with an *. The number of hits associated with a GO term is shown as n.

For the biological process category, 19 of the downregulated genes were significantly enriched and were involved in processes that ranged from response to calcium ion (GO:0051592) to regulation of stomatal closure (GO:0090333), or even long-chain fatty acid transport (GO:0015909) (**Figure 3.21A**). Once again, no significantly enriched GO terms were found in upregulated genes, and most terms had only one count (**Figure 3.21B**). It is important to mention that many of the GO terms were inferred either by phylogenetic principles, such as by biological aspect of Ancestor (IBA) or by electronic annotation (IEA).

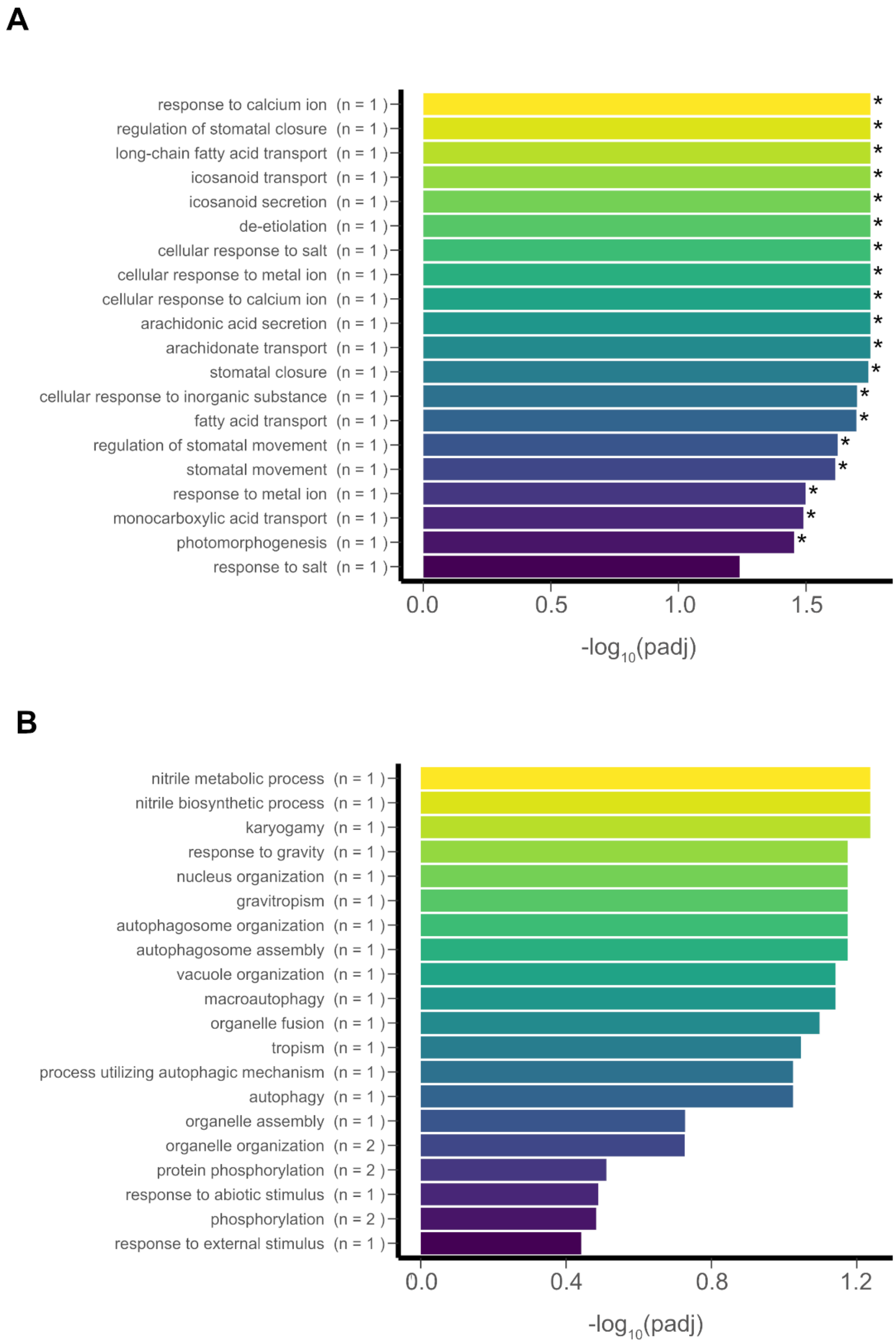


Figure 3.21− GO terms by biological process of differentially expressed parent genes. (**A**) GO terms associated with downregulated genes. (**B**) GO terms associated with upregulated genes. Significant padj-values are shown with an *. The number of hits associated with a GO term is shown as n.

4. Discussion

4.1. NURC1 has a conserved RRM

RBFs are key players that display a variety of crucial roles in ribosome biogenesis. Most of the knowledge about ribosome biogenesis and RBFs is derived from studies in humans and yeast. Nonetheless, these studies have led to the identification of hundreds of putative RBFs in plants[37–40]. NURC1 is a plant ortholog of yeast Nop15 and human NIFK RBFs. Nop15 and NIFK harbor a canonical RRM and are known to be involved in the processing and maturation of LSU rRNAs, as well as in cell cycle progression[45,47,48,57]. The RRM is one of the most abundant and highly conserved type of RNA-binding motifs[58,59]. RRM-containing proteins have been shown to be involved in RNA-related processes, such as mRNA and rRNA processing, RNA export, RNA stability, RNA editing and RNA degradation[58,136]. The plant NURC1 sequence and predicted secondary structure were analyzed (**Figure 3.1-Figure 3.2**). Sequence-wise, both termini had low conservation when compared to the orthologs in yeast and human. In contrast, the central residues that comprehend the RRM showed a higher level of conservation. Structural superimposition confirmed that the three RRM domains superimpose to a high degree, whilst the termini did not due to their disordered nature. The results also showed that NURC1 had a *bona fide* RRM, with conserved aromatic residues F105 and F108 in RNP1 and Y63 in RNP2. Typically, the side chains of aromatic residues participate in RNA binding via stacking interactions with the cognate nucleotides[136]. Additionally, acidic patches were observed in the C-terminus of NIFK and in the N-terminus of Nop15 but not in NURC1. Many RBPs possess intrinsically disordered regions characterized by repetitive sequences, such as the poly-D/E electronegative cluster[142]. In Nop15, the depletion of the electronegative cluster resulted in an increase of RNA-binding affinity[57]. Further, the Nop15 electronegative cluster and RRM were shown to interact with each other via electrostatic interactions, increasing protein stability and inhibiting nonspecific RNA binding[142]. However, the human NIFK electronegative cluster resides in the Ki67-FHA interaction domain[47], and whether it also interacts with the RRM remains to be determined. It would be interesting to further investigate why plants lost the poly-D/E acidic cluster and if there is any other cluster or motif that plays a role in structural stability and RNA binding specificity of NURC1.

4.2. NURC1 polydispersity is concentration dependent

Full-length NURC1 was successfully purified (**Figure 3.3**). To assess the homogeneity of the sample prior to structural analysis, it was characterized by native-PAGE and DLS. The native-

PAGE results indicated that NURC1$_{FL}$ samples were heterogeneous at different concentrations, characterized by multiple sharp bands accompanied by smearing (**Figure 3.5**). Moreover, the intensity of the larger bands increased with the protein concentration, suggesting that larger species were more prevalent at higher concentrations. DLS analysis confirmed the polydispersity of the NURC1$_{FL}$ samples, further corroborating a concentration-dependent broad range of particle sizes (**Figure 3.6A**). Additionally, batch SAXS measurements showed that the R_g value increased with sample concentration (**Figure 3.7**). The NURC1$_{FL}$ monomer is approximately 25 kDa, suggesting an average 2 nm radius in a spherical-like particle[143]. In contrast, the AlphaFold structure of NURC1 suggested a non-spherical shape with two flexible and elongated termini (**Figure 3.2**). These structural features were confirmed by SEC-SAXS and will be later discussed in detail (**chapter 4.4**). Accordingly, NURC1 could adopt several conformations in solution which could then result in slightly different radii. Nevertheless, the variety of sizes was too broad to be explained only by the presence of different conformational states. The presence of different oligomers/aggregates likely explains the larger sizes. Indeed, the formation of oligomers or high molecular weight species is common at high protein concentrations[144,145]. Since the samples used in all downstream analyses corresponded to the monomer peak of SEC (**Figure 3.3D**), it is possible that the oligomerization/aggregation occurred after SEC (e.g., because of freezing and thawing of samples). Nonetheless, a second run of SEC appeared to restore the monomeric state, as observed in singular SEC runs, of SAXS and MALLS experiments (**Figure 3.8A-Figure 7.13A**). The oligomers/aggregates may be classified as dissociable given that they possibly reverted to a monomeric state by manipulating the solution conditions[145,146]. In this case, seemingly due to the pressure increase induced during SEC. It is known that factors other than concentration can modulate intermolecular interactions, such as pH, ionic strength, cofactors, nucleic acids, etc[144]. Hence, to further understand the effect of these factors on NURC1 stability and polydispersity, they could be screened using DLS, SAXS and Thermal Shift Assays (TSA), for example. Also, following SEC, the samples could be monitored over time to track the occurrence of new species.

In summary, the results of native-PAGE, DLS and SAXS experiments all corroborated the polydisperse nature of NURC1$_{FL}$ in solution and the particle size distribution seemed to be dependent on protein concentration. Furthermore, this effect could be reversed during SEC runs, the pressure increase being a prime suspect as the defining factor.

4.3. Both termini are important for the stability of NURC1

In addition to full-length NURC1, three other versions were generated and successfully purified. $NURC1_{1\text{-}160}$ that lacked the predicted α-helix at the C-terminus, $NURC1_{1\text{-}140}$ that had the C-tail completely removed, and $NURC1_{53\text{-}222}$ which was devoid of the highly disordered N-terminus (**Figure 7.10**). Sample homogeneity was investigated through DLS as well (**Figure 3.6B**). First, the size distribution between $NURC1_{FL}$ and $NURC1_{1\text{-}160}$ did not show major differences, except for a broadening of the $NURC1_{1\text{-}160}$ peak, which comprised particles of smaller sizes. In contrast, $NURC1_{1\text{-}140}$ demonstrated a major broadening with several unresolved peaks. Therefore, it appeared that truncating at position 160 did not have a significant impact on sample polydispersity, while further truncation of 20 amino acids did. In yeast Nop15, complete removal of the C-terminal residues resulted in aggregation[57]. Moreover, the crystal structure of Nop15 showed that aliphatic residues L167, I173 and L176 from the C-terminus interact with the aromatic residues from both RNPs in the absence of RNA (**Figure 1.5**), suggesting that these interactions promoted Nop15 solubility[57]. NURC1 harbors several aliphatic residues between positions 140 and 160. Notably, among these only one leucine (L145) and no isoleucine residue is present. However, three valines (V141, V156, V159) residues are found. Valine shares similarities with leucine and isoleucine as a non-polar aliphatic amino acid, characterized by a shorter branched side chain: an isopropyl group. Thus, these residues may play a role in stabilizing the NURC1 protein, similar to Nop15. Mutation of these residues could help uncover if they are indeed responsible for structural stability of NURC1. Lastly, the truncation of N-terminal residues led to severe aggregation. Unlike Nop15, the N-terminus of NURC1 does not harbor an acidic patch (**4.1**). As described earlier, this patch increased not only protein stability but also the RNA binding specificity via electrostatic interactions with the electropositive sites of the RRM[142]. Therefore, if the N-terminus of NURC1 also participates in electrostatic interactions with the RRM it is not through this particular cluster. Poly-S patches in the N-terminus were observed in both Nop15 and NURC1. No reports about the role of this particular patch in Nop15 were found. The side chain of serine contains a hydroxyl group (-OH) that can be phosphorylated. Hence, the phosphate group could hypothetically compensate for the absence of the acidic patch in NURC1 by providing the missing negative charge, although its bulkier nature might pose structural challenges. This region could also fulfil other roles, such as promoting temporary protein-protein interaction[147]. This would be fitting, considering that the yeast ortholog has been shown to participate in a pre-60S complex[51]. It is challenging to draw any conclusions about the real roles of the N-terminus with the current information. Similar to $NURC1_{FL}$, sample heterogeneity of the truncated versions could be reversed by SEC, except for $NURC1_{1\text{-}140}$ that showed peaks before and after the monomer peak (**Figure 7.13A**). These results suggest that NURC1$_{1\text{-}140}$ intermediary species are irreversible and could potentially be separated for further analysis. Taking everything into account, both termini seem to be essential for the stability of the NURC1 protein, with the N-terminus exerting a more pronounced impact.

4.4. NURC1 has an elongated shape and is flexible in solution

SAXS is a powerful technique to analyze macromolecules in their quasi-native environment[148]. Parameters such as overall molecular size and weight, as well as the degree of flexibility/compactness can be obtained from the experimental scattering pattern[148,149]. Usually, a high percentage of monodispersity is required (>95%)[149] . Thus, due to its polydispersity, SEC coupled to SAXS was chosen to study the structural characteristics of NURC1$_{FL}$ in solution. The Guinier analysis of the single SEC peak resulted in a R_g value of 3.52 ± 0.01 nm (**Figure 3.8A-B**), which provided a measure of the overall size of the NURC1 protein. The linearity of the fitted experimental points suggested that the sample contained particles of similar sizes with a slightly higher molecular weight of ~33.8 kDa. Usually, unstructured proteins are characterized by large average sizes due to the presence of highly extended conformations[149]. Consequently, the overestimation of NURC1$_{FL}$ size might be explained by the presence of highly flexible and/or disordered termini, as seen in the predicted AlphaFold structure (**Figure 3.2**). Moreover, the monodispersed and monomeric nature of the sample was corroborated by SEC-MALLS measurements, where the single eluted peak yielded a molecular weight of approximately 28 kDa (**Figure 7.13A-B**). Additionally, a R_h of 2.4 nm was recorded by the in-line QLES module (**Figure 7.13C**). The R_g/R_h ratio of globular proteins is usually ~0.77, whilst for NURC1 it was around 1.5, suggesting an elongated shape. The dimensionless Kratky plot and the pair distribution function also confirmed the elongated and/or partially unfolded shape of NURC1 (**Figure 3.8C-D**). Two *ab initio* models were generated by using GASBOR and DAMMIN (**Table 7.2**). Both dummies displayed an elongated shape (**Figure 3.9**). Among four structures the AlphaFold model provided the best CRYSOL fit to the scattering data of NURC1$_{FL}$ (**Table 7.2**). However, even after using SREFLEX to obtain different conformers the χ^2 was still above one. A better fit may be achieved by using the Ensemble Method Optimization (EOM), keeping the RRM core as a rigid body while providing the sequences for both termini. Superimposition between NURC1$_{FL}$ *ab initio* models and the SREFLEX model clearly showed that the flexible termini pose a challenge for structure prediction (**Figure 3.9**), since many conformers can occur in solution and the scattering data is an average of them.

In conclusion, NURC1$_{FL}$ possesses an elongated and/or partially unfolded structure that is flexible in solution. SAXS and NMR studies on yeast Nop15 revealed that its C-terminus is highly flexible in the absence of RNA and that tumbles around the RRM[57]. In the presence of RNA, the αC helix at the C-terminus of Nop15 interacts with the ITS2 RNA, as well as with Cic1[51]. Information about the N-terminus of Nop15 is still missing. Nonetheless, flexibility seems to be intrinsic to Nop15 and NURC1, which can be underlying to the molecular functions they exert. The structural flexibility likely prevented successful crystallization of NURC1$_{FL}$, NURC1$_{1-160}$ and NURC1$_{1-140}$ since crystallization of Nop15 was only achieved by truncating the N-terminus and a portion of the C-terminus[57]. Therefore, crystallization attempts could be made with NURC1$_{53-222}$ or with a fourth version that lacks entirely or partially both termini. Crystallization, Cryo-EM and/or NMR of NURC1$_{53-222}$ would allow to confirm if the predicted α-helix at the C-terminus of NURC1 resembles the αC helix of Nop15.

4.5. NURC1 binds the ITS2 RNA and its binding is impacted by C-terminus residues

Similar to its orthologs, NURC1 is likely involved in processing and maturation of LSU rRNAs. Yeast Nop15 harbors a canonical RRM, and its RNA binding capacity has been shown[57]. Identically, NURC1 harbors an RRM, therefore the RNA binding function was expected. Consequently, the binding between the NURC1 protein and ITS2 RNA of *A. thaliana* was studied through MST. Indeed, all versions of NURC1 bound to ITS2 RNA, confirming the RNA binding ability of NURC1 (**Figure 7.15A**). Much like the Nop15 protein[57], truncation of the C-terminus right after the RRM (NURC1$_{1-140}$) resulted in a large decrease of the RNA binding affinity (**Figure 7.15A**, **Table 7.3**). In contrast, removal of the last 62 residues (NURC1$_{1-160}$) did not significantly affect the binding affinity compared to the full-length version (**Figure 7.15A**, **Table 7.3**). In Nop15, C-terminal residues are paradoxically important for the recognition of RNA, although they sequester the RNA binding residues of the RRM[57]. Also, this region comprises a series of α-helices that rearrange upon RNA binding[57]. Some level of conservation in this region was observed between NURC1 and its orthologs (**Figure 3.1**). Additionally, the AlphaFold structure of NURC1 predicted two α-helix at the C-terminus (**Figure 3.9**). Hence, NURC1 might have a similar binding behavior to yeast Nop15. Once again, high resolution methods would be necessary to determine the exact structure of the C-terminus to elucidate the molecular mechanism behind the binding between NURC1 and ITS2 RNA.

Furthermore, ITS2 RNA was screened to determine which stem-loop is bound by NURC1 (**Figure 7.15B**, **Table 7.3**). However, no major differences were found across the different ITS2 fragments. NURC1 showed higher affinity toward fragments that contained the stems I+II ($ITS2_{1-79}$) and stem III ($ITS2_{82-162}$) (**Figure 3.10A**). In contrast, $ITS2_{157-187,}$ which hypothetically harbors stem IV, was bound by NURC1 with less affinity. (**Figure 3.10A**). The predicted structures of $ITS2_{1-79}$ and $ITS2_{82-162}$ revealed some level of structural similarity, while $ITS2_{157-187}$ did not (**Figure 7.14**). Yeast Nop15 has been shown to recognize RNA structure over sequence[45,57]. Thus, these results suggest that NURC1 shares this characteristic behavior with the yeast ortholog.

4.6. NURC1 has RNA chaperone-like activity

In *Saccharomyces cerevisiae*, processing of ITS2 involves a dynamic conformational change between two models: the ring and the hairpin[56]. The disruption of the ring form resulted in the complete arrest of the 25S rRNA maturation, whereas disruption of the hairpin structure resulted only in reduced levels of the mature product[56]. Furthermore, it has been proposed that the ring form is required for recognition by the processing machinery, thus vital for a correct processing, and that the hairpin structure facilitates the efficiency of processing[56,57]. A major rearrangement of the ITS2 stem III.A in the ring conformation (the Nop15 binding site) is required for a correct transition to the hairpin form[56]. It seems that Nop15 binds and stabilizes the stem III.A holding the ITS2 in the ring conformation, which allows the assembly of other RBFs[57]. RNA chaperones are proteins that aid the RNA molecule acquire the correct and most stable structure by resolving kinetically trapped misfolded conformations[133]. This definition appears to apply to the yeast Nop15 and the same could be expected for the plant RBF NURC1. Subsequently, the chaperone activity of NURC1 was investigated (**Figure 3.11**, **Table 3.1**). NURC1 exhibited a chaperone-like activity and removal of the C-terminal residues resulted in a decrease of the annealing constant, whereas removal of the N-terminal residues slightly increased annealing activity (**Figure 3.11**, **Table 3.1**). Much like in the case of RNA binding (**chapter 4.5**), the chaperone function of NURC1 seemed to be impacted by the C-terminal tail residues. The reason why removal of the highly disordered N-terminus improved chaperone activity still needs to be explained, as there is no apparent structural similarity in this region to the Nop15 ortholog. Moreover, further studies focused on the *A. thaliana* ITS2 conformations and their stabilization by NURC1 and other plant RBFs are necessary.

4.7. *B. napus* phloem contains a diverse population of lncRNAs

The phloem is a conduit for the transport of a wide array of molecules. All classes of RNAs have been found in the phloem of several plant species, including lncRNAs[84,86–88,150]. Phloem lncRNAs have been identified and characterized in several plant species[87,115,150]. However, the lncRNA population of *B. napus* phloem remains to be explored. Hence, the identification of lncRNAs in the *B. napus* phloem was carried out. The expression profile of leaf lncRNAs was also investigated. For that, pair-end Illumina sequencing was performed.

In total, 3468 putative lncRNAs were identified of which 1142 were found in leaves and 2326 in phloem sap (**Figure 3.15**). 274 common lncRNAs were found based on the annotated parent gene IDs. Accordingly, 868 and 2052 lncRNAs were unique to leaf and phloem samples, respectively. The expression of lncRNAs is known to be tissue-specific, which could easily explain the differences between leaf and phloem populations. Tissue profiling in apple also revealed differences in the unique sets of lncRNAs, with 491 and 243 lncRNAs identified in phloem and leaves, respectively[115]. The expression of lncRNAs fluctuates with developmental stage, stress responses, or even the circadian rhythm[104,151]. Therefore, it is important to mention that leaf and phloem samples were not collected from the same plant nor at the same timepoint, which could have introduced some variation. Additionally, differences in the library preparation or execution of sequencing could be other sources of variation, since leaf and phloem samples were sequenced at different timepoints. However, this scenario seems less likely because the samples were sequenced at Novogene with the same pre-requisites and parameters. Moreover, when looking at the QC summary, the values seemed to be in the acceptable range. The sequencing depth was similar across phloem and leaf samples, and the Q30 values were above 90% (**Table 3.2**). Interestingly, mapping of leaf samples showed a high percentage of multiple mapped reads (~50%) in comparison to phloem samples (~20%). Multiple mapped reads are typically in the range of 5 to 40% of total reads mapped[152]. These are reads that align equally well at more than one location in the genome due to the presence of gene duplicates. Duplication is essential for evolution and can occur through different mechanisms such as recombination, whole genome duplication, or transposable elements[152]. *B. napus* is an allotetraploid hybrid that had its whole genome duplicated (**chapter 1.1**). Therefore, a higher percentage of multi-mapping is not entirely surprising. The difference of 30% between leaf and phloem samples is likely due to the different levels of transcriptome complexity in the two tissues. Leaves have a high intricacy of molecular processes, such as photosynthesis, that require complex and dynamic gene expression patterns when compared to the phloem. Higher complexity could explain a higher multi-mapping rate because of the presence of extensive gene families or genes with highly similar paralogs.

A substantial portion of the reads were mapped to protein-coding regions (**Figure 3.13**). This result was expected since the genome of *B. napus* has more protein-coding genes annotated than non-coding genes and, thus far, it seems that there is no available annotation for lncRNAs. The signal recognition particle (SRP) RNA was found to be ~5 times more prevalent in the phloem compared to the leaf. The SRP RNA is part of the SRP ribonucleoprotein complex and ensures the proper assembly of SRP proteins by acting as a scaffold[153]. The SRP complex targets the signal peptide at the ribosomal exit site, halting the translation of the protein until association with the SRP receptor at the endoplasmic reticulum (ER)[154]. In eukaryotes, upon release of the SRP complex from the signal peptide, translation is resumed, and the protein is translocated co-translationally to the membrane of the ER[153,154]. SRP RNA is mainly found in the cytosol[154]. Nonetheless, SRP RNA has been previously detected in pumpkin and *B. napus* phloem[84,92]. Moreover, circular SRP RNAs were recently identified in the phloem of *B. napus*[134]. Therefore, the presence of SRP RNAs in the phloem is not a new finding, but the higher incidence in phloem compared to leaf is. The experiment should be repeated with phloem and leaf samples from the same plants to confirm if the difference of SRP RNA percentage between the tissues is real and not prompted by biological differences. Components of the translation machinery have been found in the phloem, but thus far there is no evidence of active protein synthesis[84]. Furthermore, addition of phloem RNA to an *in vitro* translation reaction arrested protein synthesis[92]. The role of SRP RNA in the phloem is not yet understood and it would be interesting to see if other members of the SRP complex are also present in phloem sap.

The lncRNAs classes present in phloem and leaf tissues were explored. No major differences between phloem and leaf lncRNA classes were observed (**Figure 3.15**). Several classes were present, such as sense intronic, antisense, and sense overlapping. However, both tissues showed a high percentage of lincRNAs. Intergenic lncRNAs are transcribed from regions that do not overlap with annotated protein coding genes. This class represents over half of the lncRNA transcripts in humans and their poor conservation poses challenges in predicting their functions across species[155]. In *A. thaliana*, over 6,000 lincRNAs have been identified by a strict systematic genome wide search[156]. Another study identified ~130,000 lincRNAs in four *Brassicaceae*, including *A. thaliana* and *B. rapa*[157]. Furthermore, lincRNAs exhibited a higher prevalence compared to other lncRNA classes across various tissues and plant species[86,87,112,115,150]. Thus, lincRNAs seem to be a prevailing class not only within diverse tissues, but also across multiple species. However, only a few have been experimentally confirmed and characterized. Furthermore, the lack of consistent classification hinders the discovery of these cases in literature, given that many authors either did not categorize them as lincRNAs or grouped them collectively as “lncRNAs”[155]. Examples of experimentally characterized lincRNAs in plants include the INDUCED BY PHOSPHATE STARVATION (IPS1) and AT4 lincRNAs, both of which serve as target mimics for miRNAs[156].

The quantification of the transcripts was also performed. The FPKM distribution of transcripts and lncRNAs looked similar across replicates, which indicated consistency and reproducibility (**Figure 3.14** and **Figure 7.23**). In phloem samples, the most expressed lncRNA (TCONS_00053408) and an antisense analog (TCONS_00053411) were annotated to the ENSRNA049479067 gene. The gene is associated with a 643 bp long eukaryotic SSU rRNA that contains a single exon (**Figure 7.24**). In the Ensembl Plants database, there was no information about the function of this ribosomal gene. Moreover, the question remains as to why these lncRNAs are found in the phloem and what roles they might play in an environment where ribosome biogenesis does not occur. Another highly expressed putative lncRNA was associated with the gene GSBRNA2T00091120001. This gene encodes the BnaA08g02520D protein that is predicted to be involved in lipid metabolic processes and a phosphoric diester hydrolase activity. In leaf samples, the most abundant lncRNA (TCONS_00018173) was annotated to the GSBRNA2T00002257001 gene, which encodes the BnaC04g03080D protein, a mitogen-activated protein kinase. A novel set of lncRNAs close to the SRP RNA ENSRNA049474326 was discovered in both tissues. In phloem samples, two variants (sense and antisense) were found, while in leaves only a sense version was identified. Based on the FPKM values, the phloem antisense variant was expressed at a lower level than the sense variant. In leaves, the sense lncRNA had a value below one. This supports the finding that SRP RNAs appear at a higher percentage in the phloem than in the leaf. The SRP RNA itself can be considered a lncRNA, thus finding another lncRNA associated with it is intriguing. Even more interesting is the exon distribution of these novel lncRNAs. One exon spans a significant portion of the ENSRNA049474326 SRP RNA, while the other spans a small portion of the ENSRNA049474308 SRP RNA (**Figure 3.16**). It is possible that the novel lncRNAs have bifunctionality and regulate both genes. Another hypothesis could be that each single exon ncRNA locus is part of a larger transcriptional unit that includes multiple isoforms and are not annotated as such. In conclusion, even though many novel lncRNAs have been identified in both phloem and leaf samples, they still lack proper annotation compared to coding transcripts. Those with annotated parent genes suggest that the lncRNAs are associated with both coding and non-coding genes that are involved in diverse and distinct cellular processes.

Several structural features of the leaf and phloem lncRNAs were investigated to clarify if there were any differences between the two populations. The size distribution of lncRNAs was determined (**Figure 3.17A-B**). Both tissues displayed a similar size distribution and most transcripts had sizes between 400 and 800 nucleotides. The directionality of the transcripts was also investigated (**Figure 3.17C**). No major differences were observed between the two tissues or between sense and antisense transcripts. The third feature that was analyzed was the exon distribution (**Figure 3.17D**). Once more, no major differences between the phloem and leaf lncRNA populations were found. Most transcripts possessed two exons. Based on these results, there were no evident structural differences between the lncRNA populations of leaf and phloem samples. Some studies in different plant species, tissues and/or developmental stages have also reported similar features. For example, in many cases the size distribution of lncRNAs was not so different from the one observed in this work. Most lncRNA transcripts exhibited sizes between 200 and 800 nucleotides, not only in *B. napus*[158] but also in species like cucumber[87], apple[86,115] and carrot[150]. These studies also demonstrated that most of the lncRNAs possessed either one or two exons.

The next step was to determine if there were any differentially expressed genes. A list of 274 common genes between leaf and phloem was analyzed with the DESeq2 program. From those, 67 genes were found be significantly differential expressed in the phloem compared to leaves, and the number of up and downregulated genes was equally distributed (**Figure 3.18** and **Table 7.10**). Transcripts associated with the parent gene ENSRNA049479067 were significantly upregulated in the phloem compared to leaves, which includes the sense TCONS_00053408 and antisense TCONS_00053411 lncRNAs. As discussed above, the TCONS_00053408 transcript was the most expressed lncRNA in the phloem and its parent gene corresponds to a eukaryotic SSU rRNA. Examples of other significantly upregulated parent genes were the GSBRNA2T00157764001 and GSBRNA2T00015777001, which correspond to the protein coding genes BnaC09g23130D and BnaA09g30280D, respectively. No experimental evidence for the existence, localization and/or function of the corresponding proteins could be found. In the downregulated category, the parent genes like GSBRNA2T00105977001, ENSRNA049439209 and GSBRNA2T00003166001 were significantly downregulated in the phloem compared to leaves. The proteins coded by the two GSBRNA genes remain to be characterized, while ENSRNA049439209 is annotated as a threonine specific tRNA. Yet, another translation-related parent gene was identified in the phloem. This analysis unraveled which parent genes were differentially expressed in phloem samples from *B. napus*. However, the analysis was limited to common annotated genes between leaf and phloem samples, leaving non-annotated genes out of the picture. Further, information about the differential expression of specific lncRNAs is missing because the counts of transcripts belonging to the same parent gene were merged. The analysis could be improved by searching common transcripts based on their genomic location rather than relying on the annotation. This would allow the direct identification of differentially expressed lncRNA transcripts, eliminating the need for tracing case by case to know which transcripts were associated with the differentially expressed parent gene. Nonetheless, using only lncRNAs that had annotated parent genes permitted the next and last analysis performed in this thesis.

The GO term analysis was conducted to understand which molecular functions, biological processes and cellular components were associated with the differentially expressed lncRNAs based on their parent gene annotations. A total of 171 and 95 GO terms were found in the down- and upregulated set of parent genes, respectively. Molecular functions such as phospholipase A2 activity and glutamyl-tRNA reductase activity were found to be significantly enriched in the downregulated set (**Figure 3.19A**). In the upregulated set, most lncRNAs parent genes had terms related to binding (**Figure 3.19B**), ranging from general binding (n=7) to small molecule binding (n=2) and protein binding (n=3). Terms related to catalytic activity, such as protein kinase and peptidase activity, were also present. These molecular functions are not uncommon, as phospholipase A and other lipases[159], binding proteins (RNA, GTP, etc)[16], kinases[160], peptidases[83,161] have been previously identified in phloem exudate of different plant species. As for the glutamyl-tRNA reductase activity, it is related to an enzyme involved in the biosynthesis of a tetrapyrrole precursor and tetrapyrrole is an essential compound for the synthesis of chlorophyll[162,163]. For the cellular component, GO terms related to the nucleus and ribosome were found in the downregulated set (**Figure 3.20**). Since phloem SE are enucleated (**chapter 1.5.2**), the presence of such GO terms may initially seem odd. However, these could simply be cellular remnants, especially when considering the concurrent lack of RNase activity in the phloem sap[68]. Moreover, as previously mentioned, parts of an inactive ribosomal machinery have been found in phloem sap[84]. Ultimately, these GO terms are in fact expressed at lower levels in the phloem sap samples as compared to the leaf samples. Finally, in the biological process category many processes were significantly enriched in the downregulated gene sets, whilst there were no significant processes found in the upregulated set (**Figure 3.21**). This diversity might reflect the multifaceted roles that lncRNAs can play in diverse biological processes. These findings should be interpreted with due consideration since the GO terms were annotated either by IBA or IEA, which means that the annotation was inferred either by phylogenic relation or electronically without manual curation, respectively. Hence, none of the *B. napus* parent genes had their GO terms experimentally proven.

This analysis provided a preliminary overview of the lncRNAs populations in leaf and phloem samples from *B. napus*. While many transcripts were identified as lncRNAs, their presence in the tissues should be confirmed through other techniques, such as RT-qPCR and northern blotting, to confirm their authenticity. Furthermore, to validate the differential expression results the sequencing of lncRNAs should be repeated with leaf and phloem samples that belong to the same pool of individuals.

5. Conclusions and Future Perspectives

Two topics were explored in this thesis: ribosome biogenesis and the long non-coding transcriptome of the phloem.

The first part aimed at studying the structure and function of the RBF NURC1 in *A. thaliana* to better understand its role in plant ribosome biogenesis. NURC1 was successfully purified along with three other versions that lacked different structural features. The sequence analysis of NURC1 revealed a conserved *bona fide* RRM domain flanked by two termini. Based on the secondary structure predicted by AlphaFold both termini are likely highly flexible, and the N-terminus possesses a more disordered structure compared to the C-terminus. The structural flexibility of NURC1 was corroborated by the SEC-SAXS data obtained in this work. SEC-SAXS results also showed that the sample was monomeric and that NURC1 displayed an elongated shape in solution. SEC-MALLS results agreed with the presence of monomers in solution. However, the sample homogeneity was highly compromised when increasing the protein concentration in batch measurements as it was observed by DLS, native electrophoresis and SAXS. Moreover, complete truncation of the N- or C-terminus led to an increase in the sample polydispersity possibly due to the structural destabilization of NURC1, and this effect was more prominent in $NURC1_{53\text{-}222}$. Therefore, both termini seem to be essential for the stability of NURC1 with the N-terminus exerting a more pronounced effect. In yeast Nop15, several residues of both termini were shown to not only play a role in structural stability but also RNA binding specificity[57,142]. More specifically, Nop15 holds an electronegative cluster at the N-terminus as well as aliphatic residues at the C-terminus[57,142]. However, no such cluster was observed at the N-terminus of NURC1. As a first approach to better comprehend the roles of each terminus in the structural stabilization of NURC1, bioinformatic tools could be employed to find repetitive clusters and/or novel motifs in the sequence of NURC1. Bioinformatic tools such as ScanProsite tool[164] or Motif Scan[165] may be used, for example.

The Nop15 paralog has been shown to bind RNA and it has been proposed that it binds and holds the ITS2 in the ring conformation, preventing premature processing of the rRNA[57]. The binding capacity of NURC1 was also investigated. Certainly, NURC1 full-length bound to the *A. thaliana* ITS2 RNA with a K_d within the nanomolar range. Truncation at the C-terminus led to a decrease in binding affinity, suggesting that C-terminal residues are essential for RNA recognition and/or binding. This was even more critical when further removing residues between positions 140 and 160, indicating the importance of these residues in the stability and RNA binding capability of

NURC1. Interestingly, this region harbors several aliphatic residues but to what extent these play a similar stabilizing role as the aliphatic residues at the C-terminus of Nop15 remains to be determined. Binding by the N-terminally truncated version was observed, but it was not quantified due to fluorescence quenching at higher protein concentrations. Buffer optimization or temperature control of the samples could be adjusted to enhance solubility of the complex, facilitating the determination of the affinity constant between NURC1$_{53\text{-}222}$ and ITS2. The binding of NURC1 to different stem-loops of ITS2 was also investigated. However, it was not clear which fragment of ITS2 RNA is preferably bound by NURC1. Lastly, NURC1 also exhibited RNA chaperone-like activity, which might be intimately connected to its potential role in the stabilization of the ITS2 conformation in *A. thaliana*.

Determining the atomic structure of the NURC1 protein and of the NURC1-RNA complex through high-resolution techniques such as X-ray crystallography and/or NMR would allow to gain insightful and valuable structural information in the absence and presence of the cognate RNA. SEC-SAXS experiments with the RNP complex would also be interesting to perform since it could unravel how RNA binding would affect the flexibility and overall shape of NURC1. Future work should also include mapping of the binding site at the RNA and/or protein level through RNase/chemical footprinting- or crosslinking-based methods in tandem with MS or sequencing[166–170]. In fact, the mapping of Nop15 binding site as well of other yeast RBFs to the ITS2 was achieved by crosslinking and analysis of cDNA (CRAC) and dimethylsulphate (DMS) footprinting[46]. Finally, it would be truly interesting to explore which other RBFs interact with NURC1 and the ITS2 RNA to gain insights into the plant pre-60S ribosomal particle components. The protocol in Oeffinger *et al*. 2007[49] describes a rapid affinity purification of RNP complexes using Nop15 as bait which may be adapted to NURC1, for example.

The second part of this thesis focused on exploring the long non-coding transcriptome of *B. napus* phloem sap. Leaf samples were also analyzed to aid in the identification of phloem-specific transcripts and to obtain a tissue profile. Over 3000 putative lncRNAs were identified in phloem and leaf samples from *B. napus*. More than 200 long non-coding transcripts were annotated to common parent genes between phloem and leaves. Differential expression analysis was performed to determine if there was a difference or change in the expression levels of common parent genes between the two tissues. Furthermore, the GO terms associated with up- and downregulated parent genes were investigated. A wide variety of GO terms were found, which might reflect the versatile functions that lncRNAs play in various biological processes. Additionally, transcripts associated with the SRP RNA were found in higher percentages in phloem than in leaf samples. Hence, it would be interesting to investigate if other members of the SRP complex are also present and enriched in phloem sap. In conclusion, this analysis allowed a first look into the lncRNA profile of leaves and phloem sap of *B. napus*. However, it is critical to repeat the experiment with leaf and phloem samples coming from the same pool of individuals and collected at the same timepoint. Moreover, the authenticity of candidate lncRNAs must be experimentally proven by complementary techniques such as RT-qPCR and northern blotting, for example. The content and expression levels of coding transcripts may also be explored, and the study of lncRNA-mRNA co-expression networks might be an interesting next step. It would also be interesting to find out if any of the identified lncRNAs exhibit long-distance mobility and if they participate in RNP complexes.

6. References

1. Knapp, S. People and plants: The unbreakable bond. *Plants People Planet* **1**, 20–26 (2019).
2. Crepet, W. L. Progress in understanding angiosperm history, success, and relationships: Darwin’s abominably ‘perplexing phenomenon’. *Proc Natl Acad Sci U S A* **97**, 12939–12941 (2000).
3. Taiz, L. & Zeiger, E. Plant Physiology. in *Plant Physiology* 2 (Sinauer Associates Inc., 2010).
4. Koornneef, M. & Meinke, D. The development of Arabidopsis as a model plant. *Plant Journal* **61**, 909–921 (2010).
5. Woodward, A. W. & Bartel, B. Biology in bloom: A primer on the arabidopsis thaliana model system. *Genetics* **208**, 1337–1349 (2018).
6. SzydłOwska-Czerniak, A. Rapeseed and its Products-Sources of Bioactive Compounds: A Review of their Characteristics and Analysis. *Crit Rev Food Sci Nutr* **53**, 307–330 (2013).
7. L. Iniguez-Luy, F. & L. Federico, M. The genetics of Brassica napus. in *Genetics and Genomics of the Brassicaceae* vol. 9 291–322 (Springer, 2011).
8. Kaur, H. *et al.* The impact of reducing fatty acid desaturation on the composition and thermal stability of rapeseed oil. *Plant Biotechnol J* **18**, 983–991 (2020).
9. Wang, P., Xiong, X., Zhang, X., Wu, G. & Liu, F. A Review of Erucic Acid Production in Brassicaceae Oilseeds: Progress and Prospects for the Genetic Engineering of High and Low-Erucic Acid Rapeseeds (Brassica napus). *Front Plant Sci* **13**, (2022).
10. Allender, C. J. & King, G. J. Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers. **10**, 54 (2010).
11. Lu, K. *et al.* Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement. *Nat Commun* **10**, (2019).
12. Song, J. M. *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nat Plants* **6**, 34–45 (2020).
13. Yang, J. *et al.* The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. *Nat Genet* **48**, 1225–1232 (2016).
14. Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R. & Snowdon, R. Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci Rep* **7**, (2017).
15. Kitashiba, H. & Nasrallah, J. B. Self-incompatibility in brassicaceae crops: Lessons for interspecific incompatibility. *Breed Sci* **64**, 23–37 (2014).
16. Giavalisco, P., Kapitza, K., Kolasa, A., Buhtz, A. & Kehr, J. Towards the proteome of Brassica napus phloem sap. *Proteomics* **6**, 896–909 (2006).

17. Dinant, S. & Kehr, J. Sampling and Analysis of Phloem Sap. in *Plant Mineral Nutrients* (ed. Maathuis, F. J. M.) vol. 953 185–194 (Humana Press, 2013).

18. Ostendorp, A., Pahlow, S., Deke, J., Thieß, M. & Kehr, J. Protocol: Optimisation of a grafting protocol for oilseed rape (Brassica napus) for studying long-distance signalling. *Plant Methods* **12**, (2016).

19. Urquidi Camacho, R. A., Lokdarshi, A. & von Arnim, A. G. Translational gene regulation in plants: A green new deal. *Wiley Interdiscip Rev RNA* **11**, (2020).

20. Huang, K. C., Lin, W. C. & Cheng, W. H. Salt hypersensitive mutant 9, a nucleolar APUM23 protein, is essential for salt sensitivity in association with the ABA signaling pathway in Arabidopsis. *BMC Plant Biol* **18**, (2018).

21. Huang, C. K. *et al.* The DEAD-box RNA helicase AtRH7/PRH75 participates in pre-rRNA processing, plant development and cold tolerance in arabidopsis. *Plant Cell Physiol* **57**, 174–191 (2016).

22. Maekawa, S., Ishida, T. & Yanagisawa, S. Reduced expression of APUM24, encoding a novel rRNA processing factor, induces sugar-dependent nucleolar stress and altered sugar responses in arabidopsis thaliana. *Plant Cell* **30**, 209–227 (2018).

23. Martinez-Seidel, F., Beine-Golovchuk, O., Hsieh, Y. C. & Kopka, J. Systematic review of plant ribosome heterogeneity and specialization. *Front Plant Sci* **11**, (2020).

24. Ciganda, M. & Williams, N. Eukaryotic 5S rRNA biogenesis. *Wiley Interdiscip Rev RNA* **2**, 523–533 (2011).

25. Sáez-Vásquez, J. & Delseny, M. Ribosome biogenesis in plants: From functional 45S ribosomal DNA organization to ribosome assembly factors. *Plant Cell* vol. 31 1945–1967 Preprint at https://doi.org/10.1105/TPC.18.00874 (2019).

26. Greber, B. J. Mechanistic insight into eukaryotic 60S ribosomal subunit biogenesis by cryo-electron microscopy. *RNA* **22**, 1643–1662 (2016).

27. Tomecki, R., Sikorski, P. J. & Zakrzewska-Placzek, M. Comparison of preribosomal RNA processing pathways in yeast, plant and human cells – focus on coordinated action of endo- and exoribonucleases. *FEBS Lett* **591**, 1801–1850 (2017).

28. Sáez-Vasquez, J., Caparros-Ruiz, D., Barneche, F. & Echeverría, M. Characterization of a crucifer plant pre-rRNA processing complex. *Biochem Soc Trans* **32**, 578–580 (2004).

29. Zakrzewska-Placzek, M., Souret, F. F., Sobczyk, G. J., Green, P. J. & Kufel, J. Arabidopsis thaliana XRN2 is required for primary cleavage in the pre-ribosomal RNA. *Nucleic Acids Res* **38**, 4487–4502 (2010).

30. Samaha, H. *et al.* Identification of protein factors and U3 snoRNAs from a Brassica oleracea RNP complex involved in the processing of pre-rRNA. *Plant Journal* **61**, 383–398 (2010).

31. Palm, D. *et al.* Plant-specific ribosome biogenesis factors in Arabidopsis thaliana with essential function in rRNA processing. *Nucleic Acids Res* **47**, 1880–1895 (2019).

32. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**, 1138–1152 (2017).

33. Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A. & Gleizes, P. E. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA* **6**, 225–242 (2015).

34. Weis, B. L., Kovacevic, J., Missbach, S. & Schleiff, E. Plant-Specific Features of Ribosome Biogenesis. *Trends Plant Sci* **20**, 729–740 (2015).

35. Whittle, C. A. & Krochko, J. E. Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in brassica napus. *Plant Cell* **21**, 2203–2219 (2009).

36. Dörner, K., Ruggeri, C., Zemp, I. & Kutay, U. Ribosome biogenesis factors—from names to functions. *EMBO J* (2023) doi:10.15252/embj.2022112699.

37. Palm, D. *et al.* Proteome distribution between nucleoplasm and nucleolus and its relation to ribosome biogenesis in Arabidopsis thaliana. *RNA Biol* **13**, 441–454 (2016).

38. Pendle, A. F. *et al.* Proteomic Analysis of the Arabidopsis Nucleolus Suggests Novel Nucleolar Functions □ D. *Mol Biol Cell* **16**, 260–269 (2005).

39. Simm, S. *et al.* Identification and expression analysis of ribosome biogenesis factor co-orthologs in solanum lycopersicum. *Bioinform Biol Insights* **9**, 1–17 (2015).

40. Montacié, C. *et al.* Nucleolar proteome analysis and proteasomal activity assays reveal a link between nucleolus and 26S proteasome in A. Thaliana. *Front Plant Sci* **8**, (2017).

41. Dembowski, J. A., Kuo, B. & Woolford, J. L. Has1 regulates consecutive maturation and processing steps for assembly of 60S ribosomal subunits. *Nucleic Acids Res* **41**, 7889–7904 (2013).

42. Liu, Y. & Imai, R. Function of plant DExD/H-Box RNA helicases associated with ribosomal RNA biogenesis. *Frontiers in Plant Science* vol. 9 Preprint at https://doi.org/10.3389/fpls.2018.00125 (2018).

43. Huang, C. K. *et al.* A DEAD-box protein, AtRH36, is essential for female gametophyte development and is involved in rRNA biogenesis in arabidopsis. *Plant Cell Physiol* **51**, 694–706 (2010).

44. Kant, P., Kant, S., Gordon, M., Shaked, R. & Barak, S. Stress Response Suppressor1 and Stress Response Suppressor2, two Dead-box RNA helicases that attenuate Arabidopsis responses to multiple abiotic stresses. *Plant Physiol* **145**, 814–830 (2007).

45. Oeffinger, M. & Tollervey, D. Yeast Nop15p is an RNA-binding protein required for pre-rRNA processing and cytokinesis. *EMBO J* **22**, 6573–6583 (2003).

46. Granneman, S., Petfalski, E. & Tollervey, D. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO Journal* **30**, 4006–4019 (2011).

47. Pan, W. A. *et al.* The RNA recognition motif of NIFK is required for rRNA maturation during cell cycle progression. *RNA Biol* **12**, 255–267 (2015).

48. Lin, T.-C. *et al.* The nucleolar protein NIFK promotes cancer progression via CK1a/b-catenin in metastasis and Ki-67-dependent cell proliferation. (2016) doi:10.7554/eLife.11288.001.

49. Oeffinger, M. *et al.* Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat Methods* **4**, 951–956 (2007).

50. Jeon, Y. *et al.* The nucleolar GTPase nucleostemin-like 1 plays a role in plant growth and senescence by modulating ribosome biogenesis. *J Exp Bot* **66**, 6297–6310 (2015).

51. Zhou, D. *et al.* Cryo-EM structure of an early precursor of large ribosomal subunit reveals a half-assembled intermediate. *Protein Cell* **10**, 120–130 (2019).

52. Sanghai, Z. A. *et al.* Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature* **556**, 126–129 (2018).

53. Wu, S. *et al.* Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* **534**, 133–137 (2016).

54. Ismail, S. *et al.* Emergence of the primordial pre-60S from the 90S pre-ribosome. *Cell Rep* **39**, (2022).

55. Wu, S., Tan, D., Woolford, J. L., Dong, M. Q. & Gao, N. Atomic modeling of the ITS2 ribosome assembly subcomplex from cryo-EM together with mass spectrometry-identified protein–protein crosslinks. *Protein Science* **26**, 103–112 (2017).

56. Côté, C. A., Greer, C. L. & Peculis, B. A. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* **8**, 786–797 (2002).

57. Zhang, J., Gonzalez, L. E. & Hall, T. M. T. Structural analysis reveals the flexible C-terminus of Nop15 undergoes rearrangement to recognize a pre-ribosomal RNA folding intermediate. *Nucleic Acids Res* **45**, 2829–2837 (2017).

58. Cléry, A., Blatter, M. & Allain, F. H. T. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**, 290–298 (2008).

59. Daubner, G. M., Cléry, A. & Allain, F. H. T. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr Opin Struct Biol* **23**, 100–108 (2013).

60. Sevilem, I., Miyashima, S. & Helariutta, Y. Cell-to-cell communication via plasmodesmata in vascular plants. *Cell Adhesion and Migration* vol. 7 27–32 Preprint at https://doi.org/10.4161/cam.22126 (2013).

61. Lough, T. J. & Lucas, W. J. Integrative plant biology: Role of phloem long-distance macromolecular trafficking. *Annu Rev Plant Biol* **57**, 203–232 (2006).

62. Robert, H. S. & Friml, J. Auxin and other signals on the move in plants. *Nature Chemical Biology* vol. 5 325–332 Preprint at https://doi.org/10.1038/nchembio.170 (2009).

63. Agustí, J. & Blázquez, M. A. Plant vascular development: mechanisms and environmental regulation. *Cellular and Molecular Life Sciences* **77**, 3711–3728 (2020).

64. Lucas, W. J. *et al.* The Plant Vascular System: Evolution, Development and Functions. *J Integr Plant Biol* **55**, 294–388 (2013).

65. Myburg, A. A., Lev-Yadun, S. & Sederoff, R. R. Xylem Structure and Function. *eLS* (2013) doi:10.1002/9780470015902.a0001302.pub2.

66. Hipper, C., Brault, V., Ziegler-Graff, V. & Revers, F. Viral and cellular factors involved in phloem transport of plant viruses. *Front Plant Sci* **4**, (2013).

67. G. White, R. Cell Biology of Sieve Element–Companion Cell Complexes. in *Phloem: Molecular Cell Biology, Systemic Communication, Biotic Interactions* (eds. Thompson, G. A. & van Bel, A. J. E.) 8–29 (Wiley-Blackwell, 2012).

68. Kehr, J. & Buhtz, A. Long distance transport and movement of RNA through the phloem. *J Exp Bot* **59**, 85–92 (2008).

69. Lemoine, R. *et al.* Source-to-sink transport of sugar and regulation by environmental factors. *Front Plant Sci* **4**, (2013).

70. Slewinski, T. L., Zhang, C. & Turgeon, R. Structural and functional heterogeneity in phloem loading and transport. *Front Plant Sci* **4**, (2013).

71. Münch, E. *Die Stoffbewegungen in Der Pflanze*. (Jena: G. Fischer, 1930).

72. Ham, B. K. & Lucas, W. J. The angiosperm phloem sieve tube system: A role in mediating traits important to modern agriculture. *J Exp Bot* **65**, 1799–1816 (2014).

73. Knoblauch, M. *et al.* Testing the Münch hypothesis of long distance phloem transport in plants. (2016) doi:10.7554/eLife.15341.001.

74. Dinant, S. & Lemoine, R. The phloem pathway: New issues and old debates. *C R Biol* **333**, 307–319 (2010).

75. Fukumorita, T., Chino, M., Noziri, Y., Haraguchi, H. & Chino, M. Inorganic content in rice phloem sap. *Soil Sci Plant Nutr* **29**, 185–192 (1983).

76. Richardson, P. T., Baker, D. A. & Ho, L. C. *The Chemical Composition of Cucurbit Vascular Exudates*. *Journal of Experimental Botany* vol. 33 http://jxb.oxfordjournals.org/ (1982).

77. Hoad, G. V. Transport of hormones in the phloem of higher plants. *Plant Growth Regul* **16**, 173–182 (1995).

78. Weibull, J. & Melint, G. Free amino acid content of phloem sap from Brassica plants in relation to performance of Lipaphis erysimi (Hemiptera: Aphididae). *Ann. appl. Biol* vol. 116 417–423 (1990).

79. Riens, B., Lohaus, G., Heineke, D. & Heldt, H. W. *Amino Acid and Sucrose Content Determined in the Cytosolic, Chloroplastic, and Vacuolar Compartments and in the Phloem Sap of Spinach Leaves1*. *Plant Physiol* vol. 97 https://academic.oup.com/plphys/article/97/1/227/6087189 (1991).

80. Madey, E., Nowack, L. M. & Thompson, J. E. Isolation and characterization of lipid in phloem sap of canola. *Planta* **214**, 625–634 (2002).

81. Rodriguez-Medina, C., Atkins, C. A., Mann, A. J., Jordan, M. E. & Smith, P. M. C. Macromolecular composition of phloem exudate from white lupin (Lupinus albus L.). *BMC Plant Biol* **11**, (2011).

82. Lin, M. K., Lee, Y. J., Lough, T. J., Phimmey, B. S. & Lucas, W. J. Analysis of the pumpkin phloem proteome provides insights into angiosperm sieve tube function. *Molecular and Cellular Proteomics* **8**, 343–356 (2009).

83. Kehr, J. Phloem sap proteins: Their identities and potential roles in the interaction between plants and phloem-feeding insects. *J Exp Bot* **57**, 767–774 (2006).

84. Kehr, J. & Kragler, F. Long distance RNA movement. *New Phytologist* **218**, 29–40 (2018).

85. Sasaki, T., Chino, M., Hayashi, H. & Fujiwara, T. *Detection of Several mRNA Species in Rice Phloem Sap*. *Plant Cell Physiol* vol. 39 https://academic.oup.com/pcp/article/39/8/895/1835392 (1998).

86. An, N. *et al.* Genome-wide identification, characterization and expression analysis of long non-coding RNAs in different tissues of apple. *Gene* **666**, 44–57 (2018).

87. Zhang, Z. *et al.* Plant lncRNAs are enriched in and move systemically through the phloem in response to phosphate deficiency. *J Integr Plant Biol* **61**, 492–508 (2019).

88. Zhang, L. *et al.* Genome-Wide Identification of Long Non-Coding RNAs and Their Potential Functions in Poplar Growth and Phenylalanine Biosynthesis. *Front Genet* **12**, (2021).

89. Ham, B. K. & Lucas, W. J. Phloem-Mobile RNAs as Systemic Signaling Agents. *Annu Rev Plant Biol* **68**, 173–195 (2017).

90. Kehr, J., Morris, R. J. & Kragler, F. Long-Distance Transported RNAs: From Identity to Function mRNA: messenger RNA PD: plasmodesmata. (2022) doi:10.1146/annurev-arplant-070121.

91. Thieme, C. J. *et al.* Endogenous Arabidopsis messenger RNAs transported to distant tissues. *Nat Plants* **1**, (2015).

92. Zhang, S., Sun, L. & Kragler, F. The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation. *Plant Physiol* **150**, 378–387 (2009).

93. Zhang, W. *et al.* tRNA-related sequences trigger systemic mRNA transport in plants. *Plant Cell* **28**, 1237–1249 (2016).

94. Buhtz, A., Pieritz, J., Springer, F. & Kehr, J. Phloem small RNAs, nutrient stress responses, and systemic mobility. *BMC Plant Biol* **10**, 64 (2010).

95. Molnar, A. *et al.* Small Silencing RNAs in Plants Are Mobile and Direct Epigenetic Modification in Recipient Cells. *Science (1979)* **328**, 872–875 (2010).

96. Branco, R. & Masle, J. Systemic signalling through translationally controlled tumour protein controls lateral root formation in Arabidopsis. *J Exp Bot* **70**, 3927–3940 (2019).

97. Buhtz, A., Springer, F., Chappell, L., Baulcombe, D. C. & Kehr, J. Identification and characterization of small RNAs from the phloem of Brassica napus. *Plant Journal* **53**, 739–749 (2008).

98. Yu, Y., Zhang, Y., Chen, X. & Chen, Y. Plant noncoding RNAs: Hidden players in development and stress responses. *Annu Rev Cell Dev Biol* **35**, 407–431 (2019).

99. Kim, E. D. & Sung, S. Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. *Trends Plant Sci* **17**, 16–21 (2012).

100. Rai, M. I., Alam, M., Lightfoot, D. A., Gurha, P. & Afzal, A. J. Classification and experimental identification of plant long non-coding RNAs. *Genomics* **111**, 997–1005 (2019).

101. Zhao, Z. *et al.* Long Non-Coding RNAs: New Players in Plants. *International Journal of Molecular Sciences* vol. 23 Preprint at https://doi.org/10.3390/ijms23169301 (2022).

102. Chekanova, J. A. Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol* **27**, 207–216 (2015).

103. Budak, H., Kaya, S. B. & Cagirici, H. B. Long Non-coding RNA in Plants in the Era of Reference Sequences. *Front Plant Sci* **11**, (2020).

104. Wang, H. L. V. & Chekanova, J. A. Long noncoding RNAs in plants. *Adv Exp Med Biol* **1008**, 133–154 (2017).

105. Kazimierczyk, M., Kasprowicz, M. K., Kasprzyk, M. E. & Wrzesinski, J. Human long noncoding RNA interactome: Detection, characterization and function. *Int J Mol Sci* **21**, (2020).

106. Deng, W. *et al.* FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. *Proc Natl Acad Sci U S A* **108**, 6680–6685 (2011).

107. Kim, D. H. & Sung, S. Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Dev Cell* **40**, 302-312.e4 (2017).

108. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science (1979)* **331**, 76–79 (2011).

109. Helliwell, C. A., Robertson, M., Finnegan, E. J., Buzas, D. M. & Dennis, E. S. Vernalization-repression of Arabidopsis FLC requires promoter sequences but not antisense transcripts. *PLoS One* **6**, (2011).

110. Csorba, T., Questa, J. I., Sun, Q. & Dean, C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci U S A* **111**, 16160–16165 (2014).

111. Jeon, M. *et al.* Vernalization-triggered expression of the antisense transcript COOLAIR is mediated by CBF genes. *Elife* **12**, (2023).

112. Li, Y. *et al.* Regulation of seed oil accumulation by lncRNAs in Brassica napus. *Biotechnology for Biofuels and Bioproducts* **16**, (2023).

113. Waseem, M. *et al.* Genome-wide identification of long non-coding RNAs in two contrasting rapeseed (Brassica napus L.) genotypes subjected to cold stress. *Environ Exp Bot* **201**, 104969 (2022).

114. Tan, X., Li, S., Hu, L. & Zhang, C. Genome-wide analysis of long non-coding RNAs (lncRNAs) in two contrasting rapeseed (Brassica napus L.) genotypes subjected to drought stress and re-watering. *BMC Plant Biol* **20**, (2020).

115. Wang, D., Gao, Y., Sun, S., Li, L. & Wang, K. Expression Profiles and Characteristics of Apple lncRNAs in Roots, Phloem, Leaves, Flowers, and Fruit. *Int J Mol Sci* **23**, (2022).

116. Manalastas-Cantos, K. *et al.* ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* **54**, 343–355 (2021).

117. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).

118. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* **50**, W276–W279 (2022).

119. Sun, L. *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* **41**, (2013).

120. Kang, Y. J. *et al.* CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* **45**, W12–W16 (2017).

121. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578 (2012).

122. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* vol. 34 i884–i890 (Oxford University Press, 2018).

123. Kolberg, L. *et al.* G:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207–W212 (2023).

124. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).

125. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).

126. Robinson, J. T. *et al. Integrative genomics viewer*. *Nature Publishing Group* http://maq.sourceforge.net/ (2011) doi:10.1038/nbt0111-24.

127. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419 (2021).

128. The PyMOL Molecular Graphics System, Version 2.5.4 Schrödinger, LLC.

129. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

130. Panjkovich, A. & Svergun, D. I. SASpy: A PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics* **32**, 2062–2064 (2016).

131. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).

132. Unfried, I. & Gruendler, P. *Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from Arabidopsis thaliana*. *Nucleic Acids Research* vol. 18.

133. Rajkowitsch, L. & Schroeder, R. Coupling RNA annealing and strand displacement: A FRET-based microplate reader assay for RNA chaperone activity. *Biotechniques* **43**, 304–310 (2007).

134. Lühmann, K. L. Characterization of a phloem-mobile RNA-binding protein and the circRNA content of phloem in Brassica napus. (Universität Hamburg, 2023).

135. Chalhoub, B. *et al. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome*. vol. 345 (2014).

136. Fernandes, R. *et al.* Structural and functional analysis of a plant nucleolar RNA chaperone-like protein. *Sci Rep* **13**, (2023).

137. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

138. Byeon, I. J. L., Li, H., Song, H., Gronenborn, A. M. & Tsai, M. D. Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. *Nat Struct Mol Biol* **12**, 987–993 (2005).

139. Säftel, W., Winkler, U. & Stabenau, H. Separation of Native Basic Proteins by Cathodic, Discontinuous Polyacrylamide Gel Electrophoresis. vol. 1098 98–369 Preprint at https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_2376.pdf.

140. Trewhella, J. *et al.* 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: An update. *Acta Crystallogr D Struct Biol* **73**, 710–728 (2017).

141. McDermaid, A., Monier, B., Zhao, J., Liu, B. & Ma, Q. Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics* vol. 20 2044–2054 Preprint at https://doi.org/10.1093/bib/bby067 (2019).

142. Zaharias, S. *et al.* Intrinsically disordered electronegative clusters improve stability and binding specificity of RNA-binding proteins. *Journal of Biological Chemistry* **297**, (2021).

143. Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological Procedures Online* vol. 11 32–51 Preprint at https://doi.org/10.1007/s12575-009-9008-x (2009).

144. Kumari, N. & Yadav, S. Modulation of protein oligomerization: An overview. *Progress in Biophysics and Molecular Biology* vol. 149 99–113 Preprint at https://doi.org/10.1016/j.pbiomolbio.2019.03.003 (2019).

145. Narhi, L. O., Schmit, J., Bechtold-Peters, K. & Sharma, D. Classification of protein aggregates. *Journal of Pharmaceutical Sciences* vol. 101 493–498 Preprint at https://doi.org/10.1002/jps.22790 (2012).

146. Amin, S., Barnett, G. V., Pathak, J. A., Roberts, C. J. & Sarangapani, P. S. Protein aggregation, particle formation, characterization & rheology. *Current Opinion in Colloid and Interface Science* vol. 19 438–449 Preprint at https://doi.org/10.1016/j.cocis.2014.10.002 (2014).

147. Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Muzio, L. Lo. The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). *International Journal of Molecular Medicine* vol. 40 271–280 Preprint at https://doi.org/10.3892/ijmm.2017.3036 (2017).

148. Gräwert, M. & Svergun, D. A Beginner's Guide A beginner's guide to solution small-angle X-ray scattering (SAXS). *Biochem (Lond)* **42**, 36–42 (2020).

149. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett* **589**, 2570–2577 (2015).

150. Chialva, C., Blein, T., Crespi, M. & Lijavetzky, D. Insights into long non-coding RNA regulation of anthocyanin carrot root pigmentation. *Sci Rep* **11**, (2021).

151. Mosig, R. A. & Kojima, S. Timing without coding: How do long non-coding RNAs regulate circadian rhythms? *Seminars in Cell and Developmental Biology* vol. 126 79–86 Preprint at https://doi.org/10.1016/j.semcdb.2021.04.020 (2022).

152. Deschamps-Francoeur, G., Simoneau, J. & Scott, M. S. Handling multi-mapped reads in RNA-seq. *Computational and Structural Biotechnology Journal* vol. 18 1569–1576 Preprint at https://doi.org/10.1016/j.csbj.2020.06.014 (2020).

153. Rosenblad, M. A., Larsen, N., Samuelsson, T. & Zwieb, C. Kinship in the SRP RNA family. *RNA Biol* **6**, 508–516 (2009).

154. Schünemann, D. Structure and function of the chloroplast signal recognition particle. *Current Genetics* vol. 44 295–304 Preprint at https://doi.org/10.1007/s00294-003-0450-z (2004).

155. Ransohoff, J. D., Wei, Y. & Khavari, P. A. The functions and unique features of long intergenic non-coding RNA. *Nature Reviews Molecular Cell Biology* vol. 19 143–157 Preprint at https://doi.org/10.1038/nrm.2017.104 (2018).

156. Liu, J. *et al.* Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in arabidopsis. *Plant Cell* **24**, 4333–4345 (2012).

157. Palos, K. *et al.* Identification and functional annotation of long intergenic non-coding RNAs in Brassicaceae. *Plant Cell* **34**, 3233–3260 (2022).

158. Shen, E. *et al.* Genome-wide identification of oil biosynthesis-related long non-coding RNAs in allopolyploid Brassica napus. *BMC Genomics* **19**, (2018).

159. Benning, U. F., Tamot, B., Guelette, B. S. & Hoffmann-Benning, S. New aspects of phloem-mediated long-distance lipid signaling in plants. *Front Plant Sci* **3**, (2012).

160. Yoo, B. C., Lee, J. Y. & Lucas, W. J. Analysis of the complexity of protein kinases within the phloem sieve tube system. Characterization of Cucurbita maxima calmodulin-like domain protein kinase 1. *Journal of Biological Chemistry* **277**, 15325–15332 (2002).

161. Ogden, A. J. *et al.* Phloem exudate protein profiles during drought and recovery reveal abiotic stress responses in tomato vasculature. *Int J Mol Sci* **21**, 1–20 (2020).

162. Wang, P. & Grimm, B. Organization of chlorophyll biosynthesis and insertion of chlorophyll into the chlorophyll-binding proteins in chloroplasts. *Photosynthesis Research* vol. 126 189–202 Preprint at https://doi.org/10.1007/s11120-015-0154-5 (2015).

163. Schubert, W.-D., Moser, J., Schauer, S., Heinz, D. W. & Jahn, D. *Structure and function of glutamyl-tRNA reductase, the first enzyme of tetrapyrrole biosynthesis in plants and prokaryotes*. *Photosynthesis Research* vol. 74 http://www.hhmi.swmed.edu/ex-ternal/Doc/GI_ren-der/ (2002).

164. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, (2013).

165. Motif Scan. https://myhits.sib.swiss/cgi-bin/motif_scan.

166. Sarnowski, C. P., Bikaki, M. & Leitner, A. Cross-linking and mass spectrometry as a tool for studying the structural biology of ribonucleoproteins. *Structure* vol. 30 441–461 Preprint at https://doi.org/10.1016/j.str.2022.03.003 (2022).

167. Clarke, P. A. *RNA Footprinting and Modification Interference Analysis*. *From: Methods in Molecular Biology* vol. 118.

168. Esteban-Serna, S., McCaughan, H. & Granneman, S. Advantages and limitations of UV cross-linking analysis of protein–RNA interactomes in microbes. *Molecular Microbiology* vol. 120 477–489 Preprint at https://doi.org/10.1111/mmi.15073 (2023).

169. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* **11**, 1064–1070 (2014).

170. Nilsen, T. W. RNase footprinting to map sites of RNA–protein interactions. *Cold Spring Harb Protoc* **2014**, 677–682 (2014).

7. Supplementary

Figure 7.1− pUC57 plasmid map. Map of pUC57 showing the main characteristics. The plasmid was linearized with *HindIII* for the insertion of 45S rDNA. Image created with SnapGene.

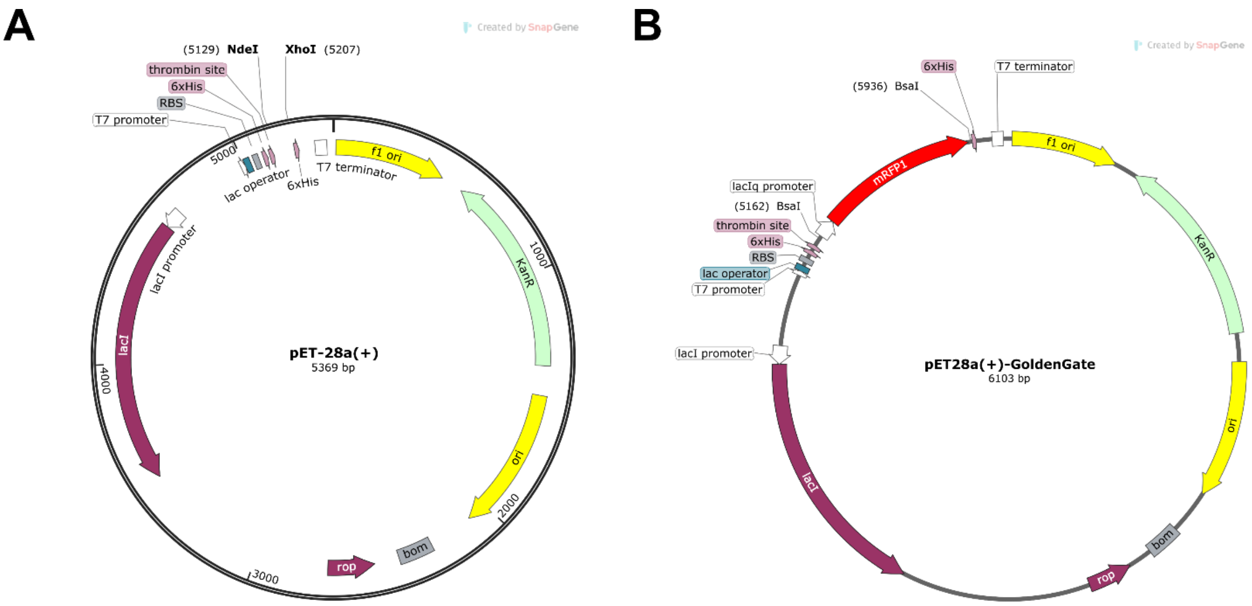


Figure 7.2− pET28a(+) plasmid maps. (**A**) Map of pET28a(+) showing the main characteristics, such as, *NdeI* and *XhoI* recognition sites. (**B**) Modified version of pET28a(+) suitable for golden gate cloning with *BsaI*. Image created with SnapGene.

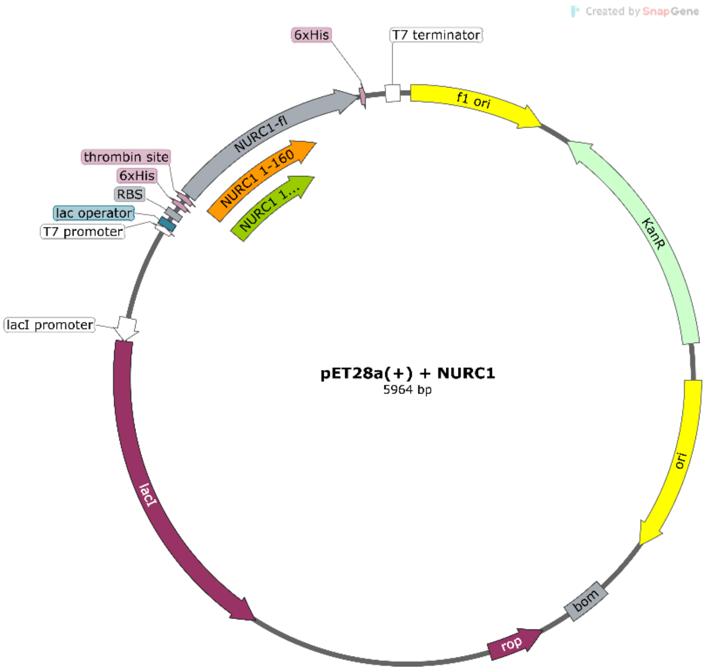


Figure 7.3– NURC1 construct map. Construct map of all NURC1 versions in pET28a(+). NURC1-fl is represented in grey, NURC1$_{1\text{-}160}$ in orange and NURC1$_{1\text{-}140}$ in green. Image created with SnapGene.

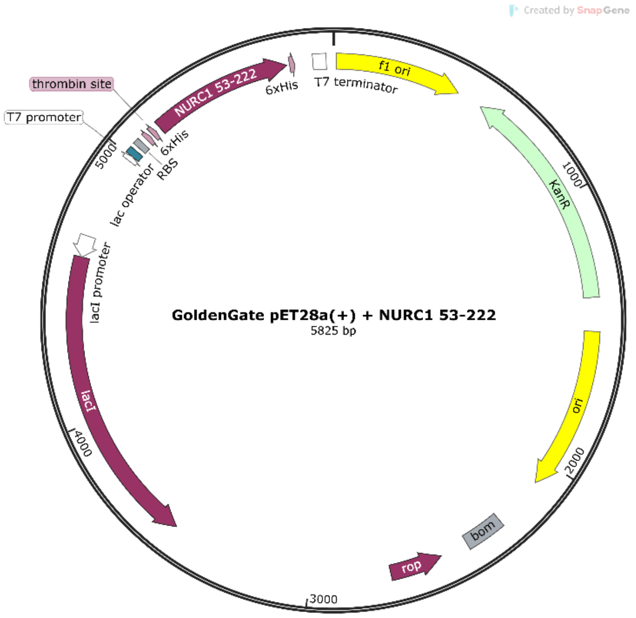


Figure 7.4– NURC1$_{53\text{-}222}$ construct map. Map of all NURC1$_{53\text{-}222}$ in pET28a(+) golden gate modified version. Image created with SnapGene.

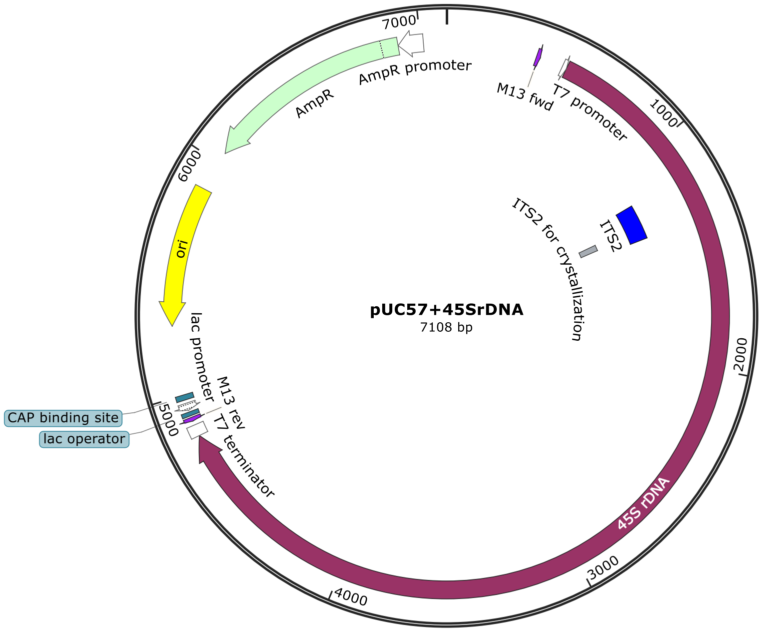


Figure 7.5− 45S rDNA construct map. Map of 45S rDNA in pUC57. Image created with SnapGene.

```r
############################# lncRNAs size distribution #############################
pacman::p_load(readxl, dplyr, tidyr, ggplot2, rtracklayer, scales)

# Gtf file w/ transcript coordinates
leaf <- readGFF("Data/Bn_Leaf_lncRNAs_Inputs/3.LncRNAIdentification/LncRNA_Filter/novel_lncRNA.gtf")

# Selecting relevant columns
leaf<- leaf |>
  select(seqid, start, end, transcript_id, gene_id)

# Calculating transcript size
leaf_size_by_transcript <- leaf |>
  group_by(transcript_id) |>
  summarise(leaf_total_exon_length = sum(end - start +1),
            leaf_num_exons = n()) #1142 transcripts


# Temporary table w/ sizes
temp1 <- leaf_size_by_transcript |>
  rename("leaf_total_exon_length"="Size")


# Grouping transcripts by size
temp1$grouped_values <- ifelse(temp1$Size <= 2000, temp1$Size, 2000 + 1)


# Plotting size distribution
leaf_size_plot <- ggplot(temp1, aes(x = grouped_values)) +
  geom_histogram(binwidth = 100, fill = "#F76C6C", color = "black") +
  scale_x_continuous(breaks = c(seq(200, 1800, by = 200), 2001),
                     labels = c(seq(200, 1800, by = 200), ">2000"),
                     expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x = "Size", y = "Count")+
  theme_bw()+
  theme(axis.text.x = element_text(size=35, angle = 45, hjust = 1),
        axis.text.y = element_text(size=35),
        axis.title = element_text(size= 40),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_line(colour = "black", linewidth = 2),
        axis.ticks.length = unit(0.3, "cm"),
        aspect.ratio=1/1)


# Saving plot in .png format
ggsave(filename = "lncRNAs/Leaf_lncRNAs_SizeDistribution.png",
       plot = leaf_size_plot,
       width = 10, height = 8,
       dpi = 600)
```

Figure 7.6− R script used for size distribution analysis. The current script was used for the size distribution analysis of leaf lncRNAs. The same process was used for the phloem lncRNAs.

```r
############################ lncRNAs strand distribution ############################
pacman::p_load(dplyr, tidyr, writexl, rtracklayer, ggplot2)

# Gtf files w/ transcript strand info
leaf_strands <- readGFF("Data/Bn_Leaf_lncRNAs_Inputs/3.LncRNAIdentification/LncRNA_Fi
lter/novel_lncRNA.gtf") |> select(strand) |> rename("strand"="Leaf")

phloem_strands <- readGFF("Data/Bn_Phloem_lncRNAs/Result_X204SC20032004-Z01-F005_Bras
sica_napus/3.LncRNAIdentification/LncRNA_Filter/novel_lncRNA.gtf") |>
select(strand) |> rename("strand"="Phloem")

# Counting sense and antisense transcripts
leaf_length_sense <- length(which(leaf_strands$Leaf=='+'))
leaf_length_antisense <- length(which(leaf_strands$Leaf=='-'))
phloem_length_sense <- length(which(phloem_strands$Phloem=='+'))
phloem_length_antisense <- length(which(phloem_strands$Phloem=='-'))

# Organizing them in a tibble
graph_table <- tibble("Leaf" = c(leaf_length_sense, leaf_length_antisense),
                      "Phloem" = c(phloem_length_sense, phloem_length_antisense),
                      "Strand" = factor(c("+", "-")))

# Normalizing percentages within each tissue
graph_table_normalized <- graph_table |>
  mutate_at(vars(Leaf, Phloem), function(x) x / sum(x) * 100) |>
  pivot_longer(cols = c(Leaf, Phloem), names_to = "Tissue", values_to = "Percentage")
|> mutate(Tissue = factor(Tissue, levels= c("Phloem", "Leaf")))

# Plotting strand distribution
strand_bar <- ggplot(graph_table_normalized, aes(x = Strand, y = Percentage, fill = T
issue)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Tissue", y = "Percentage", fill = NULL) +
  scale_fill_manual(values=c("#23305E",  "#F76C6C"))+
  theme_bw()+
  theme(legend.position = "bottom",
        text = element_text(size=40),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_line(colour = "black", linewidth = 2),
        axis.ticks.length = unit(0.3, "cm"),
        axis.ticks = element_line(linewidth = 2),
        legend.key.size = unit(2, 'cm'),
        aspect.ratio=1/1)+
  scale_y_continuous(expand = c(0, 0))

# Saving plot in .png format
ggsave(filename = "lncRNAs/lncRNAs_StrandDistribution.png",
       plot = strand_bar, width = 10, height = 8, dpi = 600)
```

Figure 7.7− R script used for strand distribution analysis. The current script was used for the strand distribution analysis of leaf and phloem lncRNAs.

```
############################## lncRNAs exon distribution ##############################
pacman::p_load(readxl, dplyr, tidyr, writexl, rtracklayer, ggplot2)

# Getting leaf exons by transcript
leaf_locations <- readGFF("Data/Bn_Leaf_lncRNAs_Inputs/3.LncRNAIdentification/LncRNA_
Filter/novel_lncRNA.gtf")

leaf_exons <- leaf_locations |>
  group_by(transcript_id) |>
  summarize(leaf_exons = n()) |> #1142
  mutate(leaf_exons = as.factor(leaf_exons))

leaf_exons_counts <-

  as_tibble(table(leaf_exons$leaf_exons), .name_repair = "universal") |>
  rename("...1" = "Class", "n" = "Leaf") |>
  rbind(tibble("Class" = 12, "Leaf" = 0)) |>
  mutate(Class = factor(Class, levels=(c(2:12))))


# Getting phloem exons by transcript
phloem_locations <- readGFF("Data/Bn_Phloem_lncRNAs/Result_X204SC20032004-Z01-F005_Br
assica_napus/3.LncRNAIdentification/LncRNA_Filter/novel_lncRNA.gtf")

phloem_exons <- phloem_locations |>
  group_by(transcript_id) |>
  summarize(phloem_exons = n()) |>
  mutate(phloem_exons = as.factor(phloem_exons))

phloem_exons_counts <-
  as_tibble(table(phloem_exons$phloem_exons), .name_repair = "universal") |>
  rename("...1" = "Class", "n" = "Phloem") |>
  mutate(Class = factor(Class, levels=(c(2:12))))


# Organizing results in a tibble
graph_table <- as_tibble(merge(phloem_exons_counts, leaf_exons_counts)) |>
  arrange(Class)

# Normalizing percentages within each tissue
graph_table_normalized <- graph_table |>
  mutate_at(vars(Leaf, Phloem), function(x) x / sum(x) * 100) |>
  pivot_longer(cols = c(Leaf, Phloem), names_to = "Tissue", values_to = "Percentage")
|> mutate(Tissue = factor(Tissue, levels= c("Phloem", "Leaf")))


# Plotting exon distribution
strand_bar <- ggplot(graph_table_normalized, aes(x = Class, y = Percentage, fill = Ti
ssue)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Exon number",
       y = "Percentage",
       fill= NULL) +
  scale_fill_manual(values=c("#23305E",  "#F76C6C"))+
  theme_bw()+
  theme(legend.position = "bottom",
```

```
        text = element_text(size=40),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_line(colour = "black", linewidth = 2),
        axis.ticks.length = unit(0.3, "cm"),
        axis.ticks = element_line(linewidth = 2),
        legend.key.size = unit(2, 'cm'),
        aspect.ratio=1/1)+
  scale_y_continuous(expand = c(0, 0))


# Saving plot in .png format
ggsave(filename = "lncRNAs/lncRNAs_ExonDistribution.png",
       plot = strand_bar,
       width = 10, height = 8,
       dpi = 600)
```

Figure 7.8− R script used for exon distribution analysis. The current script was used for the exon distribution analysis of leaf and phloem lncRNAs.

```r
####################### Differential expression of lncRNAs ##########################
pacman::p_load(DESeq2, readxl, dplyr, ReportingTools, ggplot2, EnhancedVolcano,
               writexl, tibble, rtracklayer, rlang, flextable)


# Implementing scientific format function
format_scientific <- function(x) {
  formatC(x, format = "e", digits = 2)
}


# Getting leaf and phloem counts for all transcripts
leaf_counts <- read_excel("Data/Bn_Leaf_lncRNAs_Inputs/5.Quantification/Expression/tr
anscripts.readcount.annot.xlsx") |>
  rename("Leaf1"=Input_1,
         "Leaf2"=Input_2,
         "Leaf3"=Input_3) |>
  select(-gene_name, -gene_description)

phloem_counts <- read_excel("Data/Bn_Phloem_lncRNAs/Result_X204SC20032004-Z01-F005_Br
assica_napus/5.Quantification/Expression/transcripts.readcount.annot.xlsx")|>
  select(-gene_name, -gene_description)


# Getting lncRNAS transcript_id (TCONS)
leafCP <- read_excel("Data/Bn_Leaf_lncRNAs_Inputs/3.LncRNAIdentification/LncRNA_Filte
r/Coding_potential_filter.result.xlsx")


phloemCP <- read_excel("Data/Bn_Phloem_lncRNAs/Result_X204SC20032004-Z01-F005_Brassic
a_napus/3.LncRNAIdentification/LncRNA_Filter/Coding_potential_filter.result.xlsx")

leaf_lncRNAs <- leafCP |> filter(CNCI=='noncoding' & CPC=='noncoding' & PFAM=='noncod
ing')
leaf_lncRNAs_TCONS <- leaf_lncRNAs |>
  select(transcript_id) |>
  unique() #removing duplicates

phloem_lncRNAs <- phloemCP |> filter(CNCI=='noncoding' & CPC=='noncoding' & PFAM=='no
ncoding')
phloem_lncRNAs_TCONS <- phloem_lncRNAs |>
  select(transcript_id) |>
  unique()


# Getting lncRNAs counts by transcript_id and filtering out counts<5
leaf_tracking_lncRNA_IDs <- merge(leaf_counts, leaf_lncRNAs_TCONS) |>
  filter(!Leaf1<5 & !Leaf2<5 & !Leaf3<5) |>
  relocate(gene_id) |>
  rename(Leaf_id = transcript_id)

phloem_tracking_lncRNA_IDs <- merge(phloem_counts, phloem_lncRNAs_TCONS) |>
  filter(!Phloem1<5 & !Phloem2<5 & !Phloem3<5) |>
  relocate(gene_id) |>
```

```
  rename(Phloem_id = transcript_id)


# Merging and summing isoforms
leaf_count_sums <- aggregate(. ~ gene_id, data = leaf_tracking_lncRNA_IDs [-2], FUN =
sum)
phloem_count_sums <- aggregate(. ~ gene_id, data = phloem_tracking_lncRNA_IDs[-2], FU
N = sum)


# Running DESeq2
samples_counts <-  merge(phloem_count_sums, leaf_count_sums)|>
  filter(!grepl(paste0("^", "XLOC_"), gene_id)) #removing XLOC IDs

row.names(samples_counts) <-  samples_counts$gene_id
samples_counts <- samples_counts[-1]

samples_info <- tibble(
  Sample=c("Phloem1", "Phloem2", "Phloem3", "Leaf1", "Leaf2", "Leaf3"),
  Tissue=c("phloem", "phloem", "phloem", "leaf", "leaf", "leaf"))

samples_info$Tissue <- factor(samples_info$Tissue)

sample_dds <- DESeqDataSetFromMatrix(countData = samples_counts,
                                     colData = samples_info,
                                     design = ~Tissue)
sample_dds <- DESeq(sample_dds)

sample_results <- results(sample_dds,contrast = c("Tissue", "phloem", "leaf"))


# Saving results in a table
results_table <- sample_results |> data.frame() |>
  rownames_to_column(var = "gene_id") |>
  arrange(padj) |>
  mutate_at(vars(padj), ~(signif(., 3))) |>
  mutate_at(vars(baseMean, log2FoldChange, lfcSE, stat), ~(round(., 2))) |>
  select(-stat, -pvalue)
write_xlsx(results_table, path = "Expression/DifferentialExpression/CombiningIsoforms
&Strands/DESeq2_PhloemvsLeaf_lncRNAs.xlsx")


# Volcano Plot
keyvals <- ifelse(
  sample_results$log2FoldChange < -1 & sample_results$padj < 1e-5, 'red',
  ifelse(sample_results$log2FoldChange > 1 & sample_results$padj < 1e-5, 'green',
         'grey'))
keyvals[is.na(keyvals)] <- 'black'
names(keyvals)[keyvals == 'green'] <- 'high'
names(keyvals)[keyvals == 'grey'] <- 'mid'
names(keyvals)[keyvals == 'red'] <- 'low'

lncVolcano <- EnhancedVolcano(sample_results,
                              lab = rownames(sample_results),
                              x = 'log2FoldChange',
```

```
                              y = 'padj',
                              title = '',
                              subtitle = NULL,
                              axisLabSize = 18,
                              gridlines.major = F,
                              gridlines.minor = F,
                              colCustom = keyvals,
                              selectLab = c("ENSRNA049479067",
                                            "GSBRNA2T00105977001",
                                            "ENSRNA049439209",
                                            "GSBRNA2T00157764001",
                                            "GSBRNA2T00015777001",
                                            "GSBRNA2T00001751001",
                                            "GSBRNA2T00056993001",
                                            "GSBRNA2T00074276001",
                                            "GSBRNA2T00122319001",
                                            "GSBRNA2T00030279001"),
                              xlab = noquote(expression('log'[2]*'FoldChange')),
                              ylab = noquote(expression('-log'[10]*'(padj)')),
                              labSize = 4,
                              pointSize = 4,
                              legendPosition = 'right',
                              legendLabSize = 18,
                              legendIconSize = 8,
                              drawConnectors = TRUE,
                              widthConnectors = .5,
                              colConnectors = 'black')+
  coord_fixed(ratio = .15)


# Saving plot in .png format
ggsave(filename = "Expression/DifferentialExpression/CombiningIsoforms&Strands/DESeq2_lncRNAs_PhloemvsLeaf_VolcanoPlot.png",
       plot = lncVolcano,
       width = 10, height = 8,
       dpi = 600)
```

Figure 7.9– R script used for differential expression analysis. The current script was used for the differential expression analysis of phloem vs leaf lncRNAs. Data cleaning steps are also included.

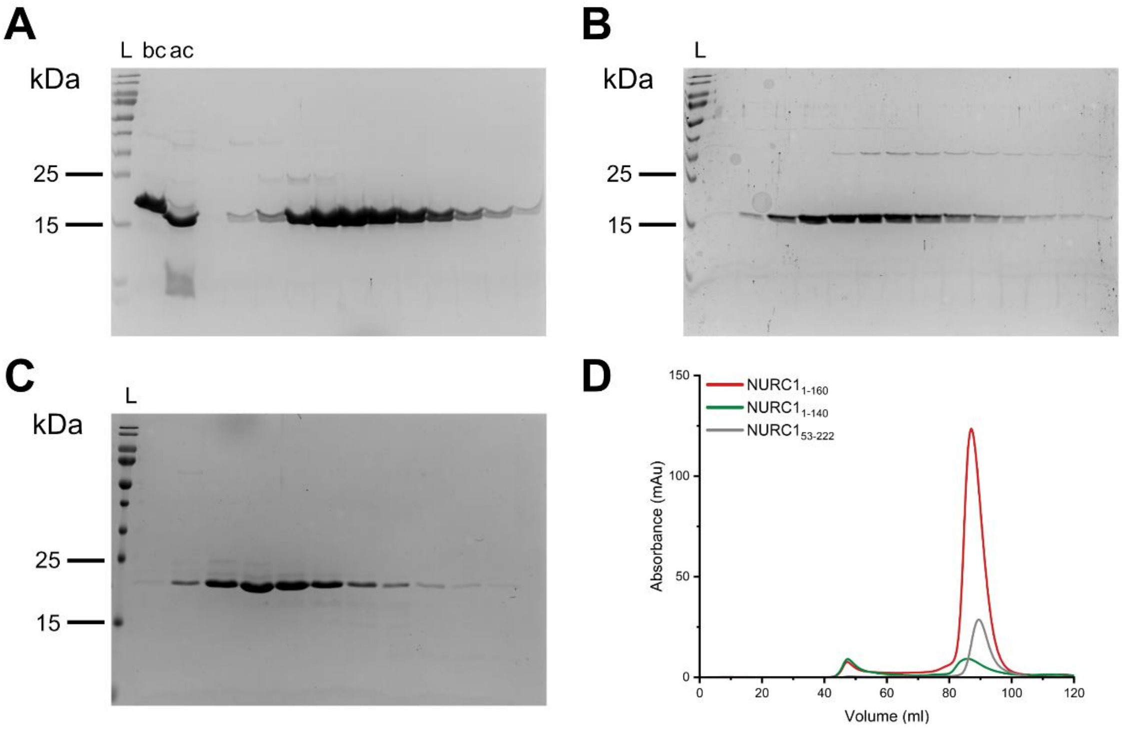


Figure 7.10− SEC results of NURC1 truncated versions. (**A**) SDS-PAGE after SEC of NURC1$_{1-160}$. Aliquots from before cleaving (bc) and after cleavage (ac) of the His-tag are shown. (**B**) SDS-PAGE after SEC of NURC1$_{1-140}$. (**C**) SDS-PAGE after SEC of NURC1$_{53-222}$. 15% acrylamide gels are shown. L: PageRuler Prestained Protein Ladder (ThermoFisher Scientific). (**D**) SEC chromatograms showing the absorbance (mAu) as a function of elution volume (ml). NURC1$_{1-160}$ is represented in red, NURC1$_{1-140}$ in green, and NURC1$_{53-222}$ in grey.

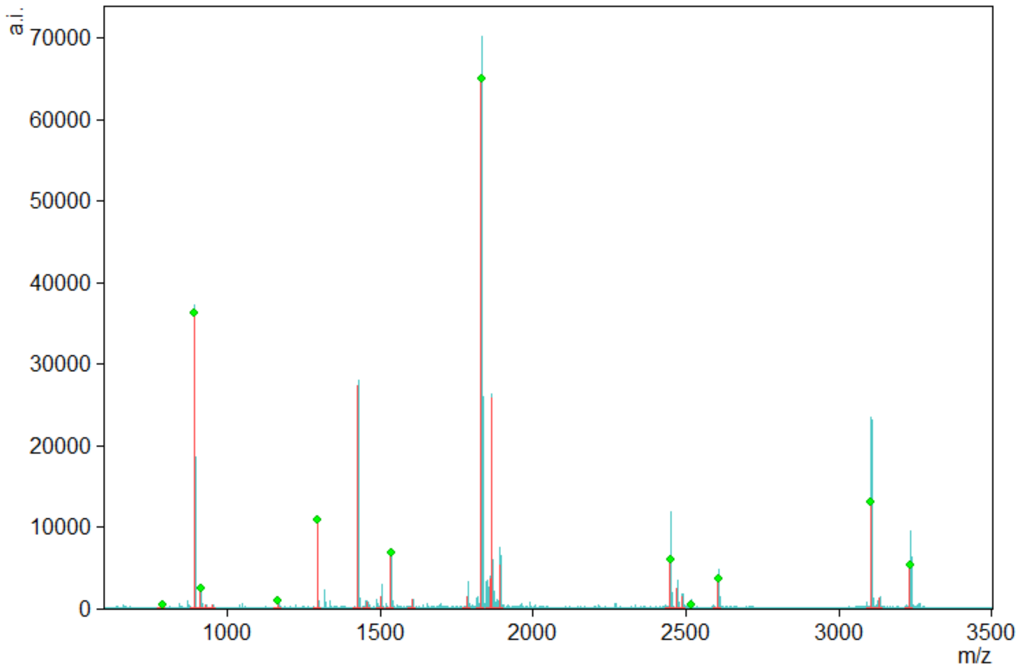


Figure 7.11− MS spectrum of NURC1$_{FL}$. MALDI-TOF spectrum of NURC1$_{FL}$. The absolute intensity is shown on the y axis, and the mass-to-charge ration (m/z) on the x axis. The green dots represent the peaks that matched NURC1$_{FL}$ peptides after trypsin digestion.

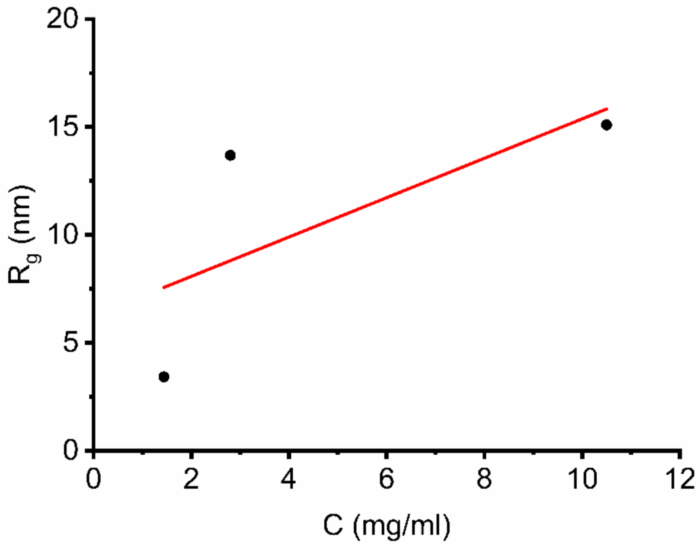


Figure 7.12− Linear regression of R_g values. R_g values resulting from the Guinier analysis of SAXS measurements in batch mode (black dots) and curve fit (red line). The fit was performed in OriginPro 2023 using the linear fit. Pearson’s R value was approximately 0.7.

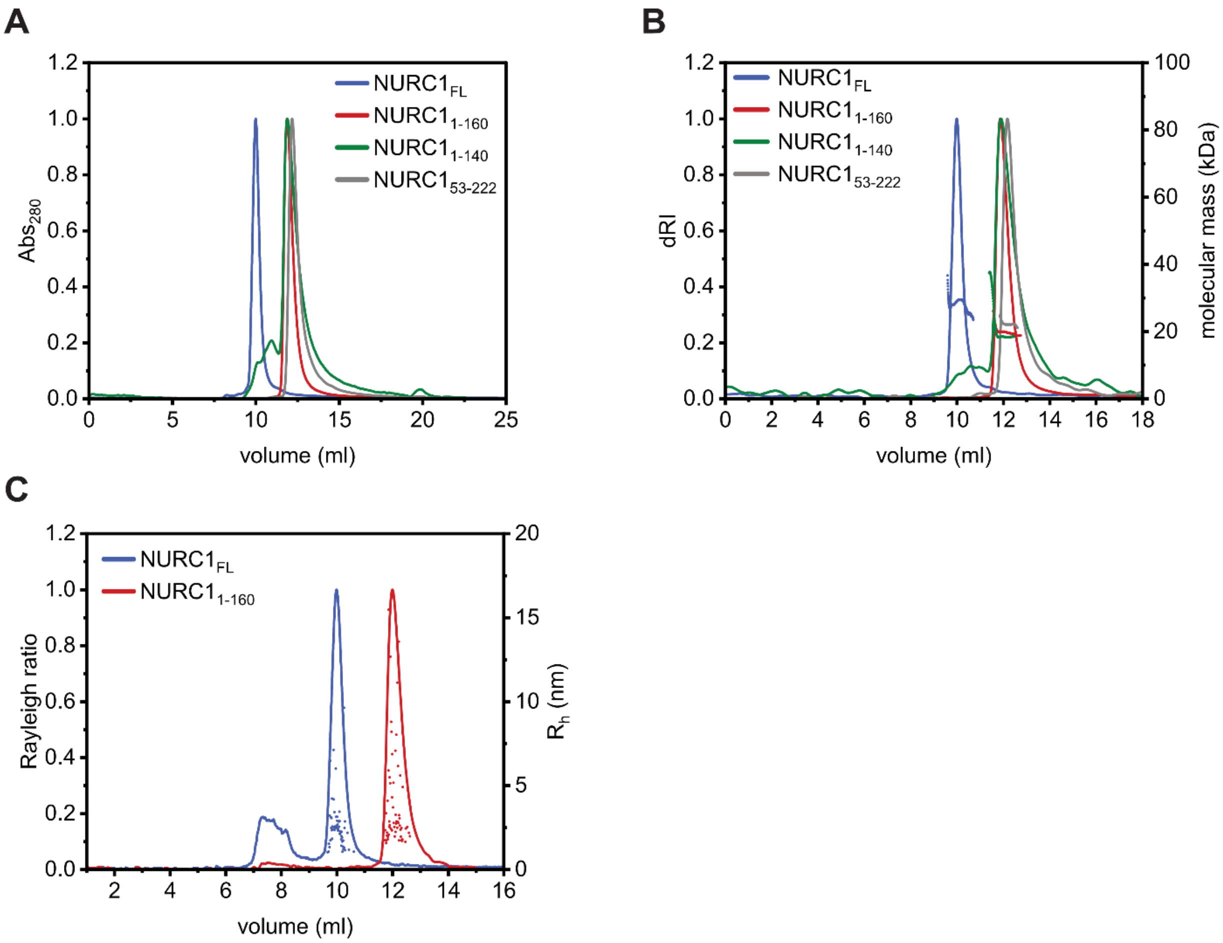


Figure 7.13− SEC-MALLS results of NURC1$_{FL}$. (**A**) SEC chromatogram of full-length and truncated versions of NURC1. (**B**) Differential refractive index (dRI) vs. elution volumes (ml) profiles of NURC1$_{FL}$ and truncated versions (full line), and their respective molar mass predictions (spheres). (**C**) Rayleigh Ratio vs. elution volume profiles of NURC1$_{FL}$ and NURC1$_{1\text{-}160}$ (full line) and their respective Rh prediction (spheres). Figure from ref.[136].

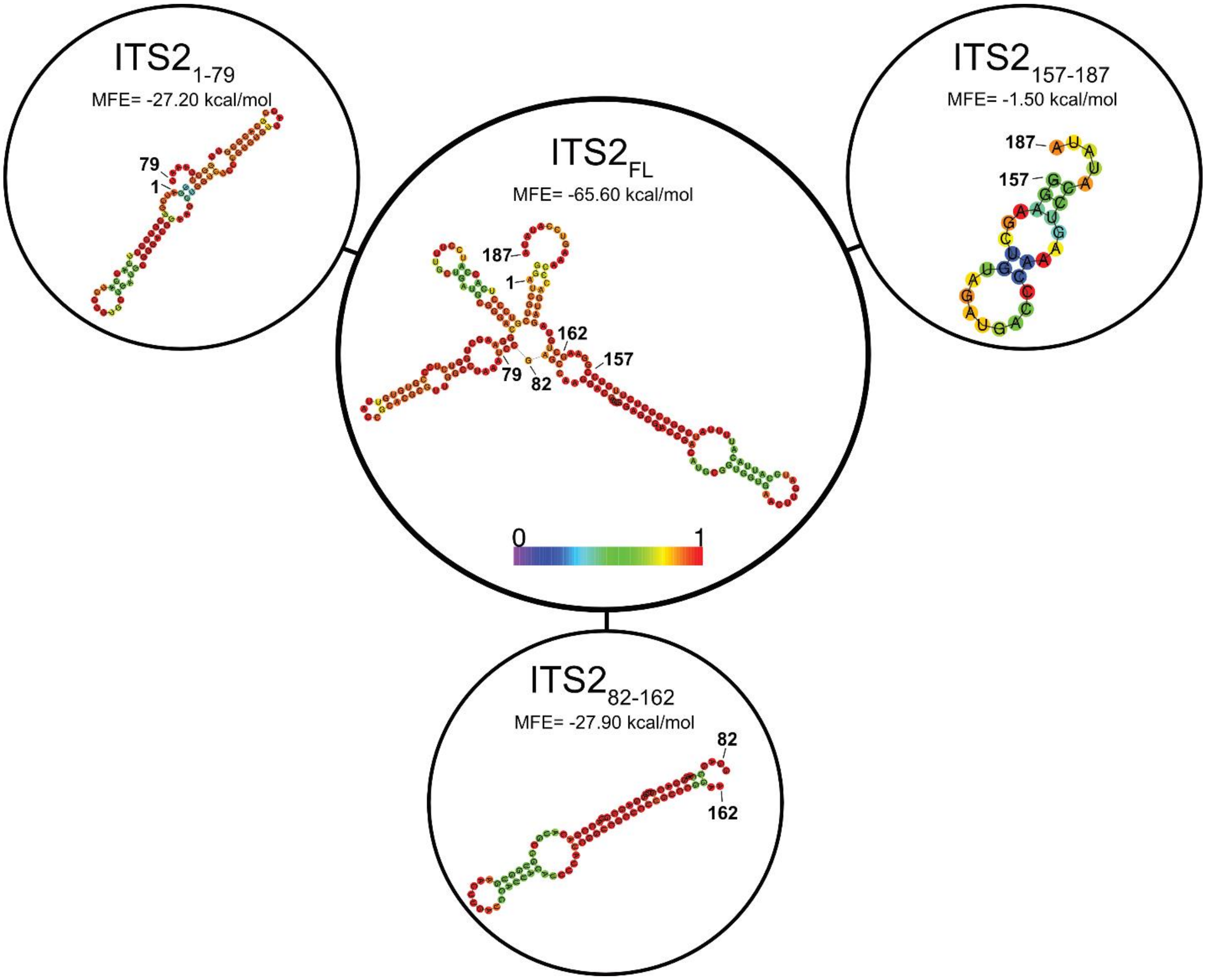


Figure 7.14− ITS2 RNA fragments structure predictions. Each fragment's minimum free energy (MFE) structures are shown. The probability of base pairs is displayed in colors. The color spectrum is blue (0) to red (1). The Vienna RNA Websuite was used to create the structures (Gruber et al., Nucleic Acids Research, Vol. 36, 2008). Figure from ref.[136].

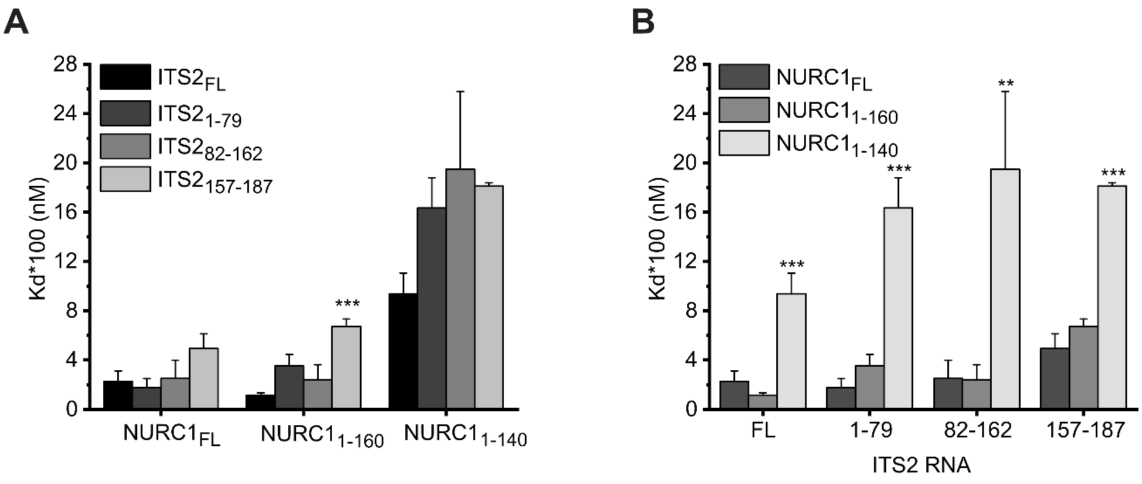


Figure 7.15− Overview of the binding between NURC1 and ITS2 versions. (**A**) K_d values comparing the binding of different truncated NURC1 proteins to the ITS2 fragments. (**B**) Binding affinities of NURC1$_{FL}$, NURC1$_{1-160}$, and NURC1$_{1-140}$ grouped by ITS2 region. The measurements were taken in triplicate, and the results were plotted as mean + standard deviation. The statistical analysis was performed in OriginPro 2021b (OriginLab) using one-way ANOVA (*p 0.05, **p 0.01 and ***p 0.001). Figure from ref.[136].

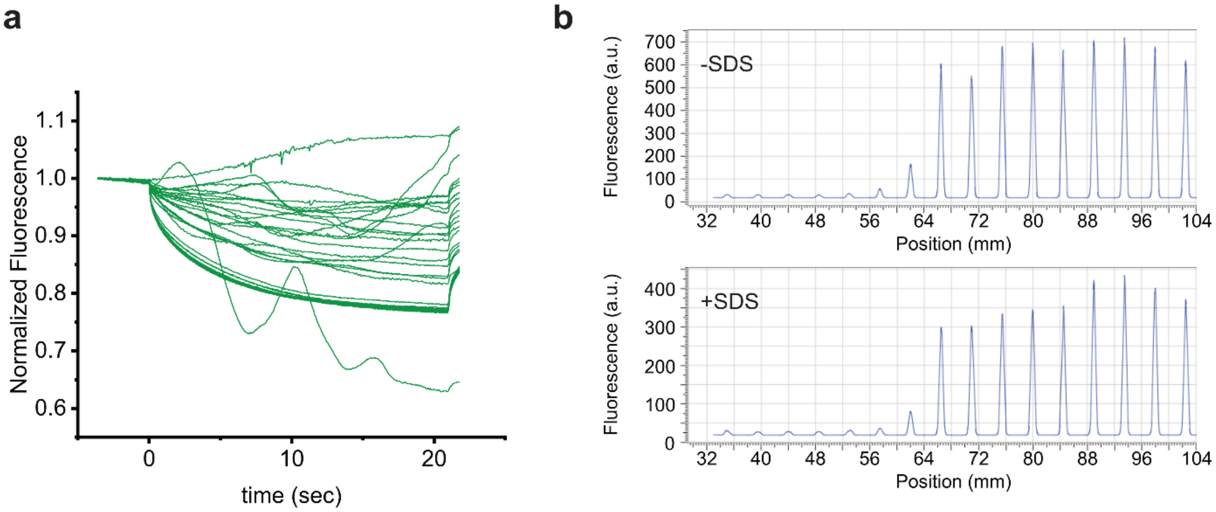


Figure 7.16− SD-test of NURC1$_{53-222}$ and ITS2$_{FL}$. (**A**) MST traces of cy5 labelled ITS2$_{FL}$ (20 nM) and NURC1$_{53-222}$ (the concentration increases from the bottom to top). (**B**) SD-test of NURC1$_{53-222}$ and ITS2 FL. Capillary scan of samples without (top) and with (bottom) SDS treatment are shown. The protein concentration increases from the right to the left. Figure from ref.[136].

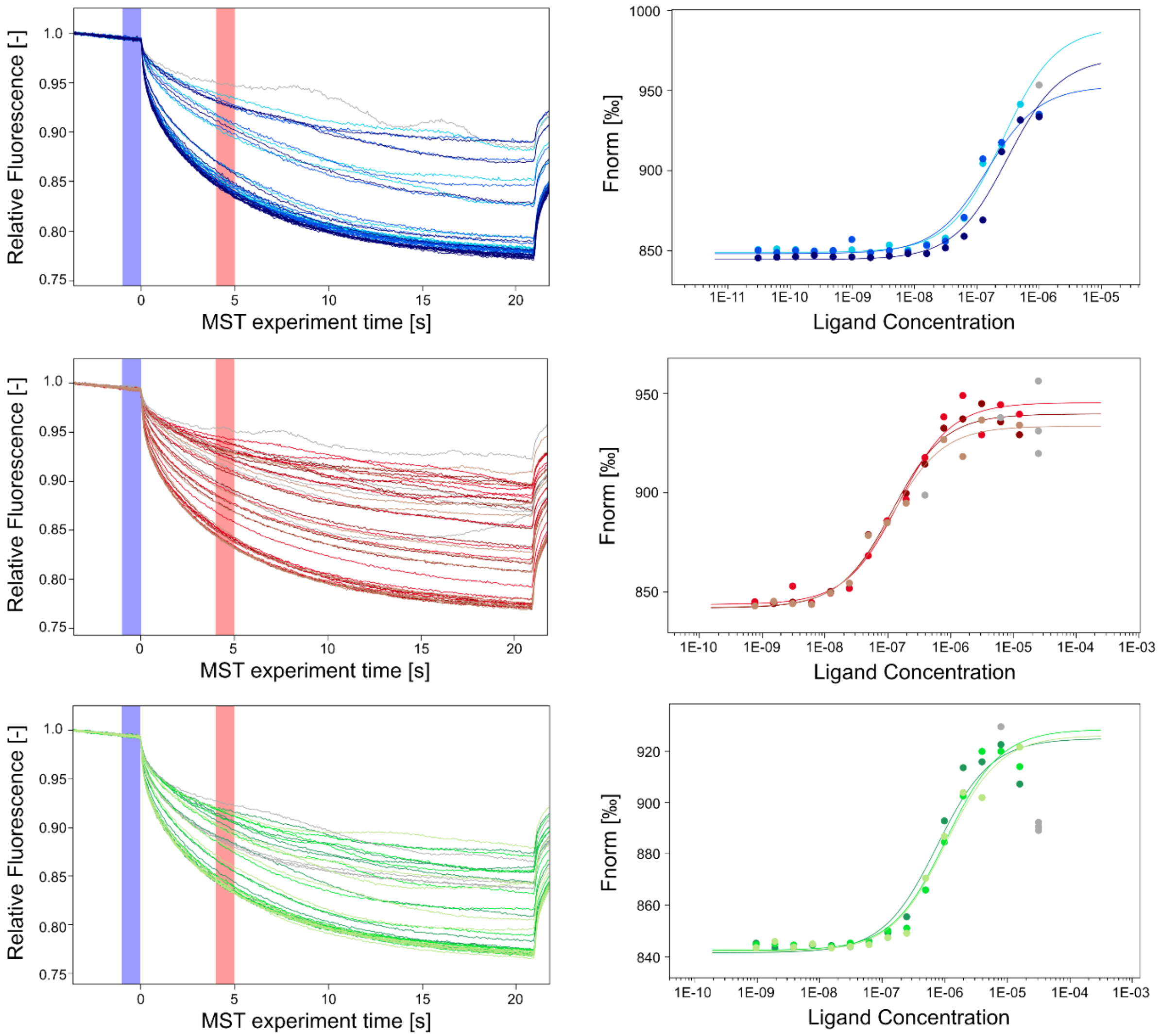


Figure 7.17− MST traces and binding curves between ITS2$_{FL}$ and NURC1 versions. NURC1$_{FL}$ is represented in blue, NURC1$_{1\text{-}160}$ in red and NURC1$_{1\text{-}140}$ in green. Triplicates are shown in different shades of the corresponding color. Data points shown in grey were not used for the calculation of K$_d$.

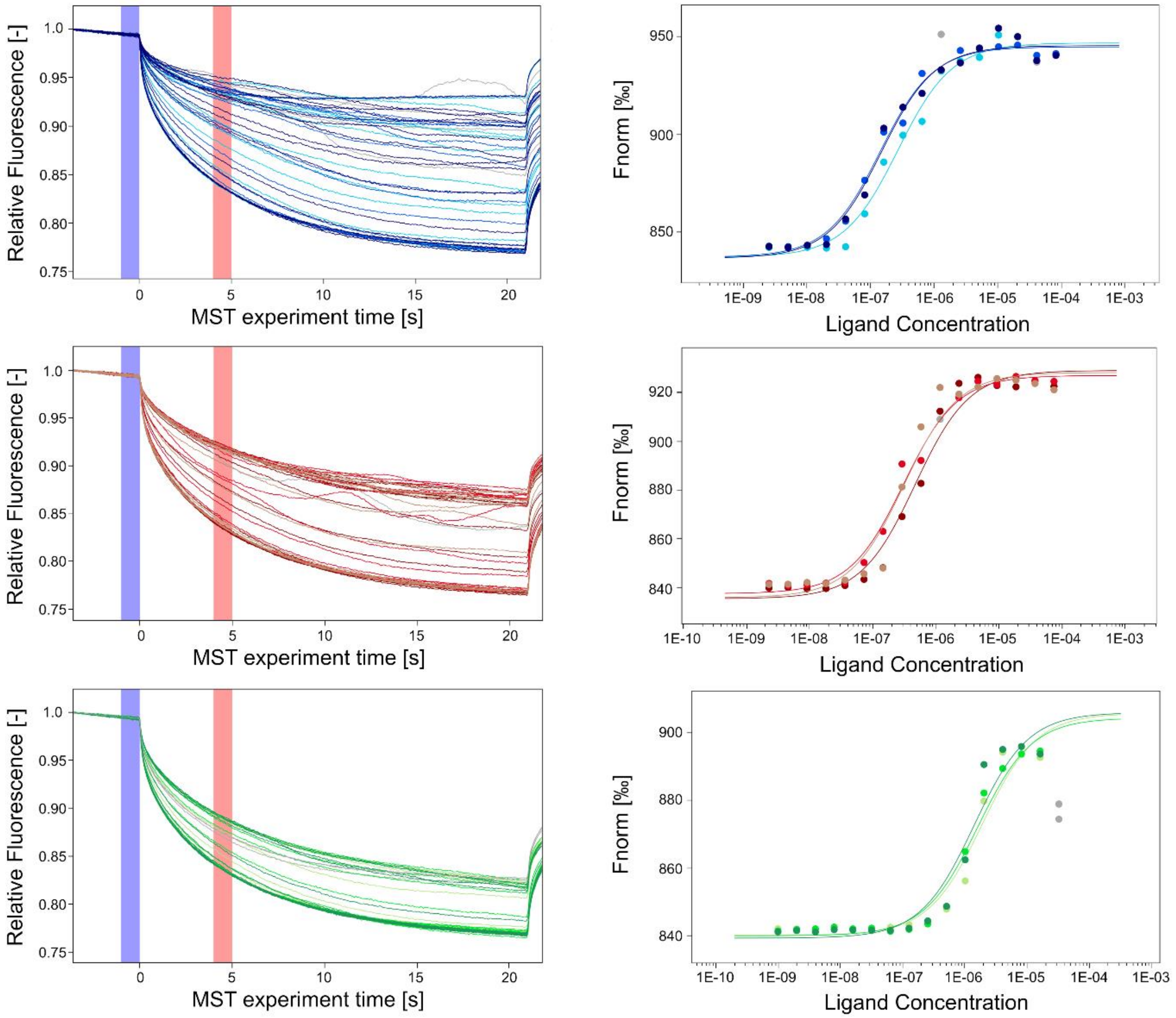


Figure 7.18− MST traces and binding curves between ITS2$_{1-79}$ and NURC1 versions. NURC1$_{FL}$ is represented in blue, NURC1$_{1-160}$ in red and NURC1$_{1-140}$ in green. Triplicates are shown in different shades of the corresponding color. Data points shown in grey were not used for the calculation of K$_d$.

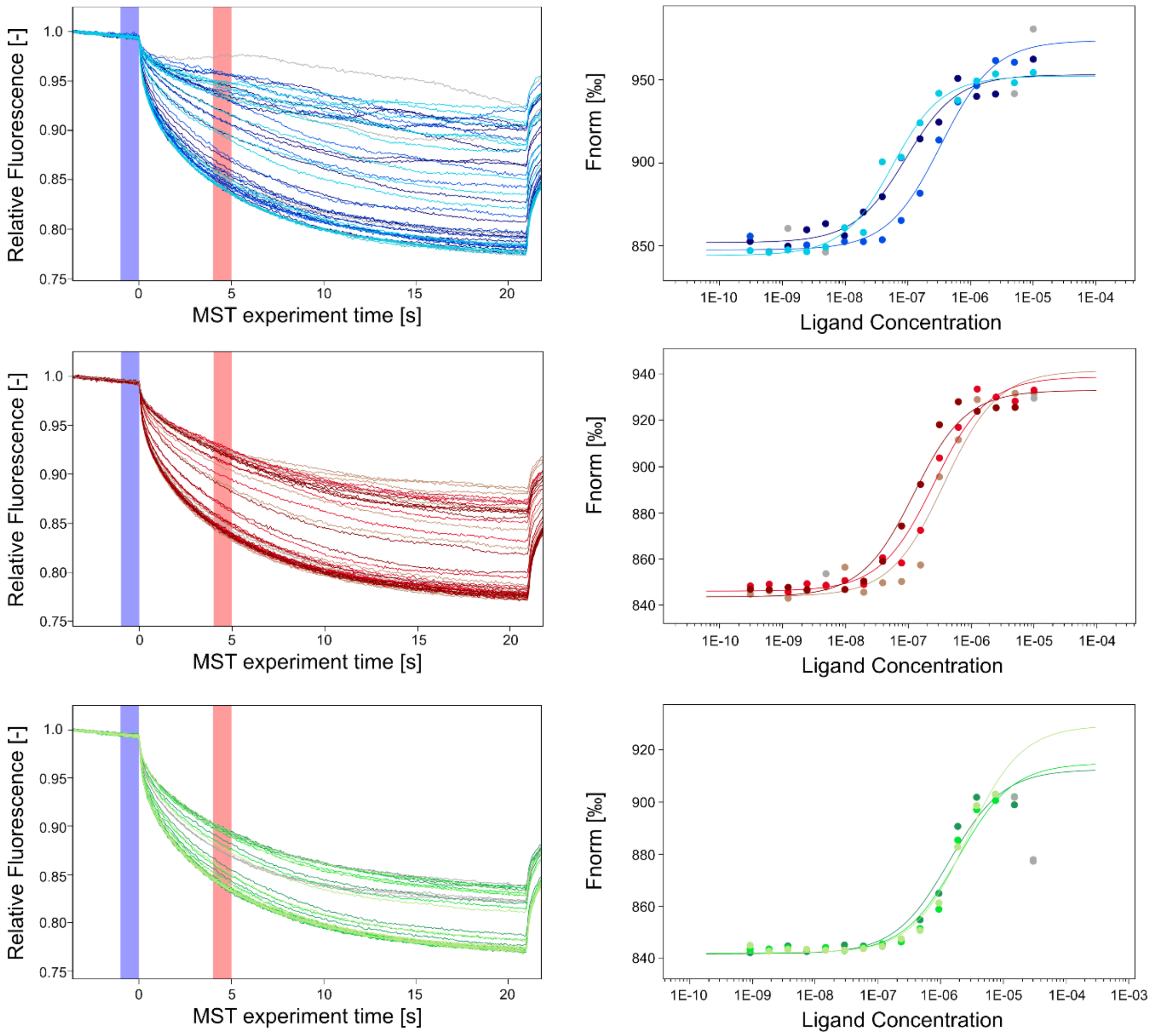


Figure 7.19− MST traces and binding curves between ITS2$_{86\text{-}162}$ and NURC1 versions. NURC1$_{FL}$ is represented in blue, NURC1$_{1\text{-}160}$ in red and NURC1$_{1\text{-}140}$ in green. Triplicates are shown in different shades of the corresponding color. Data points shown in grey were not used for the calculation of K$_d$

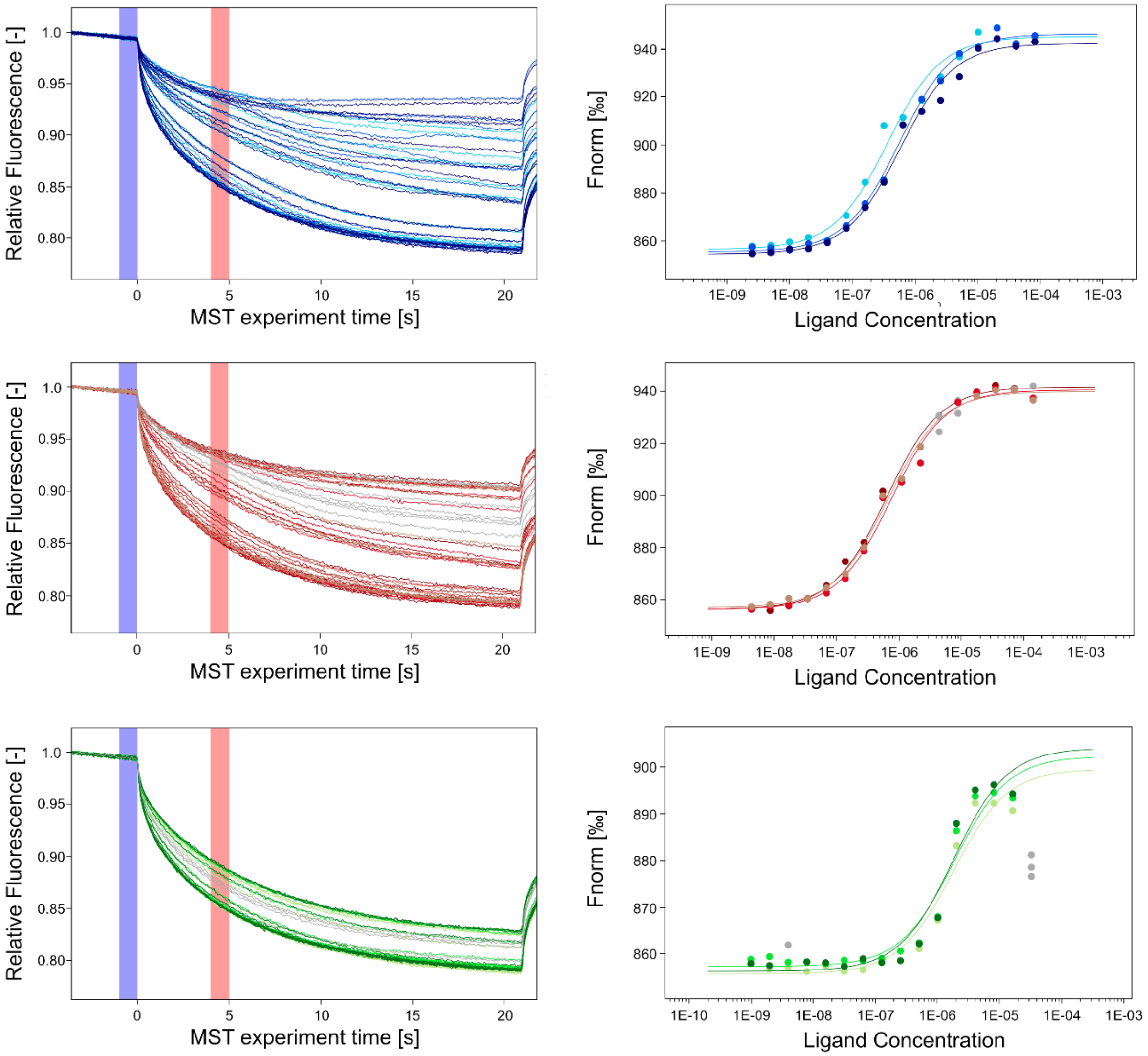


Figure 7.20− MST traces and binding curves between ITS2$_{157-187}$ and NURC1 versions. NURC1$_{FL}$ is represented in blue, NURC1$_{1-160}$ in red and NURC1$_{1-140}$ in green. Triplicates are shown in different shades of the corresponding color. Data points shown in grey were not used for the calculation of K$_d$.

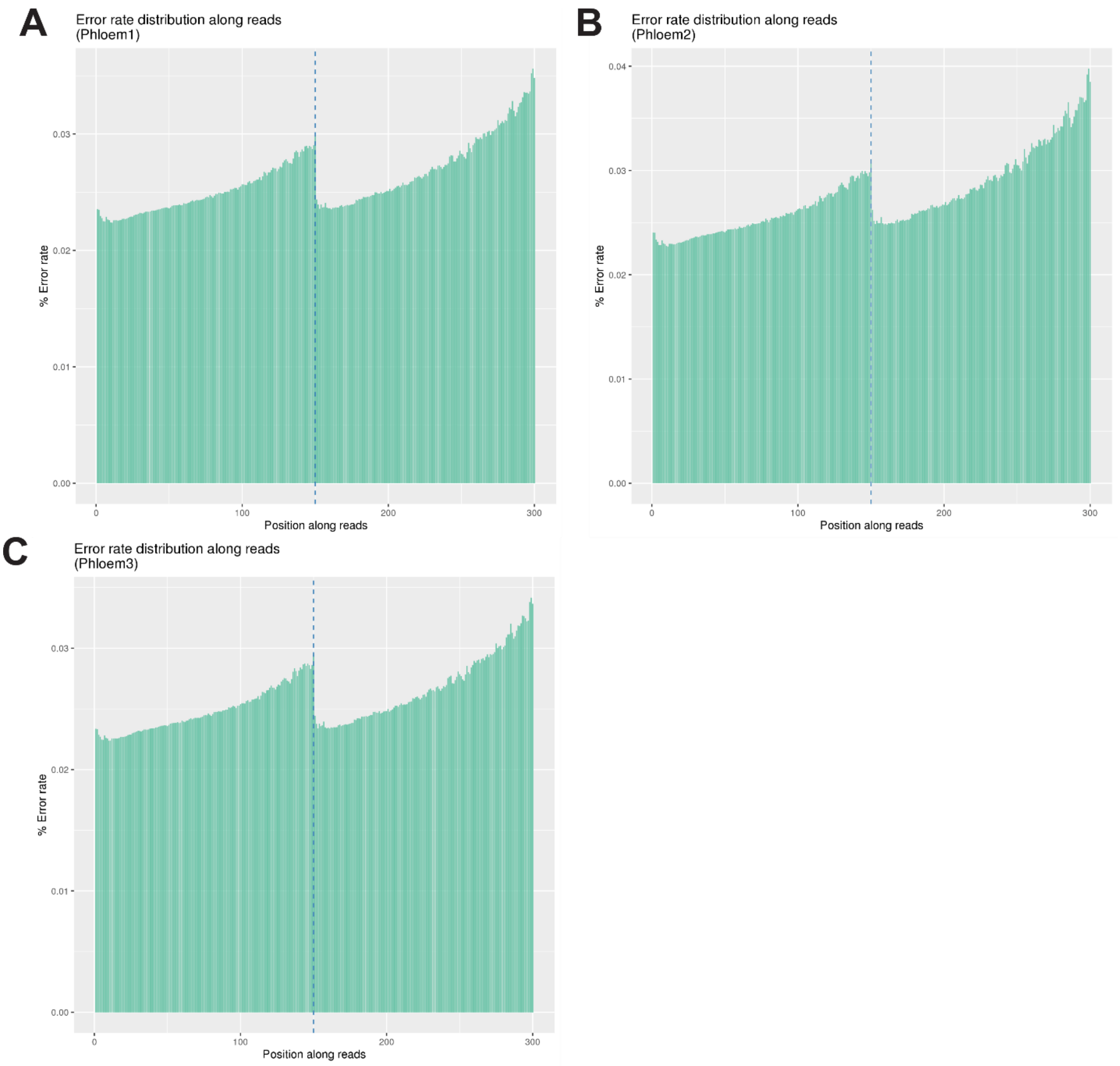


Figure 7.21− Phloem error rate distribution. The three plots show the error rate (%) of sequenced reads by the Illumina high-throughput sequencing platform across all triplicates (**A**-**C**).

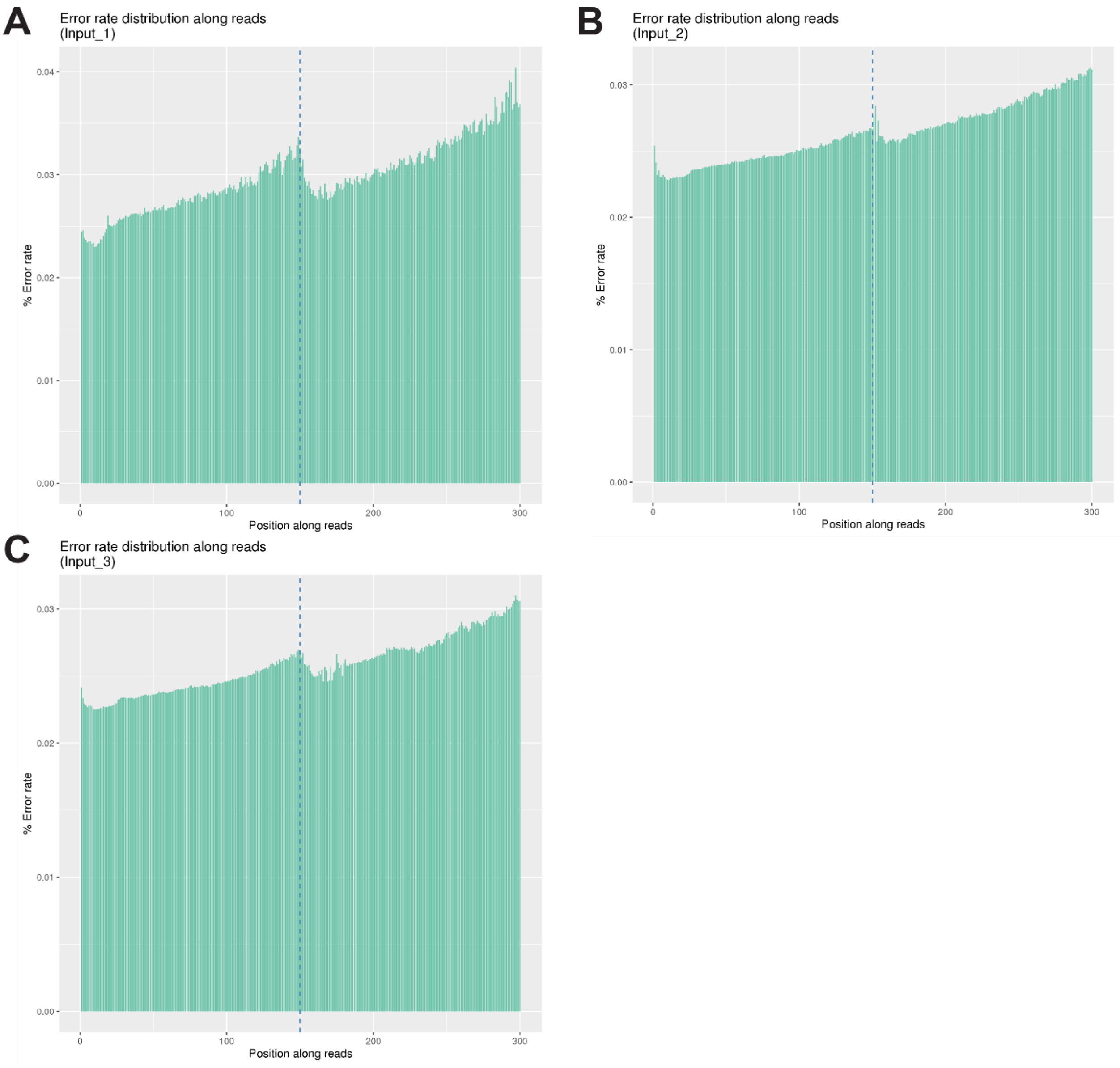


Figure 7.22− Leaf error rate distribution. The three plots show the error rate (%) of sequenced reads by the Illumina high-throughput sequencing platform across all triplicates (**A**-**C**).

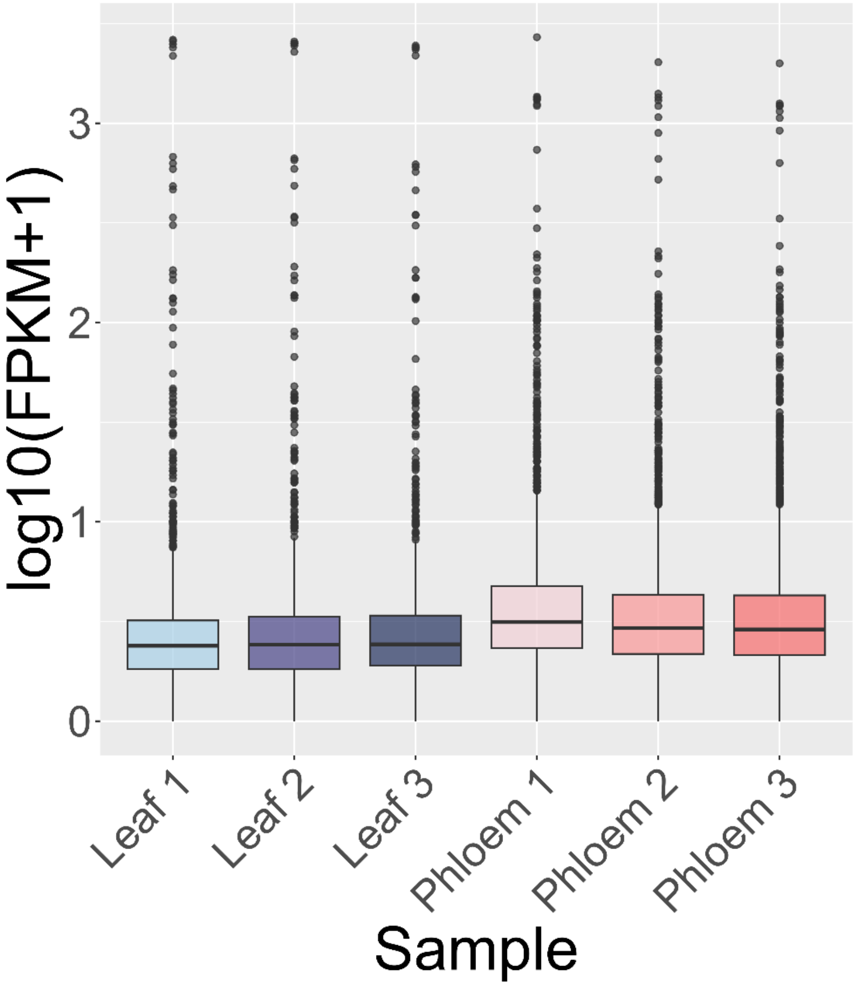


Figure 7.23− FPKM distribution of leaf and phloem lncRNAs. The FPKM distribution is represented as box plots. Leaf replicates are shown in shades of blue and phloem replicates in shades of pink.

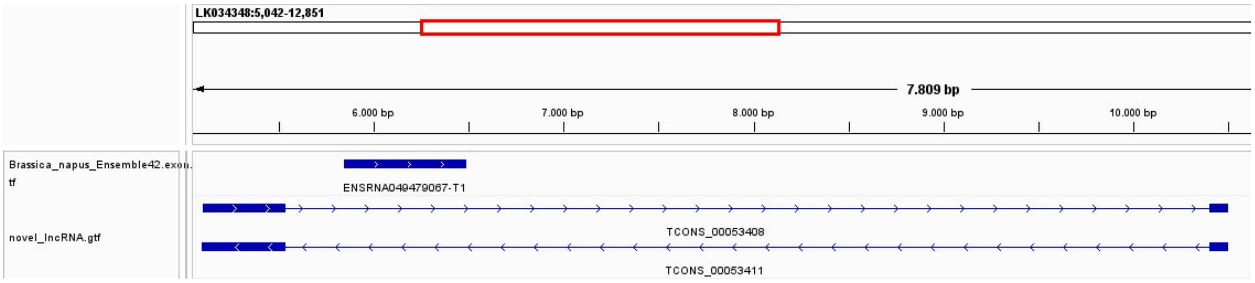


Figure 7.24− Example of two novel lncRNAs identified in the phloem. The transcripts TCONS_00053408 and TCONS_00053411 were annotated to the ENSRNA049479067 gene. The non-coding gene is annotated as a eukaryotic SSU rRNA and contains a single exon. The results were visualized with IGV[126] software.

Table 7.1- List of primers. All primers used in this work are listed below.

ID	Sequence 5´ to 3´
NURC1$_{\text{FL}}$_NdeI_fw	GATCCATATGGGGTGCCAAGGCGAAG
NURC1$_{\text{FL}}$_XhoI_rev	GATCCTCGAGGTCTTCTTCACTAAACTTGATCCTC
NURC1$_{\text{1-160}}$_fw	CAAACCAGTGGATTCGGTTCAGTAAGAGCGAAGACAACTTAATAAGG
NURC1$_{\text{1-160}}$_rev	CCTTATTAAGTTGTCTTCGCTCTTACTGAACCGAATCCACTGGTTTG
NURC1$_{\text{1-140}}$_fw	GCTTAAAGTCCATGTCATTGAACCAGAGAATTAAAAACCCAATCTGTGGAG
NURC1$_{\text{1-140}}$_rev	CTCCACAGATTGGGTTTTTAATTCTCTGGTTCAATGACATGGACTTTAAGC
NURC1$_{\text{53–222}}$_fw	TTAAGGTCTCGTTGCCCACCTCTTCAGAATAAAGCTACTGTGCTATAC
NURC1$_{\text{53-222}}$_rev	TTAAGGTCTCGAAGCTTAGTCTTCTTCACTAAACTTGATCCTCTTCGGG
45SrDNA_HindIII_fw	GACTGCAGAGGCCTGCATGCATAATACGACTCACTATAGGGAATTCCTAGTAAGCGCGAGTCATCAGC
45SrDNA_HindIII_rev	ACCATGATTACGCCACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGAATTCGTTAAGGAGCTGTTGCTTTGTTAGTG
T7p_ITS2$_{\text{FL}}$_fw	GAAATTAATACGACTCACTATAGGATCGTCGTCCCTCACCATCCTT
ITS2$_{\text{FL}}$_rev	GCCTTCGACATCTACTGGGTTTCAGGTATAT
T7p_ITS2$_{\text{1-79}}$_fw	GAAATTAATACGACTCACTATAGGATCGTCGTCCCTCACCATCCTTTGCTGATGCGGGACGGAAGCTGGTCTCCCGTGTGTTACCGCACGCGTTGGCCTAATT
ITS2$_{\text{1-79}}$_rev	AATTAGGCCAACGCGTGCGGTAACACACGGGAGACCAGCTTCCGTCCCGCATCAGCAAAGGATGGTGAGGGACGACGATCCTATAGTGAGTCGTATTAATTTC
T7p_ITS2$_{\text{82-162}}$_fw	GAAATTAATACGACTCACTATAGGAGCCAAGGACGCCTGGAGCGTACCGACATGCGGTGGTGAACTTGATCCATTACATTTTATCGGTCGCTCTTGTCCGGAA
ITS2$_{\text{82-162}}$_rev	TTCCGGACAAGAGCGACCGATAAAATGTAATGGATCAAGTTCACCACCGCATGTCGGTACGCTCCAGGCGTCCTTGGCTCCTATAGTGAGTCGTATTAATTTC
T7p_ITS2$_{\text{153-187}}$_fw	GAAATTAATACGACTCACTATAGGAAGCTGTAGATGACCCAAAGTCCATATA
ITS2$_{\text{153-187}}$_rev	TATATGGACTTTGGGTCATCTACAGCTTCCTATAGTGAGTCGTATTAATTTC

Table 7.2- SAXS data modeling results. *Ab initio* (DAMMIN and GASBOR) and CRYSOL fitting χ^2 results for the different models. Table from ref. [136].

Model	ID	χ^2
DAMMIN P1	-	1.06
GASBOR P1	-	1.56
AlphaFold	Q9LZ65 (UniProt)	11.97
i-Tasser	-	18.72
Nop15 CryoEM	3JCT, Chain OA (PDB)	55.57
Nop15 crystal - monomer	5T9P (PDB)	253.22

Table 7.3- Overview of NURC1 binding affinities. Binding affinities between different NURC1 protein versions and ITS2 RNA regions by MST. Table from ref. [136].

				Protein comparison		RNA comparison	
RNA	**Protein**	**Kd (nM)**	**SD**	**p-value**	**Fold change**	**p-value**	**Fold change**
$ITS2_{FL}$	$NURC1_{FL}$	228	83	-	1.0	-	1.0
	$NURC1_{1\text{-}160}$	116	20	4.6×10^{-1}	0.5	-	1.0
	$NURC1_{1\text{-}140}$	938	167	4.8×10^{-4}	4.1	-	1.0
$ITS2_{1\text{-}79}$	$NURC1_{FL}$	177	73	-	1.0	9.7×10^{-1}	0.8
	$NURC1_{1\text{-}160}$	353	93	4.3×10^{-1}	2.0	7.6×10^{-2}	3.0
	$NURC1_{1\text{-}140}$	1634	246	7.1×10^{-5}	9.2	2.5×10^{-1}	1.7
$ITS2_{82\text{-}162}$	$NURC1_{FL}$	251	147	-	1.0	10×10^{-1}	1.1
	$NURC1_{1\text{-}160}$	239	124	10×10^{-1}	0.9	4.9×10^{-1}	2.1
	$NURC1_{1\text{-}140}$	1948	631	3.8×10^{-3}	7.8	7.0×10^{-2}	2.1
$ITS2_{157\text{-}187}$	$NURC1_{FL}$	494	119	-	1.0	5.2×10^{-2}	2.2
	$NURC1_{1\text{-}160}$	673	60	7.0×10^{-2}	1.4	7.9×10^{-4}	5.8
	$NURC1_{1\text{-}140}$	1812	27	2.1×10^{-6}	3.7	1.2×10^{-1}	1.9

Table 7.4- Leaf mapping summary. Overview of leaf mapped reads. Read-1 refers to left reads (5' end) and read-2 to right reads (3' end).

Sample name	Leaf1	Leaf2	Leaf3
Total reads	60656346	71868818	72190676
Total mapped	56583709 (93.29%)	68539019 (95.37%)	69073734 (95.68%)
Multiple mapped	29957265 (49.39%)	35961181 (50.04%)	36558744 (50.64%)
Uniquely mapped	26626444 (43.90%)	32577838 (45.33%)	32514990 (45.04%)
Read-1	13380253 (22.06%)	16406920 (22.83%)	16336569 (22.63%)
Read-2	13246191 (21.84%)	16170918 (22.50%)	16178421 (22.41%)
Reads map to '+'	13248157 (21.84%)	16227382 (22.58%)	16204141 (22.45%)
Reads map to '-'	13378287 (22.06%)	16350456 (22.75%)	16310849 (22.59%)
Non-splice reads	16695805 (27.53%)	20208394 (28.12%)	20166057 (27.93%)
Splice reads	9930639 (16.37%)	12369444 (17.21%)	12348933 (17.11%)

Table 7.5- CuffCompare transcript classification.

Class symbol	Description
=	Complete, exact match of intron chain
c	Contain in reference (intron compatible)
k	Contained of reference (reverse containment)
m	Retained intron(s), all introns matched or retained
n	Retained intron(s), not all introns matched/covered
vv	Multi-exon with at least one junction match
e	Single exon transfrag partially covering an intron possible pre-mRNA fragment
o	Other same strand overlap with reference exons
s	Intron match on the opposite strand (likely a mapping error)
x	Exonic overlap on the opposite strand (like o or e but on opposite strand)
i	Fully contained within a reference intron
y	Contains a reference within its intron(s)
p	Possible polymerase run-on (no actual overlap)
r	Repeat (at least 50% bases soft-masked)
u	None of the above (unknown, intergenic)

Table 7.6- Phloem lncRNAs. Expression level in FPKM and annotation of phloem lncRNAs. The representation is limited to the top 20 most highly expressed lncRNAs.

transcript_id	Phloem_FPKM_Mean	gene_id	gene_name	gene_description
TCONS_00053408	2,239.68188	ENSRNA049479067	SSU_rRNA_eukarya	Eukaryotic small subunit ribosomal RNA
TCONS_00052053	1,039.20532	XLOC_044198	-	-
TCONS_00029205	407.66635	GSBRNA2T00091120001	BnaA08g02520D	BnaA08g02520D protein [Source:UniProtKB/TrEMBL;Acc:A0A078FPZ9]
TCONS_00025002	141.89797	XLOC_021198	-	-
TCONS_00041178	134.25845	XLOC_034925	-	-
TCONS_00013779	133.05062	GSBRNA2T00150017001	BnaC03g45540D	-
TCONS_00002288	125.66635	XLOC_001880	-	-
TCONS_00013331	116.20519	XLOC_011305	-	-
TCONS_00054552	113.23682	GSBRNA2T00053531001	BnaA07g37340D	BnaA07g37340D protein [Source:UniProtKB/TrEMBL;Acc:A0A078JI76]
TCONS_00047350	112.24225	XLOC_040172	-	-
TCONS_00016826	111.65617	GSBRNA2T00157764001	BnaC09g23130D	-
TCONS_00049928	108.33002	XLOC_042382	-	-
TCONS_00029870	98.30328	GSBRNA2T00001766001	BnaC01g27010D	BnaC01g27010D protein [Source:UniProtKB/TrEMBL;Acc:A0A078FXW0]
TCONS_00008124	83.65340	ENSRNA049474326	Plant_SRP	Plant signal recognition particle RNA
TCONS_00031102	83.07324	XLOC_026353	-	-
TCONS_00008210	79.50490	XLOC_006904	-	-
TCONS_00029325	68.84329	GSBRNA2T00094548001	BnaA09g20730D	BnaA09g20730D protein [Source:UniProtKB/TrEMBL;Acc:A0A078FQF0]
TCONS_00048943	63.72621	XLOC_041529	-	-
TCONS_00015161	62.53132	XLOC_012876	-	-
TCONS_00053411	55.89162	ENSRNA049479067	SSU_rRNA_eukarya	Eukaryotic small subunit ribosomal RNA

Table 7.7- Leaf lncRNAs. Expression level in FPKM and annotation of leaf lncRNAs. The representation is limited to the top 20 most highly expressed lncRNAs.

transcript_id	Leaf_FPKM_Mean	gene_id	gene_name	gene_description
TCONS_00018173	367.96200	GSBRNA2T00002257001	BnaC04g03080D	Mitogen-activated protein kinase [Source:UniProtKB/TrEMBL;Acc:A0A078FY15]
TCONS_00014756	166.79786	XLOC_012879	-	-
TCONS_00032202	61.79006	ENSRNA049449359	SSU_rRNA_eukarya	Eukaryotic small subunit ribosomal RNA
TCONS_00033144	43.61933	XLOC_028988	-	-
TCONS_00017913	41.09435	GSBRNA2T00095181001	BnaA05g10630D	BnaA05g10630D protein [Source:UniProtKB/TrEMBL;Acc:A0A078FRL3]
TCONS_00032448	39.70510	ENSRNA049479067	SSU_rRNA_eukarya	Eukaryotic small subunit ribosomal RNA
TCONS_00007644	38.94707	GSBRNA2T00147506001	BnaA06g26730D	-
TCONS_00010305	32.93892	XLOC_008995	MT54	Metallothionein-like protein LSC54
TCONS_00007965	32.81093	XLOC_006967	-	-
TCONS_00009021	31.77283	GSBRNA2T00153162001	BnaC08g06050D	-
TCONS_00028733	30.87420	XLOC_025097	-	-
TCONS_00001382	29.25289	XLOC_001183	-	-
TCONS_00024833	27.68939	XLOC_021682	-	-
TCONS_00002720	27.52079	XLOC_002334	-	-
TCONS_00033145	20.18332	XLOC_028988	-	-
TCONS_00033633	19.84136	GSBRNA2T00066951001	BnaA07g38720D	BnaA07g38720D protein [Source:UniProtKB/TrEMBL;Acc:A0A078JR33]
TCONS_00030346	18.97018	XLOC_026513	-	-
TCONS_00015096	17.44129	XLOC_013167	-	-
TCONS_00016428	17.07857	GSBRNA2T00070839001	BnaC07g03340D	BnaC07g03340D protein [Source:UniProtKB/TrEMBL;Acc:A0A078FF99]
TCONS_00021897	15.72549	GSBRNA2T00041514001	BnaC06g26300D	BnaC06g26300D protein [Source:UniProtKB/TrEMBL;Acc:A0A078GUS3]

Table 7.8- Annotation of phloem lncRNAs. Further annotation details of phloem lncRNAs. The representation is limited to the top 5 most highly expressed lncRNAs.

gene_id	gene_biotype	gene_name	class_code	transcript_biotype	nearest_ref
ENSRNA049479067	rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049479067-T1
ENSRNA049479067	rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049479067-T1
XLOC_044198	novel	XLOC_044198	u	lincRNA	
XLOC_044198	novel	XLOC_044198	u	lincRNA	
GSBRNA2T00091120001	protein_coding	BnaA08g02520D	o	sense_overlapping	CDY16515
GSBRNA2T00091120001	protein_coding	BnaA08g02520D	o	sense_overlapping	CDY16515
XLOC_021198	novel	XLOC_021198	u	lincRNA	
XLOC_021198	novel	XLOC_021198	u	lincRNA	
XLOC_034925	novel	XLOC_034925	u	lincRNA	
XLOC_034925	novel	XLOC_034925	u	lincRNA	
XLOC_034925	novel	XLOC_034925	u	lincRNA	

transcript_id	seqid	start	end	strand	exon_number
TCONS_00053408	LK034348	5103	5534	+	1
TCONS_00053408	LK034348	10407	10506	+	2
TCONS_00052053	LK033740	11333	11430	+	1
TCONS_00052053	LK033740	11480	12880	+	2
TCONS_00029205	LK032064	458083	458279	-	1
TCONS_00029205	LK032064	708451	708666	-	2
TCONS_00025002	LK031993	180791	180892	+	1
TCONS_00025002	LK031993	181624	182126	+	2
TCONS_00041178	LK032402	296634	296839	+	1
TCONS_00041178	LK032402	296954	297019	+	2
TCONS_00041178	LK032402	297454	297805	+	3

Table 7.9- Annotation of leaf lncRNAs. Further annotation details of leaf lncRNAs. The representation is limited to the top 6 most highly expressed lncRNAs.

gene_biotype	gene_name	class_code	transcript_biotype	nearest_ref
protein_coding	BnaC04g03080D	o	sense_overlapping	CDY17911
protein_coding	BnaC04g03080D	o	sense_overlapping	CDY17911
novel	XLOC_012879	u	lincRNA	
novel	XLOC_012879	u	lincRNA	
rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049449359-T1
rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049449359-T1
novel	XLOC_028988	u	lincRNA	
novel	XLOC_028988	u	lincRNA	
protein_coding	BnaA05g10630D	o	sense_overlapping	CDY17125
protein_coding	BnaA05g10630D	o	sense_overlapping	CDY17125
rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049479067-T1
rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049479067-T1
rRNA	SSU_rRNA_eukarya	o	sense_overlapping	ENSRNA049479067-T1

transcript_id	seqid	start	end	strand	exon_number	gene_id
TCONS_00018173	LK032079	424136	424156	-	1	GSBRNA2T00002257001
TCONS_00018173	LK032079	483525	483778	-	2	GSBRNA2T00002257001
TCONS_00014756	LK031984	60250	60431	+	1	XLOC_012879
TCONS_00014756	LK031984	60835	61470	+	2	XLOC_012879
TCONS_00032202	LK034093	14961	15064	+	1	ENSRNA049449359
TCONS_00032202	LK034093	15114	16305	+	2	ENSRNA049449359
TCONS_00033144	LK035195	9967	10270	-	1	XLOC_028988
TCONS_00033144	LK035195	10345	10560	-	2	XLOC_028988
TCONS_00017913	LK032071	38475	38857	-	1	GSBRNA2T00095181001
TCONS_00017913	LK032071	39018	39109	-	2	GSBRNA2T00095181001
TCONS_00032448	LK034348	4525	4614	+	1	ENSRNA049479067
TCONS_00032448	LK034348	4664	5534	+	2	ENSRNA049479067
TCONS_00032448	LK034348	10407	10444	+	3	ENSRNA049479067

Table 7.10- Differential expression analysis of leaf and phloem lncRNAs. DESeq2 partial output table showing the statistically significant differential expressed parent genes of lncRNAs of the phloem compared to the leaf. The baseMean represents the average of the normalized counts taken over all samples. The log2FoldChange between the groups, it's standard error (lfcSE), and the adjusted p-value are shown.

gene_id	baseMean	log2FoldChange	lfcSE	padj
ENSRNA049479067	23340.37	3.51	0.2	1.63E-68
GSBRNA2T00105977001	474.96	-3.75	0.21	1.11E-66
ENSRNA049439209	232.56	-3.94	0.26	1.01E-50
GSBRNA2T00157764001	1272.6	3.66	0.26	1.82E-43
GSBRNA2T00015777001	758.77	3.5	0.25	2.01E-43
GSBRNA2T00001751001	303.25	3.95	0.28	1.20E-42
GSBRNA2T00056993001	614.08	3.71	0.28	3.00E-38
GSBRNA2T00074276001	160.7	-3.59	0.28	3.57E-35
GSBRNA2T00122319001	186.24	-3.38	0.27	9.91E-35
GSBRNA2T00030279001	260.56	3.41	0.27	2.47E-34
GSBRNA2T00094548001	1671.95	2.64	0.23	2.99E-29
GSBRNA2T00021212001	928.12	2.17	0.2	2.16E-25
GSBRNA2T00094565001	262.08	3.23	0.31	1.04E-24
GSBRNA2T00123104001	206.72	-2.81	0.27	1.04E-24
GSBRNA2T00003166001	1274.35	-3.51	0.35	4.47E-22
GSBRNA2T00053301001	546.83	2.09	0.21	2.91E-21
GSBRNA2T00147506001	1272.99	-1.91	0.2	2.91E-21
GSBRNA2T00093012001	958.23	1.89	0.2	2.07E-20
GSBRNA2T00048150001	1592.53	2.28	0.24	4.69E-20
GSBRNA2T00110431001	316.33	-2.13	0.23	1.35E-19
GSBRNA2T00149899001	911.22	2.84	0.31	2.12E-19
GSBRNA2T00059499001	368.96	2.33	0.25	2.44E-19
GSBRNA2T00067049001	228.67	-2.66	0.29	3.69E-19
GSBRNA2T00001766001	1643.11	1.95	0.22	5.60E-18
GSBRNA2T00067883001	342.03	-2.82	0.34	9.69E-16
GSBRNA2T00121801001	98.73	-2.79	0.34	1.04E-15
GSBRNA2T00158295001	671.91	1.56	0.2	2.35E-14
GSBRNA2T00155075001	349.8	2.22	0.28	2.53E-14
GSBRNA2T00101872001	220.23	-2.46	0.33	5.06E-13
GSBRNA2T00129173001	356.06	1.63	0.22	7.79E-13
GSBRNA2T00033363001	299.98	-1.8	0.25	2.40E-12
GSBRNA2T00062965001	104.48	2.22	0.32	2.68E-11
GSBRNA2T00042511001	665.54	-1.3	0.19	3.27E-11

GSBRNA2T00111268001	158.47	-2.52	0.37	3.94E-11
GSBRNA2T00020051001	145.22	1.93	0.28	4.39E-11
GSBRNA2T00095210001	212.64	-1.68	0.25	1.06E-10
GSBRNA2T00118795001	614.59	-1.49	0.22	1.20E-10
GSBRNA2T00038973001	242.12	1.83	0.27	1.36E-10
GSBRNA2T00139640001	225.04	-1.81	0.27	1.74E-10
GSBRNA2T00150028001	189.33	2.52	0.39	6.11E-10
GSBRNA2T00077189001	393.54	1.34	0.21	2.28E-09
GSBRNA2T00064585001	397.94	-1.78	0.28	2.46E-09
GSBRNA2T00062302001	203.78	-1.48	0.24	3.00E-09
GSBRNA2T00030290001	132.92	1.67	0.27	4.66E-09
GSBRNA2T00082747001	91.14	-1.72	0.29	1.51E-08
GSBRNA2T00050712001	262.32	1.35	0.23	3.74E-08
GSBRNA2T00156339001	157.92	2.1	0.37	5.73E-08
GSBRNA2T00032350001	283.91	1.34	0.23	7.01E-08
GSBRNA2T00010572001	184.82	-1.33	0.24	1.13E-07
GSBRNA2T00152013001	135.56	2.08	0.37	1.13E-07
GSBRNA2T00040093001	131.14	-1.63	0.31	5.26E-07
GSBRNA2T00119433001	82.35	-1.68	0.32	7.00E-07
GSBRNA2T00121275001	87.4	-1.56	0.3	7.07E-07
GSBRNA2T00119065001	432.68	-1.01	0.19	7.16E-07
GSBRNA2T00154126001	99.72	-2.39	0.46	1.12E-06
GSBRNA2T00005568001	38.3	-2.23	0.43	1.34E-06
GSBRNA2T00130804001	178.91	1.42	0.28	1.34E-06
GSBRNA2T00121304001	155.11	1.98	0.39	1.37E-06
GSBRNA2T00061839001	186.99	-1.34	0.26	1.40E-06
GSBRNA2T00123125001	168.01	1.45	0.28	1.77E-06
GSBRNA2T00059174001	184	1.44	0.29	3.40E-06
GSBRNA2T00086485001	340.81	-1.38	0.28	3.42E-06
GSBRNA2T00146335001	89.11	1.66	0.34	4.29E-06
GSBRNA2T00004176001	160.19	1.57	0.32	5.42E-06
GSBRNA2T00098545001	83.91	-1.91	0.4	6.59E-06
GSBRNA2T00034635001	53.28	-1.82	0.38	7.62E-06
GSBRNA2T00062391001	66.08	-1.89	0.4	8.77E-06

Scientific Publication

Part of the research presented in this thesis has been published in:

Fernandes, R. *et al.* Structural and functional analysis of a plant nucleolar RNA chaperone-like protein. *Sci Rep* **13**, (2023).

Acknowledgements

I would like to thank my supervisors, Prof. Dr. Julia Kehr and Prof. Dr. Stefan Hoth, for warmly welcoming me into their teams and for providing invaluable advice and support throughout my PhD journey.

Many thanks to Prof. Dr. Richard Morris for taking the time to read my dissertation amidst his busy schedule and for sharing his expertise. Also, my sincere appreciation for his willingness to travel to Hamburg and join my defense committee.

My heartfelt thanks to Dr. Sven Falke and Dr. Melissa Graewert. I am grateful for your readiness in assisting with the many scientific experiments at DESY and for helping me out with the data analysis. It has truly been a pleasure to work with you.

My gratitude to all the members of the MPG and MPP groups who have been a part of my daily life in the past few years. It has been a pleasure working alongside each and every one of you. In particular, thanks to Dr. Steffen Ostendorp for his guidance and knowledge, and for always addressing my questions with a smile and his contagious enthusiasm for science.

To all my fellow PhD colleagues, thanks for the companionship and help, in particular, Linn von Pein, Sebastian Fricke, Tina Niemann, Mariana Motta and Stephan Portheine. Special thanks to my PhD sisters, my ride or die - Francisca Méndez and Kim Lühmann - for their friendship and unconditional support.

I would like to thank my lifelong friends, Liliana and Raul, for never letting the distance and time wear our friendship out.

My utmost appreciation goes to my parents, Lucília and João, for always believing and supporting me no matter what, and for their unconditional love. Thanks to my grandparents, Zézinha and Fernando, for sharing their life stories, advice, support, and love.

Last but not least, a special thanks to my life partner, Diogo. I am grateful for his love and friendship, for always having my back, and for constantly bringing a smile to my face.