# Functional-structural characterization of phloem-mobile protein GRP8 in *Arabidopsis thaliana*

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

Francisca Méndez-Pinochet

Hamburg, 2022

Date of Disputation: 28.04.2023

Commission:
Prof. Dr. Julia Kehr
Prof. Dr. Richard Morris
Prof. Dr. Sigrun Reumann
Dr. Magdalena Weingartner

The following research has been conducted from May 2019 through November 2022 under the supervision of Prof. Dr. Julia Kehr at the Department of Molecular Plant Genetics located in the Institute of Plant Science and Microbiology in Hamburg.

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Synergy grant ERCSyG 2018 810131 – PLAMORF).

Structural research has been performed in cooperation with the European Molecular Biology Laboratory (EMBL) located at Deutsches Elektronen-Synchrotron (DESY) in Hamburg.

Supervisor: Professor Dr. Julia Kehr

First examiner: Professor Dr. Julia Kehr
Second examiner: Professor Dr. Richard Morris

# Index

## Abstract

Long-distance signaling, is permitted via the phloem system, is enabled by the translocation of small molecules. It is well established that RNA molecules are complexed with RNA-binding proteins (RBPs) forming ribonucleoprotein complexes (RNPs) in cells. The co-occurrence of different RNAs and RBPs in phloem exudate samples suggests their involvement in conferring RNA stability and mobility. RBPs are enriched in intrinsically disordered regions (IDRs) that are characterized by a low content of bulky hydrophobic amino acids and a high proportion of small, polar and/or charged amino acids. These intrinsic disordered proteins (IDPs) are characterized by the lack of defined structure either partially or completely, including IDRs. They possess the ability to fold into different conformations according to their binding partner in a wide range of interactions of high selectivity and low affinity.

The presence of GRP8 as protein has been detected in phloem exudates of *Arabidopsis thaliana*. GRP8 structure consists of an RRM at the N-terminus and a low complexity glycine-rich domain at the C-terminus. GRP8 was studied *in vitro* in order to characterize its structure and function. For this, different approaches were performed including methods that are less common such as microscale thermophoresis (MST) and small angle x-ray scattering (SAXS). GRP8 was *in vitro* produced and purified. Protein production was achieved by the combination of several techniques that include cloning and expression of GRP8 and GRP8$^{short}$ in *E. coli* cells, protein extraction and purification. For further characterization assays, a truncated version of GRP8 was produced (GRP8$^{short}$), which only included the RRM domain. After the purification of GRP8 and GPR8$^{short}$, different assays were performed to elucidate the functionality of GRP8. RNA was extracted and isolated from phloem sap from *B. napus* and a CnBr-Sepharose GRP8-bound affinity column was performed. *B. napus* phloem sap was used instead of phloem sap from *A. thaliana* because this can only be obtained by EDTA-facilitated exudation or aphid stylectomy. These methods in *A. thaliana* can lead to poor sample quality because of contamination and low amount of phloem sap sample collection. In addition, it has been established that *B. napus* is a suitable model system for phloem sap analysis. The purity and integrity of the phloem-RNA was check by PCR and *Bioanalyzer*. Samples with contamination or RNA degradation were discarded. After conducting the CnBr-Sepharose GRP8-bound affinity column and analyzing the RNA-sequencing results, transcripts of interest were aligned through *EnsemblPlants* to find homolog transcripts in the *A. thaliana* genome. The CnBr-Sepharose GRP8-bound affinity column results showed that GRP8 was binding to a

wide range of different RNAs. After the RNA-sequencing analysis, some enriched transcripts were selected and tested by MST to confirm if
GRP8 binds them, and it was confirmed that GRP8 was binding all of them. GRP8 was also bound to UTRs of two different transcripts. In addition, a motif alignment was performed by the motif-based sequence alignment tool, nevertheless no common motifs were found, supporting that GRP8 is binding a wide range of RNAs. $GRP8^{short}$ was also tested, and it was observed that it was not binding any of the mRNAs from the RNA-sequencing analysis. Because $GRP8^{short}$ only includes the RRM, it was concluded that the glycine-rich domain plays a key role in RNA binding. GRP8 has been described as a IDP with a glycine-rich domain as a low complexity region including [G/S]Y[G/S] and RGG motifs. During protein purification of GRP8 three major hallmarks that either aggregated or degraded the protein were revealed: time length of the purification, protein concentration and environmental temperature. These appeared to be linked to the low complexity region of GRP8 and the presence of a PrLD in the glycine-rich domain. The reversibility of the aggregation was an indication of a liquid-liquid phase separation state. The PrLD behavior of GRP8 was tested by the Thioflavin T (ThT) and the liquid-liquid phase separation (LLPS) assay, where the formation of condensation droplets were observed. Nevertheless, the labeling of the protein is necessary to confirm the formation of an RNA granule. Structural characterization of GRP8 was performed by SAXS because of the IDP nature of GRP8. First, different temperatures were measured with GRP8 monomer in batch-SAXS where qualitative analysis was performed. Overall, when the environmental temperature was 5°C, aggregation indication of GRP8 was observed. Furthermore, SEC-SAXS was performed to understand the native structure of GRP8 monomer and dimer. GRP8 dimer versus GRP8 monomer results showed that the GRP8 monomer presents a more unfolded and disordered shape than GRP8 dimer. Although both proteins showed a similar behavior with the Kratky plots, a slight difference was observed between them. After the SEC-SAXS data was processed, it was proceeded to the 3D modeling of GRP8 monomer. The 3D model of GRP8 was presented and it represents the combination of different techniques including experimental data measured through SEC-SAXS and computational tools that were used to create this hybrid model of GRP8 monomer.

## Zusammenfassung

Die Signalübertragung über große Entfernungen, die über das Phloemsystem geschieht, wird durch die Translokation kleiner Moleküle ermöglicht. Es ist bekannt, dass sich RNA-Moleküle in Zellen mit RNA-bindenden Proteinen (RBPs) zu Ribonukleoprotein-Komplexen (RNPs) verbinden. Das gemeinsame Auftreten verschiedener RNAs und RBPs in Phloem-Exsudat-Proben lässt vermuten, dass diese für die Stabilität und Mobilität der RNA verantwortlich sind. RBPs sind in intrinsisch ungeordneten Regionen (IDRs) angereichert, die durch einen geringen Gehalt an sperrigen hydrophoben Aminosäuren und einen hohen Anteil an kleinen, polaren und/oder geladenen Aminosäuren gekennzeichnet sind. Diese intrinsisch ungeordneten Proteine (IDPs) zeichnen sich dadurch aus, dass ihnen eine definierte Struktur entweder teilweise oder vollständig fehlt, einschließlich der IDRs. Sie besitzen die Fähigkeit, sich in Abhängigkeit von ihrem Bindungspartner in verschiedene Konformationen zu falten, was zu einer Vielzahl von Wechselwirkungen mit hoher Selektivität und geringer Affinität führt.
Das Vorkommen von GRP8 als Protein wurde in Phloem-Exsudaten von *Arabidopsis thaliana* nachgewiesen. Die Struktur von GRP8 besteht aus einem RRM am N-Terminus und einer Glycin-reichen Domäne mit geringer Komplexität am C-Terminus. GRP8 wurde *in vitro* untersucht, um seine Struktur und Funktion zu charakterisieren. Dazu wurden verschiedene Ansätze verfolgt, darunter auch weniger verbreitete Methoden wie die Thermophorese (MST) und die Kleinwinkel-Röntgenstreuung (SAXS). GRP8 wurde *in vitro* hergestellt und gereinigt. Die Proteinproduktion wurde durch die Kombination mehrerer Techniken erreicht, die das Klonen und die Expression von GRP8 und GRP8$^{short}$ in *E. coli*-Zellen, die Proteinextraktion und die Reinigung umfassen. Für weitere Charakterisierungsversuche wurde eine verkürzte Version von GRP8 (GRP8$^{short}$) hergestellt, die nur die RRM-Domäne enthält. Nach der Reinigung von GRP8 und GPR8$^{short}$ wurden verschiedene Tests durchgeführt, um die Funktionalität von GRP8 zu klären. RNA wurde aus dem Phloem-Saft von *B. napus* extrahiert und isoliert, und es wurde eine CnBr-Sepharose-Affinitätssäule mit daran gebundenem GRP8 durchgeführt. Der Phloemsaft von *B. napus* wurde anstelle des Phloemsaftes von *A. thaliana* verwendet, da dieser nur durch EDTA-vermittelte Exsudation oder Blattlaus-Stylektomie gewonnen werden kann, die beide zu einer geringen Probenkonzentration und Kontamination führen. Darüber hinaus wurde festgestellt, dass *B. napus* ein geeignetes Modellsystem für die Phloemsaftanalyse ist. Die Reinheit und Integrität der Phloem-RNA wurde mittels PCR und Bioanalyzer überprüft. Proben mit Verunreinigungen oder RNA-Degradation wurden

aussortiert. Nach der Durchführung der CnBr-Sepharose GRP8-gebundenen Affinitätssäule und der Analyse der RNA-Sequenzierungsergebnisse wurden interessante Transkripte mit Hilfe von *EnsemblPlants* aligned, um homologe Transkripte im Genom von *A. thaliana* zu finden. Die Ergebnisse der CnBr-Sepharose-Affinitätssäule für GRP8 zeigten, dass GRP8 an ein breites Spektrum verschiedener RNAs bindet. Nach der RNA-Sequenzierungsanalyse wurden einige angereicherte Transkripte ausgewählt und mittels MST getestet, um zu bestätigen, dass GRP8 sie bindet, was für alle diese Transkripte zutraf. GRP8 wurde auch an UTRs von zwei verschiedenen Transkripten gebunden. Darüber hinaus wurde ein Motiv-Alignment mit einem motivbasierten Sequenz-Alignment-Tool durchgeführt, wobei jedoch keine gemeinsamen Motive gefunden wurden. Dies spricht dafür, dass GRP8 eine Vielzahl von RNAs bindet. GRP8$^{short}$ wurde ebenfalls getestet und es wurde festgestellt, dass es keine der mRNAs aus der RNA-Sequenzierungsanalyse bindet. Da GRP8$^{short}$ nur die RRM enthält, wurde gefolgert, dass die glycinreiche Domäne eine Schlüsselrolle bei der RNA-Bindung spielt. GRP8 wurde als IDP mit einer glycinreichen Domäne als Region geringer Komplexität beschrieben, die [G/S]Y[G/S] und RGG-Motive enthält. Bei der Proteinreinigung von GRP8 wurden drei Hauptmerkmale festgestellt, die entweder zur Aggregation oder zum Abbau des Proteins führten: Dauer der Reinigung, Proteinkonzentration und Umgebungstemperatur. Diese schienen mit der wenig komplexen Region von GRP8 und dem Vorhandensein einer PrLD in der glycinreichen Domäne zusammenzuhängen. Die Reversibilität der Aggregation war ein Hinweis auf einen flüssig-flüssigen Phasentrennungszustand. Das PrLD-Verhalten von GRP8 wurde mit dem Thioflavin T (ThT) und dem Flüssig-Flüssig-Phasentrennungs-Assay (LLPS) getestet, wobei die Bildung von Kondensationströpfchen beobachtet wurde. Dennoch ist die Markierung des Proteins notwendig, um die Bildung eines RNA-Granulums zu bestätigen. Die strukturelle Charakterisierung von GRP8 wurde aufgrund der IDP-Natur von GRP8 mittels SAXS durchgeführt. Zunächst wurden verschiedene Temperaturen mit GRP8-Monomer in Batch-SAXS gemessen und eine qualitative Analyse durchgeführt. Bei einer Umgebungstemperatur von 5°C eine Aggregation von GRP8 beobachtet. Darüber hinaus wurde eine SEC-SAXS durchgeführt, um die native Struktur von GRP8-Monomer und -Dimer zu verstehen. Die Ergebnisse des GRP8-Dimers im Vergleich zum GRP8-Monomer zeigten, dass das GRP8-Monomer eine stärker entfaltete und ungeordnete Form aufweist als das GRP8-Dimer. Obwohl beide Proteine ein ähnliches Verhalten in den Kratky-Diagrammen zeigten, wurde ein leichter Unterschied zwischen ihnen festgestellt. Nach der Verarbeitung der SEC-SAXS-Daten wurde mit der 3D-Modellierung des GRP8-Monomers fortgefahren. Das 3D-Modell von GRP8 wurde dargestellt und repräsentiert die Kombination verschiedener

Techniken, einschließlich experimenteller Daten, die mittels SEC-SAXS gemessen wurden, und rechnerischer Werkzeuge, die zur Erstellung dieses Hybridmodells des GRP8-Monomers verwendet wurden.

## Abbreviations

| | |
|---|---|
| 3D | Three dimensional |
| 3’UTR | Three prime untranslated region |
| 5’UTR | Five prime untranslated region |
| Å | Angstrom |
| A.C | Affinity chromatography |
| AGP11 | Arabinogalactan protein 11 |
| BBX14 | B-Box type zinc finger protein with a CTT domain containing protein |
| Bp | Base pair |
| BSA | Bovine serum albumin |
| CCs | Companion cells |
| cDNA | Complementary DNA |
| cds | coding sequence |
| DLS | Dynamic light scattering |
| dsRBD | Double stranded RNA binding domain |
| DNA | Desoxyribonucleic acid |
| FER | Receptor kinase Feronia |
| FIB | Fibrillin precursor protein |
| FPKM | Fragments per kilobase of exon per million mapped reads |
| GRP | Glycine rich protein |
| GRP7 | Glycine rich protein seven |
| GRP8 | Glycine rich protein eight |
| GRP8s | truncated GRP8 |
| I(0) | Initial intensity |
| IDPs | Intrinsically disordered proteins |
| IDR | Intrinsically disordered region |
| Kd | Dissociation constant |
| KH | K-homology domain |
| LLPS | Liquid-liquid phase separation |
| miNovel2 | Micro-RNA novel two |
| miNovel106 | Micro-RNA novel hundred six |
| miNovel149 | Micro-RNA novel hundred forty-nine |

| | |
|---|---|
| miRNA | Micro-RNA |
| miRNA164 | Micro-RNA hundred sixty-four |
| mRNA | Messenger RNA |
| MST | Microscale thermophoresis |
| MW | Molecular weight |
| nm | Nanometer |
| NMR | Nuclear magnetic resonance |
| nt | Nucleotide |
| OHP | One helix protein |
| P(r) | Distance distribution function |
| P-bodies | Processing bodies |
| PARCL | Phloem associated RNA-chaperone like protein |
| PD | Plasmodesmata |
| PG45 | Polygalacturonase 4 protein |
| PrLD | Prion like domain |
| pre-mRNA | Precursor messenger RNA |
| RBPs | RNA-binding proteins |
| RD22 | Responsive to dehydration 22 protein |
| Rg | Hydrodynamic radius |
| rRNA | Ribosomal RNA |
| RNA | Ribonucleic acid |
| RNPs | Ribonucleoprotein complexes |
| RRM | RNA recognition motif |
| RT | Room temperature |
| RuBisCO | Ribulose-1,5-bisphosphate carboxylase-oxygenase |
| SEC | Size exclusion chromatography |
| SEC-SAXS | Size exclusion chromatography coupled to small angle X-ray scattering |
| siRNA | Small-interfering RNAs |
| SAXS | Small X-ray scattering |
| SEs | Sieve elements |
| SG | Stress granule |
| s min | Minimum value of scattering |
| sRg | Hydrodynamic radius in relation to the scattering |
| ThT | Thioflavin T |

| | |
|---|---|
| TLS | tRNA like structures |
| tRNA | Transfer RNA |
| UV | Ultraviolet |
| Znf | Zinc finger |

Nucleotide abbreviates

| | |
|---|---|
| A | Adenine |
| C | Cytosine |
| G | Guanine |
| T | Thymine |
| U | Uracil |

FASTA format amino acid abbreviates

| | |
|---|---|
| A | Alanine (ALA) |
| B | Aspartate or asparagine (ASX) |
| C | Cysteine (CYS) |
| D | Aspartate (ASP) |
| E | Glutamate (GLU) |
| F | Phenylalanine (PHE) |
| G | Glycine (GLY) |
| H | Histidine (HIS) |
| I | Isoleucine (ILE) |
| K | Lysine (LYS) |
| L | Leucine (LEU) |
| M | Methionine (MET) |
| N | Asparagine (ASN) |
| P | Proline (PRO) |
| Q | Glutamine (GLN) |
| R | Arginine (ARG) |
| S | Serine (SER) |
| T | Threonine (THR) |
| V | Valine (V) |
| W | Tryptophan (TRP) |
| Y | Tyrosine (TYR) |

## Figures

## Tables

# 1. Introduction

## 1.1 Phloem-mobile transport

In plants, long-distance signaling is enabled via the phloem system is enabled by the translocation of small molecules such as metabolites and phytohormones as well as proteins and RNA molecules (Notaguchi, 2015). This system consists of companion cells (CCs), sieve elements (SEs), fibers, ray and phloem parenchyma cells (Van Bel, 2003). CC-SE forms a conductive complex connected by the plasmodesmata (PD), which forms membrane-lined, tunnel-like channels across the cell wall that interconnects the cytoplasm of adjacent cells (Notaguchi, 2015).

It has been hypothesized that the macromolecules found in the phloem are synthesized in the CCs and imported into the SEs via PDs and then transported to their target tissues (Turgeon and Wolf, 2009; Ostendorp *et al.*, 2017; Kehr and Kragler, 2018). One major class of macromolecules found within phloem sap are polypeptides, which already hundreds of them have been identified in the phloem sap of different plant species (Walz *et al.*, 2002; Giavalisco *et al.*, 2006; Lin *et al.*, 2009; Batailler *et al.*, 2012; Ostendorp *et al.*, 2017). In addition, potential nucleic acid binding proteins could be observed that might play a role in RNA transport. This was confirmed by Thieme and colleagues in 2015 by transcriptomic analysis of grafted plants, where the detected proteins are potentially translated from mobile endogenous mRNAs that are transported to distant tissues in *A. thaliana*. (Thieme *et al.*, 2015).

Certainly, it is well established that no naked RNA molecules exist in cells, but that they are complexed with RNA-binding proteins (RBPs). These can interact with RNAs forming ribonucleoprotein complexes (RNPs) (Ostendorp *et al.*, 2017). The co-occurrence of different RNAs and RBPs in phloem exudate samples suggests their involvement in conferring RNA stability and mobility. In addition, the presence of RNA molecules in the phloem with no translation nor RNase activity supports the long-distance mobility and delivery to distant tissues of these RNPs (Sasaki *et al.*, 1998; Doering-Saad *et al.*, 2002; Gaupels *et al.*, 2008; Zhang, Sun and Kragler, 2009). Mobile RNAs must meet certain criteria to qualify as signaling molecules including: the mobile RNA must change over time and in response to a stimulus, the mobile RNA molecule must be produced in source tissues, be present in the phloem, and leave SEs in target tissues, and lastly the mobile RNA must be functional after transport (Kehr, Morris and Kragler, 2022), therefore it is unlikely that all RNA molecules are mobile, suggesting selectivity of import into the phloem.

Over the years, studies have detected all major classes of endogenous RNAs have been detected in phloem exudates: small and large non-coding RNAs including silence-interfering RNAs (siRNAs), micro RNAs (miRNA), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and protein-encoding mRNAs (Ruiz-Medrano, Xoconostle-Cázares and Lucas, 1999; Xoconostle-Cázares *et al.*, 1999; Yoo *et al.*, 2004; Haywood *et al.*, 2005; Buhtz *et al.*, 2008; Thieme *et al.*, 2015; Zhang *et al.*, 2016; Kehr and Kragler, 2018). It is now known that specific mRNAs and small RNAs are mobile signals that coordinate development and responses to both abiotic and biotic stresses across the plant. Evidence has shown that motifs in the sequence, secondary structure, or base modifications trigger mRNA movement (Morris, 2018; Maizel *et al.*, 2020; Wang *et al.*, 2021).

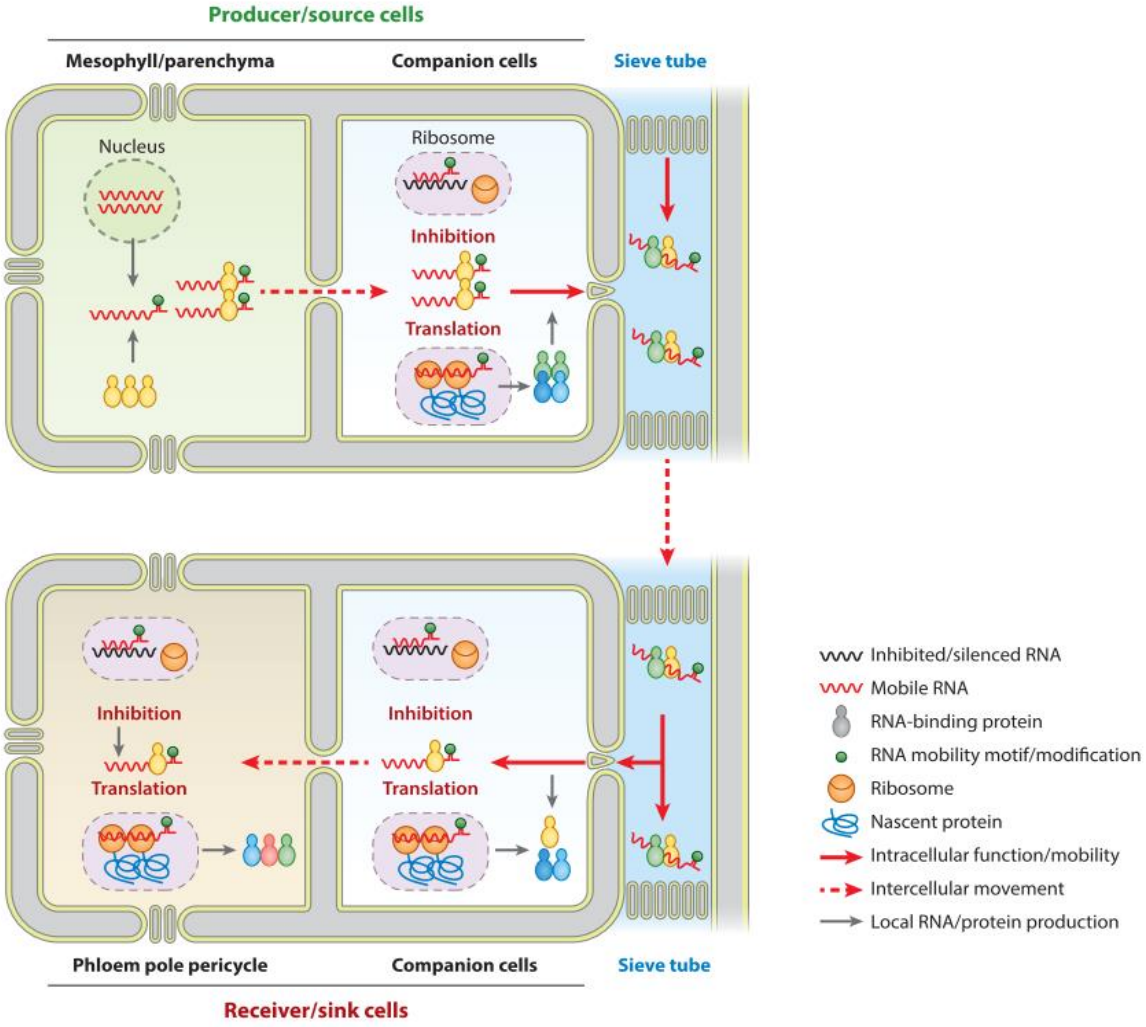



**Figure 1: Representation of the transport of RNAs and RNA-binding proteins via the phloem** (Kehr, Morris and Kragler, 2022)**.** The exported RNA-RNA-binding protein complexes are moved via the phloem bulk flow from the source cell and then unloaded in the sink cell to comply with their fate at their destination. In this example, the translation of the mRNA to protein.

RNA mobility is linked to certain characteristics including the modification in the 5-methylcytosine ($m^5C$), which is significantly enriched in mobile transcripts detected in *Arabidopsis* grafting experiments (Wang *et al.*, 2021). Besides $m^5C$, the presence of tRNA-like structures (TLS) is enriched in the population of graft-mobile transcripts in *Arabidopsis* (Zhang *et al.*, 2016; Guan *et al.*, 2017). Another possibility for regulating mRNA mobility is via interaction with specific RNA binding proteins (Ham *et al.*, 2009).

To summarize and illustrate the overall phloem transport system, figure 1 shows a schematic representation of long-distance transport of RNA and RBPs forming complexes via the phloem from sink to source cells.

Furthermore, it is important to emphasize that signaling between cells, tissues and organs is essential for multicellular organisms to coordinate their adaptive response (Notaguchi, 2015) and the further elucidation can help understand plant adaptation to their development and growth to internal and environmental changes.

## 1.2 RNA-binding proteins

RBPs are known to bind RNA and participate in forming RNPs. They have important functions in the regulation of gene expression and they play a key role in post-transcriptional processes such as mRNA transport, modulation, translation and decay (Lunde, Moore and Varani, 2007). RBPs contain domains where the RNA-protein interaction will occur. These include RNA-recognition motif (RRM), K-homology domain (KH), double-stranded RNA-binding domain (dsRBD), zinc fingers (Znf), DEAD box helicase domain, and others (Järvelin *et al.*, 2016). Most RBPs are built from versatile modular structures, with multiple repeats of few conserved domains, arranged in a variety of ways to fulfill their diverse functional requirements (Ambrosone *et al.*, 2012). Some examples are represented in figure 2 where the different RBPs are shown with their corresponding domain structure, including the glycine rich protein family represented by GR-RBP.

Probably, the best characterized domain is the RRM, which is composed of 80–90 amino acids that form a four-stranded anti-parallel β-sheet that binds to RNA (Burd and Dreyfuss, 1994; Lunde, Moore and Varani, 2007). RRMs are deeply conserved across bacteria, archaea, and eukaryotes. Their modular design also enables the rapid evolutionary adaptation of proteins to new RNA targets (Gerstberger, Hafner and Tuschl, 2014).

RBPs are enriched in intrinsically disordered regions (IDRs) that are characterized by a low content of bulky hydrophobic amino acids and a high proportion of small, polar and/or charged

amino acids (Hentze *et al.*, 2018). Several studies have shown that intrinsically disordered regions could contribute to RNA binding, moreover, the IDR can represent a multifunctional RNA-binding motif.

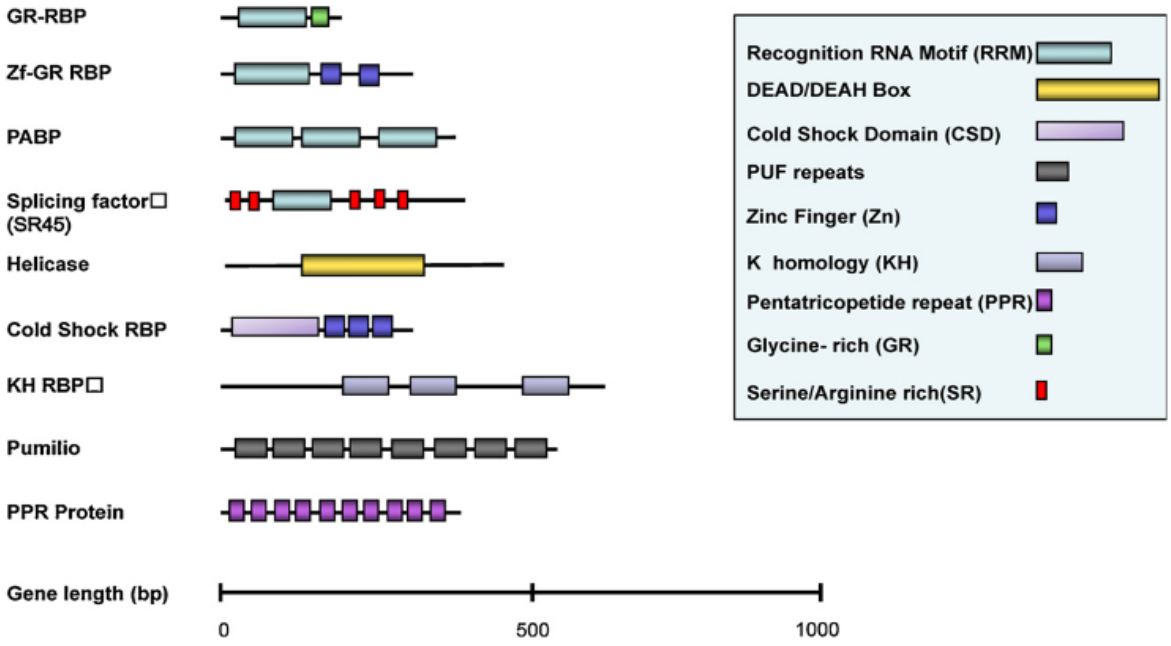



**Figure 2: Representative modular structures from RBPs classes in *A. thaliana*** (Ambrosone *et al.*, 2012)

## 1.3 Intrinsically disordered proteins

As the name states, intrinsically disordered proteins (IDPs) are characterized by the lack of defined structure either partially or completely, including IDRs. They are often enriched in charged and structure-breaking residues, almost lacking hydrophobic residues and consequently avoiding the formation of hydrophobic cores to initiate folding (Pazos *et al.*, 2013; Habchi *et al.*, 2014). In addition, protein disorder is also related to low sequence complexity (Habchi *et al.*, 2014).

Because of their not poorly-defined structure and high conformational flexibility, they possess the ability to fold into different conformations according to their binding partner in a wide range of interactions of high selectivity and low affinity (Pazos *et al.*, 2013; Habchi *et al.*, 2014). Intrinsic disorder plays a crucial role in promoting successful competition of IDP with other proteins for binding partners (Uversky, 2013) and a role for increasing molecular signaling machinery connectivity has been suggested (Pazos *et al.*, 2013).

IDPs have been demonstrated to be very sensitive to the environment (Habchi *et al.*, 2014). In plants it has been found that these proteins are involved in stress-response processes and in *A. thaliana* it has been shown that IDPs are enriched in processes related to cell cycle, signaling, DNA metabolism and RNA splicing (Pazos *et al.*, 2013). Evidently, IDPs are very important for the adaptation of plants and to their environment.

Their RNA-binding capacity can range from highly specific to nonselective and may promote protein–RNA co-folding upon their interaction with target RNAs (Järvelin *et al.*, 2016; Hentze *et al.*, 2018). Flexible regions in RBPs rich in serine and arginine [S/R] and arginine and glycine [R/G] were found to contribute to RNA-binding activities. In addition to this, there is evidence that shows certain RBPs go through phase separation when [G/S]Y[G/S] motifs are present in their structure and bound to RNA. These condensates composed of RBPs and RNAs, known as RNA granules, are formed when the concentration of a protein reaches a critical concentration that triggers condensation (Hyman, Weber and Jülicher, 2014). RNA granules have roles in RNA localization, stability, and translation (Kiledjian and Dreyfuss, 1992; Valcárcel *et al.*, 1996; Järvelin *et al.*, 2016) and they provide functional compartmentalization of biomolecules within cells (Tian, Curnutte and Trcek, 2020). Phase separation allows granules to rapidly condense and dissolve depending on the environment in which they form and enables the exchange of granule components with the granule environment (Hyman, Weber and Jülicher, 2014). These RNA granules are divided in different classes including stress granules (SG), processing bodies (p-bodies) and germ granules (only present in germ cells). SGs have been described as a triage for mRNA during cellular stress where they either store translationally silent mRNA or transfer mRNA transcripts to processing bodies (p-bodies) where they will be degraded (Kedersha *et al.*, 2000; Ray Mc Dermott *et al.*, 2002; Campos-Melo *et al.*, 2021). SGs tend to accumulate translationally repressed mRNAs and disassemble quickly upon stress removal (Wheeler *et al.*, 2016). In plants, the regulation of mRNA dynamics is essential for growth, development, and stress responses. Numerous environmental stresses trigger plant SG assembly, including high salt, heat, darkness, hypoxia, the inhibition of oxidative phosphorylation, and viral infection. The SG disassembly releases mRNA and proteins to the cytoplasm to reactivate translation and reassume cell growth and development (Jang, Jang and Wu, 2020; Maruri-López *et al.*, 2021).

Furthermore, the [G/S]Y[G/S] motif is present in a broad spectrum of RBPs and is sufficient and necessary to cause aggregation *in vitro* and *in vivo* (Han *et al.*, 2012; Kato *et al.*, 2012; Kinsella and Monk, 2012; Järvelin *et al.*, 2016). In humans, several RBPs which are linked to

diseases (e.g., neurodegenerative, cancer and diabetes), have been described to include this motif and engaging in phase separation and RNA sequestration into RNA granules (Takanashi and Yamaguchi, 2014; Diering, Maxson & Mitchell and Freeman, 2018). These all share the glycine-rich domain that exhibit low protein complexity (Wolozin, 2012) and computational analysis indicates that this motif is common among prion proteins, hence this glycine-rich domain is referred to as a prion domain (Han *et al.*, 2012).

## 1.4 Glycine-rich protein family

This protein family is characterized by the presence of a glycine-rich domain in glycine ([G]n-X) repeats. They are simple in structure, but the arrangement of their different domains allows them to be clustered them in different classes (Mangeon, Junqueira and Sachetto-Martins, no date). Figure 3 shows a schematic representation of the different classes of GRPs in plants. Because of the glycine-rich repeats, which are highly flexible, they have the ability to adapt necessary structures for their correct conformation and/or interactions (Sachetto-Martins, Franco and De Oliveira, 2000).

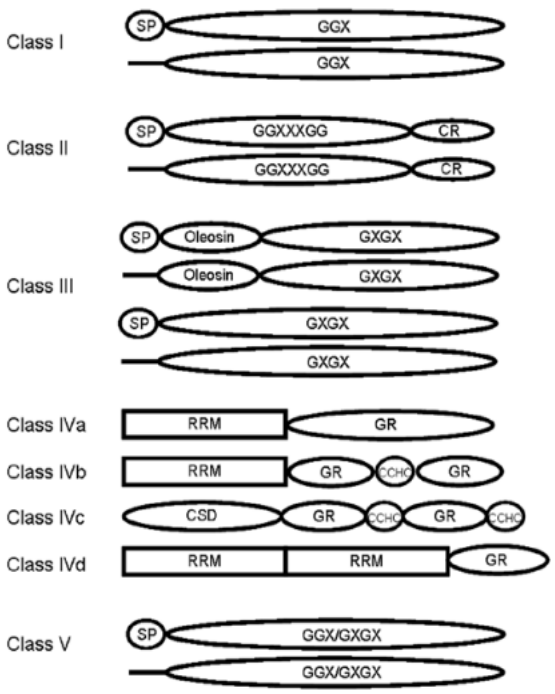



**Figure 3: Schematic representation of glycine-rich family in plants.** (Mangeon, Junqueira and Sachetto-Martins, no date)**.** SP, signal peptide. CR, cysteine-rich domain. Oleosin, oleosin-conserved domain. CCHC, zinc-finger. CSD, cold-shock domain. G and X are glycine and any amino acid, together they represent glycine-rich repeats.

Several biological processes have GRPs as players. In plants GRPs are involved in modulation of gene expression, auxin regulation, abscisic acid (ABA) modulation, water stress, circadian-rhythm, light response, cold response, wounding, organ and tissue expression pattern, flowering and post-transcriptional modulation (Mangeon, Junqueira and Sachetto-Martins, no date; Sachetto-Martins, Franco and De Oliveira, 2000).

Only Class IVa will be described here, as GRPs from this class are RBPs and are of interest of this research. They are involved in alternative splicing (AS), regulation of transcription, stomatal movement, seed, pollen and stamen development and their accumulation is regulated by the circadian clock (Czolpinska and Rurek, 2018). This last one is particularly interesting, hence the circadian clock is described as molecular networks including (interlocked) transcriptional-translational feedback loops (Schmal, Reimann and Staiger, 2013). The transcript and protein oscillation are linked; the transcript levels decline when the protein reaches a maximum (Sachetto-Martins, Franco and De Oliveira, 2000).

### 1.4.1 Glycine-rich protein 8 and 7

Although this research is focused on GRP8, it is worth noting GRP7, as they both are interacting with each other and most importantly, they are similar in structure. In figure 4 is shown a global alignment done *in silico* with the *EMBOSS Needle Pairwise Sequence Alignment* (EMBL).

```
RBG8_ARATH         1 --MSEVEYRCFVGGLAWATNDEDLQRTFSQFGDVIDSKIINDRESGRSRG     48
                       ..:|||||||||||||||:|..|:..|:|:|||||||||||||||:|||||
RBG7_ARATH         1 MASGDVEYRCFVGGLAWATDDRALETAFAQYGDVIDSKIINDRETGRSRG     50

RBG8_ARATH        49 FGFVTFKDEKAMRDAIEEMNGKELDGRVITVNEAQSRGSGGGGGGRGGSG     98
                     ||||||||||||:||||.|||::||||.||||||||||||||||.|||.|
RBG7_ARATH        51 FGFVTFKDEKAMKDAIEGMNGQDLDGRSITVNEAQSRGSGGGGGHRGGGG    100

RBG8_ARATH        99 GGYRSGGGGGYSGGGGGGYSGG-----GGGGYERRSGGYGSGGGGGGRGY    143
                     |||||||||||| ||||.|.||     |||||....|||.|.||||| .|
RBG7_ARATH       101 GGYRSGGGGGYS-GGGGSYGGGGGRREGGGGYSGGGGGYSSRGGGGG-SY    148

RBG8_ARATH       144 GGGGRREGGGYGGGDGGSYG--GGGGGW    169
                     |||.|..|||||||:||.||  ||||||
RBG7_ARATH       149 GGGRREGGGGYGGGEGGGYGGSGGGGGW    176
```

**Figure 4: Pairwise global alignment of GRP8 and 7.** RBG8_ARATH, GRP8 in *A. thaliana*. RBG7_ARATH, GRP7 in *A. thaliana*. Lines represent the same amino acids. Dots represent different amino acids. Hyphens represent gaps in the sequence.

GRP8 and 7 are circadian clock regulated proteins. They both undergo high amplitude oscillations with a peak at the end of the day and are involved in transcriptional-translational feedback loops (Schmal, Reimann and Staiger, 2013). They interact with each other negatively, auto-regulate and cross-regulate their own expressions by causing AS of their pre-mRNAs and then later decay (Schöning *et al.*, 2008).

Functionally, GRP8 has been found to be involved in flowering (Steffen, Elgner and Staiger, 2019), cold stress (Clifford D. Carpenter, Joel A. Kreps, 1994; Horvath and Olson, 1998) and resistance to pathogen attacks (Fu *et al.*, 2007; Reumann *et al.*, 2007). In addition, because of its structural similarity with GRP7, GPR8 could also be involved in other physiological processes, including roles such as salt stress, oxidative stress and water deprivation (Kyung, Yeon and Kang, 2005; Schmal, Reimann and Staiger, 2013).

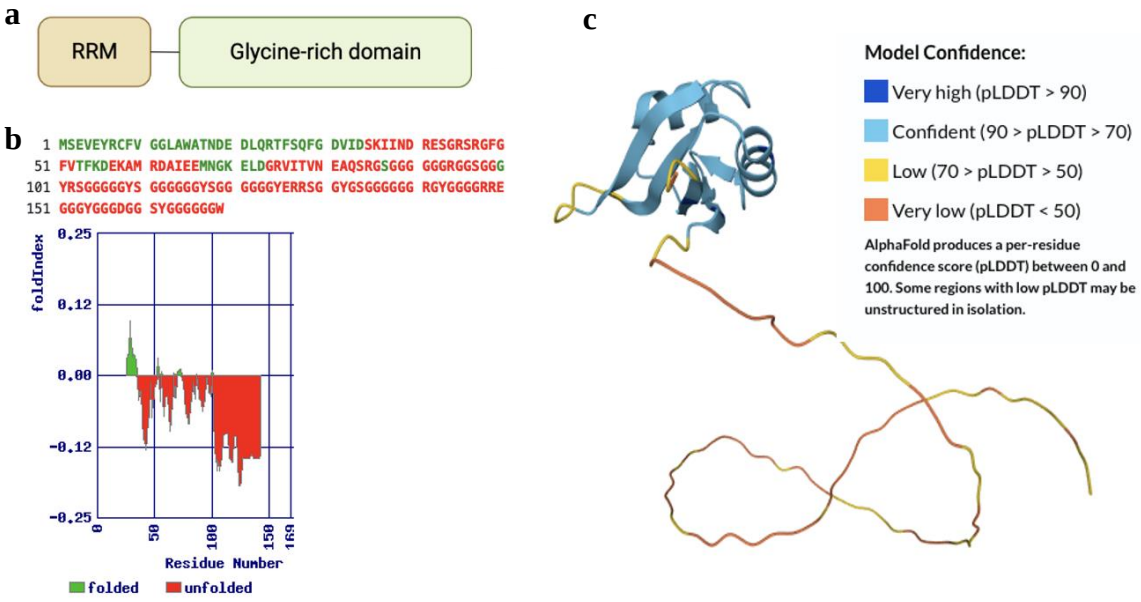



**Figure 5: Schematic representation and structure prediction of GRP8.** a) Schematic representation of the domains in GRP8. RRM is shorter because of the globular shape it folds, and the glycine-rich domain shows a more elongated shape because of its disorder. Scheme created by *BioRender*. b) Structure disorder prediction by GRP8 amino acid sequence by *FoldIndex*. Ordered/folder regions in green, disordered/unfolded regions in red. Number of disordered regions: 5, longest disordered region: 69 residues. Number of disordered residues: 122. c) 3D-structure prediction by *Alphafold*. Model confidence represented in the figure.

Structurally, as already, GRP8 belongs to the class IVa. Its structure consists of an RRM at the N-terminus and a glycine-rich domain at the C-terminus (figure 5a). It is well known that the glycine-rich domain is highly flexible and disordered. Disorder prediction (figure 5b) was calculated *in silico* with bioinformatical tool *FoldIndex* (Prilusky *et al.*, 2005) and it is visible

that GRP8 is mostly unfolded behind the RRM. Figure 6c shows an *Alphafold* 3D-model representation of GRP8. While the RRM domain is modeled with high confidence, the glycine-rich domain confidence remains low in the disordered regions.

Moreover, mRNA appears to be transported through the phloem and acts as a signaling molecule facilitated by RNA-binding proteins (Heintzen *et al.*, 1997; Xoconostle-Cázares *et al.*, 1999; Yoo *et al.*, 2004; Deeken *et al.*, 2008; Ham *et al.*, 2009; Lin *et al.*, 2009). The presence of GRP8 as protein has been detected in phloem exudates of *A. thaliana* (Batailler *et al.*, 2012). This data and the functional characteristics already described, support that GPR8 could have a role in signaling when complexed with RNA.

### 1.5.1 Characterization techniques

In this section, only two characterization techniques are described because they are less common than other methods. These techniques were chosen for the structural and functional characterization of GRP8 because they provide the information needed to answer some of the objectives that are mentioned in chapter 1.7.

### 1.5.2 Microscale thermophoresis

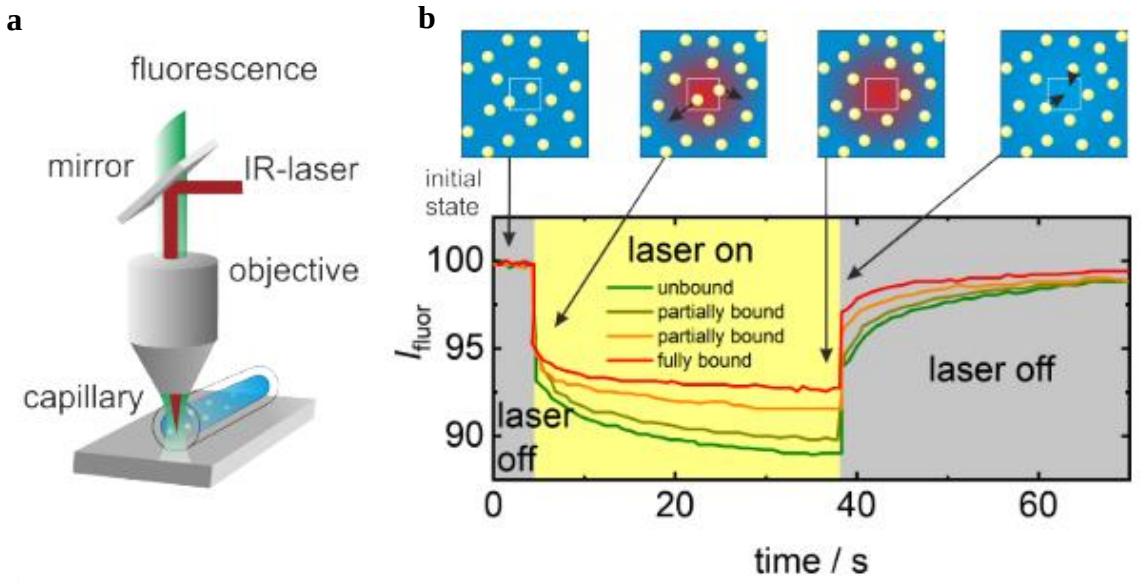



**Figure 6: Microscale thermophoresis method representation** (Lee and Wiegand, 2020)**.** a) MST setup b) Temporal dependence of the fluorescence intensity. When the heating laser is switched on and off, the fluorescent-labeled particles accumulate in the cold region. Temporal fluorescence intensity plots with the x-axis representing time [s] and the y-axis representing intensity [FU].

Microscale thermophoresis (MST) has emerged as a revolutionary method utilizing the thermophoresis effect to measure the directed movement of fluorescent molecules through a microscopic temperature gradient in microliter volumes (Jerabek-Willemsen *et al.*, 2014).
The temperature difference in solution leads to a thermophoresis depletion which under constant buffer conditions, thermophoresis probes the size, charge, and solvation entropy of the molecules (Seidel *et al.*, 2013). Figure 6 shows a schematic representation of how the thermophoresis is measured. In figure 6a, the device is represented, which consists of an epifluorescence microscope with an additional infrared laser. Figure 6b explains how the laser is only used to generate the temperature gradient into the capillary which is absorbed by the water inside of the containing solution of fluorescent molecules. This fluorescent light is then observed as a function of time through the same objective as the infrared laser. The fluorescent-labeled particles move and leave the heated area and the fluorescent intensity decays towards a plateau value, which increases when the fluorescent-labeled particles bind to a small ligand molecule. This plateau value with the ligand concentration determines the equilibrium binding constant of the reaction (Lee and Wiegand, 2020). This technique was used for determining the dissociation constant (Kd) through binding assays of GRP8 and RNAs of interest.

### 1.6.2 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) has become a powerful method for structural characterization of biomolecular disordered systems. This technique provides information about the overall conformation and structural changes of biomolecules in solution. High-throughput SAXS studies are possible because of advances in synchrotrons with data collection times within seconds, robotic sample chargers and automated data collection and analysis pipelines (Hura *et al.*, 2009).

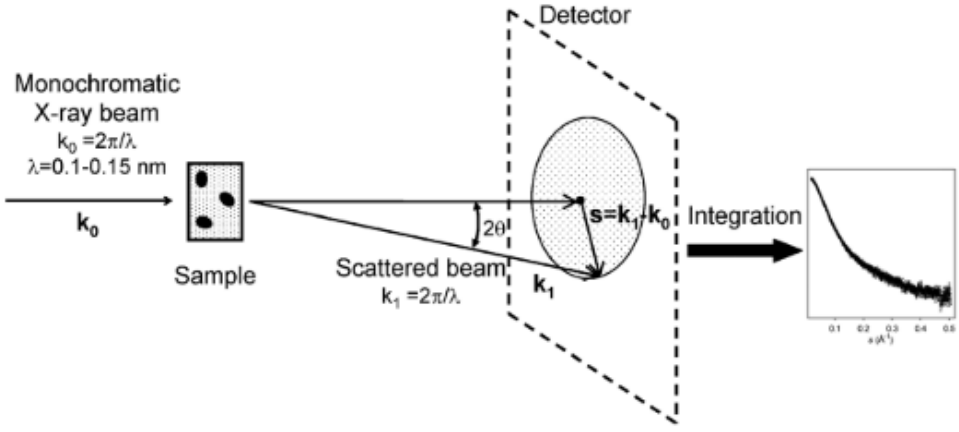



**Figure 7: Basic scheme of a SAXS experiment** (Bernadó and Svergun, 2012)**.**

Recent developments have allowed quantitative description of the conformational ensembles for flexible macromolecules in solution (Mylonas and Petoukhov, 2007). Although small-angle scattering provides information on the large-scale features, it is often possible to model proteins based on high-resolution structures of domains from computational tools or databases. Figure 7 shows a schematic representation of a basic SAXS experiment. A solution of macromolecules is placed in a cuvette or in a capillary and then exposed by a collimated monochromatic X-ray beam ($k_0$) and the intensity of the scattered beam is measured in function of the scattering angle ($2\theta$). SAXS patterns are usually recorded by a two-dimensional detector which provides a more accurate signal detection after radial averaging. The solution buffer must be also measured alone in the same manner to obtain the scattering. This is then subtracted from the macromolecular solution. The difference pattern arises from the dissolved particles and proves information about the structure (Bernadó and Svergun, 2012). SAXS data is then processed for further analysis.

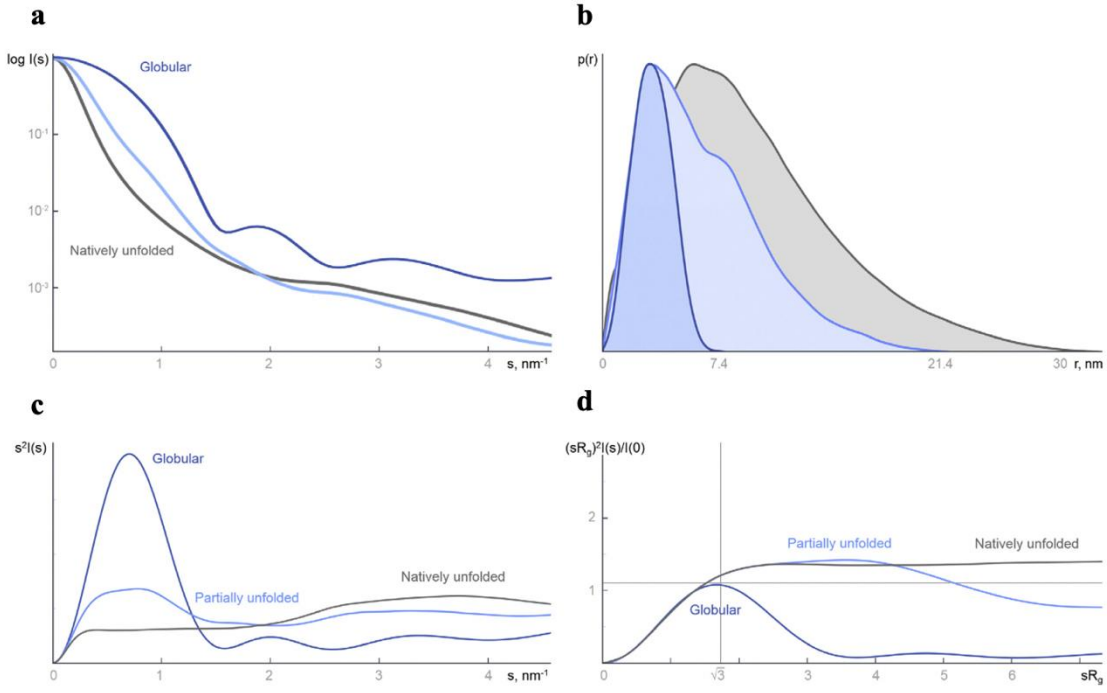



**Figure 8: Data simulated from three 60 kDa proteins** (Kikhney and Svergun, 2015) Globular protein in dark blue. 50% Unfolded in light blue. Fully disordered in gray. a) Logarithmic plot of the scattering intensity I(s) (in arbitrary units) versus s (in inverse nanometers). b) Distance distribution functions p(r) (in arbitrary units) vs. r (in nanometers). c) Kratky plot s2I(s) vs. s. (d) Normalized (or ‘‘dimensionless”) Kratky plot $(sRg)^2I(s)/I(0)$ versus sRg.

For a more comprehensive understanding of the processed SAXS data, figures 8 and 10 represent an example found in literature of the presented plots in results chapter 3.3.5. Each

plot contains relevant information that transfers either sample quality or structural characteristics: Guinier fit, and Guinier plot (Guinier analysis) reveal if the data quality is good. There are certain irregularities that can be seen when the data quality has been compromised (aggregation, radiation damage, interparticle interactions or buffer mismatch).

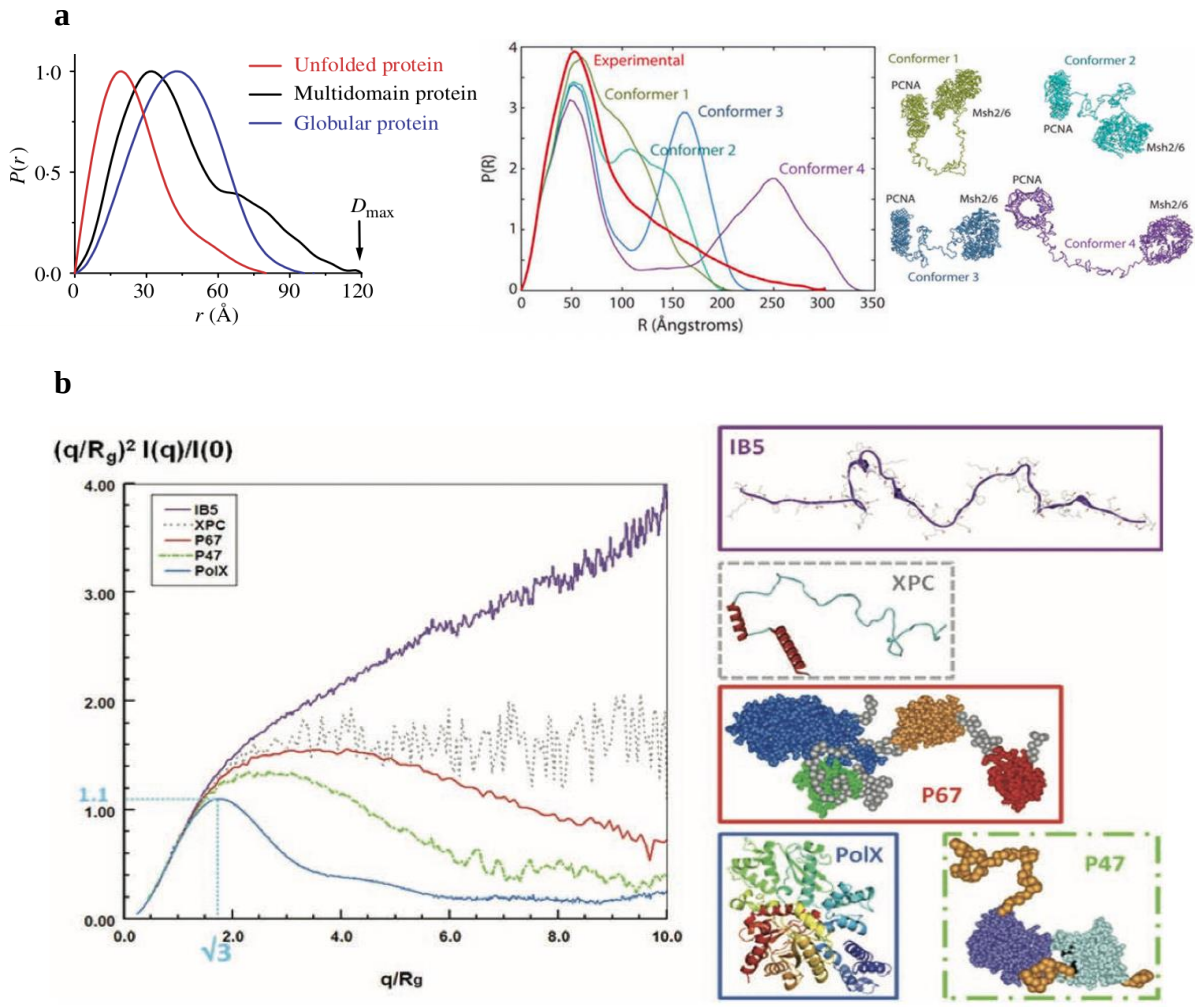



**Figure 9: Comparison of experimental and calculated distance distribution functions and dimensionless Kratky plot of flexible multidomain protein** (Putnam *et al.*, 2007; Receveur-Brechot and Durand, 2012)**.** a) P(r) functions calculated for four randomly generated models of Msh2-Msh6 linked to PCNA via random peptides with different interdomain distances reveals that no single conformer can account for the observed P(r) curve of the Msh2-Msh6-PCNA complex. The red curve with the long tail corresponds to the experimental P(r) curve of Msh2-Msh6-PCNA complex. b) Normalized Kratky plots. The scattering pattern of globular proteins in a normalized Kratky plot exhibits a bell-shaped profile with a clear maximum value of 1.104 for qRg=√3. Bell-shaped profile of a globular protein (PolX in blue line). Curve of a protein consisting of several domains tethered by linkers with rather compact conformations (p47phox in dotted green line). Extended conformations (p67phox in continuous red line). Curve of a fully disordered protein with very short elements of secondary structure (XPC in dotted gray line). Curve of a fully disordered and extended protein with short segments of polyproline repeats (salivary protein IB5 in continuous purple line). In both graphs: q=s.

Furthermore, the Guinier analysis is important for the calculation of the radius of gyration (Rg) from the scattering intensity (Zheng and Best, 2018). The Guinier analysis must range (sRg limits) for sRg min < 0.65 and sRg max ~1.3 - 1.0 (closer to 1.3 if protein is globular; closer to 1 if protein is elongated) (Svergun, 1987; Putnam *et al.*, 2007; Jacques and Trewhella, 2010). Kratky plots and pair distribution indicate flexibility, conformational state, and shape of the protein.

In addition to the flexibility and folding of the protein, distance distribution and dimensionless Kratky plots provide some extra information about the protein conformation. Figure 9a shows some examples of different experimental data found in literature where it is possible to differentiate if a protein is globular, unfolded or includes other domains (multidomain). Figure 9b shows an example of a dimensionless Kratky plot with different types of proteins, from a more globular one (PolX) to a highly disordered one (IB5). Evidently, there are several patterns to look out for when analyzing SAXS data which indicate crucial information for the characterization of GRP8.

## 1.7 Objectives

In this research, GRP8 was studied *in vitro* in order to characterize its structure and function. GPR8 is described as an RNA-binding protein, therefore it was important to determine which RNAs are binding, how well are they binding to GRP8, what kind of functions they showed and if they presented any common motifs between them. Because GRP8 includes two domains, it was important to understand their function and structure, as well as how they interact with the RNAs. GRP8 contains a natively disordered structure which is known as the glycine-rich domain. It was interesting to understand if this disordered region played a role in its functionality. Because GRP8 included regions of low complexity with specific motifs that are reported to form condensation droplets in similar proteins, it was interesting to elucidate if these motifs were predicted to be prion-like domains. These domains are known to form liquid-liquid phase separation, moreover, it was interesting to test GRP8 for the detection of condensation droplets. For the structural analysis, different approaches were performed to evaluate the structural parameters of GRP8 (including the intrinsic disorder) and determine the structure of GRP8. Lastly, the reconstruction of a hybrid 3D model of GRP8 was proposed.

## Materials

| Devices | |
|---|---|
| -20°C freezer | Liebherr, Biberach (D) |
| -80°C freezer | GFL, Burgwedel (D) |
| 4°C refrigerator | Liebherr, Biberach (D) |
| Äkta prime plus | GE Healthcare, Uppsala (S) |
| Äkta start | GE Healthcare, Uppsala (S) |
| Avegene gel documentation SLite 140S | Pacific Image Electronics, New Taipei City (TWN) |
| Avanti JXN-30 centrifuge | Beckman Coulter, Krefeld (D) |
| Balance | ABJ Kern & Sohn GmbH, Balingen (D) |
| Bioanalyzer 2100 | Agilent, Waldbronn (D) |
| BioPhotometer | Eppendorf, Hamburg (D) |
| Branson Sonifier 250 Branson Ultrasonics | Eemnes (NL) |
| Centrifuge 2K15 | Sigma, Göttingen (D) |
| Centrifuge 5417R | Eppendorf, Hamburg (D) |
| Centrifuge 5424/5424R | Eppendorf, Hamburg (D) |
| Chemidoc Touch Gel/Blot | BioRad, München (D) |
| DLS SpectroSize 300 | Xtal Concepts, Hamburg (D) |
| DLS SpectroLight 600 | Xtal Concepts, Hamburg (D) |
| DLS SpectroLight 610 | Xtal Concepts, Hamburg (D) |
| Electrophoresis Power Supply EPS 301 | Amersham Pharmacia, Uppsala (S) |
| Frac30 | GE Healthcare, Uppsala (S) |
| Gel electrophoresis separation Owl | Owl Separation Systems Inc., Portsmouth (USA) |
| Heating block OriBlock | OB-1 Bibby Scientific, Stone (UK) |
| Heraeus incubator | Hanau (D) |
| IKAMAG RCT magnetic stirrer | IKA, Staufen (D) |
| Magnetic stirrer | RSM-10HS Phoenix Instruments, Garbsen (D) |
| Mini-PROTEAN III | BioRad, München (D) |
| MiniSpin® centrifuge | Eppendorf, Hamburg (D) |
| Monolith NT.115TM | NanoTemper, München (D) |
| NanoDrop onec UV/Vis Photometer | NanoDrop products, Wilmington (USA) |

| | |
|---|---|
| Olympus MVX10 Macroscope | Evident, Hamburg (D) |
| Oryx8 protein crystallization robot | Douglas Instruments, Berkshire (UK) |
| PCR Cycler T3000 | Biometra, Göttingen (D) |
| peqTWIST Vortexer | VWR, Darmstadt (D) |
| Peristaltic pump 2232 Microperplex S | Pharmacia LKB, Uppsala (S) |
| PETRA III Beamline P12 | DESY/EMBL, Hamburg (D) |
| pH-Meter | Mettler-Toledo, Giessen (D) |
| Pilatus 2M pixel X-ray Detector | Dectris, Baden-Daettwil (CH) |
| Rotina 380R centrifuge | Hettich, Tuttlingen (D) |
| Shaker | AG, Bottmingen (CH) |
| Stratalinker 2400 UV crosslinker | Agilent, Waldbronn (D) |
| Table centrifuge | Biozym, Hessisch Oldendorf (D) |
| TECAN SPARK® | TECAN, Männedorf (CH) |
| ThermoShaker TS1 | Biometra, Göttingen (D) |
| Ultraflex III MALDI-TOF-TOF (MS) | Bruker Daltonics GmbH, Bremen (D) |
| UltrafleXtreme MALDI-TOF-TOF (MS) | Bruker Daltonics GmbH, Bremen (D) |
| Vacuum pump | Aeromat KNF, Freiburg (D) |
| Water bath | Julabo, Seelbach (D) |

Chemicals

Unless otherwise noted, all chemicals from AppliChem (Darmstadt, D), Roth (Karlsruhe, D), Sigma-Aldrich (Taufkirchen), Serva (Heidelberg, D) and Merck (Darmstadt, D).

Consumables

The plastic consumables such as pipette tips and petri dishes are from Sarstedt (Nümbrecht, D) and reaction tubes from Eppendorf (Hamburg, D).

| | |
|---|---|
| 5-methyl-CTP | Jena Bioscience, Jena (D) |
| 8-azido-ATP | Jena Bioscience, Jena (D) |
| Agilent RNA 6000 Nano kit | Agilent, Waldbronn (D) |
| Cyanine 5-UTP | Enzo Life Science, Lörrach, (D) |
| CnBr activated Sepharose | GE Healthcare, Uppsala (S) |
| Complete Protease Inhibitor | Roche, Mannheim (D) |
| CutSmart® | New England Biolabs, Frankfurt a. M. (D) |

| | |
|---|---|
| FloppyChoppy Kit | Jena Bioscience, Jena (D) |
| GelRed DNA Stain | Biotium, Hayward (USA) |
| GeneRuler™ 1 kb plus DNA ladder | Fermentas, St. Leon-Rot (D) |
| JBScreen JCSG++HTS | Jena Bioscience, Jena (D) |
| HiLoad Superdex 75 pg Gel Filtration | GE Healthcare, Uppsala (S) |
| HisTrap™ High Performance | GE Healthcare, Uppsala (S) |
| Low-binding Tubes 0.5 ml | Sarstedt, Nümbrecht (D) |
| Microtitration plate TECAN | TECAN, Männedorf (CH) |
| Monolith NT.115 Standard capillaries | NanoTemper, München (D) |
| Monolith NT.115 Premium capillaries | NanoTemper, München (D) |
| NucleoSpin gel and PCR Clean-up kit | Macherey-Nagel, Düren (D) |
| NucleoSpin Plasmid EasyPure (250) | Macherey-Nagel, Düren (D) |
| Ni-NTA Agarose | Qiagen, Hilden (D) |
| PACT premier™ Eco Screen HT 96 | Molecular Dimensions, Eching (D) |
| PageRuler prestained 10 to 180 kDa | Fermentas, St. Leon-Rot (D) |
| Parafilm | M Brand, Wertheim (D) |
| PCT pre-crystallization test | Hampton Research, Schwerte (D) |
| PD10 empty columns | GE Healthcare, Uppsala (S) |
| QS Quartz Cuvette10 mm | Hellma GmbH, Müllheim (D) |
| RNA Clean & Concentrator-25 | Zymo Research, Irvine (USA) |
| RevertAid cDNA synthesis | Thermo Scientific, Darmstadt (D) |
| Ribolock | Thermo Scientific, Darmstadt (D) |
| SpectraPor Dialysis memb. 3500MWCO | Carl Roth, Karlsruhe (D) |
| TRIzol® reagent | Invitrogen, Darmstadt (D) |
| TRIzol® Ls reagent | Invitrogen, Darmstadt (D) |
| Vivaspin Concentrator 3000MWCO | Sartorius, Göttingen (D) |
| Vivaspin Concentrator 5000MWCO | Sartorius, Göttingen (D) |
| Vivaspin Concentrator 10000MWCO | Sartorius, Göttingen (D) |

Enzymes

| | |
|---|---|
| Antarctic phosphatase (FastAP) | New England Biolabs, Frankfurt a. M. (D) |
| HindIII restriction enzyme | New England Biolabs, Frankfurt a. M. (D) |
| DNase I | Applichem, Darmstadt (D) |
| dNTP Mix | Fermentas, St. Leon-Rot (D) |

| | |
|---|---|
| Inorganic Pyrophosphatase (0.1u/µl) | New England Biolabs, Frankfurt a. M. (D) |
| Lysozyme | Applichem, Darmstadt (D) |
| NdeI restriction enzyme | New England Biolabs, Frankfurt a. M. (D) |
| Phusion High-Fidelity DNA Polymerase | Fermentas, St. Leon-Rot (D) |
| SmaI restriction enzyme | New England Biolabs, Frankfurt a. M. (D) |
| T4 DNA Ligase | Fermentas, St. Leon-Rot (D) |
| T4 DNA Polymerase | New England Biolabs, Frankfurt a. M. (D) |
| T7 RNA Polymerase | *In House* |
| Taq DNA Polymerase | Fermentas, St. Leon-Rot (D) |
| Tobacco Etch Virus (TEV) Protease | *In House* |
| Trypsin | Promega, Mannheim (D) |

| Vectors | |
|---|---|
| pET-28a | Merck Millipore, Darmstadt (D) |
| pUC57 | Thermo Scientific, Darmstadt (D) |

| Bacterial Strains | |
|---|---|
| *Escherichia coli* BL21 Gold (DE3) | Agilent, Waldbronn (D) |
| *Escherichia coli* CodonPlus (DE3) RIPL | Agilent, Waldbronn (D) |
| *Escherichia coli* XL10 Gold | Agilent, Waldbronn (D) |

| Plant Material | |
|---|---|
| *Arabidopsis thaliana* Columbiana 0 | *In House* |
| *Brassica napus* cv. Drakkar | CBGP, Madrid (E) |

| Software and databases | |
|---|---|
| Alphafold (2.0.0) | https://alphafold.ebi.ac.uk |
| ATSAS (3.0.5) | EMBL, Hamburg (D) |
| Bioanalyzer | Agilent, Santa Clara (USA) |
| BioRender | https://biorender.com |
| CHROMIXS | EMBL, Hamburg (D) |
| CRYSOL (3.0) | EMBL, Hamburg (D) |
| DAMMIF (1.1.2) | EMBL, Hamburg (D) |
| EMBL-EBI SAS | http://www.ebi.ac.uk/thornton-srv/databases/sas/ |

| | |
|---|---|
| EnsemblPlants | http://plants.ensembl.org/ |
| EOM (2.1) | EMBL, Hamburg (D) |
| ExPASy | http://www.expasy.org/ |
| flexControl | Bruker Daltonik, Bremen (D) |
| GNOM (4.6) | EMBL, Hamburg (D) |
| MASCOT server | http://www.matrixscience.com/ |
| MEME-Suite | https://meme-suite.org/ |
| miRBase | https://www.mirbase.org/ |
| mMass (5.5.0) | http://www.mmass.org/ |
| MO. Affinity analysis | NanoTemper, München (D) |
| MO.Control | NanoTemper, München (D) |
| OriginLab (7.1.2) | OriginLab, Northampton (USA) |
| PRIMUS (3.1) | EMBL, Hamburg (D) |
| SnapGene (5.2.3) | GLS biotech LLC, San Diego (USA) |
| RCSB PDB | http://www.rcsb.org/ |
| SASpy (4.1.0) | EMBL, Hamburg (D) |
| SREFLEX | EMBL, Hamburg (D) |
| Unicorn Control | GE Healthcare, Uppsala (S) |
| UniProt protein databank | http://www.uniprot.org/ |
| i-Tasser | https://zhanggroup.org/I-TASSER/about.html |
| TECAN | TECAN, Männedorf (CH) |
| PyMOL (2.5) | Schrödinger, New York (USA) |
| DLS SpectroSize 300 | Xtal Concepts, Hamburg (D) |
| DLS SpectroLight 600 | Xtal Concepts, Hamburg (D) |

Frequently used solutions and buffers

| | |
|---|---|
| LB medium | 10g/L tryptone, 5g/L yeast extract, 10g/L NaCl |
| LB + agar | 10g/L tryptone, 5g/L yeast extract, 10g/L NaCl, 12g/L agar |
| TB medium | 12g/L tryptone, 24g/L yeast extract, 5g/L glycerol |
| 10x TB salts | 23.1g/L $KH_2PO_4$, 125.4g/L $K_2HPO_4$ |
| ZY | 10g/L tryptone, 5g/L yeast extract |
| 20x NPS | 0.5 M (NH4)2SO4, 1 M KH2PO4, 1 M Na2HPO4 |
| 50x 5052 | 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose |

| | |
|---|---|
| 6x SDS loading buffer | 375 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 100 mM DTT, 0.01% bromophenol blue |
| 10x SDS running buffer | 30 g/L Tris, 144 g/L glycine, 10 g/L SDS |
| Colloidal Coomassie | 0.02% (w/v) CBB G-250, 5% (w/v) aluminum sulfate (14-18)-hydrate, 10% (v/v) ethanol, 2% (v/v) ortho-phosphoric acid (85%) |
| Colloidal Coomassie distain | 10% (v/v) ethanol, 2% (v/v) ortho-phosphoric acid (85%) |
| 50x TAE buffer | 2 M Tris, 5.71% acetic acid, 50 mM EDTA |
| 6x DNA loading die | 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF |
| MST buffer | 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mg/ml BSA, 0.1% Tween-20, 1 mM DTT |

Sequencing, DNA and RNA oligonucleotides

All oligonucleotides were ordered from Eurofins (Ebersberg, D). Plasmid/DNA sequencing was performed by Eurofins (Ebersberg, D) or Microsynth (Göttingen, D). RNA sequencing was performed by Novogene (Cambridge, UK)

## Methods

### Plant Material

*A. thaliana Col. 0* was grown in a growth chamber at 24°C with a 10-hour light period with 55% relative air humidity.

*B. napus Drakkar* was grown in a greenhouse at 18°C with a 16-hour light period with 55% relative air humidity.

### Production of *E. coli* competent cells

The production of the chemically competent cells was performed following the method from Inoue and colleagues (Inoue, Nojima and Okayama, 1990), in a modified form. Initially, 200 ml LB medium were used with the selection antibiotic, 34 µl/ml chloramphenicol or 25 µg/ml spectinomycin. Inoculation with bacterial cells from bacterial colonies grown in LB was done. The culture was incubated overnight at 24°C shaking at 170 rpm until the optical density ($OD_{600nm}$) of 0.6 - 0.8 was reached. The culture was divided and placed into four Falcon tubes

of 50 ml each and kept on ice for 10 minutes. Afterwards, it was centrifuged at 3000 x g for 10 minutes at 4°C. The pellets were collected and resuspended with 16 ml of TB solution (10 mM PIPES, pH 6.7, 250 mM KCl, 15mM $CaCl_2$, 55 mM $MnCl_2$), kept 10 minutes on ice and then centrifuged at 3000 x g for 10 minutes at 4°C. The pellets were resuspended in 4 ml TB solution, treated with 280 µl of dimethyl sulfoxide (DMSO) and kept on ice for 20 minutes. 200 µl aliquots were frozen in liquid nitrogen and later stored at -80°C.

**cDNA synthesis for GRP8 and RNA synthesis**

Total plant RNA was isolated with *Plant total RNA isolation kit* (Avegene) in accordance with the manufacturer's instructions with leaf material. cDNA synthesis was performed.

**Table 1: cDNA synthesis from *A. thaliana* RNA**

| **20 µl reaction** | **Program** |
|---|---|
| 11µl RNA from *A. thaliana* | 42°C x 1 hour |
| 1 µl $(oligo\ dT)_{18}$ Primer (0.5 µg/µl) | 70°C x 5 min |
| 2 µl 5x reaction buffer | |
| 1 µl Ribolock RNAase inhibitor (40 U/µl) | |
| 2 µl 10mM dNTPs | |
| 1 µl reverse transcriptase (20 U/µl) | |
| 2 µl $ddH_2O$ | |

**Cloning of GRP8 in pet28a vector**

From the cDNA synthesis of GRP8, gene amplification was done by PCR with Phusion polymerase and GC buffer. Primer and constructs design were planned with *SnapGene* software.

**Table 2: Primers for GRP8 construct**

| **Name** | **sequence** |
|---|---|
| At4g39260_fw | TTAACATATGGAAAACCTGTATTTTCAGGGCATGTCTGAAGT TGAGTACCGGTG |
| At4g39260_rev | AATTAAGCTTCTCTATCTTTGATTACCAGCCGCC |

**Table 3: Phusion polymerase PCR**

| **50 µl reaction** | | **Program** | |
|---|---|---|---|
| 10µl 5x reaction buffer GC | denaturing | 98°C x 30s | |
| 1 µl 10 mM dNTP | denaturing | 98°C x 15s | 30 cycles |
| 1 µl 10 mM forward primer | annealing | 65°C x 30s | |
| 1 µl 10 mM reverse primer | elongation | 72°C x 25s (1kb/30s) | |
| 1 µl Phusion DNA polymerase (2U/µl) | final elongation | 72°C x 5min | |
| 2 µl cDNA | | | |
| 34 µl $ddH_2O$ | | | |

Vector including GRP8 is represented in the following map after ligation.

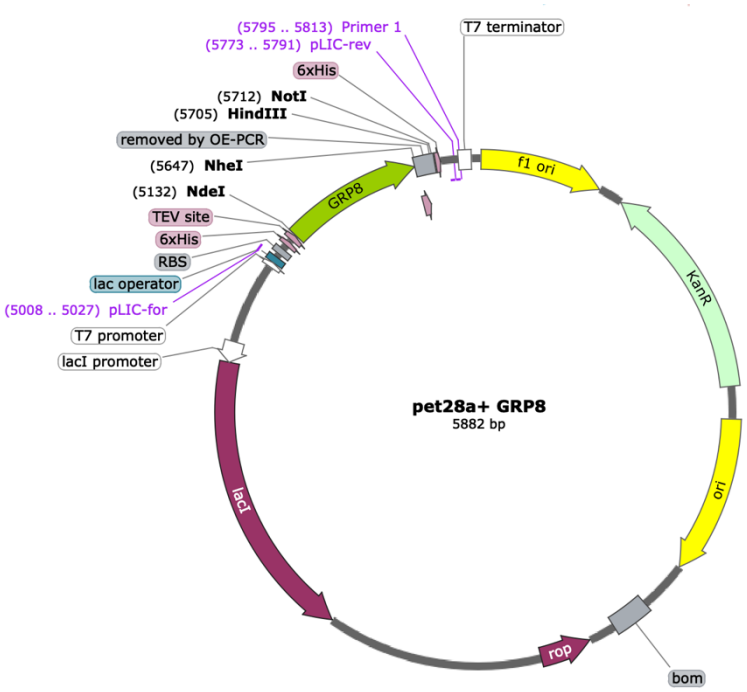



**Figure 10: Map of the pet28a + GRP8 vector by *SnapGene*.** Vector sites are indicated in the scheme.

## Restriction enzyme digestion and ligation

Restriction enzyme digestion was done for GRP8 fragments and pNH vector with both NdeI and HindIII with the same PCR program settings. After 1 hour digestion, 5 µl of Antarctic

phosphatase (FastAP) and 6 µl of FastAP buffer were added for dephosphorization of the DNA 5'- and 3'- overhangs. After inactivation both products were checked through gel electrophoresis. 1 % TAE-agarose gel with gel red stain was prepared and run for 30 minutes at 100 V, and then the PCR products were cleaned with *NucleoSpin gel and PCR Clean-up kit* (Macherey-Nagel). Ligation was left ongoing overnight at room temperature.

**Table 4: Restriction enzyme digestion**

| **50 µl reaction** | **Program** | |
|---|---|---|
| 5 µl Restriction Enzyme | Digestion | 37°C x 1 hour |
| 5 µl PCR fragment / vector | FastAP treatment | 37°C x 15 min |
| 10 µl CutSmart® buffer | Inactivation | 80°C x 20 min |
| 30 µl $ddH_2O$ | | |

**Transformation in *E. coli* cells**

Ligation products were transformed into chemically competent XL10-Gold cells. 10 µl of the ligation mix was transferred into 200 µl of XL10 cells and it was kept on ice for 30 minutes. Then the cells were heat-shocked at 42°C for 30 seconds. 800 µl of LB medium was added, and it was left incubating at 37°C for 2 hours shaking at 170 rpm. The cells were centrifuged at 3000 x g for 1 minute at 4°C. Supernatant was discarded and the pellet was resuspended in 200 µl of LB medium. The resuspension was plated in LB-agar medium with kanamycin (100 µg/ml) selection antibiotic and the plates were left incubating at 37°C overnight.

**Colony screening**

To ensure the correct insertion of the ligated vector and plasmid, a colony screening was performed with the colonies that grew overnight on the plate. Three to five colonies were picked with the pipette tip and used as template for the T7 primer-standardized colony screening. PCR products were checked by gel electrophoresis. 1 % TAE-agarose gel with *GelRed* stain was prepared and ran for 30 minutes at 100 V. Positive colonies were cleaned with *NucleoSpin gel and a PCR Clean-up kit* (Macherey-Nagel). To ensure that the cleaned-up plasmids included the insertions, plasmids were sent for sequencing.

**Table 5: Colony screening PCR**

| **50 µl reaction** | | **Program** | |
|---|---|---|---|
| 5µl 10x reaction buffer | denaturing | 95°C x 2min | |
| 2 µl 25 mM $MgCl_2$ | denaturing | 95°C x 30s | 25 cycles |
| 2 µl 10 mM T7 forward primer | annealing | 56°C x 30s | |
| 2 µl 10 mM T7 reverse primer | elongation | 68°C x 1kb/min | |
| 1 µl 10 mM dNTP | final elongation | 68°C x 5min | |
| 2 µl Taq polymerase (20 U/µl) | | | |
| 1 colony | | | |
| 36 µl $ddH_2O$ | | | |

## Site-directed mutagenesis

GRP8 was truncated into a shorter version without the glycine-rich domain. For this, primers were designed on *SnapGene* software and GRP8 plasmid was used as a template.

**Table 6: Primers for GRP8 truncation**

| **Name** | **sequence** |
|---|---|
| GRP8_short_fw | GAACGAGGCTCAGTCGAGAGGTAGCTGAGGTGGCGGAG GAGGCCGTGGTGGAA |
| GRP8_short_rev | TTCCACCACGGCCTCCTCCGCCACCTCAGCTACCTCTCG ACTGAGCCTCGTTC |

**Table 7: Site-directed mutagenesis PCR**

| **50 µl reaction** | | **Program** | |
|---|---|---|---|
| 10 µl 5x reaction buffer GC | denaturing | 98°C x 5min | |
| 1 µl 10 mM dNTP | denaturing | 98°C x 10s | 16 cycles |
| 2.5 µl 10 mM GRP8s forward primer | annealing | 65°C x 30s | |
| 2.5 µl 10 mM GRP8s reverse primer | elongation | 72°C x 1kb/30s | |
| 5-10 ng GRP8 template (final concentration) | final elongation | 72°C x 10min | |

| |
|---|
| 1 µl Phusion DNA polymerase (20 U/µl) |
| Up to 50 µl dd$H_2O$ |

After the PCR, the product was digested with 1 µl of DpnI and incubated for 1 hour at 37°C. The digested product was then transformed into XL10-Gold cells following the previously described protocol and checked through gel electrophoresis 1% TAE-agarose gel with *GelRed.* Plasmids from positive colonies were isolated with *NucleoSpin gel and PCR Clean-up kit* (Macherey-Nagel). To ensure the cleaned-up plasmids included the insertions, plasmids were sent to *Eurofins* for sequencing.

**SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)**

This technique allows for the separation of proteins by size, therefore in protein purification it is essential to identify the protein of interest and check its purity. This was done according to the protocol established by Laemmli (Laemmli, 1970). Due to the small size of GRP8 and $GRP8^{short}$, gels with 15% acrylamide were prepared with 0.75mm width. Samples for the gel electrophoresis were prepared with 10 µl of sample and 2 µl of 1x loading buffer, then heated for 5 minutes at 95°C and then loaded onto the gel. 1.5 µl of *PageRuler prestained 10 to 180 kDa* was used as a ladder. The electrophoresis chamber Mini-PROTEAN III was filled with SDS-running buffer and the gel was run for 1 hour at 150 V and 60 mA. The gel was then washed with dd$H_2O$ for 5 minutes three times and stained with Colloidal Coomassie stain (Dyballa and Metzger, 2009) for around 1-2 hours and then destained for 20-30 minutes.

**Table 8: SDS gel preparation**

| **15 % gel** | **up to 5 ml** | **stacking gel** | **up to 1 ml** |
|---|---|---|---|
| $H_2O$ | 1.1 ml | $H_2O$ | 0.68 ml |
| 30 % acrylamide mix | 2.6 ml | 30 % acrylamide mix | 0.17 ml |
| 1.5 M Tris (pH 8.8) | 1.3 ml | 1 M Tris (pH 6.8) | 0.13 ml |
| 10 % SDS | 50 µl | 10 % SDS | 10 µl |
| 10 % APS | 50 µl | 10 % APS | 10 µl |
| TEMED | 2 µl | TEMED | 1 µl |

**Protein Production**

For *in-vitro* studies, it is essential to produce large amounts of pure protein. The production of proteins is an extensive process that requires many steps and optimizations to get to the desired yield. Due to its instability in low temperatures, the GRP8 protein production process was carried at 20°C. GRP8$^{short}$ was produced at 4°C.

**Protein expression test**

Two *E. coli* strains were used for the expression test of GRP8 and GRP8$^{short}$. In the table below these are described.

2 µl of plasmid was added to 200 µl of cells, kept on ice for 2 minutes, then heat-shocked for 30 seconds, kept on ice again for 2 minutes, added 400 µl of LB medium and 200 µl was plated. Lastly, the plate was incubated overnight at 37°C shaking at 170 rpm. The next day, 2 ml ZY-autoinduction medium (ZY medium, 1x NPS, 1x5052, 1 mM $MgSO_4$, 100 µg/ml kanamycin) in a 15 ml Falcon tube was prepared with the overnight culture and was inoculated. The culture was incubated overnight at 37°C shaking at 170 rpm. The next day the cultures were centrifuged at 3000 x g at room temperature. The pellets were then resuspended in 20 µl $ddH_2O$ and then prepared for SDS-PAGE. The protocol proceeded as described before with the same parameters. After Colloidal Coomassie staining, overexpression of the cultures should be visible. The stained SDS-gel was taken for MALDI-TOF/TOF MS analysis for protein identification, where the bands of interest were excised and analyzed following the *Peptide Mass Fingerprint* method.

**Table 9: *E. coli* strains used for GRP8 and GRP8$^{short}$ expression**

| *E. coli* strain | genotype | properties |
|---|---|---|
| BL21-gold (DE3) | *E. coli* B F– ompT hsdS($r_B$- $m_B$-) dcm+ Tet$^r$ gal λ(DE3) endA Hte | IPTG/lactose inducible<br>High efficiency of protein expression under T7 promoter. |
| CodonPlus RIPL (DE3) | *E. coli* B F- ompT hsdS($r_B$-$m_B$-) dcm+ Tet$^r$ gal λ(DE3) endA Hte [*argU proL* Cam$^r$] [*argU ileY leuW* Strep/Spec$^r$] | IPTG/lactose inducible. For organisms with high AT or GC- rich genomes. Contains extra codons of argU, ileY and leuW. |

**Buffer Screening, lysis and solubilization**

Different buffers were tested to solubilize GRP8. The buffer testing was carried as described in Lindwall et al., 2000. Pellets from expressed cells were washed (10 mM Tris, pH 8.5, 100 mM NaCl and 1 mM EDTA) and then centrifuged 12.000 x g at room temperature. The supernatant was discarded, and the pellets were resuspended in 1 ml of different buffers (Lindwall *et al.*, 2000) and incubated with lysozyme (final concentration 50 μg/ml) for 5 minutes at room temperature. The cells were then disrupted by sonication (40 % energy output, 30% duty cycle) and then left resting for 10 minutes. After, the lysate was centrifuged for 10 minutes at 16000 x g at room temperature. 10 μl of the supernatant was sampled for SDS-PAGE, as previously described. After Colloidal Coomassie staining, the soluble protein in the adequate buffer should be visible.

**Protein purification**

In this part of the protein production, different purification approaches were included in different steps to ensure high yield and purity of the proteins in a larger scale.

**Large scale expression**

Due to the instability of GRP8, the expression medium was changed to TB with 3-hour IPTG induction.

1.2 liters of TB medium (1 x TB, 1 x TB salts and 100 μg/ml kanamycin) were prepared and divided into three 2-liter flasks containing 400 ml medium each. Each flask was inoculated with 600 μl of pre-plated cells and incubated at 37°C shaking at 170 rpm until reaching an OD of 0.8, then the cultures were induced with 400 μl IPTG (1 M) and left shaking at 170 rpm at 37°C for 3 hours. After, the cultures were centrifuged at 7500 x g for 30 minutes at 20°C. The pellets were stored at -20°C or directly used for further purification.

$GRP8^{short}$ was produced with ZY-autoinduction medium. 1.2 liters of ZY medium (ZY medium, 1x NPS, 1x5052, 1 mM $MgSO_4$, 100 μg/ml kanamycin) were prepared and divided into three 2-liter flasks containing 400 ml medium each. Each flask was inoculated with 600 μl of pre-plated cells and incubated at 37°C shaking at 170 rpm for 2 hours for a growth boost, then the cultures were moved to a 24°C room and left shaking at 170 rpm overnight. After, the cultures were centrifuged at 7500 x g for 30 minutes at 4°C. The pellets were stored at -20°C or directly used for further purification.

### Lysis and solubilization

Pellets were resuspended in 50 ml of lysis buffer. The resuspended pellets were transferred into a 100 ml glass beaker and then a small bunch of DNase and ribonuclease A were added with a scalpel. 1 tablet of the Complete Protease Inhibitor EDTA free (Roche) was added. 500 µl AEBSF (final concentration 1 mM) and 500 µl lysozyme (1 g/ml stock) was added. The beaker was left stirring for GRP8 at 20°C for 25 minutes and for GRP8$^{short}$ at 4°C for 45 minutes. The cells were then disrupted by sonication (50 % energy output, 50% duty cycle) eight times with 30 seconds sonication and 20 seconds rest. The lysate was then centrifuged at 7500 x g for 30 minutes at 20°C or 4°C, respectively, the supernatant was then filtered into a 50 ml Falcon tube with a 0.45 µM filter.

**Table 10: Buffers used for solubilization of GRP8 and GRP8$^{short}$**

| buffer | reagents |
|---|---|
| GRP8 lysis buffer | 50 mM HEPES, pH 8.0, 200 mM KCl, 5-10% glycerol, 1 mM DTT, 30 mM imidazole, 1 mM PMSF, 1 mM AEBSF |
| GRP8$^{short}$ lysis buffer | 50 mM HEPES, pH 8.0, 500 mM NaCl, 5-10% glycerol, 1 mM DTT, 30 mM imidazole, 1 mM PMSF, 1 mM AEBSF |

### Affinity chromatography (A.C)

GRP8 and GRP8$^{short}$ include an N-terminal 6xHis-tag. The soluble protein was loaded onto the *ÄKTA Start* where it was loaded onto a 5 ml *HisTrap™ HP* column. The A.C was run with a flow rate of 1.0 ml/minute with a pressure limit of 1.0 MPa and an equilibration volume of 5 CV.

**Table 11: Buffers used for affinity chromatography of GRP8 and GRP8$^{short}$**

| buffer | reagents |
|---|---|
| GRP8 elution buffer | 50 mM HEPES, pH 8.0, 300 mM KCl, 5-10% glycerol, 1 mM DTT, 1 M imidazole |
| GRP8$^{short}$ elution buffer | 50 mM HEPES, pH 8.0, 500 mM NaCl, 1 % glycerol, 1 mM DTT, 1 M imidazole |

The protein was eluted with different fractions using a gradient from 0% to 100% of imidazole with lysis buffer and elution buffer. The protein elution fractions were identified by the chromatogram of the ÄKTA system and then these fractions were checked through SDS-PAGE as previously described. After the Colloidal Coomassie staining, clean fractions are chosen for further purification.

**Dialysis and 6xHis-Tag removal**

GRP8 and GRP8$^{short}$ are designed to include a TEV protease cleavage after the 6xHis-Tag. Fractions containing the purified protein were pooled and loaded into a 3500 MWCO *SpectraPor* membrane with 200 µl TEV protease and left stirring in the dialysis buffer. Dialysis was done for GRP8 at room temperature for 1.5 hours and for GRP8$^{short}$ at 4°C overnight.

**Table 12: Dialysis buffers for GRP8 and GRP8$^{short}$**

| buffer | reagents |
|---|---|
| GRP8 dialysis buffer | 25 mM HEPES, pH 7.0, 300 mM KCl, 1 mM DTT |
| GRP8$^{short}$ dialysis buffer | 25 mM HEPES, pH 7.0, 300 mM NaCl, 1 mM DTT |

**Reverse nickel column**

A reverse nickel column in batch mode was performed for the removal of the remaining of 6xHis-tag with GRP8. 1 ml of *Ni-NTA Agarose* (Qiagen) beads were loaded into *PD10 empty columns* (GE healthcare) and then washed with dialysis buffer (5x CV; 2.5 ml). The dialyzed protein solution was loaded into the column and the flowthrough was collected for concentration and further purification. 1mM of AEBSF was added into the protein solution before the next step.

**Concentration**

Concentration steps were done 2 times during purification. First, after reverse nickel column step and before loading into the ÄKTA system for gel filtration. Second, after gel filtration for reaching the desired concentration of protein.

**Table 13: Concentrators used for GRP8 and GRP8s**

| protein | concentration step | concentrator |
|---|---|---|
| GRP8 | first | Vivaspin 10000 MWCO |
| | second | Vivaspin 5000 MWCO |
| GRP8s | first | Vivaspin 5000 MWCO |
| | second | Vivaspin 3000 MWCO |

### Gel filtration / size exclusion chromatography

After concentration, the protein solution was loaded into the ÄKTA system onto a *HiLoad Superdex 75 pg Gel Filtration column* (GE healthcare) attached. The A.C was run with a flow rate of 0.5 ml/minute with 0.5 MPa of pressure limit and an equilibration volume of 1 CV. The protein was eluted and identified by the chromatogram of the ÄKTA system and then these fractions were checked through SDS-PAGE as previously described. After Colloidal Coomassie staining, clean fractions were chosen for concentration and then protein aliquots were frozen in liquid nitrogen and stored at -80°C.

**Table 14: Gel filtration buffers for GRP8 and GRP8s**

| buffer | reagents |
|---|---|
| GRP8 SEC buffer | 25 mM HEPES, pH 7.0, 100 mM KCl, 1 mM DTT, 1 mM PMSF, 1 mM AEBSF |
| GRP8s SEC buffer | 25 mM HEPES, pH 7.0, 300 mM NaCl, 1 mM DTT, 1 mM PMSF, 1 mM AEBSF |

### Proteolysis

Limited proteolysis was performed with a *FloppyChoppy kit* (Jena Bioscience) according to the manufacturer's instructions, in which the flexible areas of GRP8 were removed. Digested fragments were then run through an SDS-PAGE as previously described and after Colloidal Coomassie staining, the gel was taken for Mass Spectrometry analysis for identification of the digested sites.

### MALDI-TOF/TOF mass spectrometry

Matrix assisted laser desorption/ionization – time-of-flight mass spectrometer was used for protein identification and amino acid modifications identification through *Peptide Mass Fingerprint* method.

**Peptide mass fingerprint**

Following (Walz *et al.*, 2002) protocol, SDS-gel containing the bands of interest for protein identification or amino acid modifications identifications were excised and transferred into a low-binding 0.5 ml tube (Sarstedt). The piece of gel was washed with 200 µl 50 mM $NH_4HCO_3$ for 5 minutes at room temperature. The solution was discarded and 100 µl 100 % acetonitrile (ACN) and 100 µl 50 mM $NH_4HCO_3$ was added for destaining until the color was transparent at room temperature. The solution was vortexed and discarded, 100 µl of ACN was added for 10 minutes at room temperature, then ACN was discarded, and the tube was left open for drying under a hood. 20 µl of trypsin (0.01 µg/l in 50 mM $NH_4HCO_3$) was added, and was left digesting for 1.5 hours at 37°C. After digestion, 1 µl of the sample was transferred into *AnchorChip 600/384* and mixed with 1 µl TA30 (30:70[v/v] ACN. 0.1% TFA). 0.5 µl *Peptide Calibration Standard* (Bruker Daltronik) was added onto the calibration spots. The chip was then left on top of a 37°C block until dried and then 1 µl of HCCA matrix (1.4 mg/ml HCCA in 85% ACN, 15% $H_2O$ and 0.1% TFA in 1 mM $NH_4H_2PO_4$) was added to the sample and the calibration spots. The chip was left on top of a 37°C block until dried.

Analysis was carried out with *MALDI-TOF/TOF-MS Ultraflex III* (Bruker Daltronik) in reflection mode. First, the equipment was calibrated with a mass-to-ratio (m/z) detection range from 600 to 4000 Da and with a *Peptide Calibration*. Samples were ionized with 20 % laser power and at least 1000 shots collected, and the spectrum was analyzed with the *Mmass* software. Peaks identified through the software were then run through the *MASCOT* server and run through *NCBI* and *SwissProt* databases with a peptide tolerance of 0.3, a variable modification in methionine oxidation, a miscleavage of 1 and the species was restricted to *Viridiplantae* or *A. thaliana*. As a result, proteins were successfully identified if they had a significant *MASCOT*-score and a sequence coverage of at least 25 %. Proteins (< 20 kDa; GRP8 and $GRP8^{short}$) also had to have at least five peptides clearly assigned.

**Phloem sap extraction and RNA isolation**

Phloem sampling was carried out as described in (Pahlow *et al.*, 2018) from *B. napus* cv. Drakkar. Phloem samples were frozen in liquid nitrogen and stored at -80°C.

RNA isolation was performed with TRIzol® Ls, per the manufacturer's instructions. After the aqueous phase separation, RNA precipitation was performed with *RNA Clean & Concentrator-25* (Zymo Research) following the manufacturer's instructions. RNAs were checked through the *Agilent 2100 Bioanalyzer system* (Agilent).

**Table 15: cDNA synthesis from phloem sap RNA**

| **20 µl reaction** | **Program** |
|---|---|
| 1 µl 25 µg RNA from phloem sap *(B. napus)* | 42°C x 1 hour |
| 1 µl (oligo dT)$_{18}$ Primer (0.5 µg/µl) | 70°C x 10 min |
| 4 µl 5x reaction buffer | |
| 1 µl Ribolock RNAase inhibitor (40 U/µl) | |
| 2 µl 10mM dNTPs | |
| 1 µl reverse transcriptase (20 U/µl) | |
| 10 µl dd$H_2O$ | |

**Table 16: Primers for purity check of phloem sap RNA**

| **probe** | **sequence** |
|---|---|
| TrxH | reverse: ATGGCCTGAACATCGAACTC |
| | forward: CTCAAGGCAGCCAAAGAATC |
| RuBisCO | reverse: CCGGGTACTCCTTCTTGCAT |
| | forward: TTCACCGGCTTGAAGTCATC |
| PCP | reverse: TTCCTTAATGGCCTCAGTGG |
| | forward: TCAGAACTGGAGCTTCAACG |

Purity control was performed by RT-PCR following (Pahlow *et al.*, 2018) from the RNA isolation performed previously. cDNA synthesis was done with reverse transcriptase.

After the cDNA synthesis, PCR with Phusion polymerase was performed with primers for thioredoxin h (TrxH), rubisco small chain (RuBisCO) and pollen coat protein BRU77666 (PCP). PCR preparations and settings followed by Phusion polymerase protocol. High resolution of automated electrophoresis of RNA. The *Agilent 2100 Bioanalyzer system* (Agilent) was used for total RNA analysis with the *RNA 6000 Nano kit* and *Eukaryote total RNA Nano assay* software. The analysis was performed following the manufacturer's instructions.

**CnBr-Sepharose affinity column**

Binding of the protein to the Sepharose beads was done with GRP8 and GRP8$^{short}$ in two different columns. 50 mg of the Sepharose beads were washed three times and incubated with 300 µl 1 mM HCl, after incubation the beads were centrifuged at 3000 x g at 4°C for 30 seconds and the supernatant was removed. In the last centrifugation step, the supernatant was not completely removed, and the beads were transferred to a mini centrifugation tube. The supernatant was then centrifuged at 700 x g for 30 seconds at 4°C. 500 µg of protein was added with 1 mg dextran sulfate and 250 µl coupling buffer (100 mM $NaHCO_3$ pH 8.3, 500 mM NaCl). The beads were incubated rolling overnight at 4°C. After, the column was centrifuged at 700 x g for 30 seconds at 4°C for 1 minute and then the column was washed twice with 300 µl coupling buffer. 400 µl 0.1M Tris-HCl pH 8.0 were added and left incubating rolling overnight at 4°C. Then, the column was centrifuged at 700 x g for 1 minute at 4°C and 25 µg of Phloem-RNA was loaded into the column. Beads were washed two times with 250 µl assay buffer (25 mM HEPES pH 7.5, 150 mM sodium acetate). 25 µg phloem RNA, previously isolated, was added to the column together with 200 µl assay buffer and then left incubating rolling for 15 minutes at room temperature. The column was centrifuged at 700 x g for 30 seconds at 4°C and the flowthrough was kept on ice. The beads were washed five times with 250 µl assay buffer and centrifuged at 700 x g for 30 seconds at 4°C. Elution proceeded with three different salt concentrations to generate a gradient and separate highly enriched RNAs.

**Table 17: Elution buffers for CnBr-Sepharose affinity column**

| elution buffer | enrichment | reagents |
|---|---|---|
| elution 1 | low | 25 mM HEPES pH 7.5, 250 mM sodium acetate |
| elution 2 | medium | 25 mM HEPES pH 7.5, 500 mM sodium acetate |
| elution 3 | high | 25 mM HEPES pH 7.5, 2 M sodium acetate |

After the elution, sodium acetate was added into all probes to a concentration of at least 0.3 M to ensure the adequate RNA precipitation. Then, 100 % EtOH was added in proportion, three times of the probe volume. The RNA was precipitated overnight at -20°C. The precipitated RNA was then centrifuged at full speed (16000 x g) for 25 minutes at 4°C. After, the supernatant was discarded and 1 ml, 70 % EtOH was added to the RNA pellet. The pellet was vortexed and then centrifuged at 7500 x g for 5 minutes at 4°C. The supernatant was then discarded, and the pellet was left drying under a hood. 35 µl of $ddH_2O$ were added and the

RNA was heated for 10 minutes at 60°C. The RNA samples then were analyzed through the *Agilent 2100 Bioanalyzer system* (Agilent). Samples that were complied with the requirements were then sent for RNA-sequencing (Novogene).

## RNA-sequencing

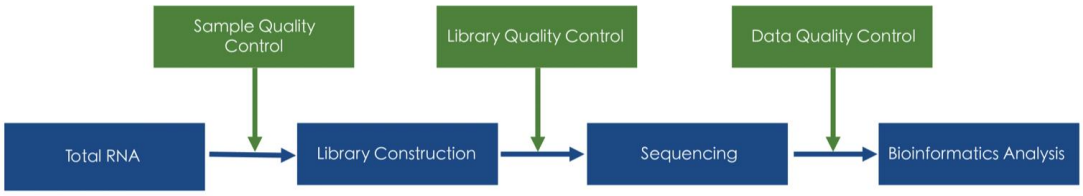



**Figure 11: RNA-sequencing project pipeline.** Courtesy of *Novogene Co, Ltd*.

RNA-sequencing analysis was performed by *Novogene Co, Ltd.* Transcriptome analysis assists to study the identification of genes that are expressed differentially in distinct sample populations. In this case, phloem RNA versus RNA-bound to GRP8 were sent for sequencing analysis. RNA sequencing was carried out as follows in figure 11.

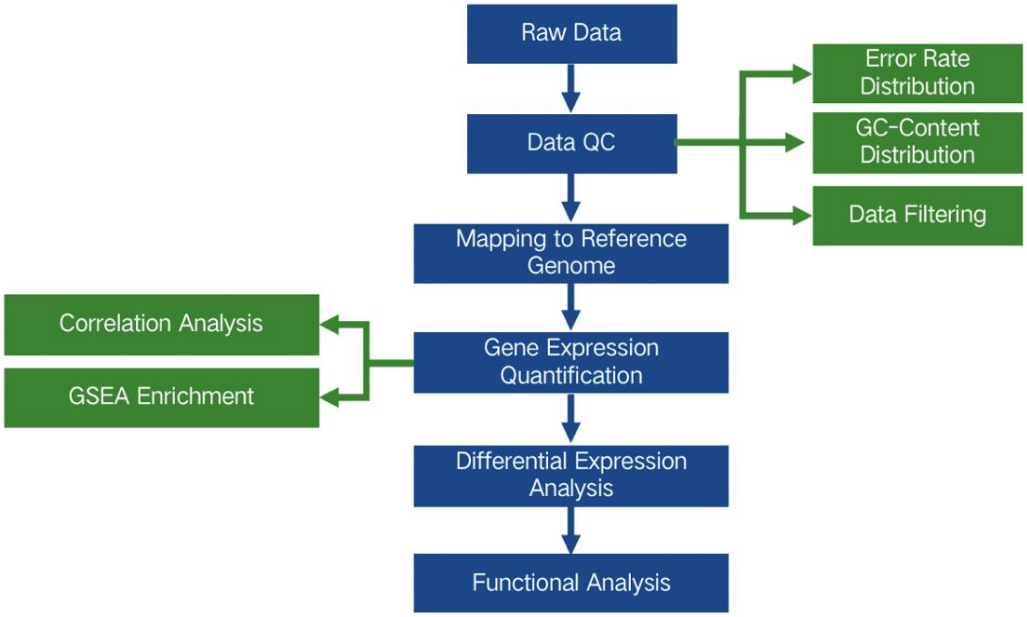



**Figure 12: Bioinformatics analysis workflow.** Courtesy of *Novogene Co, Ltd*. (Modified for GRP8 analysis)

**Microscale thermophoresis**

MST was performed for quantitative analysis of protein-RNA interactions in solution. RNAs were synthesized and labeled Cy5 for thermophoresis analyses. The biophysics behind the MST analysis are described by (Jerabek-Willemsen *et al.*, 2014)(Dijkman *et al.*, 2014)(Duhr and Braun, 2006) a local change of temperature induced by an infrared laser that will trigger a change in molecule concentration, and this can be quantified by the Soret coefficient (ST).

$$\frac{chot}{ccold} = \exp -STx\Delta T$$

These changes in the thermophoresis of a fixed fluorescent molecule (target, T) in a constant buffer will reflect changes in the entropy of the solution due to the binding of the non-fluorescent partner (ligand, L). Therefore, the measurement of this at different concentrations of ligand allows the quantification of the binding affinity through the calculation of the constant of dissociation (Kd) (Jerabek-Willemsen *et al.*, 2014).

$$Kd = [L]\ x(\frac{[T]}{[LT]})$$

The smaller the Kd, the tighter the ligand is bound to the target. Moreover, a smaller Kd translates to a higher binding affinity.

The MST analysis was performed with the *Monolith NT.115™* (Nanotemper) system. First, the preparation of the assay was done according to the *M.O control software* (Nanotemper). The *Monolith NT.115 Standard capillaries* (Nanotemper) were used for the assays. If proteins showed adhesion to the walls of the capillaries, then *Monolith NT.115 Premium capillaries* (Nanotemper) were used. For the MST assay, at least 20 nM of the labeled target RNA was used. The ligand (GRP8/GRP8s) was prepared in a series of dilutions from 0 μM to 40 μM. The MST power was set to 40 % and the excitation power was set automatically and varied in a range from 40 % to 100 %. All measurements were carried at room temperature. Raw data was analyzed with the *M.O Affinity analysis software* (Nanotemper) and then results were plotted with *OriginLab* software.

**gDNA isolation for RNA synthesis**

gDNA isolation from *A. thaliana* was performed with leaf material. First, the material was grounded in liquid nitrogen and then TRIzol® was added, following the manufacturer's

instructions. After the aqueous phase separation, RNA precipitation was performed with *RNA Clean & Concentrator-25* (Zymo Research) following the manufacturer's instructions.

*In vitro* RNA synthesis with T7-induced transcription for MST analysis

The RNA T7 transcription was based on (Cazenave and Uhlenbeck, 1994). Sequences for RNA synthesis were designed to include the T7 sequence promoter (5'-GAA ATT AAT ACG ACT CAC TAT A-3') and two extra guanines (GG).

**RNA synthesis of short RNAs**

Oligonucleotide templates of the RNAs were ordered for RNA synthesis. After the incubation overnight, 10 µl of 10 U DNase were added and incubated at 37°C for 30 minutes. To stop the reaction, 5 µl of 500 mM EDTA were added. RNA was purified with *RNA Clean & Concentrator-25* (Zymo Research) following the manufacturer's instructions. After elution, 1 µl RNA was checked through the *Agilent 2100 Bioanalyzer system* (Agilent). RNA was then frozen in liquid nitrogen and stored at -80°C.

**Table 18: RNA synthesis of short RNAs**

| **100 µl reaction** | **Program** |
|---|---|
| 10 µl 10x reaction buffer | 37°C x overnight |
| 10 µl NTPs (20 mM each) | |
| 2 µl T7 RNA-polymerase (3500 u) | |
| 1 µl Ribolock RNAase inhibitor (40 U/µl) | |
| 2 µl Pyrophosphatase (0.5 u/ml) | |
| 10 µl 100 mM DTT | |
| 0.5 µl Cy5 | |
| 10 pmol template | |
| Up to 100 µl dd$H_2O$ | |

**RNA synthesis of long RNAs**

gDNA from *A. thaliana* was extracted as previously described. PCR amplification was performed with Phusion Polymerase. To ensure that enough product was amplified, 100 µl of this reaction was prepared.

Total RNA isolation was performed from leaf and flower material for cDNA synthesis as previously described. PCR products from cDNA synthesis and gDNA were checked through gel electrophoresis. 1 % TAE-agarose gel with *GelRed* stain was prepared and ran for 30 minutes at 100 V. After, the band from the agarose gel containing the PCR product was excised and extracted with the *NucleoSpin gel and a PCR Clean-up kit* (Macherey-Nagel).

**Table 19: Phusion PCR for long RNAs**

| **100 µl reaction** | | **Program** | |
|---|---|---|---|
| 20 µl 5x reaction buffer GC | denaturing | 98°C x 5min | |
| 2 µl 10 mM dNTP | denaturing | 98°C x 10s | 16 cycles |
| 2.5 µl forward primer | annealing | 60°C x 30s | |
| 2.5 µl reverse primer | elongation | 70°C x 1kb/30s | |
| 2 µl gDNA (5-100 ng) | final elongation | 70°C x 10min | |
| 1 µl Phusion DNA polymerase (2U/µl) | | | |
| 70 µl $ddH_2O$ | | | |

**Table 20: SmaI cut ligation with T4 ligation**

| **20 µl reaction** | **Program** |
|---|---|
| 2 µl 10x T4 ligase buffer | 16°C x 16 hours |
| 1 µl SmaI (2000 U/ml) | |
| 1 µl T4 ligase (100.000 CEU) | |
| 1 µl 10 mM ATP | |
| 100 – 200 ng pUC57 | |
| PCR product (3-5:1; product:pUC57) | |
| Up to 20 µl $ddH_2O$ | |

pUC57 was ligated with the PCR product with SmaI restriction enzyme and T4 ligase. The final concentration of PCR product and pUC57 vector was 3-5:1. After 16 hours of incubation at 16°C, transformation of XL10-gold cells was performed. 10 µl of the ligation product was added to 100 µl of XL10-gold cells and left on ice for 20 minutes. The cells were heat shocked

at 42°C for 45 seconds, then 200 µl of LB medium was added and left incubating at 37°C for 1 hour shaking at 170 rpm. The transformed cells were plated on LB and carbenicillin and left incubating at 37°C overnight. Next day, colonies were picked and checked through the colony screening protocol previously described and a 1 % TAE-agarose gel with *GelRed* stain was prepared and run for 30 minutes at 100 V. Positive colonies were cleaned with *NucleoSpin Plasmid EasyPure* (Macherey-Nagel). To ensure plasmids included the desired sequence, plasmids were sent for sequencing. PCR amplification of the plasmid with Phusion polymerase was performed following the previously described method. In this step, only 1 µl of plasmid was used. After this, the PCR product was used as a template for RNA synthesis. After the incubation, 10 µl of 10 U DNase were added and incubated at 37°C for 30 minutes. To stop the reaction, 5 µl of 500 mM EDTA were added. RNA was purified with *RNA Clean & Concentrator-25* (Zymo Research), per the manufacturer's instructions. After elution, 1 µl RNA was checked through the *Agilent 2100 Bioanalyzer system* (Agilent). RNA was then frozen in liquid nitrogen and stored at -80°C.

**Table 21: RNA synthesis of long RNAs**

| **100 µl reaction** | **Program** |
|---|---|
| 10 µl 10x reaction buffer | 37°C x 2 hours |
| 10 µl NTPs (20 mM each) | |
| 10 µl T7 RNA-polymerase (*in house*) | |
| 1 µl Ribolock RNAase inhibitor (40 U/µl) | |
| 5 µl pyrophosphatase (0.5 u/ml) | |
| 10 µl DMSO (10% final volume) | |
| 0.5 µl Cy5 (10 mM stock) | |
| 5 pmol template | |
| Up to 100 µl dd$H_2O$ | |

**UV crosslinking**

Synthesis of miRNA164 was performed. 5 µl of *miRNA164* forward primer and 5 µl of miRNA164 reverse primer and this was heated at 95°C for 5 minutes and then it was left to cool down at room temperature for 15 minutes. After the annealing, reverse transcription was performed. miRNA164 was added with GRP8 in a 2 or 3:1 ratio with crosslinking buffer (100

mM Tris-HCl pH 7.5, 1.5 M NaCl and 10 mM $MgCl_2$) and it was left incubating for 15 minutes at room temperature.

**Table 22: *miRNA164* sequence**

| name | sequence |
|---|---|
| miRNA164_T7 forward | taatacgactcactataggGGGTGGAGAAGCAGGGCACGTGCA |
| miRNA164_T7 reverse | CTGCTTCTCCACCCcctatagtgagtcgtatta |

After incubation, the samples were crosslinked under a UV light for 12 minutes. When crosslinking was done, the samples were checked through an SDS-PAGE as previously described, followed by Colloidal Coomassie staining. The gel was taken for Mass Spectrometry analysis for identification of the digested sites.

**Table 23: *miRNA164* synthesis for crosslinking**

| 100 µl reaction | Program |
|---|---|
| 10 µl 10x reaction buffer | 37°C x 3 hours |
| 2 µl 100 mM GTP | |
| 2 µl 100 mM CTP | |
| 2 µl 100 mM UTP | |
| 20 µl 20 mM 8-acido-ATP | |
| 5 µl T7 RNA-polymerase (*in house*) | |
| 5 µl pyrophosphatase (0.5 u/ml) | |
| 2 µl miRNA164 annealed | |
| 1 µl Ribolock | |
| Up to 100 µl $ddH_2O$ | |

**Thioflavin-T Assay**

Thioflavin T (ThT) test was performed with GRP8 and $GRP8^{short}$. Samples were prepared to have a final concentration of 50 mM of protein and a final volume of 100 µl, in triplicates. For better uniformity, a sterilized glass sphere 2.85-3.45 mm was added to the 96-well

microtitration plate (TECAN). When GRP8 with *miRNA164* was tested, a ratio of 2:1 in concentration was calculated. ThT and SEC buffer were added as usually instructed.

**Table 24: Sample preparation ThT assay**

| **100 µl sample** |
|---|
| 1 µl 1 mM Thioflavin T |
| 50 µM protein |
| Up to 100 µl SEC buffer of protein (+1mM DTT) |

After sample preparation, the 96-well microtitration plate was gently shaken by hand for uniformity and then was sealed with a transparent sealing tape. Lastly, the plate was inserted and analyzed with the *TECAN SPARK® system* (TECAN).

**Dynamic light scattering (DLS)**

DLS was performed to determine the monodispersity of the protein samples. These measurements were done before SAXS and crystallographic experiments. Depending on the availability, different DLS devices were used. Nevertheless, all devices were from the same manufacturer (X-tal Concepts) and were set with the same parameters. The protein samples were centrifuged at 20°C (GRP8) and GRP8$^{short}$ (4°C) for 10 minutes at 20.000 x g before transferring to the quartz dust-free DLS cuvette. Depending on the device, samples between 2 – 15 µl were placed in the cuvette. The measurements were 10 – 20 seconds long and repeated 10 – 25 times at 20°C. The sample was irradiated with a red-light laser (660 nm, 100mW laser diode power) with a scattering angle of 90° or 142°.

$$D = \frac{k_B T}{6\pi\eta R_H}$$

$D$ = translational diffusion coefficient ($m^2/s$)

$k_B$ = Boltzmann constant ($m^{2*}kg/Ks^2$)

$T$ = temperature (K)

$\eta$ = viscosity (Pa*s)

$R_H$ = hydrodynamic radius (m)

For the determination of the hydrodynamic radius ($R_h$), the Stokes-Einstein equation is used. This was automatically calculated with the DLS software (X-tal Concepts), assuming that the protein is globular.

**Small angle X-Ray scattering (SAXS)**

All experiments with SAXS were performed at the Petra III Beamline P12 *BioSAXS* with the EMBL team located at DESY in Hamburg. Before all the experiments, 3 mM DTT was added in all samples for radiation damage. Then, samples were centrifuged at 20.000 x g for 30 minutes at 20°C.

All samples were suspended in the same buffer solution (25 mM HEPES 7.0, 100 mM KCl, 1 mM DTT). For the subtraction of this, the buffer without the protein was also measured through batch-mode SAXS and SEC-SAXS. Bovine serum albumin (BSA) was used as a standard measurement and with the information obtained, it was possible to calculate the estimated molecular weight (MW) of the GPR8 protein samples.

$$MW_{sample} = \frac{MW_{BSA}\ x\ I_{(0)sample}}{I_{(0)BSA}}$$

$MW_{(sample)}$ = molecular weight of the protein sample

$MW_{(BSA)}$ = molecular weight of BSA

$I_{(0)\ sample}$ = scattering intensity of the protein sample at 0° angle.

$I_{(0)\ BSA}$ = scattering intensity of BSA at 0° angle.

BSA results were normalized and fitted into a dimensionless equation. For normalization, the values were divided for the maximum value of the data set. For the dimensionless fitting, values were calculated with the following equation:

$$\frac{I_{(s)}}{I_{(0)}}\ x\ (sRg)^2$$

$I_{(s)}$ = experimental scattered intensity

$I_{(0)}$ = scattering intensity of the protein at 0° angle.

$s$ = scattering vector

$Rg$ = radius of gyration

**Table 25: Sample data-collection information**

| **SAXS data-collection** | |
|---|---|
| Instrument/data processing | PETRA III Beamline P12 BioSAXS at DESY, Hamburg. PILATUS 6M detector |
| Wavelength (nm) | 0.099984 |
| Sample-to-detector distance (m) | 3 |
| Absolute scaling method | Relative scattering of pure water |
| Monitoring for radiation damage | Frame comparison |
| Exposure time (s/frame) | Batch – 0.1<br>SEC-SAXS – 0.0245 |
| Sample configuration | Monomer – P1<br>Dimer – P2 |
| Sample T (°C) | Batch – 20, 10, 5<br>SEC-SAXS - 20 |

**Batch-mode SAXS**

Four experiments were carried in batch-mode SAXS. The first two were to measure GRP8 dimer and monomer. As a point of comparison, BSA was measured in the same sample buffer (25 mM HEPES 7.0, 100 mM KCl, 1 mM DTT). The last experiment included a GRP8 monomer measured at three different temperatures (5°C, 10°C, 20°C).

**Table 26: protein samples for batch-mode SAXS**

| **sample** | **symmetry** | **concentration** | **T(°C)** |
|---|---|---|---|
| GRP8 mono | P1 | 2.2 mg/ml | 5 |
| GRP8 mono | P1 | 2.2 mg/ml | 10 |
| GRP8 mono | P1 | 2.2 mg/ml | 20 |
| BSA | P1 | 1.64 mg/ml | 20 |

**SEC-SAXS**

GRP8 monomer and dimer protein samples were measured through SEC-SAXS.

**Table 27: protein samples for SEC-SAXS**

| sample | symmetry | concentration | T(°C) |
|---|---|---|---|
| GRP8 mono | P1 | 3.6 mg/ml | 20 |
| GRP8 dimer | P2 | 7.8 mg/ml | 20 |

### SAXS data analysis

For the experimental data processing, all data was pre-analyzed in collaboration with the EMBL team at DESY in Hamburg and then analyzed with the *ATSAS (3.0.5) software* (Manalastas-Cantos *et al.*, 2021). SEC-SAXS frames were visualized and selected through the *CHROMIXS* (Panjkovich and Svergun, 2018). *PRIMUS* (Konarev *et al.*, 2003) was used for the analysis of the one-dimension data obtained through SAXS (both methods). Through this program, it was possible to determine the quality of the data (signal-to-noise ratio) and possible aggregation presence.

**Table 28: Software used for data-processing**

| Software | | |
|---|---|---|
| SAXS data reduction | | |
| Extinction coefficient estimate | ProtParam | Gasteiger et al., 2005 |
| Calculation of delta p and v values | ATSAS 3.0.5 | Manalastas-Cantos et al., 2021 |
| Selection of SEC-SAXS frames | CHROMIXS | Konarev et al., 2003 |
| Basic analyses – Guinier, P(r), Vp | PRIMUS | Panjkovich & Svergun, 2018 |
| | GNOM | Quérouil et al., 2015 |
| Shape modeling | DAMMIF | Franke & Svergun, 2009 |
| | CRYSOL | Panjkovich & Svergun, 2016a |
| | SREFLEX | Svergun et al., 2015 |
| | EOM | Tria et al., 2015 |
| 3D graphic model representation | PyMOL | |
| | SASpy | Panjkovich & Svergun, 2016b |

Assuming that the particle size was small, the radius of gyration (Rg) could be calculated through the Guinier analysis. In this last version of *ATSAS (3.0.5) software*, *PRIMUS* also

includes the possibility of calculating the distance distribution function $P_{(r)}$ by *GNOM* (Quérouil *et al.*, 2015) program already incorporated. MW was also calculated through *PRIMUS*.

For the *ab initio* creation of 3D models, *DAMMIF* was used (Franke and Svergun, 2009). The *ATSAS online server* (https://www.embl-hamburg.de/biosaxs/atsas-online/) was used for running the experimental data in different programs for improving the 3D modeling of the protein samples. *EOM* (Tria *et al.*, 2015) was used to evaluate protein data with its amino acid sequence. *CRYSOL* and *SREFLEX* (Svergun, Barberato and Koch, 2015; Panjkovich and Svergun, 2016a) were used to compare the experimental data with *Alphafold* database (Jumper *et al.*, 2021) for 3D structure prediction of the protein and *i-TASSER* server (Yang *et al.*, 2014) for 3D structure prediction and analog proteins that are already structurally characterized. The models were then optimized in *PyMOL* (Schrödinger) and fitted with the *SASpy* plugin (Panjkovich and Svergun, 2016b).

**Liquid-liquid phase separation assay**

To induce phase separation, different concentrations of GRP8 monomer, dimer and GRP8s between 1 µM and 50 µM were incubated with 1 µM of Cy3-labeled 21R-RNA (Cy3-5′-ACUGCUAGAGAUUUUCCACAU-3) in phase separation buffer (50 mM Tris-HCl 7.5, 150 mM NaCl) for 10 minutes at room temperature. Afterwards, PEG3350 10% (w/v) was added for an incubation period of 15 minutes at room temperature. Samples were documented on *Olympus MVX10* Macroscope, equipped with GFP and RFP filter. Samples were then selected for incubation on ice and then centrifuged for 5 minutes at 4°C at 21000 x g. Supernatant was discarded and pellet with RNA-protein complex was resuspended on 4 µl of separation buffer. Samples were then re-documented on a macroscope.

**Crystallographic trials**

For high resolution studies of GRP8, crystallographic trials were attempted in collaboration with the crystallography department at DESY in Hamburg. GRP8 was concentrated to 7.8 mg/ml and then checked through DLS with different optimized buffers. In addition, 1 mM TCTP was added instead of DTT. GRP8 then was tested with two different crystallization screening kits, *JBScreen JCSG++HTS* (Jena Bioscience) and *PACT premier™ Eco Screen HT 96* (Molecular Dimensions) with a total of 192 different conditions. GRP8 and the buffers were distributed with a 1:1 ratio and injected for microbatch screening by *Oryx8 protein*

*crystallization robot* (Douglas Instruments) using the sitting drop method. The plates were then stored in an incubator at 20°C and were checked regularly for crystal-nuclei formation across a period of two weeks.

## 3. Results

The characterization of GRP8 was divided into two parts: functional and structural. Protein production included steps from cloning until obtaining the purified protein. For further characterization assays, a truncated version of GRP8 was produced (GRP8$^{short}$) which only included the RRM domain (figure 13). This truncation was introduced in the last serine [S] before the glycine-rich domain.

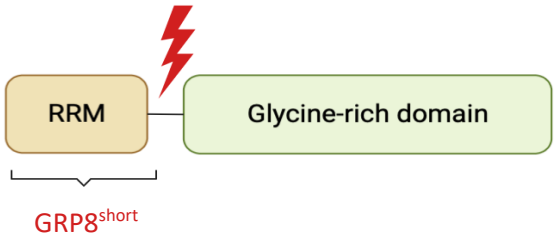



**Figure 13: Truncation of GRP8 into GRP8$^{short}$.** GRP8 was truncated into a shorter version only including the RRM domain. Red lightening indicates where the truncation was introduced.

Figure 14 shows a schematic representation of the overall workflow to produce and functionally and structurally characterize GRP8 and GRP8$^{short}$. This workflow is the result of many optimization steps added, which are shown in dotted arrows.

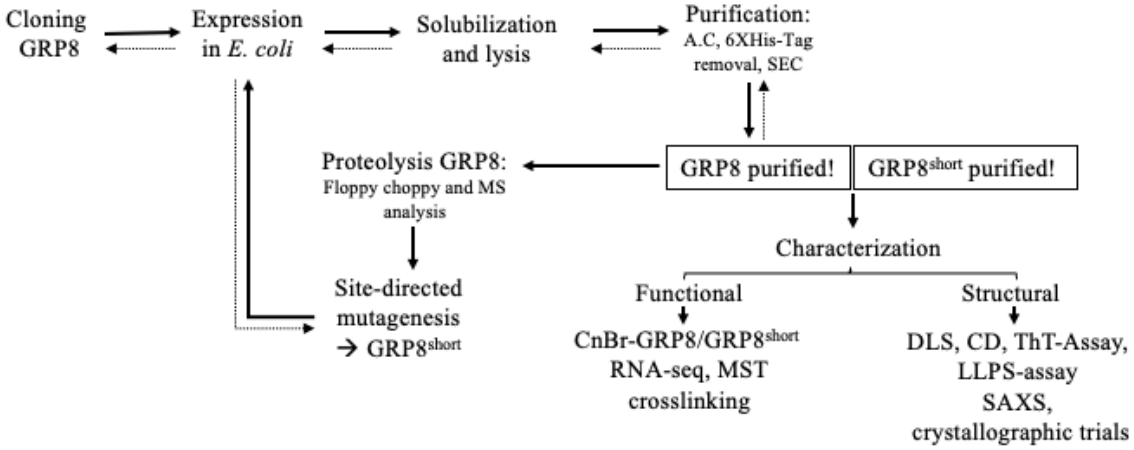



**Figure 14: Schematic representation of the workflow of GRP8 and GRP8$^{short}$ characterization.** Solid arrows represent workflow direction. Dotted arrows represent possible troubleshooting steps. After obtaining GRP8 and GRP8 short, characterization analysis was divided in two directions. A.C, affinity chromatography. 6XHis-Tag, polyhistidine tag. CnBr-GRP8/s, CnBr-activated Sepharose bound to GRP8 or GRP8s. RNA-seq, RNA sequencing. MST, microscale thermophoresis. DLS, dynamic light scattering. ThT-assay, thioflavin assay. SAXS, small angle X-ray scattering.

### 3.1 Protein production

Protein production was achieved by the combination of several techniques that include cloning and expression of GRP8 and GRP8$^{short}$ in *E. coli* cells, their disruption for the extraction of the proteins and finally the purification of GRP8 and GRP8$^{short}$.

From the expression (figure 15a), it was possible to observe that GRP8 is present, nevertheless there are a lot of unwanted *E. coli* proteins. To produce highly pure samples in large quantities different purification approaches were used. The first purification step was affinity chromatography (A.C), where the protein binds to the column on the 6xHis-tag side. As it is an affinity column, other proteins from *E. coli* can still bind to it, therefore it was necessary to move on with more purification steps (figure 15b). After A.C, there were still a lot of unwanted proteins. After the digestion of the 6His-tag with TEV protease, a reverse nickel column was performed for the separation of GRP8 undigested, which was discarded. Size exclusion chromatography (SEC) was performed as a last purification step. Figure 15c shows SDS-PAGE gel after SEC where GRP8 is observed in a pure solution state.

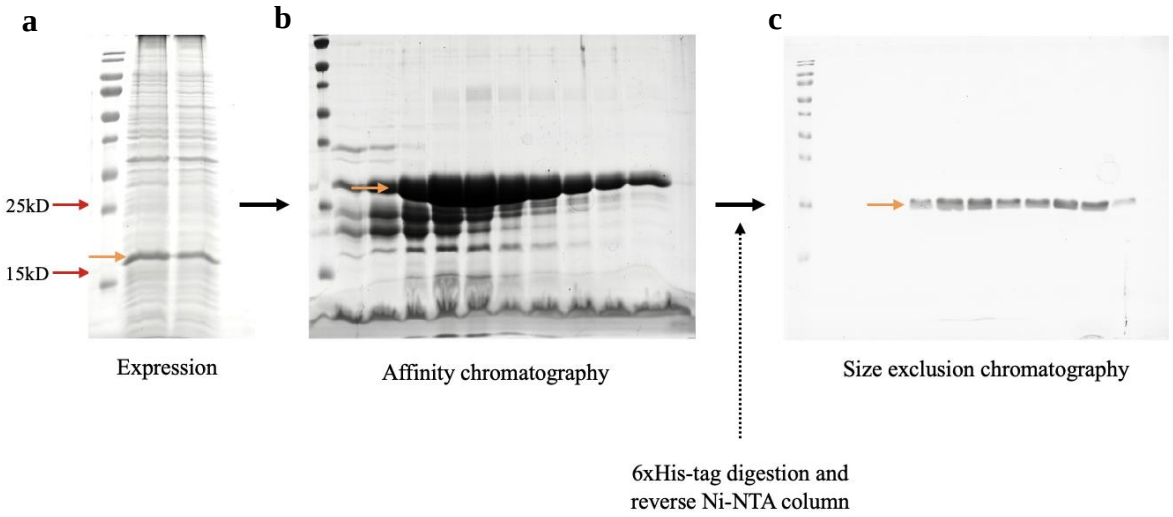



**Figure 15: Expression and purification of GRP8.** a) SDS-PAGE gel after expression. The red arrows show the reference ladder kDa. The orange arrow shows GRP8 at around 16 kDa. b) SDS-PAGE gel after affinity chromatography. The orange arrow shows the GRP8 presence at around 16kDa. c) SDS-PAGE gel after size exclusion chromatography. The orange arrow shows the GRP8 presence at around 16kDa. The black arrows show the workflow direction of the protein purification. The dotted arrow shows the extra purification steps that are not shown by SDS-PAGE gels.

After SEC, it was observed through the chromatogram that there were two forms of GRP8 present. Figure 16 shows the SEC chromatogram of the GRP8 gel filtration. Since SEC is a method for particle separation by size, it was assumed that the first peak contained the dimer and the second contained the monomer. Dimer and monomer forms were later confirmed by dynamic light scattering (DLS). This was relevant for some structural studies that were performed allowing their comparison.

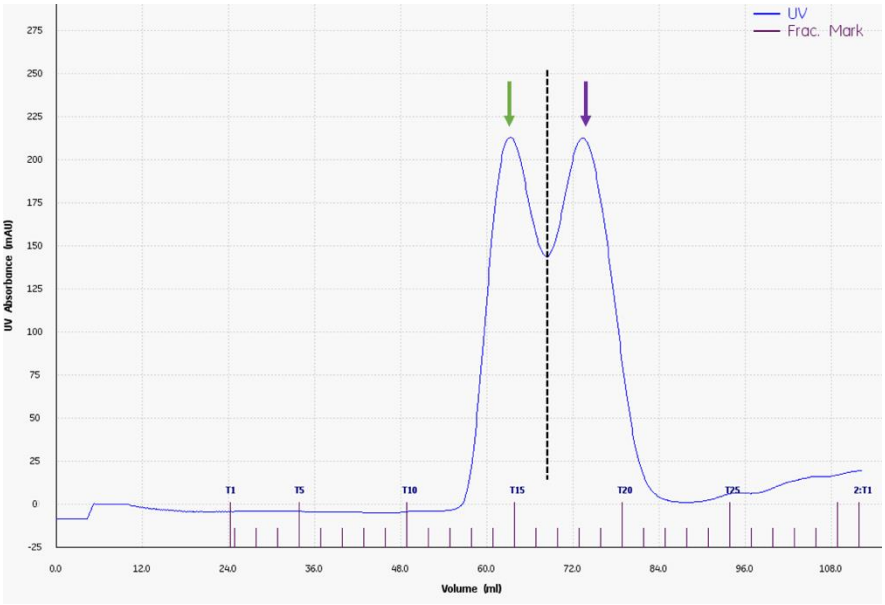



**Figure 16: SEC chromatogram from GRP8 gel filtration.** UV light line represented in blue. Fractions represented with the T and collection number. The dotted line separates different sample populations. The green arrow shows GRP8 dimer. The purple arrow shows GRP8 monomer. The x-axis shows the volume of the column run (ml). The y-axis shows the UV absorbance (mAU).

For the purification of GRP8$^{short}$ (figure 17), the same process was performed. The first purification step was the expression of GRP8$^{short}$ (Figure 17a). After A.C, GRP8$^{short}$ is observed to be present in the SDS-PAGE gel (figure 17b).

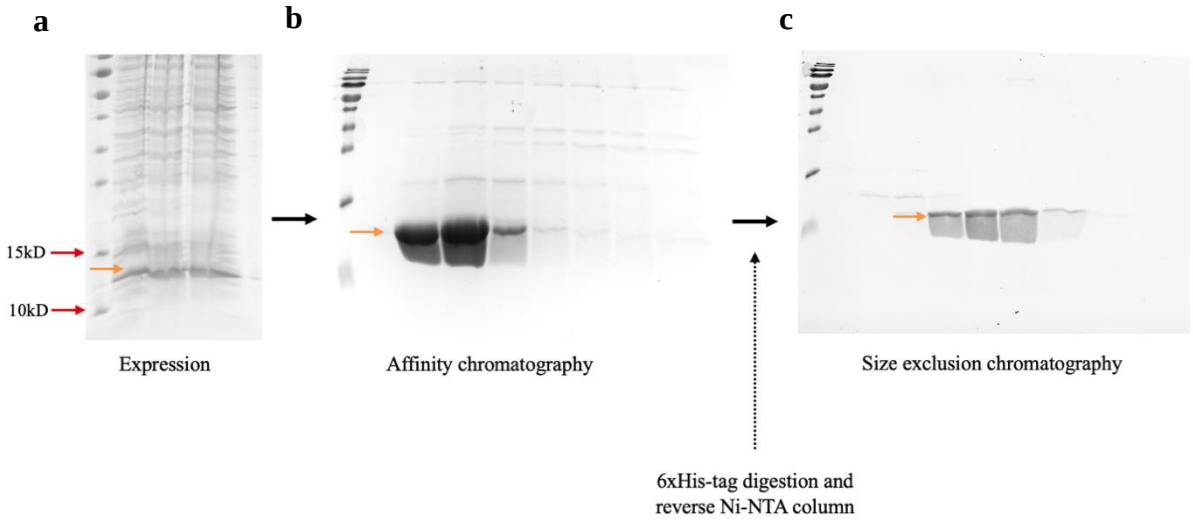



**Figure 17: Expression and purification of GRP8s.** a) SDS-PAGE gel after expression. The red arrows show the reference ladder kDa. The orange arrow shows GRP8$^{short}$ at around 12 kDa. b) SDS-PAGE gel after affinity chromatography. The orange arrow shows the GRP8 presence at around 12kDa. c) SDS-PAGE gel after size exclusion chromatography. The orange arrow shows the GRP8 presence at around 10kDa after the removal of the 6xHistag. The black arrows show the workflow direction of the protein purification. The dotted arrow shows the extra purification steps that are not shown by SDS-PAGE gels.

After the digestion of the 6His-tag with TEV protease, a reverse nickel column was performed for the separation of GRP8$^{short}$ undigested, which was discarded. Size exclusion chromatography (SEC) was performed as a last purification step. Figure 17c shows SDS-PAGE gel after SEC where GRP8$^{short}$ is observed in a pure solution state.

## 3.2 Functional characterization

After the purification of GRP8 and GPR8$^{short}$, different assays were performed to elucidate the functionality of GRP8. As mentioned previously, GRP8 is a phloem-mobile RNA-binding protein. Therefore, the first question to answer was: which RNAs is GRP8 binding? For this, GRP8 was first immobilized to the CnBr-Sepharose column. The beads were then checked through SDS-PAGE and appeared to have GPR8 present. Since the RNAs of interest are the ones present in the phloem, extraction and isolation of phloem RNA was performed. This RNA was then run through the GRP8-bound column and then eluted for the identification of these enriched RNAs. The identification was done through RNA-sequencing, comparing the input phloem RNA to the eluted fraction. After the RNA-sequencing was analyzed, a few transcripts were chosen for further testing by microscale thermophoresis (MST), to confirm that they are bound by GRP8. In addition, GRP8$^{short}$ was also tested to understand the glycine-rich relevance.

### 3.2.1 Phloem-RNA quality control

Isolation of phloem-RNA from *B. napus* plants was performed. The quality of phloem samples was verified by PCR according to Pahlow et al., 2018. RuBisCO (rubisco small unit), Thioredoxin h (TrxH) and pollen coat protein (PCP) transcripts were checked in each sample to confirm sample purity. Figure 18 shows two 3% agarose TAE gels including all the phloem samples tested for quality control. PCP transcripts that were not present in any of the samples, which indicates that there was no contamination from flowering plants. RuBisCO transcripts should not be present in phloem, however, because of the puncturing performed in the *B. napus* stems to obtain the phloem, a slight contamination could occur. Therefore, it is important to choose samples where this contamination is kept to a minimum. TrxH transcripts should be detectable in all phloem samples (Pahlow *et al.*, 2018). Samples 10, 11 and 16 were discarded for RuBisCo contamination.

For further quality check, each sample was run through the *Bioanalyzer* (figure 19). Electropherograms should show a clean RNA without DNA contamination or the presence of ribosomal RNAs. Samples 5 and 6 were discarded for showing electropherograms containing

degraded RNA (sample 6) and contamination (sample 5). Therefore, out of a total of twenty-two phloem RNA samples, five samples were discarded.

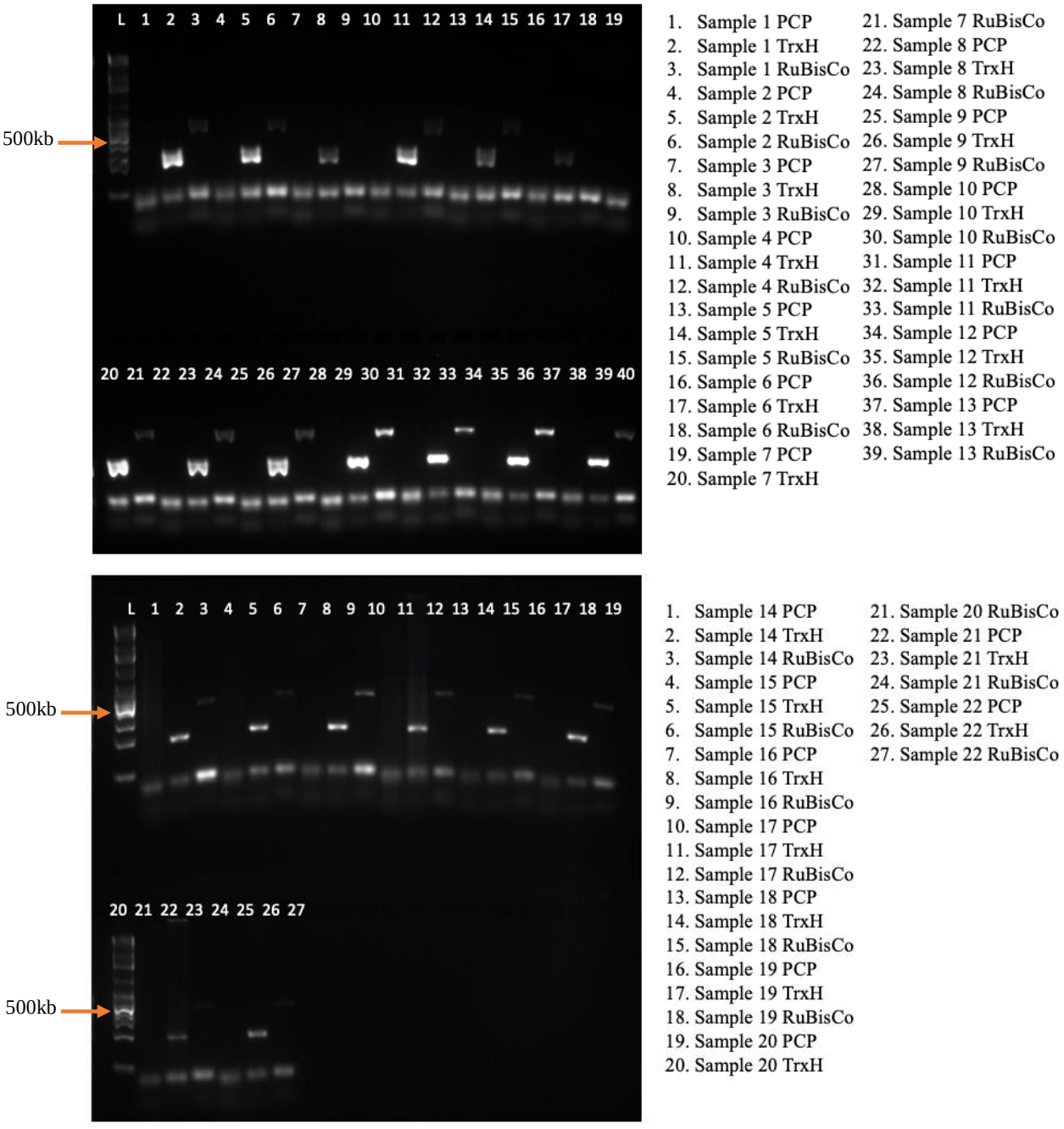



**Figure 18: 3% agarose TAE gels for phloem-RNA quality control.** Phloem-RNA samples were tested for pollen coat protein (PCP), thioredoxin h (TrxH) and rubisco small subunit (RuBisCO). Each sample was tested for the different transcripts is represented with a number indicated on the right side of the figure. The transcript presence is indicted by band in gel. The ladder is represented by the letter L and the orange arrow shows the 500 kb in ladder.

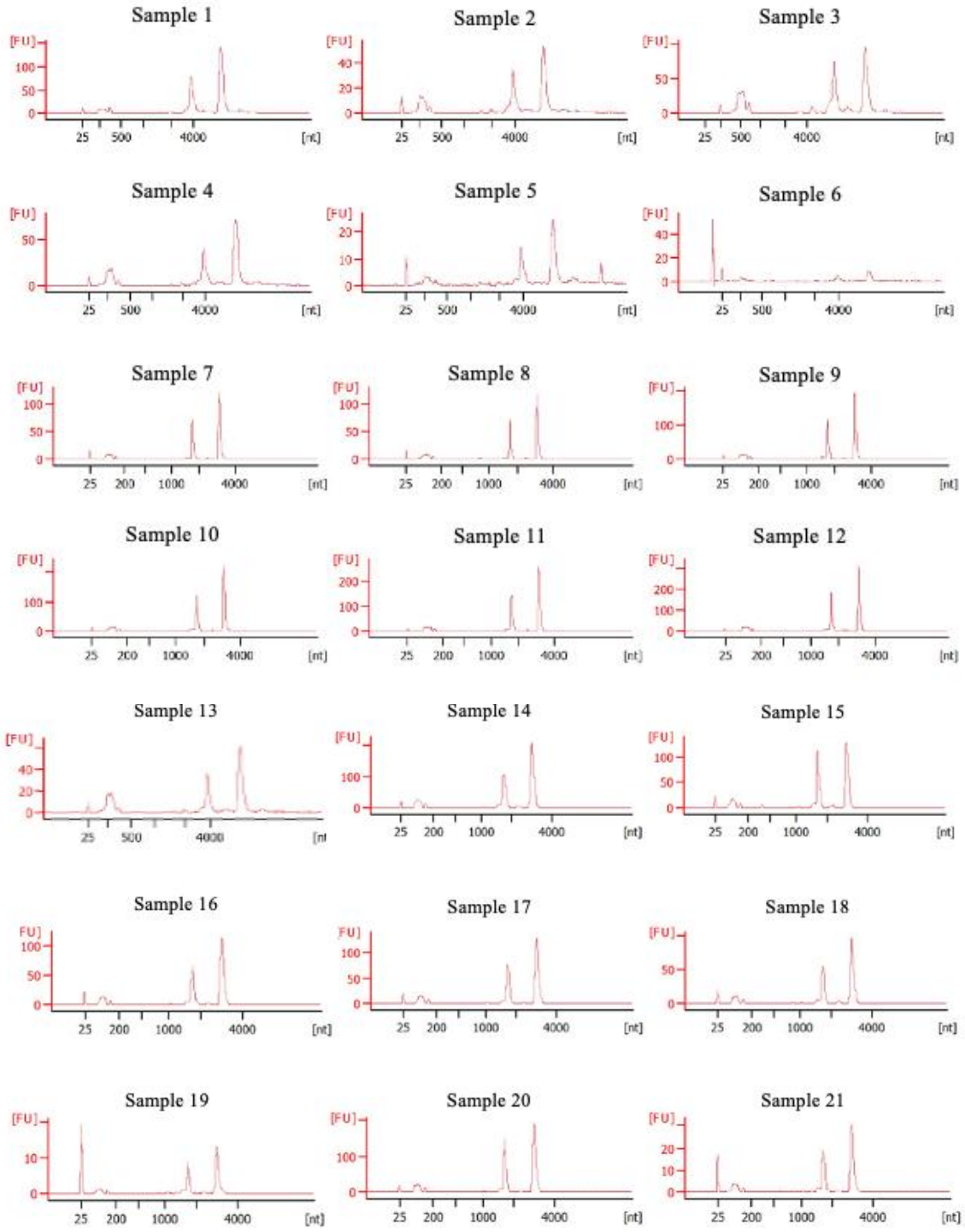



**Figure 19: *Bioanalyzer* electropherogram from phloem-RNA samples.** Samples are indicated with their corresponding number. The x-axis represents the nucleotides [nt]. The y-axis represents the fluorescence units [FU].

### 3.2.2 CnBr-Sepharose GRP8/GRP8$^{short}$ bound affinity column

After the RNA extraction and isolation from phloem sap, a CnBr-Sepharose GRP8-bound affinity column was performed. The input phloem-RNA that was used included samples previously checked for contaminants. In figure 20, it is possible to observe the different steps of this assay. When first comparing input phloem RNA with the washes, it is possible to see that after wash number 3, most of the RNAs must be gone, therefore the elution fractions should only include the RNAs bound to GRP8. The different elution fractions show interesting results versus the input phloem RNA, where the shape of the electropherogram changes. Particularly, for the RNAs between 500 nucleotides to 1000 nucleotides. This suggests that even though GRP8 shows a wide range of different RNA binding capacity, still there is selectivity in between those. In addition, input phloem RNA with the flowthrough electropherograms is very similar in shape, on the contrary it is possible to distinguish the difference in abundance of RNAs by comparing the Y axis of both. The [FU] in the flowthrough is too small to even appear in the Y axis.

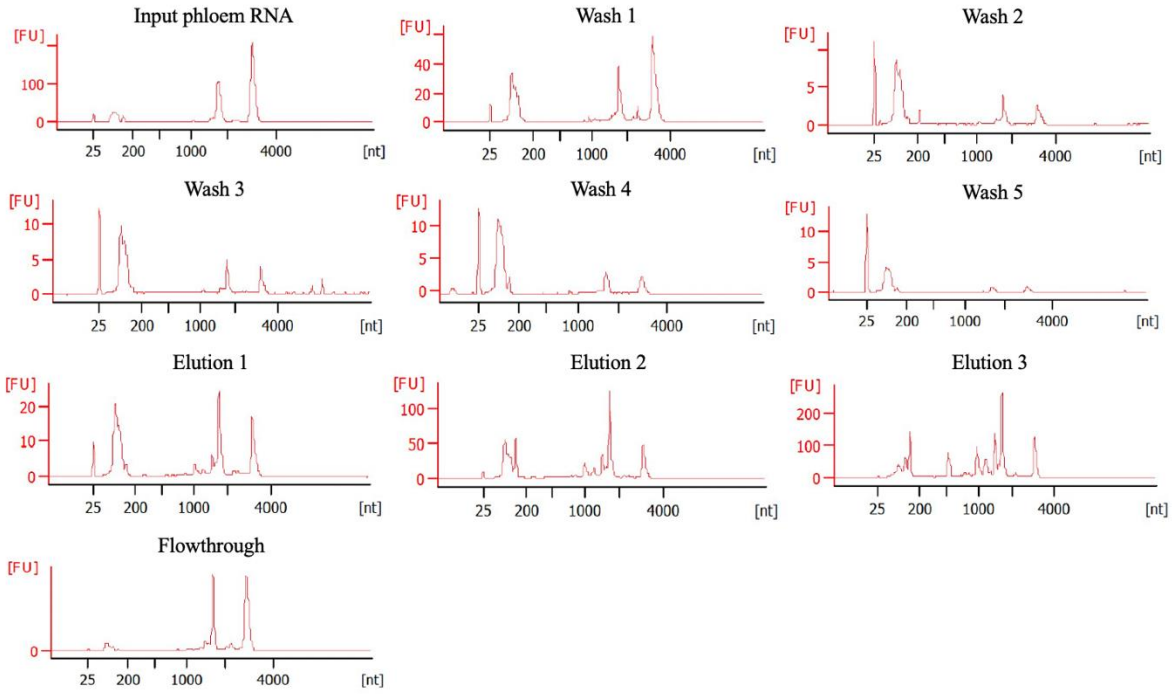



**Figure 20: *Bioanalyzer* electropherograms from the CnBr-Sepharose GRP8-bound affinity column.** Different fractions include Input phloem-RNA, wash 1-5, elution 1-3 and flowthrough. Elution 1, 250 mM sodium acetate. Elution 2, 500 mM sodium acetate. Elution 3, 2 M sodium acetate. The x-axis represents the nucleotides [nt]. The y-axis represents the fluorescence units [FU].

Moreover, a CnBr-Sepharose GRP8$^{short}$-bound affinity column was also tested to understand the relevance of the glycine-rich domain. In Figure 21, GPR8$^{short}$ elution (2M sodium acetate)

is shown in comparison to two replicates of GRP8 elution number three (2M sodium acetate). Overall, GRP8$^{short}$ possesses a different binding capacity than GRP8. Only small RNAs are found to be bound to GRP8s in comparison to GRP8. This supports the functionality of the glycine rich domain in relation to the RNA-binding ability. GRP8 binding ability to RNAs is severely compromised when the RRM is present without the glycine rich domain.

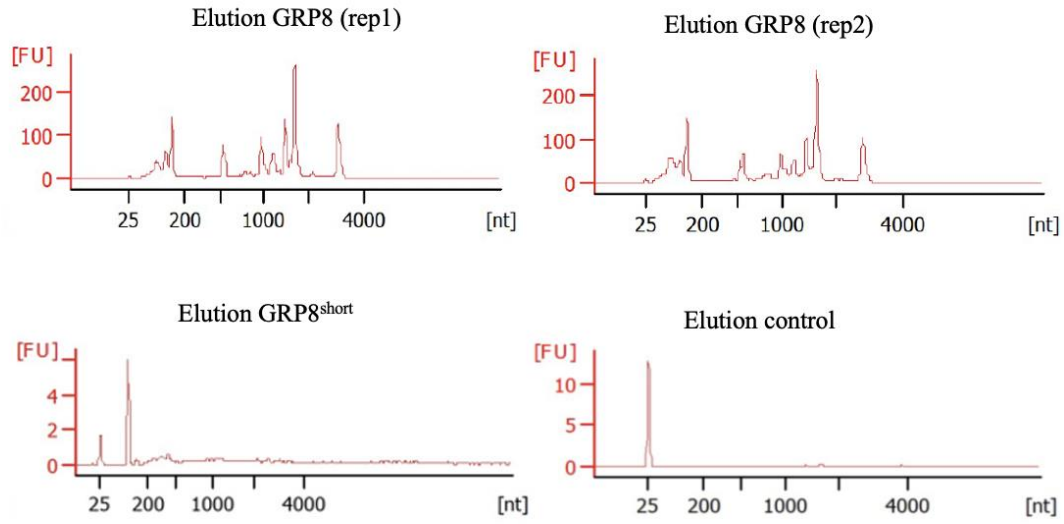



**Figure 21: *Bioanalyzer* electropherograms of elution fractions of GRP8 vs GRP8$^{short}$.** Comparison of GRP8 elution number three (2M sodium acetate) fraction replicate one and two (rep1, rep2), GRP8$^{short}$ elution fraction (2M elution acetate) and elution control fraction with CnBr-Sepharose column only with beads. The x-axis represents the nucleotides [nt]. The y-axis represents the fluorescence units [FU].

In addition, an elution control (figure 21) is shown. This only contains the CnBr-Sepharose beads without a protein bound and it phloem-RNA was run through the column as well. The elution contained no RNA bound.

### 3.2.3 RNA-sequencing analysis

After the CnBr-Sepharose GRP8-bound affinity column, an input sample and two replicates from the elution 3 were sent for sequencing. Transcriptome analysis was conducted for the identification of genes that are differentially expressed in the phloem input versus the ones bound to GRP8. Figure 22a represents a Venn diagram of the number of genes uniquely enriched in each sample with an overlapping region showing the number of genes that are co enriched in both samples. It was observed that there are over thirty-six thousand genes found in both samples.

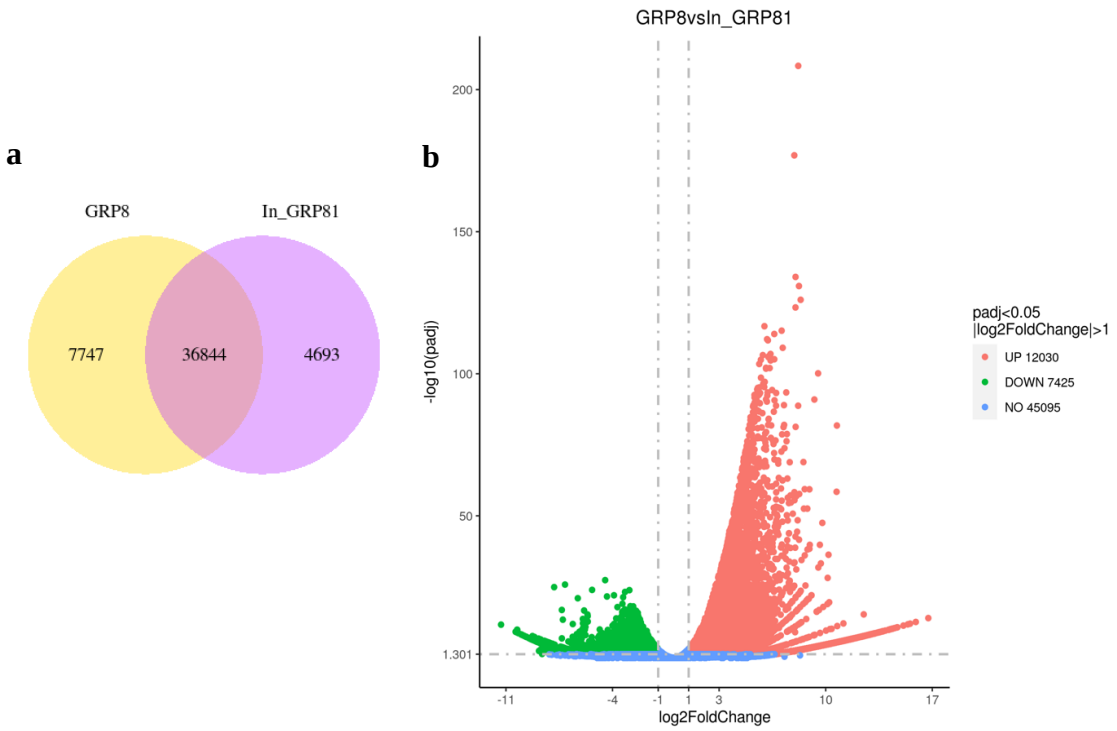



**Figure 22: Venn diagram from RNA-sequencing of GRP8 enriched transcripts versus input phloem-RNA and volcano plot of GRP8 versus input phloem-RNA differential analysis. Courtesy of *Novogene*.** a) The GRP8 enriched transcripts are represented in yellow. In_GRP81, the input phloem RNA is represented in purple. Overlapping region represents transcripts found in both samples. b) The x-axis shows the fold change in gene expression between the two samples. The y-axis shows the statistical significance of the differences. Red dots represent up-regulation genes. Green dots represent down-regulation genes. Blue dashed line indicates the threshold line for the differential gene screening criteria.

After the gene expression was quantified, statistical analysis of the enriched RNAs was performed to screen the different genes whose expression levels are significantly different in these two samples. In this case, GRP8 upregulated genes versus the phloem input genes represent the genes that are significantly enriched. The threshold was calculated with the log2fold change and the p-value adjusted. In figure 22b, volcano plot is shown to infer the overall distribution of the differentially expressed genes between GRP8 and input phloem RNA. It is visible that there are over twelve thousand significantly enriched genes in the GRP8 sample. From the genes that were enriched, it was also taken in consideration the FPKM value (fragments per kilobase of exon model per million mapped reads) (Trapnell *et al.*, 2010) sample versus input. This value is used for normalizing counts for paired-end RNA-sequencing data (Trapnell *et al.*, 2010; Zhao *et al.*, 2021). Accordingly, several transcripts of interest were selected for further functional analysis. Table 29 shows the transcripts of interest with their corresponding values and their function *in planta*. In addition, two genes that were found to not be enriched were selected for further testing.

**Table 29: Genes of interest identified through RNA-sequencing analysis.**

| Transcript of interest | Log2fold change | p-value | p-adj | FPKM (vs. input) | Function |
|---|---|---|---|---|---|
| At5g25610 (RD22) | 6.08 | 3.04E-106 | 8.07E-103 | 471.29 (vs. 6.59) | Responsive to dehydration 22. Response to abscisic acid, response to desiccation, response to salt stress. Enables mRNA binding. |
| At1g64370 (PARCL) | 9.50 | 3.09E-104 | 7.21E-101 | 537.42 (vs. 0.70) | Phloem associated RNA chaperone like. mRNA binding. |
| At5g02120 (OHP) | 5.80 | 1.10E-84 | 7.80E-82 | 196.95 (vs. 3.35) | One helix protein. Acts upstream of photosystem II assembly, response to high light intensity. Involved in circadian rhythm. |
| At4g04020 (FIB) | 6.40 | 6.13E-81 | 3.77E-78 | 140.51 (vs. 1.57) | Fibrillin precursor protein. Involved in photoinhibition, photoprotection, response to abscisic acid and response to cold. |
| At1g68520 (BBX14) | 7.17 | 1.30E-65 | 3.52E-63 | 136.22 (vs. 0.90) | B-Box type zinc finger protein with a CTT domain containing protein. Involved in regulation of DNA-templated transcription. |
| At3g01700 (AGP11) | -7.26 | 1.49E-15 | 3.52E-14 | 0.36 (vs. 49.24) | Arabinogalactan protein 11. Encodes an arabinogalactan protein that is expressed in pollen, pollen sac or pollen tube. |
| At1g02790 (PG45) | -8.63 | 1.42E-3 | 4.20E-2 | 0.12 (vs. 40.44) | Polygalacturonase 4. Involved in carbohydrate metabolic process. Encodes an exopolygalacturonase. |

### 3.2.4 Microscale thermophoresis

After the identification of the genes/transcripts of interest, these were synthesized *in vitro* for binding assays by MST.

**Table 30: Genes or transcripts of interest Kds measured through MST with GRP8 or GPR8$^{short}$.** Transcripts with UTRs represented by 3'UTR, 5'UTR or 3'+5'UTRs.

| Gene/transcript of interest | Protein | Kd (μM) |
|---|---|---|
| *miRNA164* | GRP8 | 0.079 ± 0.067 |
| | GRP8$^{short}$ | 0.165 ± 0.053 |
| *RD22* | GRP8 | 6.63 ± 1.91 |
| | GRP8$^{short}$ | No binding |
| *PARCL* | GRP8 | 35.23 ± 1.80 |
| | GRP8$^{short}$ | No binding |
| *OHP* | GRP8 | 11.13 ± 6.52 |
| | GRP8$^{short}$ | No binding |
| *FIB* | GRP8 | 18.50 ± 4.01 |
| | GRP8$^{short}$ | No binding |
| *BBX14* | GRP8 | 16.90 ± 2.79 |
| | GRP8$^{short}$ | No binding |
| *AGP11* | GRP8 | 26.46 ± 6.93 |
| | GRP8$^{short}$ | No binding |
| *PG45* | GRP8 | 21.37 ± 7.17 |
| | GRP8$^{short}$ | No binding |
| *GRP7* | GRP8 | 3.88 ± 0.62 |
| | GRP8$^{short}$ | No binding |
| *GRP8* | GRP8 | 19.83 ± 6.71 |
| | GRP8$^{short}$ | No binding |
| *miNovel2* | GRP8 | 0.296 ± 0.23 |
| | GRP8$^{short}$ | 0.415 ± 0.13 |
| *miNovel106* | GRP8 | 0.240 ± 0.08 |
| | GRP8$^{short}$ | No binding |
| *miNovel149* | GRP8 | 2.34 ± 1.75 |

| | | |
|---|---|---|
| | GRP8$^{short}$ | No binding |
| *3'UTR-GRP7* (miRNA only UTR) | GRP8 | 0.018 ± 0.01 |
| | GRP8$^{short}$ | 0.400 ± 0.22 |
| *5'UTR-OHP* | GRP8 | 3.42 ± 2.20 |
| | GRP8$^{short}$ | No binding |
| *5'+3'UTR-OHP* | GRP8 | 7.40 ± 3.07 |
| | GRP8$^{short}$ | No binding |

*miRNA164* and other miRNAs (*miNovel2, miNovel106, miNovel149*) were synthesized. This was done according to the literature, where it was described that GRP8 binds small RNAs (Yan *et al.*, 2020) and they enriched for GRP7 RNA-sequencing analysis done in our lab by *Kim Lühmann*. Furthermore, GRP7 and GRP8 from *A. thaliana* mRNAs were also produced. It is known that because of the interlocked feedback loop that GRP7 and GRP8 undergo, they both bind each other's mRNAs as well as their own mRNAs (Schöning *et al.*, 2008). In addition, it is described that the binding of GRP7/GPRP8 to its own pre-mRNA occurs to a 3'UTR region, which was also included in the list (*3'UTR-GRP7*, includes only the miRNA sequence of the UTR) (Staiger *et al.*, 2003).

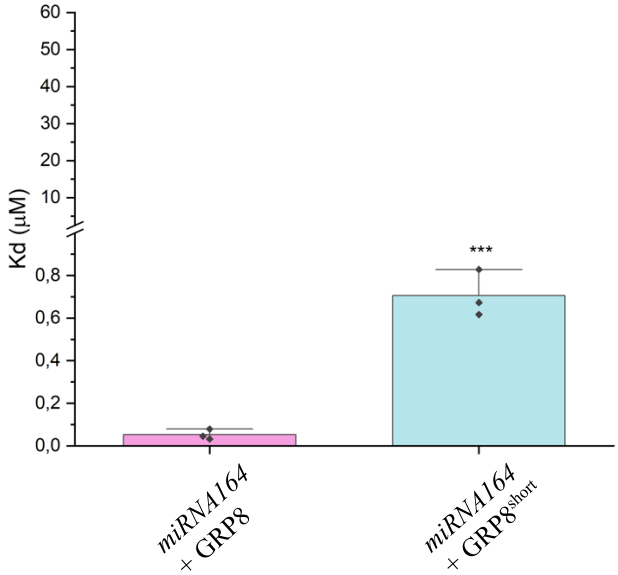



**Figure 23: Kds from *miRNA164* bound to GRP8 or GRP8$^{short}$**. The *miRNA164* bound to GRP8 is represented in pink. The *miRNA164* bound to GRP8$^{short}$ is represented in light blue. Jittered individual points represent individual Kd replicates. The colored bar represents mean Kd values. The lined bar represents error bars. (*), (**) and (***) represent significant difference between binding affinities of two samples with p-values of 0.05, 0,01 and 0.001 correspondingly. Samples were compared with *miRNA164* values. The significance was calculated through one-way ANOVA (Tukey and Bonferroni tests). The x-axis shows the sample name. The y-axis shows Kds in micromolar. The double line in between the y-axis represents a Kd break between the values 0.8 and 2.5 μM.

Moreover, to understand the functionality of the C-terminus of GRP8, GRP8$^{short}$ was also tested by MST. Table 20 shows the tested RNAs through MST with GRP8 or GRP8$^{short}$ and their binding affinities (Kds).

*miRNA164* was first chosen because it showed a very low Kd and therefore a high affinity when tested by MST. As previously explained, *miRNA164* was also of a transcript of interest because it was found to be enriched with GRP7 RNA-sequencing. Figure 23 shows GRP8 and GRP8$^{short}$ both bind *miRNA164* but with a highly significant difference between both binding affinities. Although GRP8$^{short}$ did not include the C-terminus, it still bound miRNA16 with high affinity. Because of this, it was important to test further small RNAs to see if there was a consensus in this with all miRNAs. Therefore, other miRNAs were found to be enriched in the RNA-sequencing analysis of GRP7.

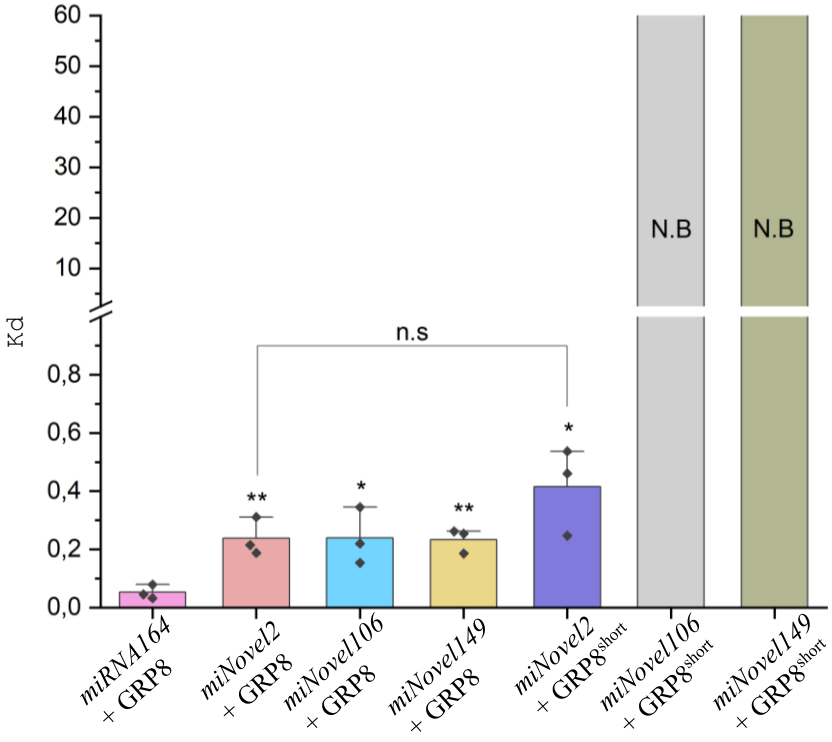



**Figure 24: Kds from other miRNAs and GRP8$^{short}$**. The *miRNA164* bound to GPR8$^{short}$ is represented in pink. The *miNovel2* bound to GRP8 is represented in salmon. The *miNovel106* bound to GRP8 is represented in light blue. The *miNovel149* bound to GRP8 is represented in mustard. The *miNovel2* bound to GRP8$^{short}$ is represented in violet. The *miNovel106* bound to GRP8$^{short}$ is represented in gray. The *miNovel149* bound to GRP8$^{short}$ is represented in olive green. Jittered individual points represent individual Kd replicates. Colored bar represents mean Kd values. The lined bar represents error bars. N.B stands for no binding. (*), (**) and (***) represent a significant difference between binding affinities of two samples with p-values of 0.05, 0,01 and 0.001, respectively. Samples were compared to *miRNA164* unless indicated otherwise. Significance was calculated through one-way ANOVA (Tukey and Bonferroni tests). The x-axis shows the sample name. The y-axis shows Kds in micromolar. The double line in between the y-axis represents a Kd break between the values 0.8 and 2.5 μM.

Figure 24 shows the other miRNAs tested with GRP8 and GRP8$^{short}$. *miRNA164* was chosen since it had the highest affinity (Table 30). When comparing the *miRNA164* versus the other miRNAs tested, a significant difference between binding affinities could be observed. Interestingly, GRP8$^{short}$ showed no binding detectable to two miRNAs tested, showing that GRP8$^{short}$ was not binding all small RNAs. Moreover, when comparing the Kds of *miNovel2* bound to GRP8 and *miNovel2* bound to GRP8$^{short}$, there was no significant difference.

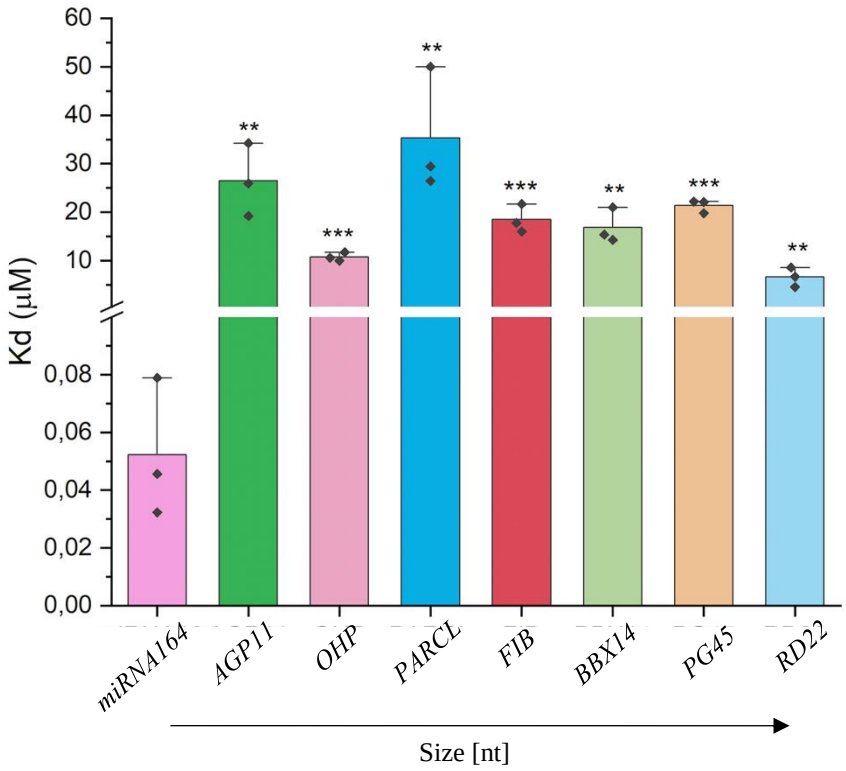



**Figure 25: Kds from different RNAs and GRP8 from RNA-sequencing vs. *miRNA164* tested with GRP8 by MST.** The *miRNA164* is represented in pink, size: 21 nucleotides. *AGP11* is represented in dark green, size: 411 nucleotides. *OHP* is represented in light pink, size: 490 nucleotides. *PARCL* is represented in dark blue, size: 537 nucleotides. *FIB* is represented in red, size: 1386 nucleotides. *BBX14* is represented in light green, size: 1419 nucleotides. *PG45* is represented in beige, size: 1545 nucleotides. *RD22* is represented in light blue, size: 2187 nucleotides. Jittered individual points represent individual Kd replicates. The colored bar represents mean Kd values. The lined bar represents error bars. (*), (**) and (***) represent a significant difference between binding affinities of two samples with p-values of 0.05, 0,01 and 0.001 correspondingly. Samples were compared with *miRNA164* values. Significant difference was calculated through one-way ANOVA (Tukey and Bonferroni tests). The x-axis shows the sample name. The y-axis shows Kds in micromolar. The double line in between the y-axis represents a Kd break between the values 0.8 and 2.5 µM.

From CnBr-Sepharose GRP8-bound affinity column and RNA-sequencing results, figure 25 shows the different RNAs that were selected and tested by MST with GRP8 including the two RNAs that were not enriched (*PG45* and *AGP11*). In this case, GRP8$^{short}$ was also tested with

these RNAs, but binding was not detected with any of them (table 30). However, GRP8 shows binding with all the RNAs tested, including the ones that were selected for not being enriched. This again shows the wide range of different RNAs that GRP8 binds. Regarding the size of these RNAs, they vary from 411 bp (*AGP11*) to 2187 bp (*RD22*) and correspond to mRNAs (supplementary table 2). Results show that GRP8 does not exclusively bind small RNAs. This is also observed in figure 25, where the RNAs are ordered from small to large nucleotide sizes and with no trend for binding the smaller mRNAs. Nevertheless, when comparing the binding affinities with miRNA164 (21 nucleotides) versus these larger RNAs, the difference ranges from significant to highly significant. It is worth mentioning that these are mRNAs including only the coding sequence. This will be discussed later, however these Kds are considerably important.

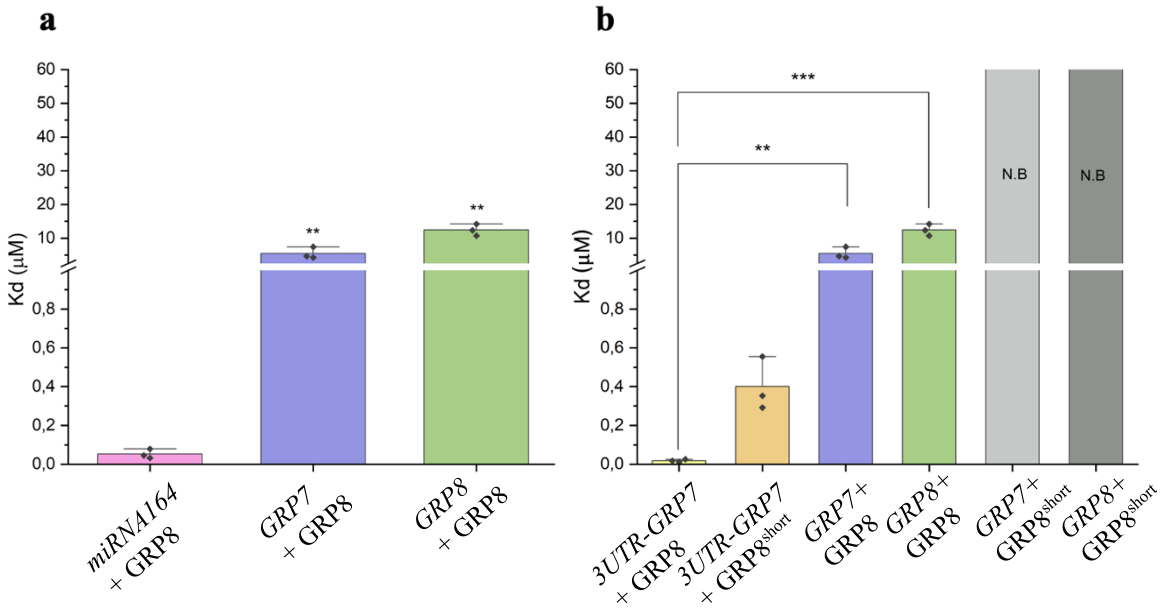



**Figure 26: Kds from GRP7 and GRP8 RNAs bound to GRP8.** a) *miRNA164*, *GRP7* and *GRP8* RNA bound to GRP8. *miRNA164* is represented in pink. *GRP7* mRNA is represented in violet. *GRP8* mRNA is represented in light green. Samples were compared with *miRNA164* values. b) comparison of *3'UTR-GPR7* bound to GRP8 or GRP8$^{short}$, *GRP7* and *GRP8* mRNAs bound to GRP8 or GRP8$^{short}$. *3'UTR-GRP7* bound to GPR8 is represented in yellow. *3'UTR-GRP7* bound to GRP8s is represented in orange. *GRP7* mRNA bound to GRP8 is represented in violet, *GRP8* mRNA bound to GPR8 is represented in light green. *GRP7* bound to GRP8$^{short}$ is represented in light gray. *GRP8* mRNA bound to GPR8$^{short}$ is represented in dark gray. Jittered individual points represent individual Kd replicates. The colored bar represents mean Kd values. The lined bar represents error bars. N.B stands for no binding. (*), (**) and (***) represent a significant difference between binding affinities of two samples with p-values of 0.05, 0,01 and 0.001 correspondingly. Samples were compared as indicated. Significant difference was calculated through one-way ANOVA (Tukey and Bonferroni tests). The x-axis shows the sample name. The y-axis shows Kds in micromolar. The double line in between the y-axis represents a Kd break between the values 0.8 and 2.5 µM.

When *GRP8* and *GRP7* mRNAs were tested, it was also observed that they were binding to GRP8 but not GRP8$^{short}$. When comparing these binding affinities to *miRNA164* there was a significant difference. *3'UTR-GRP7* was tested with both GRP8 and GRP8$^{short}$ and found to be bound with a high binding affinity in the nanomolar range (figure 26). This confirms that the section of the *3'UTR-GRP7* of the pre-mRNA of GRP7 is binding to GRP8. When comparing the binding affinities of the *3'UTR-GRP7* to mRNAs of *GRP8* and *GRP7* there was a significant difference.

As an effort to test the mRNA including their UTRs, it was intended to synthesize the RNAs from the affinity column *in vitro*. Only one gene (figure 27) was possible to test in this form including only the coding sequence (cds), including one UTR (*5'UTR-OHP*) and with both UTRs (*5'+3'UTRs-OHP*). In this case the difference in Kds when including 5'UTR was significant, suggesting that GPR8 is binding to the 5'UTR of *OHP*. On the other hand, when including both UTRs, the difference was significant. Evidently there is an effect in binding affinity when the UTRs are included, however, when the overall difference of Kds was compared, it was still in the micromolar range, and therefore not considerably different (Figure 24).

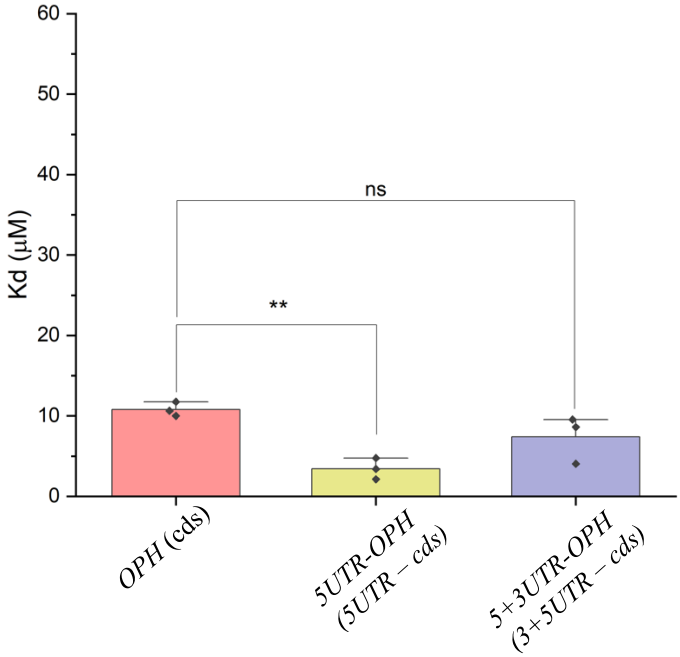



**Figure 27: Kds from *OHP* cds, *OHP* including 5'UTR and *OHP* including 3' and 5' UTRs bound to GRP8.** *OHP* cds is represented in coral pink. The *5'UTR-OHP* is represented in yellow. The *5'+3'UTR-OHP* is represented in violet. Jittered individual points represent individual Kd replicates. The colored bar represents mean Kd values. The lined bar represents error bars. (*), (**) and (***) represent a significant difference between binding affinities of two samples with p-values of 0.05, 0,01 and 0.001 correspondingly. ns stands for not significant. Samples were compared as indicated. Significant difference was calculated through one-way ANOVA (Tukey and Bonferroni tests). The x-axis shows the sample name. The y-axis shows Kds in micromolar.

A motif alignment was performed by the motif-based sequence alignment tool MEME-Suite (Bailey *et al.*, 2015). This was done to find common motifs in the mRNAs tested that could suggest a binding preference for GRP8. No significant common motifs were identified in the sequences (supplementary figure 6 and 7). In addition, it was described in literature that there is a preference for GRP7 to bind sequences rich in U and G (Staiger *et al.*, 2003). To corroborate this assumption for GRP8, the ratios of G/U of the different RNAs used in MST were calculated. Ratios can be found in the supplementary table 5. mRNAs were found to have similar ratios and around fifty percent of G/U. When comparing the ratios of miRNAs, they varied from fifty four to eighty percent of G/U. Interestingly, the ones binding to GRP8 and GRP8s where both had different ratios of G/U: *miRNA164* has a very high ratio of G (fifty percent) and very low ratio of U (eight percent), *miNovel2* has a very high content of U (fifty percent) and G (thirty one percent) and *3'GRP7-UTR* includes high ratios of U (fifty three percent) and G (twenty six percent). This would be the only difference with those not binding with GRP8$^{short}$ (*miNovel106* and *miNovel149*).

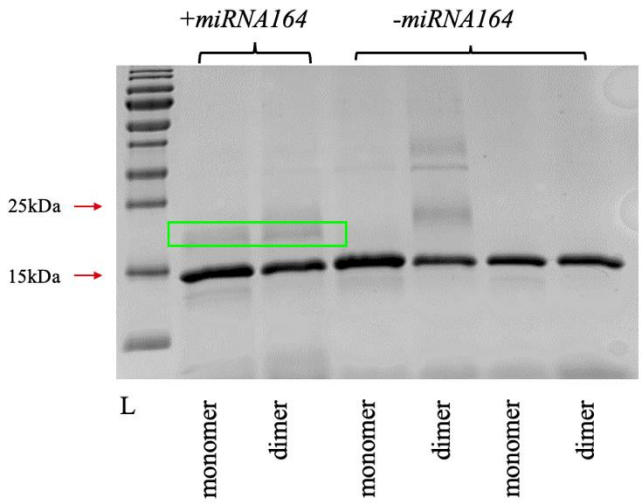
### 3.2.5 Crosslinking

Crosslinking is a method that takes advantage of the ability of a photoreactive group to trigger the formation of a covalent bond between the protein of interest and RNA upon UV-radiation (Luo and Reed, 2003). For this, *8-azido-ATP* was used to label a confirmed close interaction partner *miRNA164,* to detect the interaction between the miRNA and GRP8.

GRP8 monomer and dimer were crosslinked under UV light with *miRNA164*. The crosslinked protein samples with *miRNA164,* as well as the protein alone were loaded onto an SDS-PAGE 15% gel for the detection of the protein-RNAs interactions (figure 28a). MALDI-TOF-MS was performed to identify the peptides that were buried/missing in the analysis versus the full-length protein. The covalent binding of GRP8 and *miRNA164* was blocked by the trypsin digestion sites and these protein fragments can be identified with the MALDI-TOF based on a distinct m/z ratio. Because the trypsin sites are blocked, peaks can be compared to a non-crosslinked control sample (in this case the protein alone) and the missing peaks should indicate where the *miRNA164* was binding. It was assumed that the buried peptides meant that they were engaged in the interaction with the RNA. Figure 28b shows the amino acids in red that were identified to be interacting with the *miRNA164*. GRP8 monomer and dimer were both combined into a consensus between both measurements. As it is shown, not only the RRM

domain but the glycine-rich domain also is involved in RNA-binding when GRP8 is interacting with the *miRNA164*.

**a**




**b**

GRP8 monomer

| | | | | | |
|---|---|---|---|---|---|
| 1 | MSEVEYRCFV | GGLAWATNDE | DLQRTFSQFG | DVIDSKIIND | RESGRSRGFG |
| 51 | FVTFKDEKAM | RDAIEEMNGK | ELDGRVITVN | EAQSRGSGGG | GGGRGGSGGG |
| 101 | YRSGGGGGYS | GGGGGGYSGG | GGGGYERRSG | GYGSGGGGGG | RGYGGGGRRE |
| 151 | GGGYGGGDGG | SYGGGGGGW | | | |

GRP8 dimer

| | | | | | |
|---|---|---|---|---|---|
| 1 | MSEVEYRCFV | GGLAWATNDE | DLQRTFSQFG | DVIDSKIIND | RESGRSRGFG |
| 51 | FVTFKDEKAM | RDAIEEMNGK | ELDGRVITVN | EAQSRGSGGG | GGGRGGSGGG |
| 101 | YRSGGGGGYS | GGGGGGYSGG | GGGGYERRSG | GYGSGGGGGG | RGYGGGGRRE |
| 151 | GGGYGGGDGG | SYGGGGGGW | | | |

Consensus

| | | | | | |
|---|---|---|---|---|---|
| 1 | MSEVEYRCFV | GGLAWATNDE | DLQRTFSQFG | DVIDSKIIND | RESGRSRGFG |
| 51 | FVTFKDEKAM | RDAIEEMNGK | ELDGRVITVN | EAQSRGSGGG | GGGRGGSGGG |
| 101 | YRSGGGGGYS | GGGGGGYSGG | GGGGYERRSG | GYGSGGGGGG | RGYGGGGRRE |
| 151 | GGGYGGGDGG | SYGGGGGGW | | | |

**Figure 28: crosslinking of *miRNA164* and GRP8 monomer, dimer, and their consensus.** a) 15% SDS-PAGE gel with GRP8 monomer and dimer crosslinked with *miRNA164* and two negative controls including only GRP8 monomer and dimer crosslinked without *miRNA164*. The green rectangle represents bands with the GRP8 monomer/dimer and *miRNA164* complex. b) Amino acids are represented by FASTA sequence format. Order number in the amino acid sequence is on the left. The amino acids marked in red represent the amino acids that are involved in the RNA-binding interaction. The amino acids marked in black represent the amino acids that are not involved in RNA-binding interaction. The blue highlighted region represents the RNA-recognition motif. The green highlighted region represents the glycine-rich domain.

## 3.3 Structural characterization

Different approaches were used and combined to understand and characterize the structure of GRP8, including not only experimental work but also databases to reconstruct a 3D model of the protein.

### 3.3.1 Dynamic light scattering

This technique was used for the confirmation of the presence of GRP8 monomer, dimer and GRP8$^{short}$ in solution for further characterization analysis. Dynamic light scattering (DLS) is a low-resolution technique that through mathematical calculations estimates an hydrodynamic size average of the particles present in solution (Schmidt, 2010). DLS can be used for several types of applications regarding protein characterization, particularly in this case it was used to determine the protein polydispersity, free of aggregates in a pure solution without other particles aside from the protein. It is important to state that the hydrodynamic radius (Rg) is an approximation value assuming the particle is a globular one, therefore if the protein shape is elongated, the Rg will then be larger.

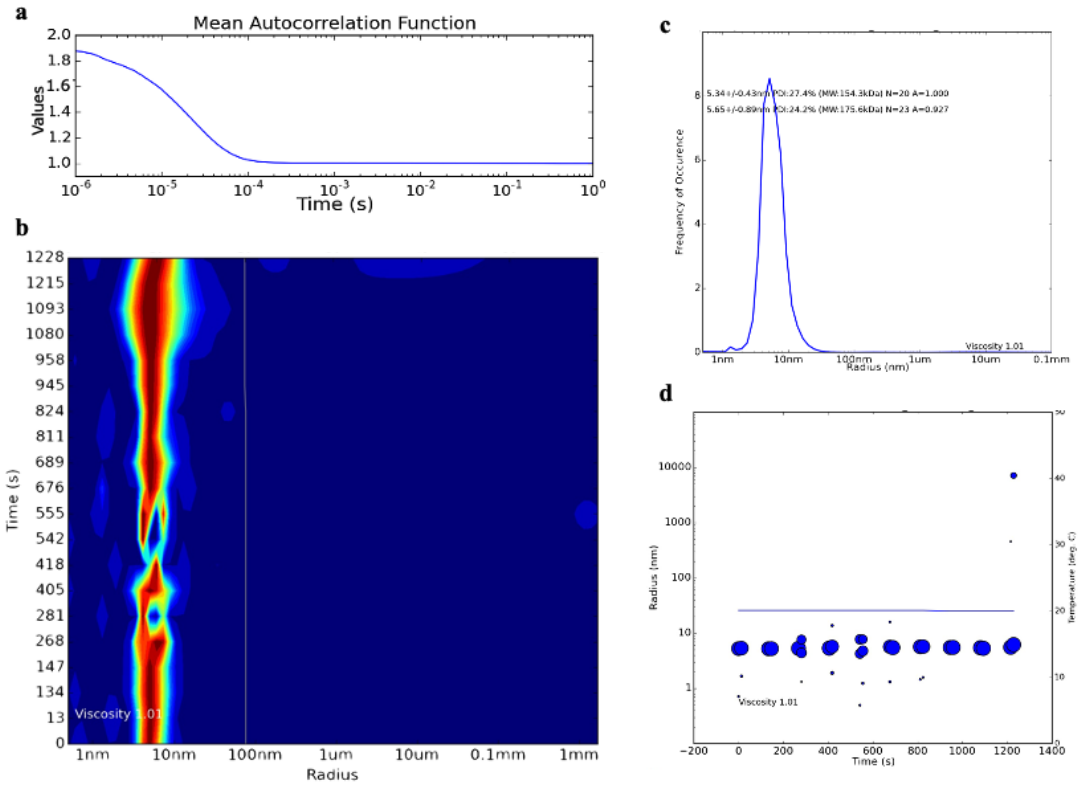



**Figure 29: DLS of GRP8 monomer.** a) Autocorrelation function. The x-axis time. The y-axis function value. b) DLS radius distribution heat map. Red color represents the distribution of the GRP8 monomer. c) Histogram of the GRP8 monomer. The x-axis radius (nm). The y-axis frequency of occurrence. d) Radius plot. The x-axis time (s). The y-axis radius (nm). DLS by *DLS SpectroSize 300* system (Xtal Concepts)

Considering that GRP8 is a highly flexible protein with a large glycine-rich region, it was assumed that the protein Rg would then show a larger value.

In figures 29, 30 and 31, show GRP8 monomer, GRP8 dimer and GRP8$^{short}$ correspondingly. Figures 29a, 30a, 31a it is possible to see the autocorrelation function which indicates intensity of the fluctuating light scattering signal over time. This function is used for further particle size calculations. Although there is a lot of information in these plots, the focus was to see: first in figures 29b, 30b and 31b that the heatmap shows a distinctive population of particles at the same size, which can then be confirmed by the population histogram in figures 29c, 30c and 31c with one clear peak. Finally, to confirm the uniformity of the radius, particles were measured in different time frames. Therefore, it can be concluded that the particles present in each sample correspond to GRP8 monomer, GRP8 dimer and GRP8$^{short}$ samples in a stable and highly pure state in solution. In the matter of sizes, the difference is observed especially when it is compared to GRP8 full length and GRP8$^{short}$. There is a difference between GRP8 monomer and dimer, however, this will be further described in the discussion section.

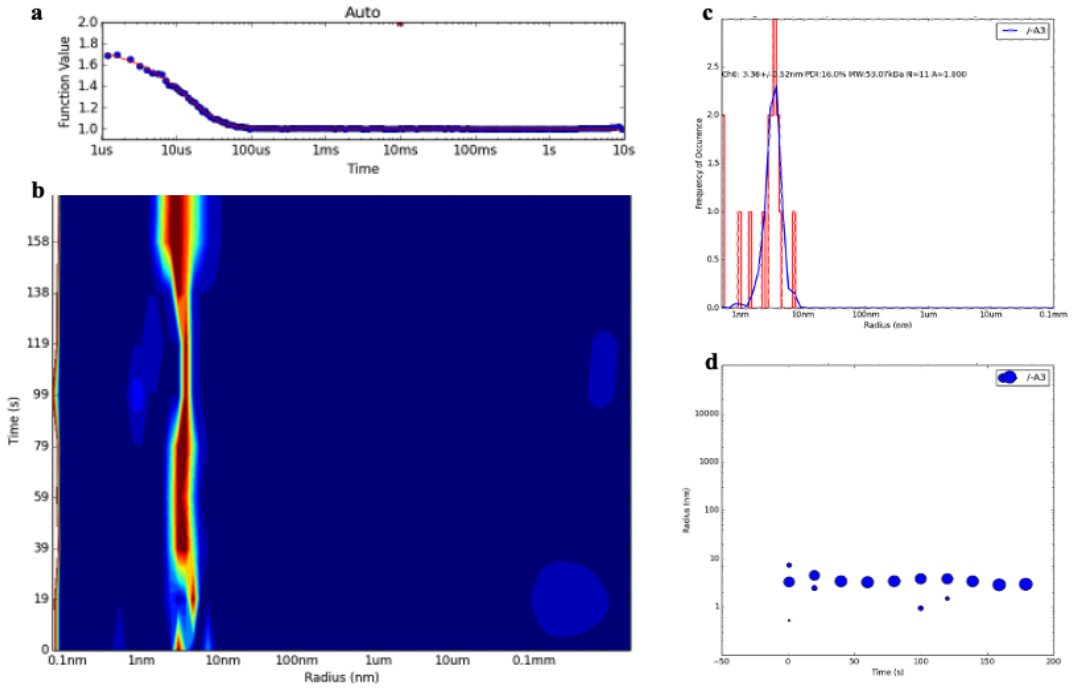



**Figure 30: DLS of GRP8 dimer.** a) Autocorrelation function. The x-axis time. The y-axis function value b) DLS radius distribution heat map. Red color represents the distribution of the GRP8 dimer. c) Histogram of the GRP8 dimer. The x-axis radius (nm). The y-axis frequency of occurrence. d) Radius plot. The x-axis time (s). The y-axis radius (nm). DLS by *DLS SpectroSize 601* system (Xtal Concepts).

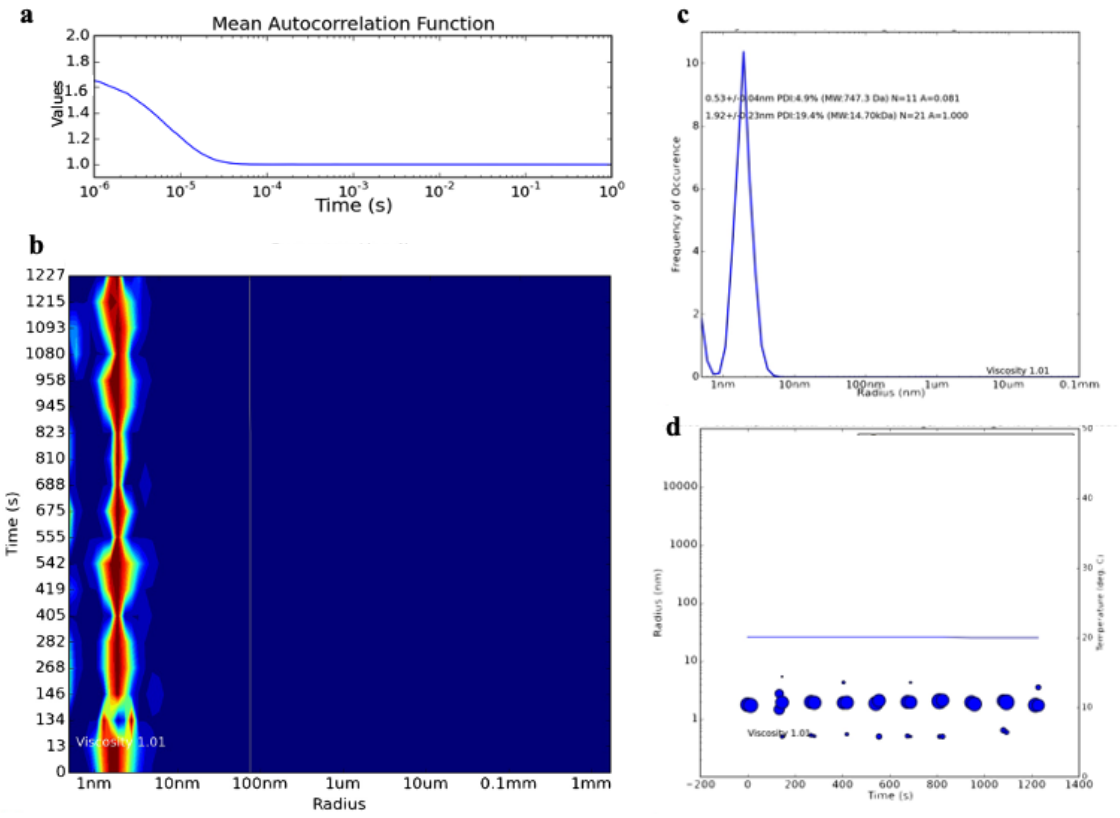



**Figure 31: DLS of GRP8$^{short}$.** a) Autocorrelation function. The x-axis time. The y-axis function value. b) DLS radius distribution heat map. Red color represents the distribution of the GRP8$^{short}$. c) Histogram of the GRP8 monomer. The x-axis radius (nm). The y-axis frequency of occurrence. d) Radius plot. The x-axis time (s). The y-axis radius (nm). DLS by *DLS SpectroSize 601* system (Xtal Concepts).

## 3.3.2 Crystallographic trials

Crystallography was attempted as an effort to elucidate the structure of GRP8 in a high resolution. Different buffer conditions were tested for initiating crystallographic trials. As previously described, DLS has different applications and in this section was used to find the best buffer match to initiate crystallographic trials. Initially, the SEC buffer where GRP8 monomer was suspended was already showing promising results, with a clear monomeric distribution. In addition, 1 mM TCEP was added for better stabilization. Two sets of crystallographic screening kits (around 200 conditions) were used and left at 20°C and incubated for two weeks, unfortunately no crystals or nucleation was formed. This could be explained because of the C-terminus glycine rich part, which is very flexible and obstructs the condensation of the protein into a solid state. Figure 32 shows DLS results with the new buffer containing GRP8 monomer. Results show a clear particle population corresponding to a stable

GRP8 monomer resuspended in a pure buffer. Although the crystallographic trials were unsuccessful, optimizations and further characterization methods will be later discussed.

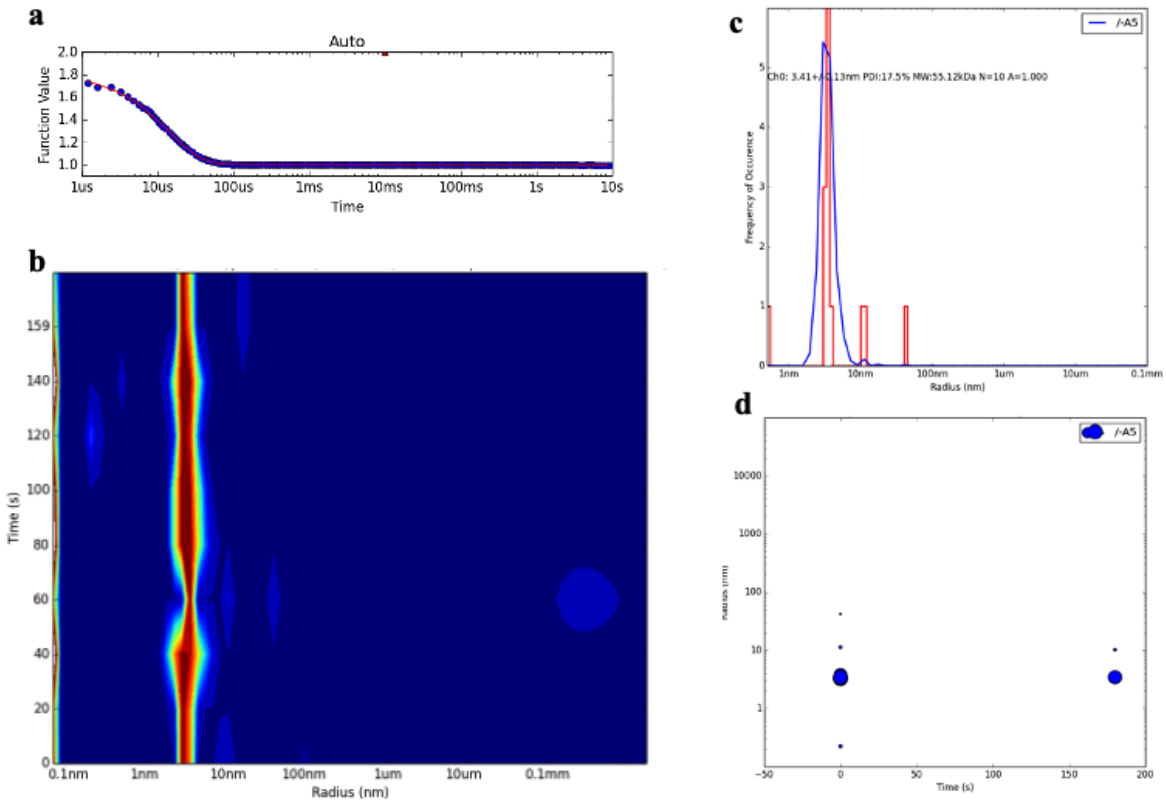



**Figure 32: DLS of GRP8 monomer in 25 mM HEPES 100 mM KCl 1 mM TCEP.** a) Autocorrelation function. The x-axis time. The y-axis function value. b) DLS radius distribution heat map. Red color represents the distribution of the GRP8 monomer. c) Histogram of the GRP8 monomer. The x-axis radius (nm). The y-axis frequency of occurrence. d) Radius plot. x-axis time (s). The y-axis radius (nm). DLS by *DLS SpectroSize 600* system (Xtal Concepts)

### 3.3.3 Thioflavin-T assay

ThT is known to bind amyloid-like fibrils and when excited, it produces a fluorescence signal upon the binding (Biancalana & Koide, 2010; Xue et al., 2017). ThT has been also used for detecting LLPS condensation droplets for proteins containing prion-like domains (PrLD) in their low complexity regions (Fernandez-Gomez *et al.*, 2019; Kaur *et al.*, 2019; Ghosh and Zhou, 2020; Pantoja-Uceda *et al.*, 2021). Thioflavin-T (ThT) assay was performed for the determination of GRP8 and its PLD behavior in forming LLPS (liquid-liquid phase separation) droplets. This was done by testing only GRP8 monomer, GRP8 dimer and GRP8$^{short}$ samples and comparing them in addition to miRNA164. The aim was to understand if there was an indication of GRP8 forming condensation droplets, engaging in LLPS when RNA was added.

For better resolution of the experiment, buffer subtraction was integrated in each of the plots according to the samples.

```
1   MSEVEYRCFVGGLAWATNDEDLQRTFSQFGDVIDSKIINDRESGRSRGFGFVTFKDEK
58  AMRDAIEEMNGKELDGRVITVNEAQSRGSGGGGGGRGGSGGGYRSGGGGGYSGGGGG
172 GYSGGGGGGYERRSGGYGSGGGGGGRGYGGGGRREGGGYGGGDGGSYGGGGGGW

1   MSEVEYRCFVGGLAWATNDEDLQRTFSQFGDVIDSKIINDRESGRSRGFGFVTFKDEK
58  AMRDAIEEMNGKELDGRVITVNEAQSRGSGGGGGGRGGSGGGYRSGGGGGYSGGGGG
172 GYSGGGGGGYERRSGGYGSGGGGGGRGYGGGGRREGGGYGGGDGGSYGGGGGGW
```

**Figure 33: Amino acid sequence of GRP8 in FASTA format.** First sequence shows RGG motifs (Highlighted in purple). Second sequence shows a consensus with the RGG motif highlighted in purple and aggregation motif [G/S]Y[G/S] is highlighted in yellow. Numbers indicate amino acid count.

GRP8 includes specific motifs with glycine or serine followed by tyrosine and then again a glycine or serine [G/S]Y[G/S], these are known to engage in dynamic liquid-liquid separation and formation of hydrogels (Han *et al.*, 2012; Kato *et al.*, 2012). Figure 33 shows the amino acid composition of GRP8 with these motifs highlighted in yellow.

Apart from this motif, the positioning of an arginine [R] next to a glycine [G] was found to contribute to RNA-binding (Järvelin *et al.*, 2016). The glycine repeats surrounding the arginine residues (RRG motif) have an important role in orienting these positively charged residues for the RNA-protein interactions (Hentze *et al.*, 2018). Figure 33 shows the RGG motifs present in the glycine-rich domain and it shows the consensus with the [G/S]Y[G/S] motif and RGG motif. In addition, PrLD prediction by *PLAAC* (Lancaster *et al.*, 2014) was done with the GRP8 sequence. Figure 34a shows that GRP8 is predicted to have a PrLD present at the glycine-rich C-terminus. Figure 34b shows the PrLD in comparison to the disordered prediction (*FoldIndex*) and the prediction where the PrLD is present in the amino acid sequence of GRP8.

Results from the thioflavin assay showed that when compared the GRP8 monomer, GRP8 dimer and GRP8$^{short}$, only GRP8 monomer shows significant aggregation. As expected, GRP8$^{short}$ shows no aggregation whatsoever, and GRP8 dimer shows very low fluorescence, indicating some aggregation (Figure 35a). When the *miRNA164* was added, both GRP8 monomer and GRP8 dimer showed that the overall interaction with RNA and protein induced a higher fluorescence and therefore it could be assumed that more condensation was present (Figure 35b and c). The difference is even more evident when comparing the fluorescence of GRP8 dimer alone with GRP8 dimer with *miRNA164*.

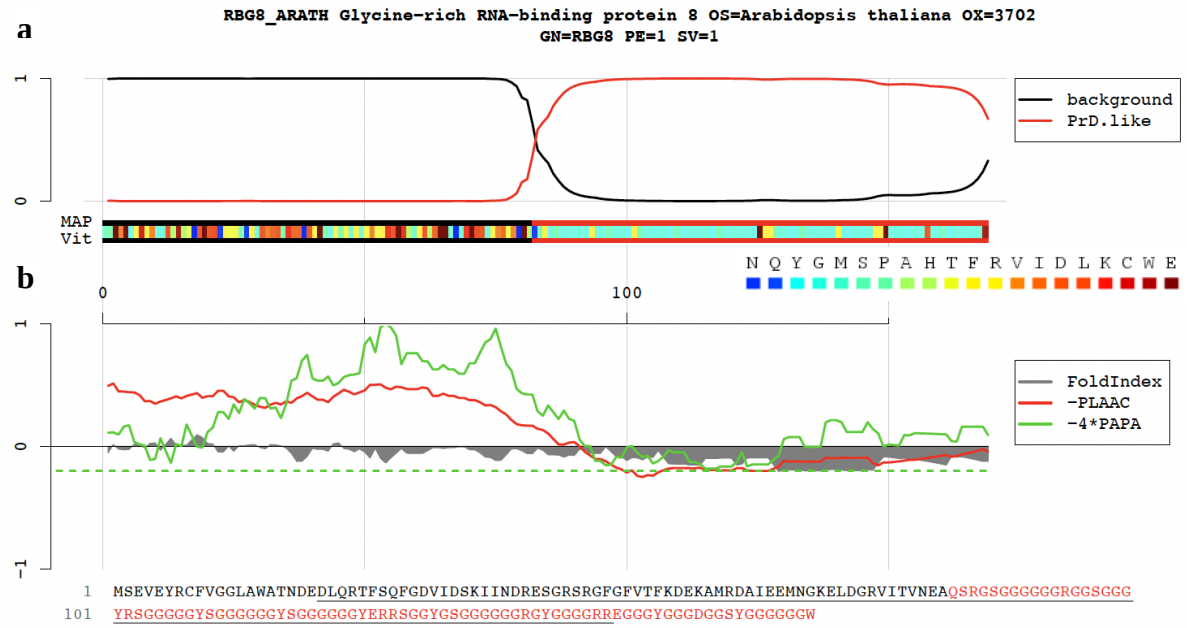



**Figure 34: Prion-like domain prediction of GRP8 by *PLAAC*.** a) Prion-like domain prediction shows that there is a prion-like domain present at the glycine-rich C-terminus. The map represents each amino acid color coded. b) Prion-like domain prediction in comparison with the disordered prediction (*FoldIndex*). c) GRP8 FASTA sequence. The letters marked in red represent the sequence where the prion-like domain is predicted.

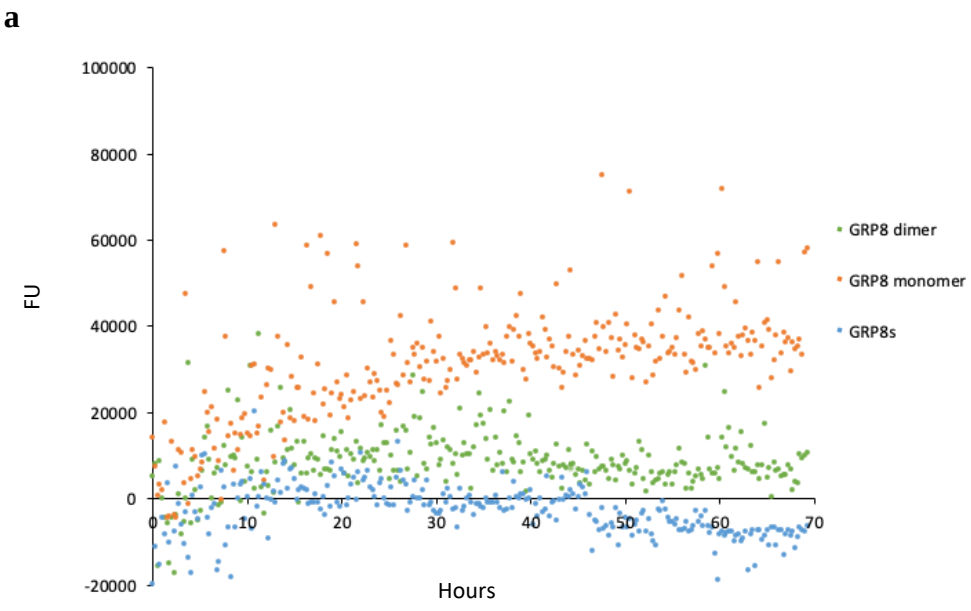



**Figure 35a: Thioflavin assay comparison with GRP8 monomer, dimer and GRP8$^{short}$ and GRP8 monomer/dimer bound to *miRNA164*.** a) Condensation comparison between GRP8 monomer, dimer and GRP8$^{short}$. GRP8 dimer is represented in green. GRP8 monomer is represented in orange, GRP8$^{short}$ is represented in blue.

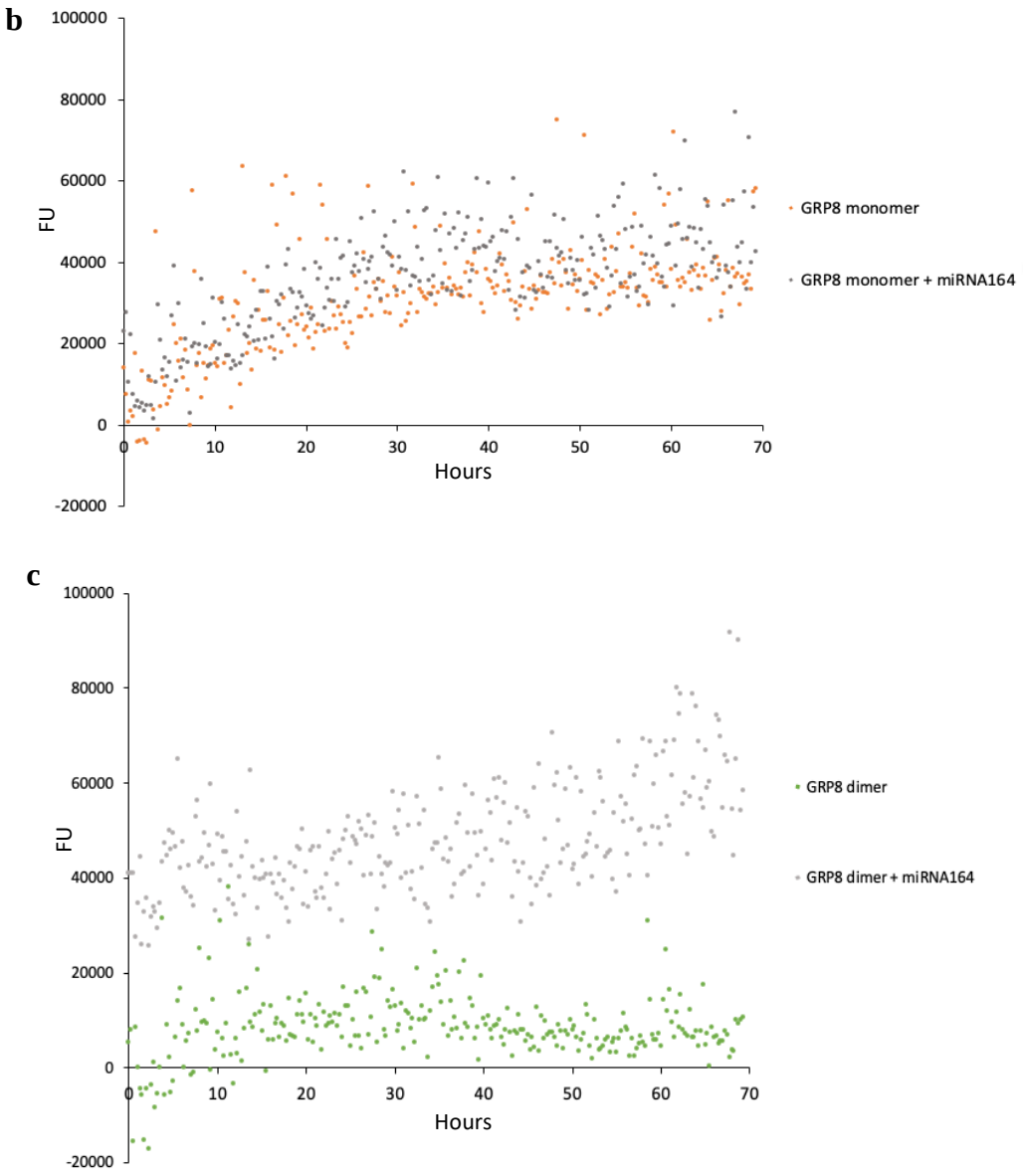



**Figure 35b and c: Thioflavin assay comparison with GRP8 monomer, dimer and GRP8$^{short}$ and GRP8 monomer/dimer bound to *miRNA164*.** b) Condensation comparison between GRP8 monomer and GRP8 monomer bound to *miRNA164*. GRP8 monomer is represented in orange. GRP8 bound to *miRNA164* is represented in gray. c) Condensation comparison between GRP8 dimer and GRP8 dimer bound to *miRNA164*. GRP8 dimer is represented in green. GRP8 bound to *miRNA164* is represented in gray. The x-axis represents time in hours. The y-axis represents fluorescence units.

## 3.3.4 Liquid-liquid phase separation assay

As previous results showed, there was condensation formed with GRP8 monomer and GRP8 dimer when the protein was incubated with RNA and no condensation with GRP8$^{short}$. For this reason, further characterization analysis was performed to support this liquid-liquid phase separation in condensation droplets.

Poly(ethylene glycol) (PEG) can be used to induce LLPS and protein precipitation (Psimadas *et al.*, 2012; Wang *et al.*, 2014). The advantage of using PEG is that it does not significantly alter the native interaction between protein molecules (Curtis *et al.*, 2002).

Figure 36 shows 1 µM, 50 µM of GRP8$^{short}$ and PEG3350 10% (w/v) incubated with RNA at room temperature (figure 36a) and 4°C (figure 36b). The extra 4°C incubation was added because LLPS phase separation is often modulated by temperature, which can affect inter and intramolecular interactions that drive LLPS (Urry *et al.*, 1992; Sanulli and Narlikar, 2021), which will be discussed further in chapter 4.4. The last step included centrifugation, which relies on the principle that upon LLPS formation, light and dense phases can be separated by simple centrifugation (Sanulli and Narlikar, 2021) and the dense phase is sedimented (Taratuta *et al.*, 1990). As expected, nothing was detected in any of the GRP8$^{short}$ samples analyzed. These included concentration between 1 µM – 50 µM GRP8$^{short}$ incubated with RNA and PEG3350 10% (w/v) at 4°C and room temperature. After, samples were all centrifuged, but no droplets were detected.

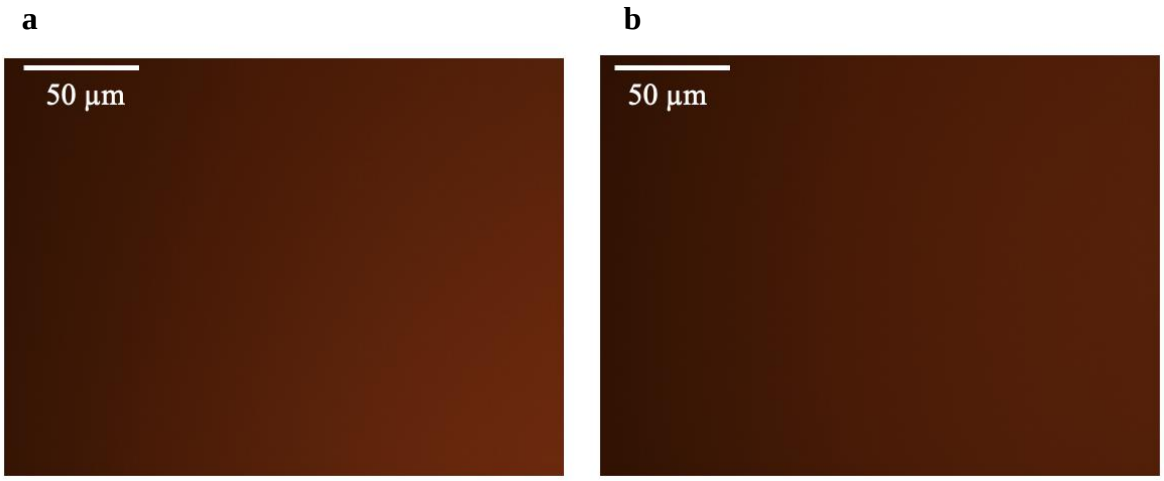



**Figure 36: Macroscopic capture of GRP8$^{short}$ and RNA induced liquid-phase separation.** a) 50 µM, PEG3350 10% (w/v) GRP8$^{short}$ and RNA at room temperature. b) 50 µM and PEG3350 10% (w/v) GRP8$^{short}$ and RNA incubated at 4°C and centrifuged at 21000 x g for 5 minutes. The white bar indicates the macroscope scale.

Figure 37 shows GRP8 dimer with the same steps and conditions as GRP8$^{short}$. No condensation was detected until the concentrations of GRP8 dimer increased to 25 µM (figure 37b), which are shown with gray arrows. All the samples were then incubated at 4°C for a short time. When the GRP8 dimer was and then analyzed under the macroscope, it was observed with condensate droplets, only when the concentration of GRP8 dimer was 25 µM. Figure 37c condensation droplets are shown with gray arrows. These droplets were then centrifuged at 4°C to pull them together and in figure 37d it is possible to observe them in a brighter resolution.

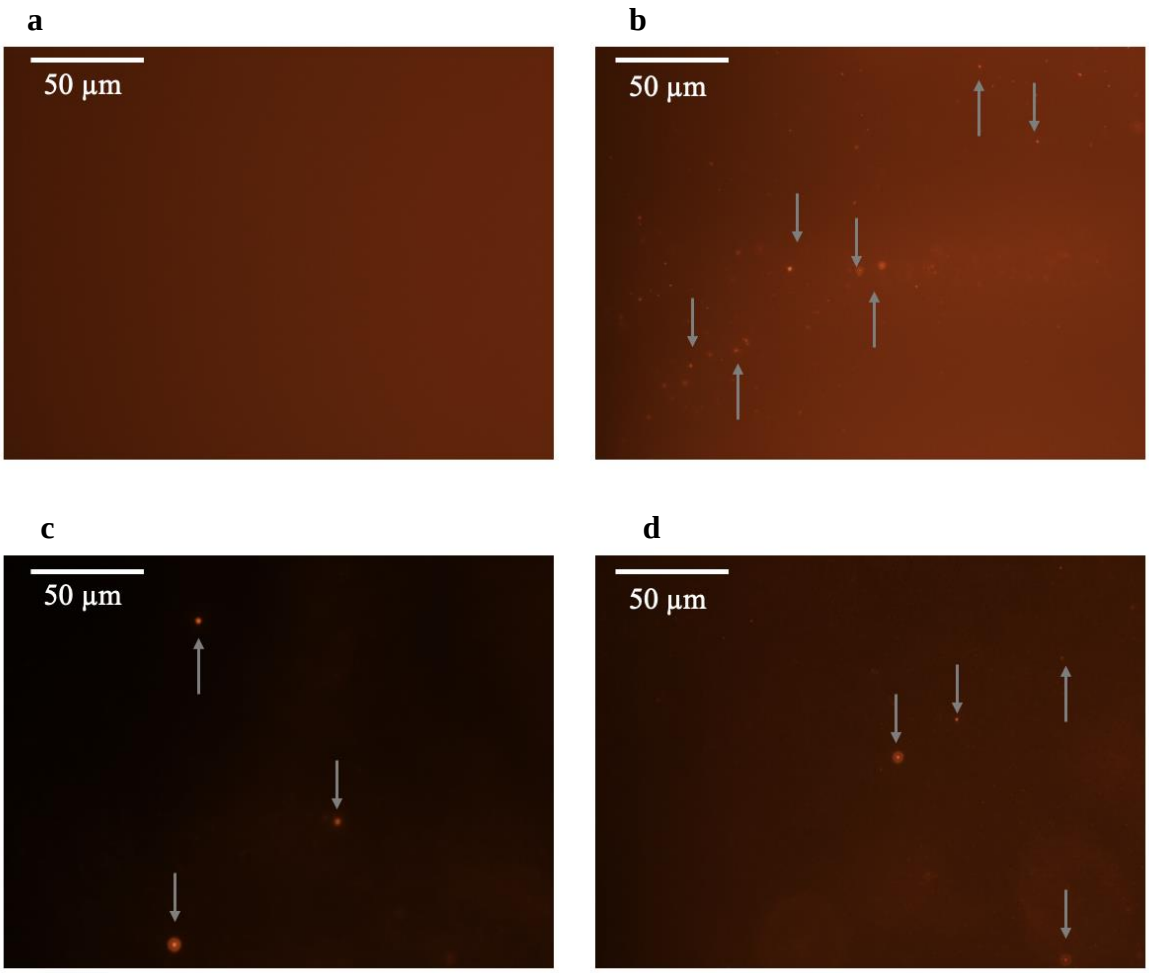



**Figure 37: Macroscopic capture of GRP8 dimer and RNA induced liquid-phase separation.** a) 1 µM, PEG3350 10% (w/v) GRP8 dimer and RNA at RT. b). 25 µM, PEG3350 10% (w/v) GRP8 dimer and RNA at room temperature c) 25µM, PEG3350 10% (w/v) GRP8 dimer and RNA incubated at 4°C. d) 25 µM, PEG3350 10% (w/v) GRP8 dimer and RNA incubated at 4°C and centrifuged at 21000 x g for 5 minutes. The gray arrows point at the condensation droplets. The white bar indicates macroscope scale.

When GRP8 monomer was analyzed, it was interesting to see that this concentration dependency threshold was lower. Figure 38a shows GRP8 monomer 1 µM at room temperature where there were no condensates formed. When the concentration of GRP8 monomer increases to 10 µM at room temperature, it is possible to observe that there is a large amount of condensation droplets formed (figure 38b). When the concentration of GRP8 monomer was increased to 50 µM at room temperature, it was possible to see a large amount of condensation droplets (figure 38c). In addition, GRP8 monomer at concentrations of 10 µM to 50 µM were incubated at 4°C for a short time. Figure 38d shows that GRP8 monomer 50 µM at 4°C, but there was not a large difference in the droplet amount between these concentrations at 4°C and RT. The last step was to centrifuge these condensation droplets to pull them together and in

figure 38e and 38f it was possible to observe that these large number of droplets decreases but the size of them are larger (~20nm).

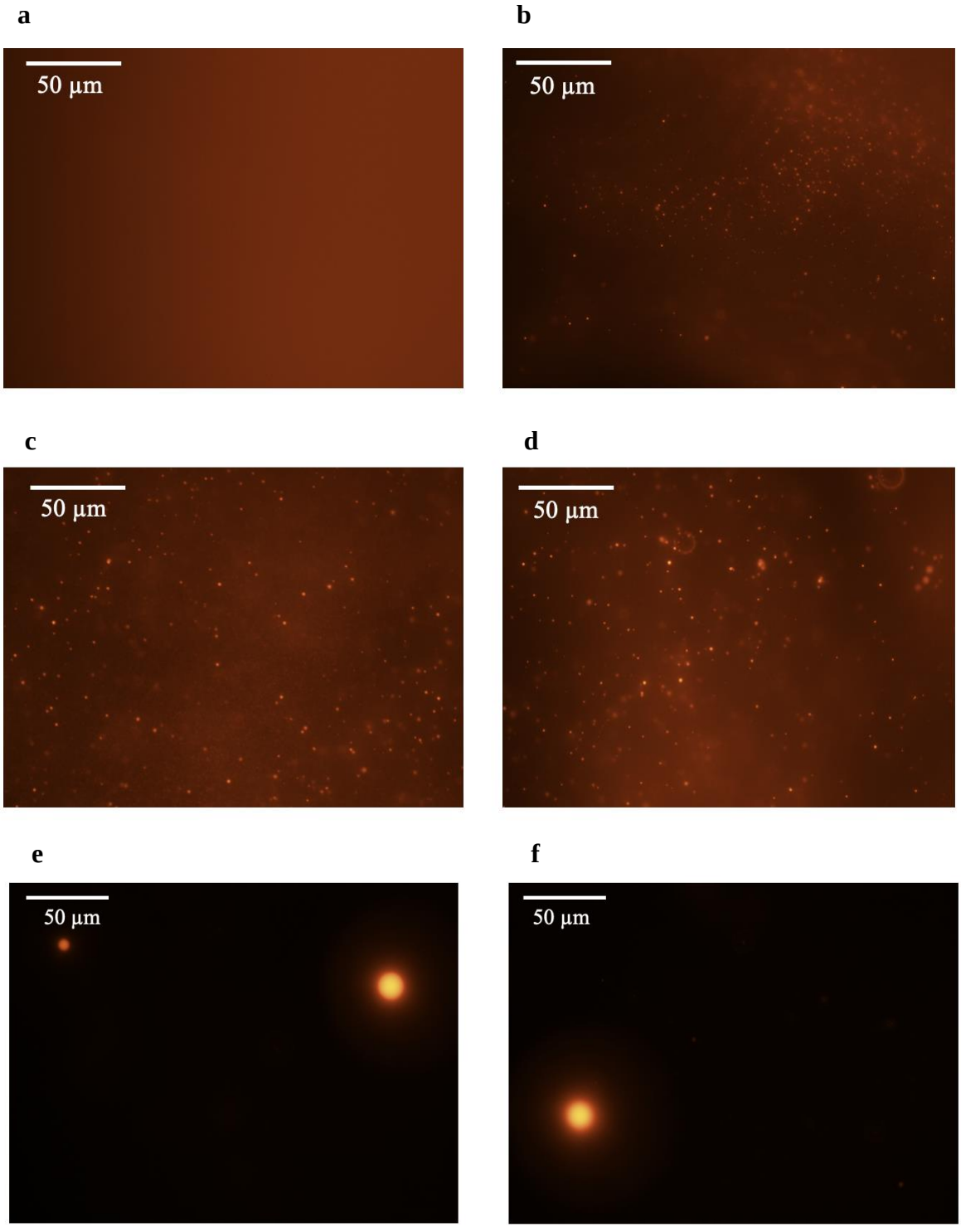



**Figure 38: Macroscopic capture of GRP8 monomer and RNA induced liquid-phase separation.** a) 1 µM, PEG3350 10% (w/v) GRP8 dimer and RNA at RT. b) 10 µM, PEG3350 10% (w/v) GRP8 dimer and RNA at RT. c) 50 µM, PEG3350 10% (w/v) GRP8 dimer and RNA at RT. d) 50 µM, PEG3350 10% (w/v) GRP8 dimer and RNA incubated at 4°C. f) 50 µM, PEG3350 10% (w/v) GRP8 dimer and RNA incubated at 4°C and centrifuged at 21000 x g for 5 minutes. The white bar indicates macroscope scale.

### 3.3.5 Small angle X-ray scattering

As a matter of further structural characterization, SAXS was performed. SAXS is a solution-state small angle X-ray scattering which was used in this case to obtain the information of GRP8 shape and structure, flexibility, intrinsic disorder, and aggregation behavior. This technique was chosen because of its advantage to characterize disordered systems. For IDPs, SAXS is the method most often used because it is faster, requires less material and usually provides more precise experimental data (Bernadó and Svergun, 2012).

In this section, SAXS experiments were divided into two main studies: Batch SAXS for different temperature measurements and SAXS coupled with a SEC column. In both experiments, BSA was used as a control protein for referential parameters and structure comparison. Some extra information from SAXS measurements can be found in supplementary tables 3 and 4.

#### 3.3.5.1 Batch SAXS for different temperature measurements

**Table 31: Structural parameters measured in batch-SAXS**

| **Structural parameters** | | | | |
|---|---|---|---|---|
| | GRP8 monomer | GRP8 monomer | GRP8 monomer | BSA |
| **Guinier analysis** | 20°C | 10°C | 5°C | |
| I(0) ($cm^{-1}$) | 0.013 ± 5.4e-05 | 0.013 ± 6.9e-05 | 0.01 ± 0.00015 | 0.052 ± 5.3e-05 |
| Rg (Å) | 23.25 ± 0.19 | 22.94 ± 0.01 | 23.65 ± 0.49 | 29.27 ± 0.05 |
| s min ($Å^{-1}$) | 0.00 | 0.00 | 0.00 | 0.00 |
| sRg max ($Å^{-1}$) | 1.24 | 1.28 | 1.30 | 1.29 |
| Fidelity (*AutoRg)* | 0.21 | 0.51 | 0.30 | 0.32 |
| sRg limits | 0.45 – 1.24 | 0.59 – 1.28 | 0.77 – 1.30 | 0.26 – 1.29 |
| **MW** | | | | |
| from I(0) (Da) | 15319 | 13142 | 11733 | 63370 |
| from Vc (Da) | 10566 | 10002 | 7685 | 66832 |
| from MoW (Da) | 9872 | 9167 | 6976 | 71192 |
| from Bayesian-inference | 9500 | 9500 | 7300 | 63872 |
| **P(r)** | | | | |
| I(0) ($cm^{-1}$) | 0.013 ± 5.4e-05 | 0.013 ± 6.9e-05 | 0.01 ± 0.00015 | 0.052 ± 5.3e-05 |
| Rg (Å) | 22.22 ± 0.01 | 22.45 ± 0.01 | 20.82 ± 0.01 | 22.46 ± 0.05 |
| D max (Å) | 66.4 | 67.2 | 63.8 | 98.7 |
| s range | 0.00 – 0.344 | 0.00 – 0.349 | 0.00-0.338 | 0.00 – 0.273 |

| | | | | | |
|---|---|---|---|---|---|
| $Chi^2$ | | 1.108 | 1.096 | 1.058 | 1.055 |
| Total estimate | quality | 0.84 | 0.82 | 0.83 | 0.88 |

Because low temperatures became a major limitation for GRP8 purification the GRP8 monomer was measured in different temperature conditions at 20°C, 10°C and 5°C for further characterization of this behavior. Table 31 shows all the structural parameters measured in this experiment. These values were then represented in plots for information inference.

In figure 39, measurements from GRP8 monomer at 20°C are represented. Guinier analysis shows that the quality of the sample is good and not aggregated (figure 39a). Scattering curve shows a partially folded protein (figure 39b). Initial Kratky plot shows that GRP8 is flexible and partially unfolded (figure 39d). When normalizing the values and comparing them with BSA, it is possible to observe that GRP8 is a multidomain protein (figure 39e). Dimensionless Kratky plot indicates that GRP8 has an extended conformation with a fully disordered side and a short-folded region (figure 39f), which correspond to the glycine-rich part and the RRM domain respectively.

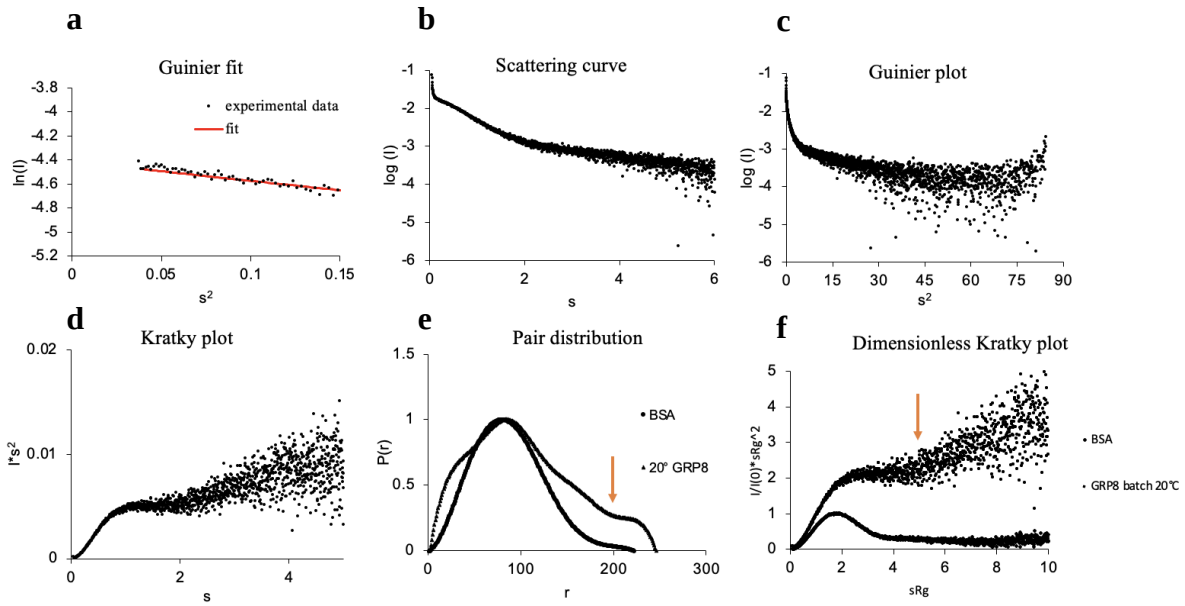



**Figure 39: Measurements from batch mode SAXS of GRP8 monomer at 20°C.** a) Guinier fit. fit in red. The x-axis represents $s^2$. The y-axis represents ln(I). b) Scattering curve. The x-axis represents s. The y-axis represents log(I). c) Guinier plot. The x-axis represents $s^2$. The y-axis represents log(I). d) Kratky plot. The x-axis represents s. The y-axis represents $I*s^2$. e) Pair distribution normalized. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. x-axis represents r. The y-axis represents P(r). f) Dimensionless Kratky plot. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents sRg. The y-axis represents $I/I(0)*sRg^2$.

In figure 40, measurements from GRP8 monomer at 10°C are represented. Guinier analysis shows that the quality of the sample is good and not aggregated (figure 40a), but this time the sRg limits are getting closer to the sRg maximum limit. Scattering curve shows a partially folded protein (figure 40b). An initial Kratky plot shows that GRP8 is flexible and partially unfolded, almost natively unfolded (figure 40d). Nevertheless, when normalizing the values, it is still possible to see the short-folded region still present (figure 40f). With the distance distribution it is possible to observe that GRP8 is a multidomain protein, but with a wider curve versus GRP8 monomer 20°C. These last two plots (40e and f) might suggest the extension of GRP8 in a more unfolded and flexible state.

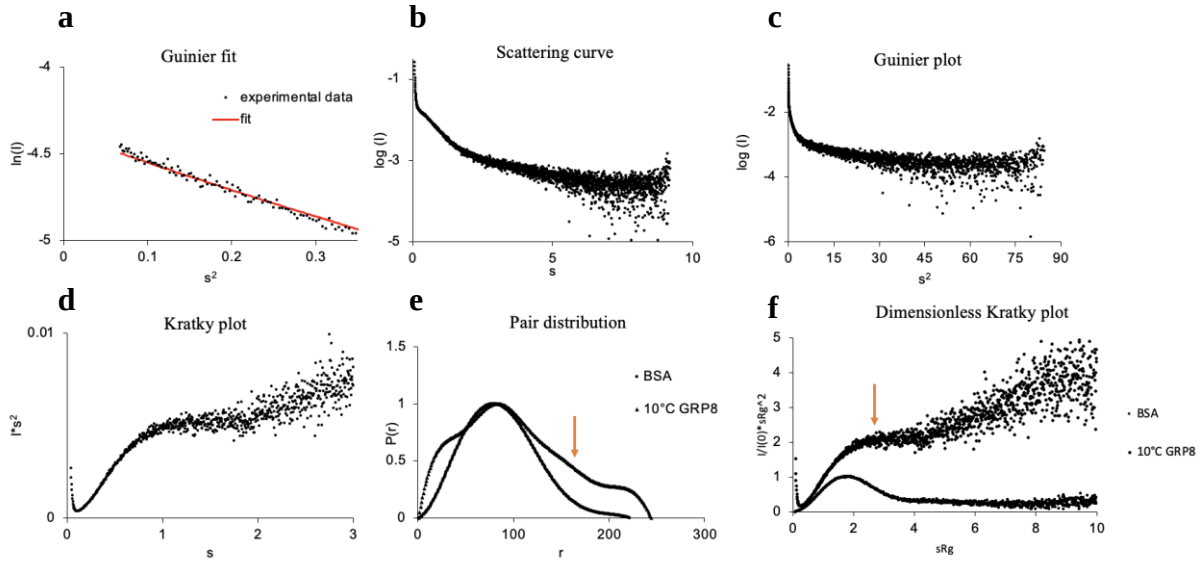



**Figure 40: Measurements from batch mode SAXS of GRP8 monomer at 10°C.** a) Guinier fit. fit in red. The x-axis represents $s^2$. The y-axis represents ln(I). b) Scattering curve. The x-axis represents s. The y-axis represents log(I). c) Guinier plot. The x-axis represents $s^2$. The y-axis represents log(I). d) Kratky plot. The x-axis represents s. The y-axis represents $I*s^2$. e) Pair distribution normalized. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents r. the y-axis represents P(r). f) Dimensionless Kratky plot. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents sRg. The y-axis represents $I/I(0)*sRg^2$.

In figure 41, measurements from GRP8 monomer at 5°C are represented. Guinier analysis shows that the quality of the sample is not good and aggregated (figure 41a), where the sRg minimum is above the limit (0.77). Scattering curve shows a partially folded protein, almost completely unfolded (figure 41b). An initial Kratky plot shows that GRP8 is highly flexible and almost natively unfolded (figure 41d). When normalizing the values, the short-folded region is not present anymore (figure 41f). Nevertheless, with the distance distribution it is possible to observe that GRP8 is a multidomain protein, but with a wider curve versus GRP8

monomer 20°C and 10°C. These last two plots (figure 41 e and f) might suggest the extension of GRP8 in an even more unfolded and flexible state, almost completely disordered.

.

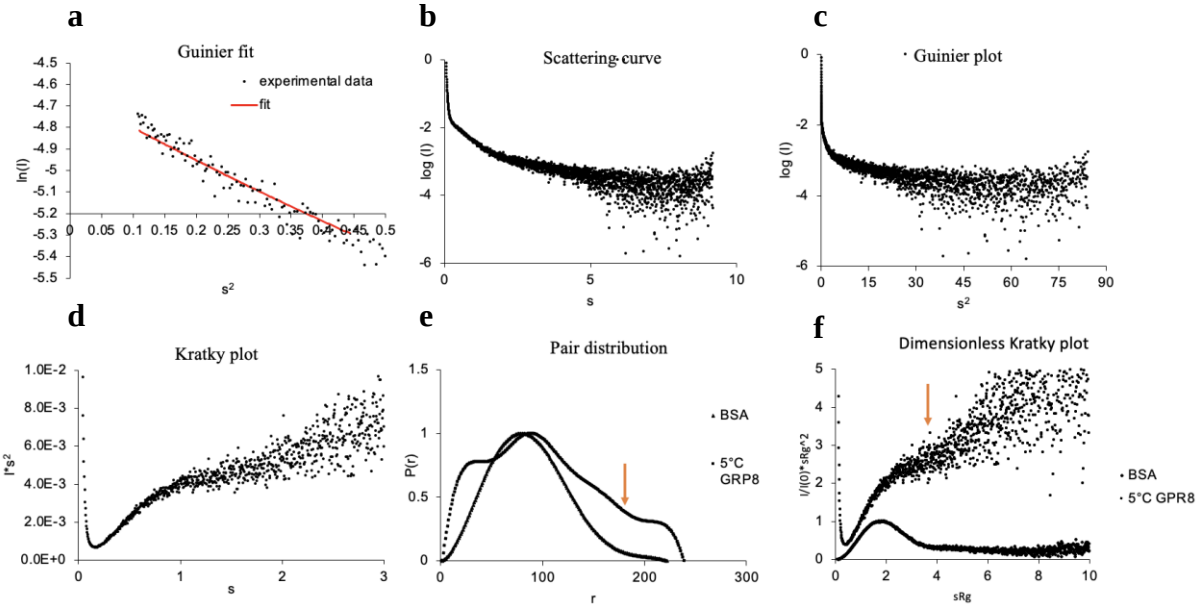



**Figure 41: Measurements from batch mode SAXS of GRP8 monomer at 5°C.** a) Guinier fit. fit in red. The x-axis represents $s^2$. The y-axis represents ln(I). b) Scattering curve. x-axis represents s. The y-axis represents log(I). c) Guinier plot. The x-axis represents $s^2$. The y-axis represents log(I). d) Kratky plot. The x-axis represents s. The y-axis represents $I*s^2$. e) Pair distribution normalized. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents r. The y-axis represents P(r). f) Dimensionless Kratky plot. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents sRg. The y-axis represents $I/I(0)*sRg^2$.

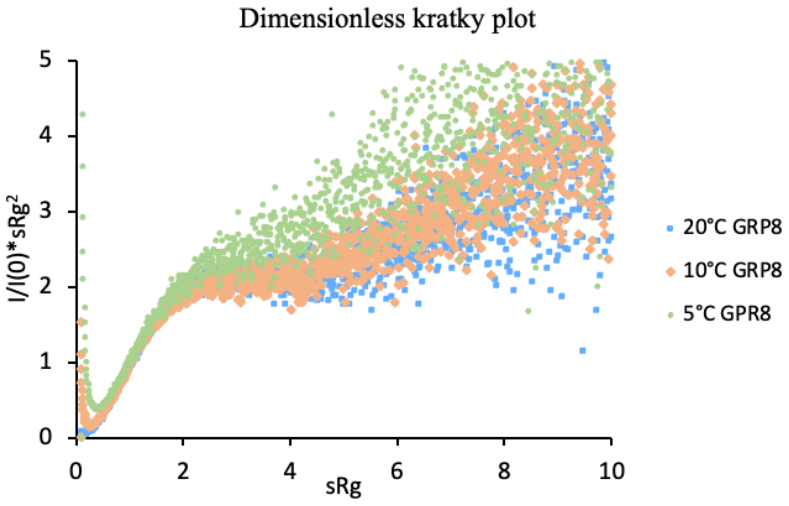



**Figure 42: Dimensionless Kratky plot from batch mode SAXS of GRP8 monomer at 20°C, 10°C and 5°C.** The 20°C GRP8 monomer is represented in blue. The 10°C GRP8 monomer is represented in orange. The 5°C GRP8 monomer is represented in green. The x-axis represents sRg. The y-axis represents $I/I(0)*sRg^2$.

Figure 42 shows a comparison through the dimensionless Kratky plots of GRP8 monomer at the three different temperatures measured. It is possible to observe that there is a change in shape and disorder, supporting that when induced to cold GRP8 monomer becomes more unfolded and fully disordered.

### 3.3.5.2 SEC-SAXS

**Table 32: structural parameters measured in SEC-SAXS**

| **Structural parameters** | | |
|---|---|---|
| **Guinier analysis** | GRP8 monomer 20°C | GRP8 dimer 20°C |
| I(0) ($cm^{-1}$) | 1547.39 ± 4.73 | 2792.85 ± 6.12 |
| Rg (Å) | 22.28 ± 0.14 | 28.50 ± 0.12 |
| s min ($Å^{-1}$) | 0.00 | 0.00 |
| sRg max ($Å^{-1}$) | 1.24 | 1.24 |
| Fidelity (*AutoRg)* | 0.11 | 0.69 |
| sRg limits | 0.28 – 1.24 | 0.26 – 1.24 |
| **MW** | | |
| from I(0) (Da) | 21531 | 43385 |
| from Vc (Da) | 17652 | 42236 |
| from MoW (Da) | 19242 | 46528 |
| from Bayesian-inference | 18675 | 43775 |
| **P(r)** | | |
| I(0) ($cm^{-1}$) | 1547.39 ± 4.73 | 2792.85 ± 6.12 |
| Rg (Å) | 22.63 ± 0.01 | 28.04 ± 0.01 |
| D max (Å) | 68.2 | 80.4 |
| s range | 0.00 – 0.343 | 0.00 – 0.281 |
| $Chi^2$ | 1.714 | 1.446 |
| Total quality estimate | 0.87 | 0.90 |

For further structural characterization, SEC-SAXS was performed. In this case, GRP8 monomer and dimer were measured to compare their structures and GRP8 monomer data was then analyzed and processed for the 3D modeling. SEC-SAXS was chosen over the batch mode because it has been proven to improve the resolution for disordered systems through the reduction of sample heterogeneity (Graewert *et al.*, 2020). Moreover, in these experiments the same criteria as batch SAXS apply for the understanding of the values and plots.

In figure 43, measurements from GRP8 monomer are represented. The first two plots (figure 43a and b) represent the SEC profiles before performing SAXS. The Rg and MW values are

represented respectively. Guinier analysis shows that the quality of the sample is good and not aggregated (figure 43c). Scattering curve shows a partially folded protein (figure 43d). Initial Kratky plot shows that GRP8 monomer is flexible and partially unfolded (figure 43f). When normalizing the values, it is possible to observe that GRP8 monomer is a multidomain protein (figure 43g). Dimensionless Kratky plot indicates that GRP8 monomer has an extended shape and is partially unfolded when compared with BSA (figure 43h). These results are in accordance with the GRP8 monomer structure considering the RRM and the glycine-rich domains.

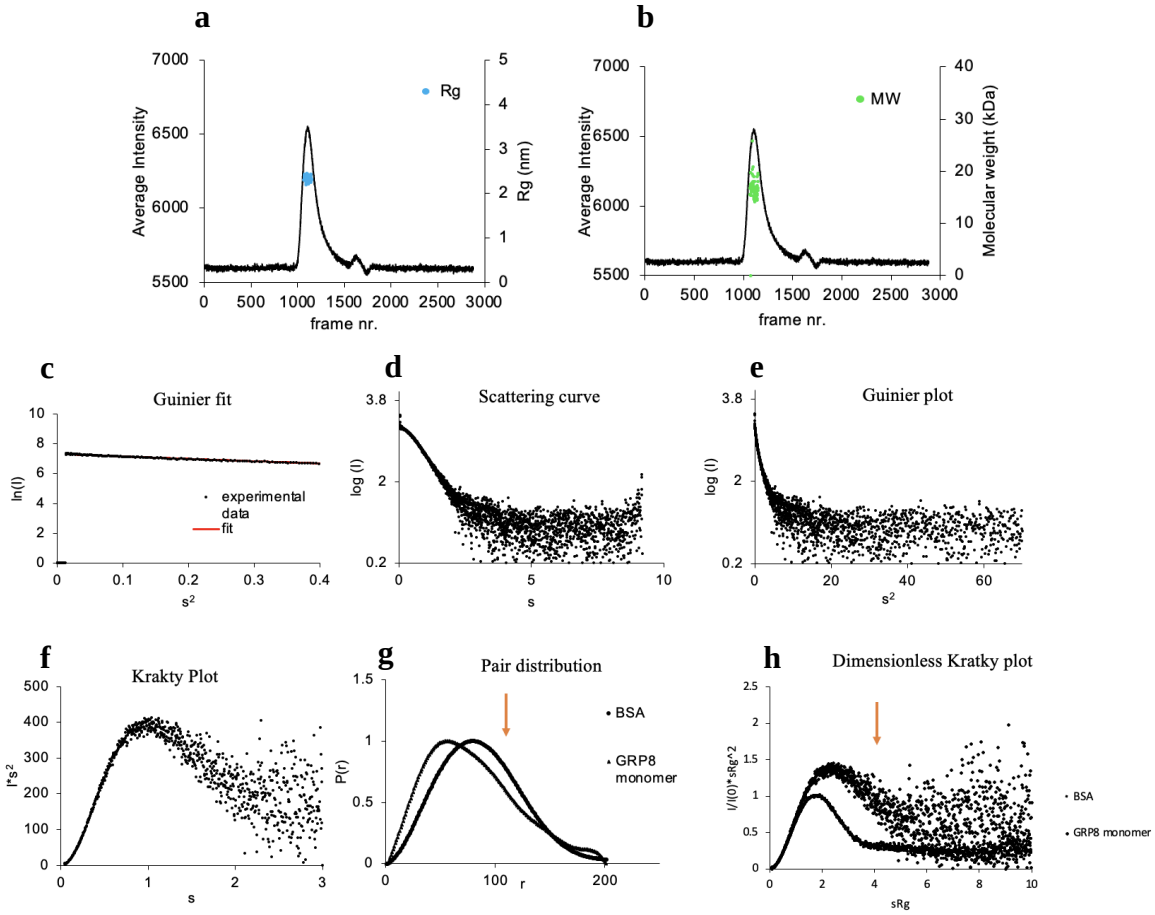



**Figure 43: Measurements from SEC-SAXS of GRP8 monomer.** a) the SEC chromatogram and Rg. The Rg is represented in blue. The x-axis represents frame number. The y-axis represents average intensity. The z-axis represents Rg (nm). b) The SEC chromatogram and MW. The MW is represented in green. The x-axis represents frame number. The y-axis represents average intensity. The z-axis represents MW (kDa). c) Guinier fit. fit in red. The x-axis represents $s^2$. The y-axis represents ln(I). d) Scattering curve. The x-axis represents s. the y-axis represents log(I). e) Guinier plot. The x-axis represents $s^2$. The y-axis represents log(I). f) Kratky plot. The x-axis represents s. y-axis represents $I*s^2$. g) Pair distribution normalized. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents r. The y-axis represents P(r). h) Dimensionless Kratky plot. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents sRg. The y-axis represents $I/I(0)*sRg^2$.

In figure 44, measurements from GRP8 dimer are represented. The first two plots (figure 44a and b) represent the SEC profiles before performing SAXS. The Rg and MW values are represented respectively. Guinier analysis shows that the quality of the sample is good and not aggregated (figure 44c). Scattering curve shows a partially folded protein (figure 44d). An initial Kratky plot shows that GRP8 dimer is flexible and partially unfolded (figure 44f). When normalizing the values, it is observed that GRP8 dimer a less sharp but still a multidomain curve (figure 44g). Dimensionless Kratky plot indicates that GRP8 dimer has an extended shape and is partially unfolded when compared with BSA (figure 44h).

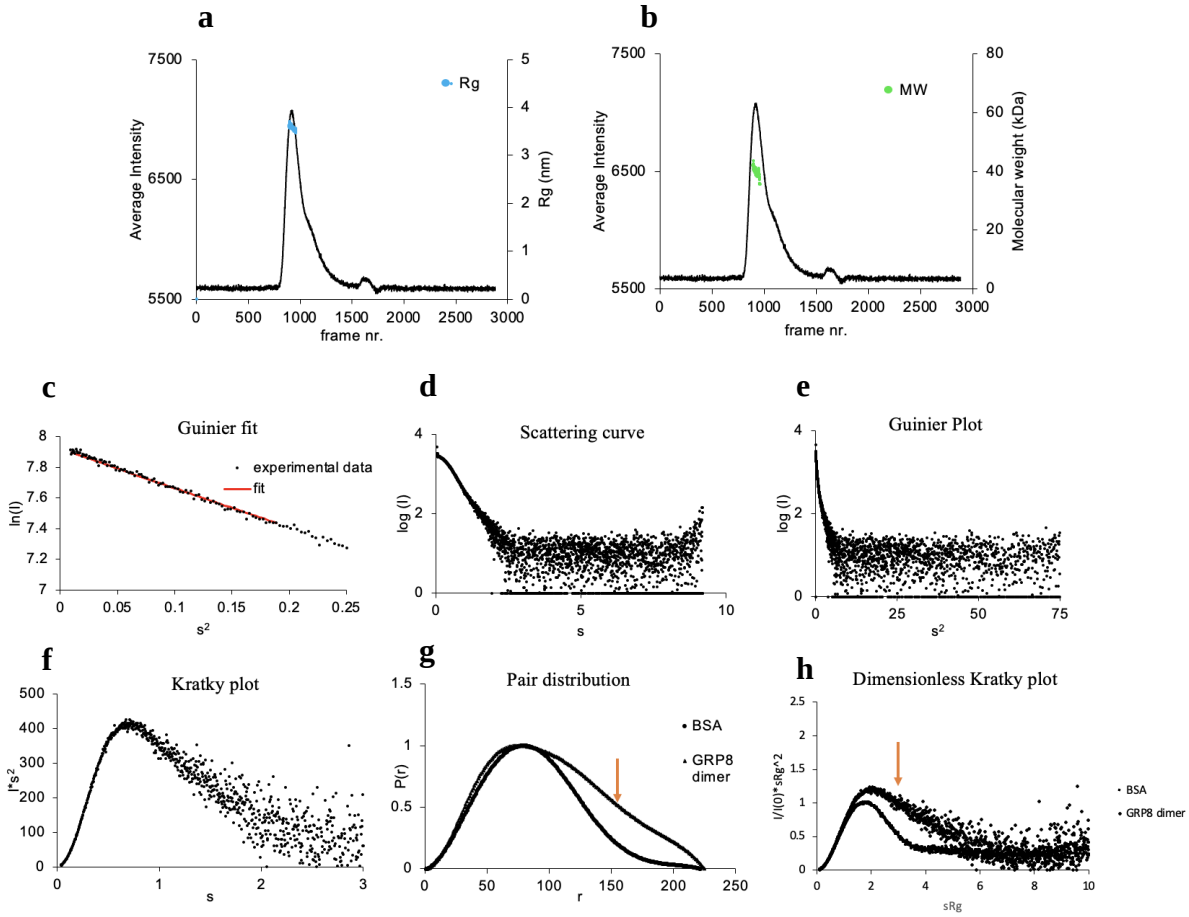



**Figure 43: Measurements from SEC-SAXS of GRP8 dimer.** a) The SEC chromatogram and Rg. The Rg is represented in blue. The x-axis represents frame number. The y-axis represents average intensity. The z-axis represents Rg (nm). b) The SEC chromatogram and MW. The MW is represented in green. The x-axis represents frame number. The y-axis represents average intensity. The z-axis represents MW (kDa). c) Guinier fit. fit in red. The x-axis represents $s^2$. The y-axis represents ln(I). d) Scattering curve. The x-axis represents s. The y-axis represents log(I). e) Guinier plot. The x-axis represents $s^2$. The y-axis represents log(I). f) Kratky plot. The x-axis represents s. The y-axis represents $I*s^2$. g) Pair distribution normalized. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents r. The y-axis represents P(r). h) Dimensionless Kratky plot. BSA is a globular control for comparison. The orange arrow shows the GRP8 curve. The x-axis represents sRg. The y-axis represents $I/I(0)*sRg^2$.

When comparing GRP8 monomer and dimer, molecular weights are consistent with their corresponding values (for monomer 16 kDa; for dimer 32kDa), considering that SAXS is a low-resolution tool. Figure 44 shows comparison of normalized plots of GRP8 monomer and dimer with BSA as a reference globular protein. Pair distribution shows that there is a difference in the conformation of GRP8 when dimerized, where the multi domain curve is less visible (figure 45a). Dimensionless Kratky plot shows that GRP8 monomer is more flexible and disordered with a more elongated and unfolded shape than GRP8 dimer (figure 45b).

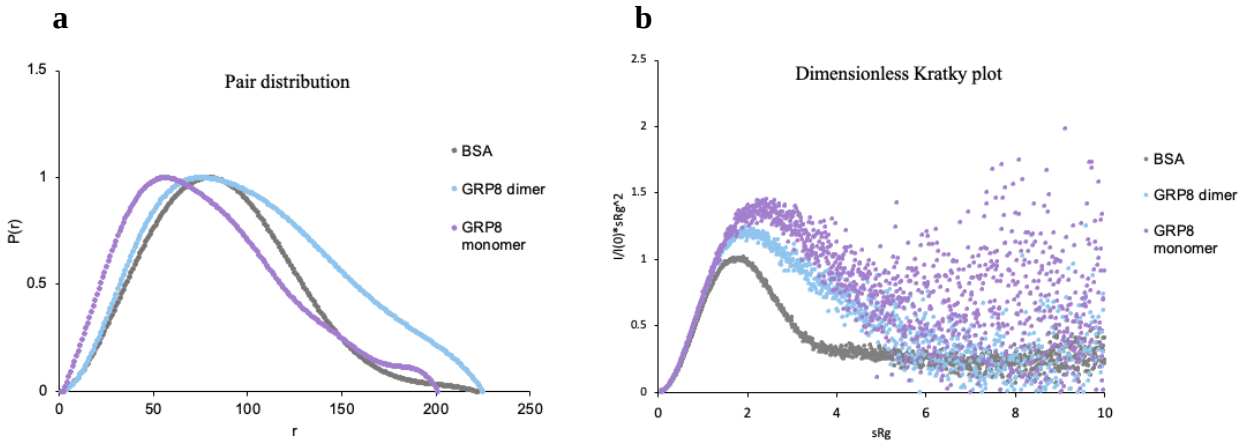



**Figure 45: Shape and flexibility comparison between GRP8 monomer and dimer.** a) Pair distribution normalized. BSA is a globular control for comparison. x-axis represents r. y-axis represents P(r). h) Dimensionless Kratky plot. BSA is a globular control for comparison. x-axis represents sRg. y-axis represents $I/I(0)*sRg^2$. In both graphs: BSA represented in gray. GRP8 monomer represented in light blue. GRP8 dimer represented in violet.

### 3.3.6 3D structure modeling

SAXS data analysis and processing was performed for GRP8 monomer for the modeling of its 3D structure. First, from the *ATSAS software* (EMBL) it was generated a *DAMMIF* model, which is an *ab-initio* shape model directly calculated from the data processed through automated algorithms in the software (Franke and Svergun, 2009). After, different databases were used to find available models to reconstruct a more refined model *in silico* of GRP8 monomer, including *Alphafold* (Jumper *et al.*, 2021) and *i-Tasser* (Yang *et al.*, 2014). When running *i-Tasser*, results showed five different models that were created by the database and five different analog models from already characterized proteins. These models were then analyzed through the *ATSAS software*, which contains several options of programs for model fitting from experimental data and available known models: *CRYSOL* (Svergun, Barberato and Koch, 2015), *SREFLEX* (Panjkovich and Svergun, 2016a) and *EOM* (Tria *et al.*, 2015). Table

33 represents the models that were evaluated. Chi squared values were calculated and used as threshold values, and ideally, they should be as small as possible. *Alphafold* model showed large Chi squared values, therefore the model did not fit to the experimental data. When comparing the *i-Tasser* models and analogs, these last ones all failed the *SREFLEX* processing. The *CRYSOL* values for the analogs were still too large. For the models generated from *i-Tasser*, *CRYSOL* values were still large, however *SREFLEX* showed promising lower Chi squared values. From all the models that were generated in combination of *SREFLEX* and *i-Tasser*, new generated models were processed with their own Chi squared value. Interestingly, all models fitted with similar Chi squared values and with the same reference model from *SREFLEX*. (cc.03). Therefore, model cc.03 from *SREFLEX* was then selected for the 3D structure modeling. Lastly, when using the *EOM* program, the Chi squared value was too large for further consideration.

**Table 33: Databases used for 3D modeling with their Chi² score.**

| database | model | analysis | Chi² SEC-SAXS | reference SREFLEX | Chi² |
|---|---|---|---|---|---|
| Alphafold | - | CRYSOL | 20.43 | | |
| | | SREFLEX | 7.13 | cc.03 | 1.7 |
| i-Tasser | Model 1 | CRYSOL | 5.7 | | |
| | | SREFLEX | 1.70 - 2.24 | cc.03 | 1.77 |
| | Model 2 | CRYSOL | 5.53 | | |
| | | SREFLEX | 1.73 - 2.04 | cc.03 | 1.95 |
| | Model 3 | CRYSOL | 7.29 | | |
| | | SREFLEX | 1.81 - 2.04 | cc.03 | 1.67 |
| | Model 4 | CRYSOL | 5.72 | | |
| | | SREFLEX | 1.61 - 1.94 | cc.03 | 1.69 |
| | Model 5 | CRYSOL | 7 | | |
| | | SREFLEX | 1.66 - 1.92 | cc.03 | 1.79 |
| | Analog 1 | CRYSOL | 5.5 | | |
| | | SREFLEX | failed | | |
| | Analog 2 | CRYSOL | 4.55 | | |
| | | SREFLEX | failed | | |
| | Analog 3 | CRYSOL | 5.13 | | |
| | | SREFLEX | failed | | |
| | Analog 4 | CRYSOL | 5.78 | | |
| | | SREFLEX | failed | | |
| | Analog 5 | CRYSOL | 11.49 | | |
| | | SREFLEX | failed | | |

| analysis | Chi² SEC-SAXS |
|---|---|
| EOM | 3.03 |

3D modeling of GRP8 monomer was done through *PyMOL*. First, the experimental data was loaded into the program through the *DAMMIF* model. Afterwards, the *in silico*-created model with *SREFLEX* and *i-Tasser* was loaded into the program. The *SASpy plugin* (EMBL) was used for the improvement of GRP8 model. *SASpy* is a tool that allows to overlay the experimental model generated from the SAXS data and the models predicted from databases, creating a more refined hybrid model (Panjkovich and Svergun, 2016b). Figure 45 shows the 3D structure reconstruction. The model shows a superimposed GRP8 model including the results from databases, including the *SREFLEX* and *i-Tasser* model cc.03 represented in pink and the SAXS data processed model represented in gray.

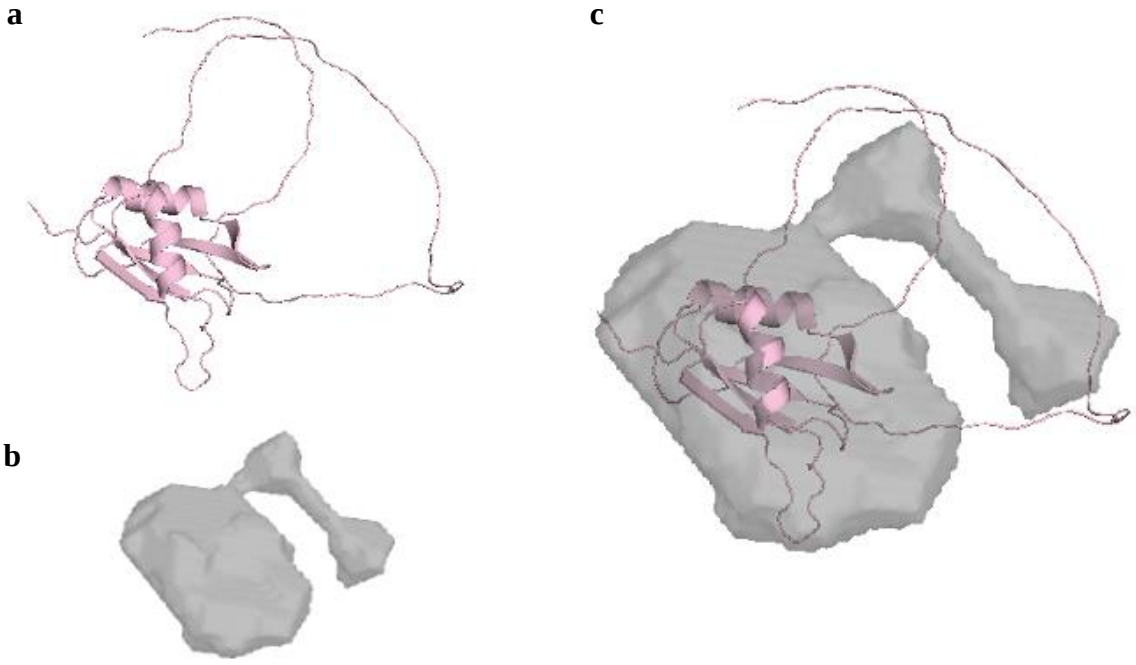



**Figure 46: GRP8 3D structure.** a) shows the GRP8 structure from alignment prediction in pink. b) shows theGRP8 shape from SEC-SAXS data processing. c) shows a superposed image including the alignment prediction and the SEC-SAXS data processing by *SASpy* in *PyMOL*.

# 4. Discussion

## 4.1 Phloem-RNA sampling and approaches for functional studies

*Arabidopsis thaliana* is the major model organism in plants for research purposes because of its small genome, short life cycle and transformation efficiency (Meyerowitz and Pruitt, 1985; Meyerowitz, 1989; Chen *et al.*, 2004). The availability of a whole genome sequence in *Arabidopsis* provides unique opportunities for genome-based systems biology approaches and it enables the use of post-genomic tools such as proteomics in its full capacity (Wienkoop, Baginsky and Weckwerth, 2010). Despite *A. thaliana* has not much economical relevance, *Arabidopsis* species share recent common ancestry with many species of significant economic importance, including *Brassica* species. Studies involving comparative genetics and physical mapping of specific chromosome segments have shown largely conserved in *Arabidopsis* and *Brassica* families, but with some disruption in gene content in deletions and insertions (Paterson *et al.*, 2001). Comparative genetics has evolved as a powerful tool for uncovering the processes and rate of genome evolution and for allowing the transfer of genetic resources between species (Parkin *et al.*, 2005).

Phloem samples from *A. thaliana* can only be obtained by EDTA-facilitated exudation or aphid stylectomy (Dinant *et al.*, 2010; Tetyuk, Benning and Hoffmann-Benning, 2013). These both can lead to very low sample amount collection and EDTA-facilitated exudation is contamination susceptible, therefore, could not represent the actual phloem composition (Kovalskaya *et al.*, 2014; Pahlow *et al.*, 2018). Therefore, phloem sap from *B. napus* was sampled to produce reliable and sufficient phloem-RNA. It has been established that *B. napus* is a suitable model system for phloem sap analysis with larger sample amounts and higher purity that can be collected (Pahlow *et al.*, 2018)

The purity and integrity of the phloem-RNA was check by PCR and *Bioanalyzer*. The results showed that some samples were contaminated with RuBisCO (figure 18) and therefore these samples were discarded. Some of the samples that showed a slight band in RuBisCO were still included. Because of the use of the needle to puncture the plants, it is hard to avoid some minimum level of contamination, even when the first exudate phloem drop was wiped for removing contaminated phloem. Although some of the samples did not show or showed very little contamination, when looking at the *Bioanalyzer* results in figure 19, it was observed that some samples were contaminated (sample 5) as there was a peak present after the 4000 nt. Sample 6 showed no RNA integrity, as the peaks are almost nonvisible. The x-axis in some of

the samples was moved to the right, which most likely was caused by the *Bioanalyzer* system when reading the marker, nevertheless the sample quality was showed to be good. Moreover, samples that passed the contamination and purity check were selected for the CnBr-Sepharose GRP8-bound affinity column. For disclosure purposes, the input RNA analyzed by RNA-sequencing to compare the elution fractions from the CnBr-Sepharose GRP8 affinity column was not the same as the one loaded into the column (because of a situation at *Novogene*), which could cause difference of transcripts found in common between input and elution samples and/or difference in contamination levels. Nevertheless, the new input sample that was sent for RNA-sequencing was also previously checked for contamination and degradation (chapter 3.2.1). In addition, it was searched for contamination transcripts including RuBisCo when the RNA-sequencing data was analyzed. No contamination transcripts were found to be enriched in input and GRP8-elution samples, which was a good indication of sample purity. Moreover, several chlorophyll a-b binding protein transcripts were found to be present in input and GRP8-elution samples. This transcript has been found in phloem saps of different plant species (Doering-Saad *et al.*, 2006; Omid *et al.*, 2007; Ruiz-Medrano *et al.*, 2007; Buhtz *et al.*, 2008). Further elucidation of this transcript remains necessary; therefore, it would be interesting to test this RNA with GPR8 by MST to confirm binding interaction.

After conducting the CnBr-Sepharose GRP8-bound affinity column and analyzing the RNA-sequencing results, transcripts of interest were aligned through *EnsemblPlants* (Yates *et al.*, 2022) to find homolog transcripts in the *A. thaliana* genome. Transcripts from *B. napus* bound to GRP8 showed all over eighty percent identity to the *A. thaliana* ones. The enriched transcripts that were selected for MST testing are showed in table 29.

## 4.2 GRP8 binds a wide range of RNAs

The CnBr-Sepharose GRP8-bound affinity column results (figure 20) showed that different RNA sizes are found in the elution fractions. The elution number three, which was eluted with the highest salt concentration, also showed a wide RNA size profile. The RNA profiles from the input phloem-RNA and flowthrough looked similar in the electropherograms, on the contrary the RNA profile of the elution number three shows a different RNA profile in the electropherogram, also showing an enrichment of RNAs with sizes between 500 to 2000 nt. This was supported by the sizes of the transcripts found through RNA-sequencing analysis (table 29, supplementary table 1). The CnBr-Sepharose was performed in duplicates which are

showed in figure 21, and it is observed that the eluted RNA profiles of the highly enriched RNAs in elution three are almost identical.

After the RNA-sequencing analysis, the transcripts selected were tested by MST to confirm that GRP8 binds them. In addition, other candidates mentioned in chapter 3.2.3 were also tested including small RNAs (*miRNA164, miNovel2, miNovel106, miNovel149,3'UTR-GRP7*) and other mRNAs (*GRP8* and *GRP7*). Although the small RNAs were not found to be highly enriched to GRP8, there were still tested because they were found to be enriched to GRP7 and since GRP8 and 7 share sequence similarity, it was interesting to see if there was a difference in binding between them. It has been previously described that GRP8 binds the mRNAs of GRP8 and 7 (Schöning *et al.*, 2008), nevertheless they were not found to be enriched in the RNA-sequencing analysis. This could be explained because these transcripts are known to be at their peak in the end of the day (Schmal, Reimann and Staiger, 2013) and the phloem extractions from plants were performed at noon.

Evidently, GRP8 can bind different interaction partners. This was demonstrated by MST experiments confirming that GRP8 binds a wide range of RNAs, including small RNAs (figures 23, 24 and 26) and mRNAs (figures 25, 26, and 27). In addition, two transcripts that were not enriched in the RNA-sequencing analysis were tested and found to be bound to GRP8 by MST. This might be an indication of GRP8 ability to bind many different RNAs because GRP8 even can bind *in vitro* transcripts that were not enriched.

As previously described, GRP8 binds the mRNAs of GRP7 and GRP8 that were already discussed. GRP7 has been demonstrated that binds the 3'UTR of its own mRNA, which again was interesting to test to compare with GRP8. Figure 26b shows the Kds of GRP8 and GRP8$^{short}$ when tested with the *3'UTR-GPR7* (only the miRNA sequence) by MST. This was found to bind GRP8 and GRP8$^{short}$, but GPR8$^{short}$ was bound to the *3'UTR-GPR7* with lower affinity. Because the only difference between them is the glycine-rich region, this might suggest that this domain assist the RNA-binding of GRP8. Nevertheless, it is observed that when mRNAs of *GRP8* and *GRP7* are tested with GRP8$^{short}$, no binding is detected. Furthermore, it would be interesting to test GRP8 and GRP8$^{short}$ with the mRNAs of *GRP8* and *GRP7* including the 3'UTR. It has been described that the 3'-UTRs of mRNA serve as templates for miRNA binding that regulates the turnover and/or function of the mRNA (Yang *et al.*, 2009; Singh *et al.*, 2012). This turnover and regulation function has been reported with GRP8 and GRP7 both autoregulate and reciprocally cross-regulate each other by binding their pre-mRNAs and promoting unproductive splicing through non-mediated decay pathway (supplementary figure 1) (Staiger *et al.*, 2003; Schöning *et al.*, 2008).

When the mRNAs from the CnBr-Sepharose GRP8-bound affinity column and RNA sequencing analysis were synthesized *in vitro*, it was attempted to produce them from cDNA templates including 3' and 5'UTRs. The only transcript where this was possible was with the *OHP*. The rest of the transcripts of interested only included the coding sequence unless indicated. Figure 27 shows the difference in binding affinities of when GRP8 is tested by MST with *OHP* (coding sequence), *5'UTR-OHP* and *3'+5'UTR-OHP*. The Kd of GRP8 bound to *OHP* is statistically significant different when the 5'UTR is present, which might indicate that GRP8 binds to the 5'UTR of the *OHP* mRNA. There is a higher affinity for GRP8 when the 5' UTR is present, however, when comparing the overall Kd values, they are still in the micromolar range and there are still higher than the Kds of GRP8 tested with small RNAs. It has been described that the 5'-UTRs are also involved in the stabilization of mRNA and the regulation of gene expression (Singh *et al.*, 2012) and they can incorporate RBPs binding sites (Araujo *et al.*, 2012). It would be interesting to test the current transcripts of interest with their UTRs to check if there is a correlation between GRP8 binding UTRs and to test if their Kds decrease in a similar manner as it was observed with *OHP*.

For further comprehension of the RNAs that were tested by MST, a motif alignment was performed by the motif-based sequence alignment tool MEME-Suite (Bailey *et al.*, 2015) and no significant common motifs were identified in the sequences (supplementary figure 6 and 7), supporting that GRP8 is binding a wide range of RNAs.

In the literature it was described that there is a preference for GRP7 to bind sequences rich in U and G (Staiger *et al.*, 2003). Results showed in supplementary table X showed first that the was no preference for mRNAs bound to GRP8. The ones that were bound by $GRP8^{short}$ were found to either show high content or G or G and U together (supplementary table X) but no U alone. Because $GRP8^{short}$ only includes the RRM, it could be speculated that the RRM prefers a very high content of G or G/U for binding, however further testing remains necessary to confirm this. It would be interesting to test more sequences rich in G or G/U and sequences that are not by MST.

## 4.3 Glycine-rich domain plays a key role in RNA-binding

The first indication that the glycine-rich domain has an important RNA-binding role was observed when comparing the CnBr-Sepharose GRP8-bound and $GRP8^{short}$-bound affinity columns. In figure 21 elution fractions of GRP8-bound and $GRP8^{short}$-bound affinity columns are showed, and it is observed that only small RNAs were bound to $GRP8^{short}$, suggesting the

functionality of the protein was compromised. This was then confirmed when GRP8$^{short}$ was not binding any of the mRNAs from the RNA-sequencing analysis (table 30), suggesting one more time that the glycine rich-part is necessary for the functionality of GRP8.

In addition, miRNA164 was tested by MST with GRP8 and GRP8$^{short}$ and there was a significant difference in their binding affinities. For understanding where the *miRNA164* binds, UV-crosslinking was done (chapter 3.2.4). This is showed in figure 28 where the GRP8 sequence represented in amino acids and the ones marked in red are interacting with *miRNA164*. It is possible to see that the miRNA164 interacts with both glycine-rich domain and the RRM, which could explain the difference in Kds.

When testing other miRNAs (figure 24 and table 30) by MST, it was interesting to see that two of them were not binding with GRP8$^{short}$. This suggests that these miRNAs most probably bind to the glycine-rich domain. In further effort to confirm this, it would be interesting to test these miRNAs (*miNovel106 and miNovel149*) with GRP8 by UV-crosslinking.

Some individual RRMs can bind to RNA with great selectivity, but multiple domains are often needed because the number of nucleotides that are recognized by an individual RRM is generally too small to define a unique binding sequence (Auweter, Oberstrass and Allain, 2006; Lunde, Moore and Varani, 2007). This is where the glycine-rich domain plays a key role in RNA-binding. GRP8 has been described as a IDP (figure 5) with a glycine-rich domain as a low complexity region including [G/S]Y[G/S] and RGG motifs (figure 32). Protein disorder is known to allow proteins that bind to different interaction partners with a wide selectivity and affinity range (Pazos *et al.*, 2013). The RGG motifs have been described to contribute and even account for the RNA-binding affinity which is driven by the flexibility of the glycine repeats surrounded by arginine residues (Järvelin *et al.*, 2016; Hentze *et al.*, 2018). Therefore, the glycine-rich domain is essential for the binding capacity of GRP8.

## 4.4 Protein purification of GRP8 revealed three major hallmarks

Protein purification was a long and tedious process (figure 14). Troubleshooting became a big part of producing GRP8. When purifying, degradation and aggregation were a constant problem throughout the entire process, sometimes only happening at the very end when concentrating the protein. Many extra steps were incorporated for achieving that GRP8 was obtained in a stable form (figure 15c), including performing the whole process at RT. Interestingly, when GRP8$^{short}$ was purified, the process was accomplished quickly (figure 17).

It was assumed that the glycine-rich domain was causing degradation and aggregation hence it is the only difference between these two proteins.

As previously stated, the glycine-rich domain is a highly disordered region (figure 5b) which is related to low sequence complexity. It has been described that IDPs are very sensitive to their environment including changes in temperature, pH and salt concentrations can lead to changes in the strength of the hydrophobic and intra-molecular interactions (Habchi *et al.*, 2014). When understanding this, it was consistent that GRP8 would bring so many extra stabilization steps. Furthermore, three different conditions were identified to cause the aggregation or degradation throughout the GRP8 purification. First, the time length of the purification process. When the overall process was taking longer, more degradation was found. Because of that, the whole process had to be done in almost one day. The first day included all purification steps but the last GRP8 concentration. When comparing this with $GRP8^{short}$, the whole purification process was performed in three days, and no degradation was observed. Second, high concentrations of GRP8 monomer drove to aggregation. As previously described, after SEC two populations of GRP8 were observed (figure 16). When concentrating GRP8 dimer, no aggregation was observed at any reached concentration, however it was still difficult to reach high concentrations and the maximum amount concentrated was to 3.3 mg/ml. On the other hand, GRP8 monomer was only possible to archive a maximum of 3 mg/ml and then the concentration would fall, suggesting its aggregation and decay in concentration. This last situation was not reported with GRP8 dimer. Lastly, the environmental temperature. If GRP8 was placed in cold at any step of the process, it would immediately aggregate. The aggregation was visible in as milk-like solution, which was reversible by centrifugation. Luckily, this could be avoided by performing the entire process at RT.

It has been described that many prion-like proteins are RNA-binding, that the information for their prion behavior is contained in the disordered low complexity region and that their self-interaction at high protein concentration promotes conformational conversion into a prion/condensate state (Franzmann and Alberti, 2019). The reversibility of the aggregation was an indication of a liquid-liquid phase separation state (Guo, Shi and Wang, 2021). In addition, it has been reported that *Arabidopsis thaliana* proteins including a PrLD forms reversible liquid droplets in response to temperature changes *in vitro* (Jung *et al.*, 2020; Xu *et al.*, 2022). Overall, these hallmarks that have been revealed throughout the protein purification process appear to be linked to the low complexity region of GRP8 and the presence of a PrLD in the glycine-rich domain.

## 4.5 GRP8 and liquid-liquid phase separation

The PrLD behavior of GRP8 was first tested by the ThT assay (chapter 3.3.3). When comparing the results from figure 35a, it is observed that there is no condensation occurring with GRP8$^{short}$ which could be explained by the missing glycine-rich c-terminus. When comparing GRP8 dimer and monomer in figure 35a, it is observed that GRP8 monomer has a higher fluorescence and a bigger increase of fluorescence in the first 20 hours versus GRP8 dimer showed lower fluorescence with a lower increase in the first 20 hours, appearing more stable. This could be speculated as GRP8 dimer forming less condensation than GRP8 monomer and it might be caused by GRP8 interacting with itself forming a more stable form. It was interesting to test GRP8 exposed to RNA by the ThT because of the RGG-containing regions, which are described to be RNA-binding regions and involved in LLPS (Chong, Vernon and Forman-Kay, 2018). In this case, *miRNA164* was chosen because it showed high affinity to GRP8 by MST (table 30). Figure 35b showed GRP8 monomer compared to GRP8 monomer exposed to *miRNA164*. There is a slight increase in fluorescence when the GRP8 monomer is exposed to the RNA, which might indicate a higher condensation state of GRP8 monomer/*miRNA164*. Figure 35c showed GRP8 dimer compared to GRP8 dimer exposed to *miRNA164.* There is an increase in fluorescence when GRP8 is exposed to the RNA and the initial fluorescence shows a higher starting point. This might indicate a higher condensation state of GRP8 dimer/*miRNA164*. Because of the intrinsic disorder nature of GRP8 and its PrLD, it might be that GRP8 dimer is recruiting the RNA and driving the formation of GRP8 condensation. This could be also the case with GRP8 monomer/*miRNA164.* As an example, this RNA recruiting increasing protein condensation has been reported in another PrLD protein, FUS (Fused in Sarcoma protein) (Davis *et al.*, 2021).

For further understanding of GRP8 forming condensation droplets, a LLPS assay was performed. A labeled-RNA was used for the detection of protein-bound to RNA under a macroscope. Figure 36 showed GRP8$^{short}$ bound to the RNA and no condensation droplets were detected at any of the conditions measured. This was expected as is missing the PrLD. Figure 37 shows condensation droplets of GRP8 dimer bound to the RNA when the concentration increases to 25 µM at RT (figure 36b). Because LLPS is often modulated by temperature (Sanulli and Narlikar, 2021), GRP8 dimer was also tested at 4°C to compare if the cold was inducing more condensation, as it was reported with GRP7 upon cold induction (Xu *et al.*, 2022), however no relevant difference was observed. Centrifuged samples including condensation droplets of GRP8 dimer and RNA were detected in samples showed in figure

37c, but no major differences in size were reported. Figure 38 is showed GRP8 monomer bound to the RNA and condensation droplets were already observed when the concentration reached to 10 µM at RT (figure 38b). When the concentration of GPR8 monomer increased to 50 µM at RT (figure 38c), the number of condensates appeared to be slightly higher and the size a slightly bigger. This might indicate that an increase in the concentration of GRP8 monomer increases the amount of condensation droplets. As the concentration of macromolecules in the solution is increased to the solubility limit the interactions between the macromolecules will become stronger than the interactions between the macromolecules and the solvent, and as a result, this solution will gain propensity to LLPS (Weber and Brangwynne, 2015). Figure 38d shows 50 µM of GRP8 monomer at 4°C, where it is observed a slight increase in size of the condensation droplets. This might suggest that cold temperatures also induce the accumulation of condensation droplets. Figures 38e and f are observed centrifuged condensation droplets showed in 38d, where their size is noticeably bigger. Moreover, it could be speculated that GRP8 bound to RNA forms LLPS and might form an RNA granule. To confirm the presence of an RNA granule, labeling of GRP8 is necessary for the detection of the RNA inside of the condensation droplet with GRP8. It would be interesting to report if GRP8 is forming LLPS without the RNA, however the labeling of the protein is also necessary.

It has been described that RNA can not only drive LLPS via electrostatic interactions, but repetitive intermolecular base pairing can also achieve multivalency and thus drive the formation of clusters *in vitro* and *in vivo* (Jain and Vale, 2017; Tong *et al.*, 2022). Phase separation and RNA is a two-way regulation. This can recruit RNA by enriching the protein which can bind to RNA and realize the directed transcription of RNA (resolving). On the other hand, the asymmetric distribution of RNA can guide the local translation of proteins, achieve a high local concentration of specific proteins, and provide conditions for the occurrence of phase separation (assembling) (Guo, Shi and Wang, 2021). Furthermore, it would be interesting to test how the RNA is resolving or assembling GRP8 LLPS and if there are different ways to assemble and resolve these speculated RNA-granules. It has been reported that modifications in the PrLD including the phosphorylation of the tyrosine sites in the PrLD inhibits the LLPS formation (Xu *et al.*, 2022), and therefore it would be interesting to test if the condensation formation is abolished by the modification of these sites in GRP8.

## 4.7 Protein disorder and structural characterization

Complementary approaches with low- and high-resolution techniques are considerably important due to the dynamic nature of IDPs (Kosol *et al.*, 2013). Therefore, it was attempted to crystallize GRP8 monomer for high resolution structural characterization (chapter 3.3.2). GRP8$^{short}$ was not considered for crystallization trials hence it only incorporates the RRM, which has been described as a conserved domain across bacteria, archaea and eukaryotes (Gerstberger, Hafner and Tuschl, 2014) and has been previously characterized in other proteins. Moreover, figure 32 showed the GRP8 monomer stable conformation before crystallographic trials by DLS scan. The crystallization trials GRP8 were unsuccessful which could be attributed to its large C-terminus is highly disordered, which prevented GRP8 from forming crystals. Because IDPs do not crystallize and have diverse conformers, traditional experimental methods such as crystallization can hardly capture their conformation ensemble and just provide average structural characters of IDPs (Mu, Pan and Chen, 2021). Other methods might be more suitable for a high-resolution characterization like co-crystallization of RNA-protein complexes (Chen and Pollack, 2016) or nuclear magnetic resonance (NMR). Co-crystallization was thought to work in the case that the disordered part of GRP8 becomes structured after the interaction with RNA. Because PEG is a major precipitation ingredient in crystallography (Bonneté, 2007) and has been described as an LLPS inductor, the co-crystallization of GRP8 and RNA would likely not work. NMR can be used to monitor the dynamic behavior of a protein at a multitude of specific sites and can potentially be used to study the dynamical nature of long IDRs including the entire disordered protein. Whereas X-ray crystallography involves less fluctuating structures, and thus IDRs found in X-ray structures tend to be limited to relatively short segments (Torchia, 2007; Jacques and Trewhella, 2010; Ota *et al.*, 2013) which commonly leads to missing residues. Interestingly, when searching for protein homologs already characterized of GRP8 in the *Protein Data Bank* (PDB), most of them have been characterized in NMR, which suggests that this could be a great tool to further characterize GRP8.

## 4.8 GRP8 structure

### 4.8.1 Cold induction in SAXS-batch mode

SAXS remains one of the few structural techniques that can probe macromolecular architecture and dynamics without size limitation under native solution conditions (Brosey and Tainer, 2019). For IDPs, SAXS is the method most often used because is fast method that requires less material and usually provides more precise experimental data, and it is a well suited tool to rapidly monitor large structural perturbations in proteins upon environmental changes

(Bernadó and Svergun, 2012). In the first SAXS experiment (chapter 3.3.5.1), different temperatures were tested to understand the structural behavior of GRP8 upon cold induction. Batch-SAXS experiments troubled in having buffer mismatch for buffer subtraction, resulting in negative scattering values. This explains the negative values in the plots presented. Nevertheless, qualitative analysis could be performed by observing the overall curve behavior of the plots. When compared the three different temperatures measured with GRP8 monomer, it was observed that when the GRP8 monomer was cold treated, it formed aggregation through the Guinier analysis plot (figure 41a) and sRg limit values (table 31). The sRg limits of the GRP8 monomer measured at 5°C were 0.77 to 1.30 which was above the limit, and it is an indication of protein aggregation (Svergun, 1987; Putnam *et al.*, 2007; Jacques and Trewhella, 2010). It was clear that when comparing all the pair distribution plots, GRP8 monomer was showing a multidomain shape (figures 39e, 40e and 41e) but when the temperature dropped, the extension of the curve was larger, as well as the Kratky plots were showing GRP8 monomer becoming more flexible and disordered (figure 42). Although GRP8 monomer with RNA was not tested in SAXS but only with the LLPS assay, it would be interesting to test the interaction between them. Experiments with PrLD proteins and nucleic acid interactions have been reported (Lima *et al.*, 2006; Silva *et al.*, 2008; Marques *et al.*, 2009; Matos, *et al.*, 2019), nevertheless complementary techniques such as NMR and CryoEM were also used and might be necessary for the characterization of these interactions.

### 4.8.2 GRP8 monomer and dimer native structure

To understand the native structure of GRP8 monomer and dimer, both were tested by SEC-SAXS. This technique has been well established hence it reduces the effects on sample heterogeneity including the often unpreventable formation of higher oligomeric species or aggregation over the time for sample preparation (Graewert *et al.*, 2020), specially for IDPs that are known to be unstable. In addition, buffer subtraction was successful this time, which allowed the qualitative and quantitative analysis the samples.

GRP8 monomer and dimer where first measured by DLS to confirm that the proteins were present in monomeric and pure solution (figure 29 and 30). The samples measured by DLS were not the same as the ones used for SAXS experiments. For them, the quality check was performed by the *EMBL* after handling the samples. The MWs estimated by DLS scans showed bigger sizes which might indicate an elongated protein. GRP8$^{short}$ is the only sample that showed the correct MW (figure 31), which might indicate that the elongation comes from the

glycine-rich region. Elongation is given by the presence of highly extended conformations of the protein, and unstructured/disordered proteins are characterized by large average sizes compared to globular proteins (Bernadó and Svergun, 2012). Moreover, MWs calculated by SEC-SAXS showed a more accurate estimation confirming the presence of GRP8 monomer and dimer forms. P(r) distribution plots revealed a difference between GRP8 monomer and dimer, showing a clear multidomain shape curve for GRP8 monomer versus for GRP8 dimer showed a less sharp multidomain curve (figure 43g, 44g and 45a). GRP8 dimer versus GRP8 monomer results showed that the GRP8 monomer presents a more unfolded and disordered shape than GRP8 dimer (figure 43h, 44h and 45b). Although both proteins showed a similar behavior with the Kratky plots, it is observed this slight difference between them (figure 45b). Moreover, after the SEC-SAXS data was processed, it was proceeded to the 3D modeling of GRP8 monomer.

## 4.9 3D model of GRP8

Because of the previously described problem with the buffer subtraction with batch SAXS, only SEC-SAXS data could be used for the modeling of GRP8.

IDPs can sample an astronomical number of conformations. The scattering profile of an IDP is the average of all those arising from the conformations that the protein adopts in solution (Bernadó and Svergun, 2012). Although traditionally considered a low-resolution technique, high-resolution differences in macromolecular conformations can be reliably detected by quantitative comparison of X-ray scattering profiles or SAXS-constrained modeling (Brosey and Tainer, 2019). Data from complementary techniques can be a powerful source of restraints to greatly improve the confidence in the uniqueness of best-fit models, as small-angle scattering provides information on the large-scale features, it is often possible to model proteins based on high-resolution structures of domains, either of the protein in question or from structural homologues (Jacques and Trewhella, 2010). *Alphafold* and *i-Tasser* were used to predict the 3D structure, only the created models from *i-Tasser* where matching. Because *Alphafold* is an artificial intelligence (AI) structure prediction based on deep learning, it is not enough for IDP structure prediction (Strodel, 2021), specially because of its low confidence in the low complexity regions of GRP8. When using the models from *i-Tasser*, the fitting was promising showing better values. This could be explained because this tool uses previously characterized proteins that are found in the PDB. Different tools available were used for this modeling process including *CRYSOL*, *SREFLEX* and *EOM* to match the processing data and reconstruct

GRP8 through computational approaches to generate a hybrid refined model. *SREFLEX* was the only one to match the data given. Because the method starts from a given conformation of the protein, which does not necessarily agree with the SAXS data, the structure is partitioned into pseudo-domains either using structural classification databases or automatically from the protein dynamics to explore the conformational space of high-resolution models and refine the agreement with the experimental SAXS data (Panjkovich and Svergun, 2016a).

The final model showed in figure 46, represents the combination of different techniques including experimental data measured through SEC-SAXS and computational tools that were used to create this hybrid model of GRP8 monomer. Although the model might be inconclusive because of the nature of GRP8 and its IDP behavior, it proved that the known *Alphafold* model is far from being accurate. Further structural characterization of GRP8 remains necessary to archive a more precise model at the local atomic level with high-resolution techniques, for example with NMR. In the meantime, this purposed GRP8 model shows the most accuracy than the ones created only by computational tools or by comparing homolog proteins.

## 5. Conclusion and outlook

GRP8 structure contains an N terminus with a RRM and a C-terminus with a glycine-rich domain. The last one is a highly disordered and low complexity region which was attributed to cause aggregation and degradation while performing its purification. GRP8 was sensitive to the environment and throughout the process of protein purification three major hallmarks were revealed: time length of the purification, protein concentration, and environmental temperature. Although the glycine-rich domain brought so many troubleshooting steps, it was accounted for having a key role in the functionality of GRP8. When comparing GRP8 and GRP8$^{short}$ binding capacities were severely compromised when the glycine-rich domain was missing. This was supported by GRP8$^{short}$ only binding small RNAs and none of the mRNAs tested by MST and showed when performed the CnBr-Sepharose affinity column. In addition, GRP8 was observed to bind different interaction partners. *In silico* analysis showed that there were no common motifs between the RNAs tested by MST, suggesting that GRP8 is binding non-selectively while binding a wide range of RNAs. However, MST analysis shows that GRP8 still binds with a significant higher affinity to smaller RNAs. Furthermore, it was observed that GRP8 was binding UTRs of two transcripts. The 5'UTRs have been described to incorporate a potential RNA binding sites and are also involved in the stabilization of mRNA and the regulation of gene expression (Araujo *et al.*, 2012; Singh *et al.*, 2012), and the 3'UTRs have been described as a repository of regulatory elements for mRNA stability, intracellular localization and translation (Szostak and Gebauer, 2013). GRP8 was observed to bind with a very high affinity to the *3'UTR-GPR7* miRNA which might indicate its involvement in translational control. This is supported by GRP8 involved in the regulation of the turnover of its own transcript and protein concentration, as well as the cross-regulation of GRP7 concentration (Staiger *et al.*, 2003; Schöning *et al.*, 2008). It would be interesting to test further RNAs with their UTRs. For example, the UTRs of the RNAs that were already tested by MST and showed to be enriched to GRP8 by the CnBr-Sepharose GRP8 bound affinity column and RNA-sequencing analysis. Another interesting thing would be to test further small RNAs with GRP8$^{short}$, to confirm if the RRM prefers sequences rich in G or G/U and sequences. From the RNA-sequencing analysis, there were found several chlorophyll a-b binding protein transcripts present in input and GRP8-elution samples. Further elucidation of this transcript, including binding studies, would be interesting since it has been shown to be present in the sap of different plant species. In addition, it would be interesting to perform UV-crosslinking with GRP8 and other RNAs, to observe where the interaction occurs in GRP8.

As previously described, GRP8 structure includes a low complexity region with a glycine-rich domain. These regions are known to promote phase separation into protein-rich liquid-like droplets (Nott *et al.*, 2015). Because of the composition of the low complexity region of GRP8 including a RGG and a [G/S]Y[G/S] motif, a PrLD prediction was calculated. This resulted in a possible prion-like region contained in the glycine-rich domain. This was interesting to test hence the elucidation of PrLD are emerging protein plant studies. Therefore, ThT and LLPS assays were performed as an effort to report this. Through these assays it was observed that there was indication of condensation formation. Moreover, further testing remains necessary. Next steps should include the labeling of GRP8 for the confirmation of the formation of an RNA granule and to test GRP8 itself forming condensation droplets including different conditions such as different temperatures and protein concentration. In addition, it would be interesting to test the RNA-GRP8 interaction and LLPS formation, if the RNA is resolving or assembling GRP8 LLPS and if there are different ways to assemble and resolve these speculated RNA-granules. The interaction GRP8 and RNA could also be structurally characterized by SAXS complemented with higher resolution methods such as NMR and CryoEM (Lima *et al.*, 2006; Silva *et al.*, 2008; Marques *et al.*, 2009; Matos *et al.*, 2019; Fitzpatrick and Saibil, 2019). In addition, It has been reported that phosphorylation of the PrLD inhibits the LLPS formation, therefore it would be interesting to test if the condensation is abolished by this modification. As a future perspective, because GRP8 has been detected phloem sap of *A. thaliana*, it would be interesting to elucidate the movement and formation of these speculated RNA-granules to understand their function *in planta*. Since there is no translation occurring in the phloem, it might be that GRP8 is involved in long-distance signaling, forming RNP complexes and these are translocated to their destination through the phloem. Chaperone assays with GRP8 bound to RNA might be also a first start to understand this.

Structurally, results showed that GRP8 is a multidomain protein with a partially unfolded and disordered shape. It was observed that the elongation of GRP8 was most-likely attributed to the glycine-rich domain. The overall structural analysis and predictions performed to describe GRP8 showed that the glycine-rich domain is a flexible and disordered region which presumably gives the IDP character. This region was accounted to promote the RNA-binding and to form condensation droplets, which has previously described as a common characteristic for certain IDPs that are RNA-binding and contain PrLDs in their structure. Moreover, a hybrid model of GRP8 monomer was proposed by combination of the SAXS processed data and computational tools. This model is more accurate that the one existing and calculated by

*AlphaFold*. Because of the IDR present, it is hard to predict the 3D structure. Nevertheless, NMR was proposed as a high-resolution tool to complement and refine this new hybrid model of GRP8 proposed into a more elegant one

# 6. Literature

Ambrosone, A. *et al.* (2012) “Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints,” *Plant Science*, 182(1), pp. 12–18. doi: 10.1016/j.plantsci.2011.02.004.

Araujo, P. R. *et al.* (2012) “Before it gets started: Regulating translation at the 5′; UTR,” *Comparative and Functional Genomics*, 2012. doi: 10.1155/2012/475731.

Auweter, S. D., Oberstrass, F. C. and Allain, F. H. T. (2006) “Sequence-specific binding of single-stranded RNA: Is there a code for recognition?,” *Nucleic Acids Research*, 34(17), pp. 4943–4959. doi: 10.1093/nar/gkl620.

Bailey, T. L. *et al.* (2015) “The MEME Suite,” *Nucleic Acids Research*, 43(W1), pp. W39–W49. doi: 10.1093/nar/gkv416.

Batailler, B. *et al.* (2012) “Soluble and filamentous proteins in Arabidopsis sieve elements,” *Plant, Cell and Environment*, 35(7), pp. 1258–1273. doi: 10.1111/j.1365-3040.2012.02487.x.

Van Bel, A. J. E. (2003) “The phloem, a miracle of ingenuity,” *Plant, Cell and Environment*, 26(1), pp. 125–149. doi: 10.1046/j.1365-3040.2003.00963.x.

Bernadó, P. and Svergun, D. I. (2012) “Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering,” *Molecular BioSystems*, 8(1), pp. 151–167. doi: 10.1039/c1mb05275f.

Biancalana, M. and Koide, S. (2010) “Molecular mechanism of Thioflavin-T binding to amyloid fibrils,” *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(7), pp. 1405–1412. doi: 10.1016/j.bbapap.2010.04.001.

Bonneté, F. (2007) “Colloidal approach analysis of the marseille protein crystallization database for protein crystallization strategies,” *Crystal Growth and Design*, 7(11), pp. 2176–2181. doi: 10.1021/cg700711a.

Brosey, C. A. and Tainer, J. A. (2019) “Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology,” *Current Opinion in Structural Biology*, 58, pp. 197–213. doi: 10.1016/j.sbi.2019.04.004.

Buhtz, A. *et al.* (2008) "Identification and characterization of small RNAs from the phloem of Brassica napus," *Plant Journal*, 53(5), pp. 739–749. doi: 10.1111/j.1365-313X.2007.03368.x.

Burd, C. G. and Dreyfuss, G. (1994) "Conserved structures and diversity of functions of RNA-binding proteins," *Science*, 265(5172), pp. 615–621. doi: 10.1126/science.8036511.

Campos-Melo, D. *et al.* (2021) "The Integral Role of RNA in Stress Granule Formation and Function," *Frontiers in Cell and Developmental Biology*, 9(May), pp. 1–19. doi: 10.3389/fcell.2021.621779.

Cazenave, C. and Uhlenbeck, O. C. (1994) "RNA template-directed RNA synthesis by T7 RNA polymerase," *Proceedings of the National Academy of Sciences of the United States of America*, 91(15), pp. 6972–6976. doi: 10.1073/pnas.91.15.6972.

Chen, Y. and Pollack, L. (2016) "SAXS studies of RNA: structures, dynamics, and interactions with partners," *Wiley Interdisciplinary Reviews: RNA*, 7(4), pp. 512–526. doi: 10.1002/wrna.1349.

Chen, Z. J. *et al.* (2004) "The development of an Arabidopsis model system for genome-wide analysis of polyploidy effects," *Biological Journal of the Linnean Society*, 82(4), pp. 689–700. doi: 10.1111/j.1095-8312.2004.00351.x.

Chong, P. A., Vernon, R. M. and Forman-Kay, J. D. (2018) "RGG/RG Motif Regions in RNA Binding and Phase Separation," *Journal of Molecular Biology*, 430(23), pp. 4650–4665. doi: 10.1016/j.jmb.2018.06.014.

Clifford D. Carpenter, Joel A. Kreps, and A. E. S. (1994) "Cold Treatment and an Endogenous Circadian Rhythm '," *Plant physiology*, 104, pp. 1015–1025.

Curtis, R. A. *et al.* (2002) "Protein-protein interactions in concentrated electrolyte solutions: Hofmeister-series effects," *Biotechnology and Bioengineering*, 79(4), pp. 367–380. doi: 10.1002/bit.10342.

Czolpinska, M. and Rurek, M. (2018) "Plant glycine-rich proteins in stress response: An emerging, still prospective story," *Frontiers in Plant Science*, 9(March), pp. 1–13. doi: 10.3389/fpls.2018.00302.

Davis, R. B. *et al.* (2021) "FUS oncofusion protein condensates recruit mSWI/SNF chromatin remodeler via heterotypic interactions between prion-like domains," *Protein Science*, 30(7), pp. 1454–1466. doi: 10.1002/pro.4127.

Deeken, R. *et al.* (2008) “Identification of Arabidopsis thaliana phloem RNAs provides a search criterion for phloem-based transcripts hidden in complex datasets of microarray experiments,” *Plant Journal*, 55(5), pp. 746–759. doi: 10.1111/j.1365-313X.2008.03555.x.

Diering, Maxson & Mitchell and Freeman (2018) *乳鼠心肌提取 HHS Public Access*, *Physiology & behavior*. doi: 10.1007/978-3-319-29073-7.

Dijkman, P. M. *et al.* (2014) “Interactions under Previously Challenging Conditions,” *Methods*, 59(3), pp. 301–315. doi: 10.1016/j.ymeth.2012.12.005.Microscale.

Dinant, S. *et al.* (2010) “Phloem sap intricacy and interplay with aphid feeding,” *Comptes Rendus - Biologies*, 333(6–7), pp. 504–515. doi: 10.1016/j.crvi.2010.03.008.

Doering-Saad, C. *et al.* (2002) “Use of aphid stylectomy and RT-PCR for the detection of transporter mRNAs in sieve elements,” *Journal of Experimental Botany*, 53(369), pp. 631–637. doi: 10.1093/jexbot/53.369.631.

Doering-Saad, C. *et al.* (2006) “A phloem-enriched cDNA library from Ricinus: Insights into phloem function,” *Journal of Experimental Botany*, 57(12), pp. 3183–3193. doi: 10.1093/jxb/erl082.

Duhr, S. and Braun, D. (2006) “Why molecules move along a temperature gradient,” *Proceedings of the National Academy of Sciences of the United States of America*, 103(52), pp. 19678–19682. doi: 10.1073/pnas.0603873103.

Dyballa, N. and Metzger, S. (2009) “Fast and sensitive colloidal Coomassie G-250 staining for proteins in polyacrylamide gels,” *Journal of Visualized Experiments*, (30), pp. 2–5. doi: 10.3791/1431.

Fernandez-Gomez, F. *et al.* (2019) *Myotonic Dystrophy: an RNA Toxic Gain of Function Tauopathy?*, *Advances in Experimental Medicine and Biology*. doi: 10.1007/978-981-32-9358-8_17.

Fitzpatrick, A. W. and Saibil, H. R. (2019) “Cryo-EM of amyloid fibrils and cellular aggregates,” *Current Opinion in Structural Biology*, 58, pp. 34–42. doi: 10.1016/j.sbi.2019.05.003.

Franke, D. and Svergun, D. I. (2009) “DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering,” *Journal of Applied Crystallography*, 42(2), pp. 342–346. doi: 10.1107/S0021889809000338.

Franzmann, T. M. and Alberti, S. (2019) “Protein phase separation as a stress survival strategy,” *Cold Spring Harbor Perspectives in Medicine*, 9(6). doi: 10.1101/cshperspect.a034058

.

Fu, Z. Q. *et al.* (2007) “A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity,” *Nature*, 447(7142), pp. 284–288. doi: 10.1038/nature05737.

Gaupels, F. *et al.* (2008) “Adaptation of aphid stylectomy for analyses of proteins and mRNAs in barley phloem sap,” *Journal of Experimental Botany*, 59(12), pp. 3297–3306. doi: 10.1093/jxb/ern181.

Gerstberger, S., Hafner, M. and Tuschl, T. (2014) “A census of human RNA-binding proteins,” *Nature Reviews Genetics*, 15(12), pp. 829–845. doi: 10.1038/nrg3813.

Ghosh, A. and Zhou, H.-X. (2020) “Fusion Speed of Biomolecular Condensates,” *bioRxiv*, 56, p. 2020.06.22.164897. Available at: https://www.biorxiv.org/content/10.1101/2020.06.22.164897v1%0Ahttps://www.biorxiv.org/content/10.1101/2020.06.22.164897v1.abstract.

Giavalisco, P. *et al.* (2006) “Towards the proteome of Brassica napus phloem sap,” *Proteomics*, 6(3), pp. 896–909. doi: 10.1002/pmic.200500155.

Graewert, M. A. *et al.* (2020) “Adding Size Exclusion Chromatography ( SEC ) and,” *Crystals*, 10(11), pp. 975–993. Available at: https://www.mdpi.com/2073-4352/10/11/975?utm_source=researcher_app&utm_medium=referral&utm_campaign=RESR_MRKT_Researcher_inbound.

Guan, D. *et al.* (2017) “PlaMoM: A comprehensive database compiles plant mobile macromolecules,” *Nucleic Acids Research*, 45(D1), pp. D1021–D1028. doi: 10.1093/nar/gkw988.

Guo, Q., Shi, X. and Wang, X. (2021) “RNA and liquid-liquid phase separation,” *Non-coding RNA Research*, 6(2), pp. 92–99. doi: 10.1016/j.ncrna.2021.04.003.

Habchi, J. *et al.* (2014) “Introducing protein intrinsic disorder,” *Chemical Reviews*, 114(13), pp. 6561–6588. doi: 10.1021/cr400514h.

Ham, B. K. *et al.* (2009) “A polypyrimidine tract binding protein, pumpkin RBP50, forms the basis of a phloem-mobile ribonucleoprotein complex,” *Plant Cell*, 21(1), pp. 197–215. doi: 10.1105/tpc.108.061317.

Han, T. W. *et al.* (2012) "Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies," *Cell*, 149(4), pp. 768–779. doi: 10.1016/j.cell.2012.04.016.

Haywood, V. *et al.* (2005) "Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development," *Plant Journal*, 42(1), pp. 49–68. doi: 10.1111/j.1365-313X.2005.02351.x.

Heintzen, C. *et al.* (1997) "AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in Arabidopsis thaliana," *Proceedings of the National Academy of Sciences of the United States of America*, 94(16), pp. 8515–8520. doi: 10.1073/pnas.94.16.8515.

Hentze, M. W. *et al.* (2018) "A brave new world of RNA-binding proteins," *Nature Reviews Molecular Cell Biology*, 19(5), pp. 327–341. doi: 10.1038/nrm.2017.130.

Horvath, D. P. and Olson, P. A. (1998) "Cloning and characterization of cold-regulated glycine-rich RNA-binding protein genes from leafy spurge (Euphorbia esula L.) and comparison to heterologous genomic clones," *Plant Molecular Biology*, 38(4), pp. 531–538. doi: 10.1023/A:1006050208670.

Hura, G. L. *et al.* (2009) "Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)," *Nature Methods*, 6(8), pp. 606–612. doi: 10.1038/nmeth.1353.

Hyman, A. A., Weber, C. A. and Jülicher, F. (2014) "Liquid-liquid phase separation in biology," *Annual review of cell and developmental biology*, 30, pp. 39–58. doi: 10.1146/annurev-cellbio-100913-013325.

Inoue, H., Nojima, H. and Okayama, H. (1990) "High efficiency transformation of Escherichia coli with plasmids," *Gene*, 96(1), pp. 23–28. doi: 10.1016/0378-1119(90)90336-P.

Jacques, D. A. and Trewhella, J. (2010) "Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls," *Protein Science*, 19(4), pp. 642–657. doi: 10.1002/pro.351.

Jain, A. and Vale, R. D. (2017) "RNA phase transitions in repeat expansion disorders," *Nature*, 546(7657), pp. 243–247. doi: 10.1038/nature22386.

Jang, G. J., Jang, J. C. and Wu, S. H. (2020) "Dynamics and functions of stress granules and processing bodies in plants," *Plants*, 9(9), pp. 1–11. doi: 10.3390/plants9091122.

Järvelin, A. I. *et al.* (2016) "The new (dis)order in RNA regulation," *Cell Communication and*

*Signaling*, 14(1). doi: 10.1186/s12964-016-0132-3.

Jerabek-Willemsen, M. *et al.* (2014) "MicroScale Thermophoresis: Interaction analysis and beyond," *Journal of Molecular Structure*, 1077, pp. 101–113. doi: 10.1016/j.molstruc.2014.03.009.

Jumper, J. *et al.* (2021) "Highly accurate protein structure prediction with AlphaFold," *Nature*, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.

Jung, J. H. *et al.* (2020) "A prion-like domain in ELF3 functions as a thermosensor in Arabidopsis," *Nature*, 585(7824), pp. 256–260. doi: 10.1038/s41586-020-2644-7.

Kato, M. *et al.* (2012) "Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels," *Cell*, 149(4), pp. 753–767. doi: 10.1016/j.cell.2012.04.017.

Kaur, T. *et al.* (2019) "Molecular crowding tunes material states of ribonucleoprotein condensates," *Biomolecules*, 9(2), pp. 1–17. doi: 10.3390/biom9020071.

Kedersha, N. *et al.* (2000) "Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules," *Journal of Cell Biology*, 151(6), pp. 1257–1268. doi: 10.1083/jcb.151.6.1257.

Kehr, J. and Kragler, F. (2018) "Long distance RNA movement," *New Phytologist*, 218(1), pp. 29–40. doi: 10.1111/nph.15025.

Kehr, J., Morris, R. J. and Kragler, F. (2022) "Long-Distance Transported RNAs: From Identity to Function," *Annual Review of Plant Biology*, 73, pp. 457–474. doi: 10.1146/annurev-arplant-070121-033601.

Kikhney, A. G. and Svergun, D. I. (2015) "A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins," *FEBS Letters*, 589(19), pp. 2570–2577. doi: 10.1016/j.febslet.2015.08.027.

Kiledjian, M. and Dreyfuss, G. (1992) "Primary structure and binding activity of the hnRNP U protein: Binding RNA through RGG box," *EMBO Journal*, 11(7), pp. 2655–2664. doi: 10.1002/j.1460-2075.1992.tb05331.x.

Kinsella, M. and Monk, C. (2012) "基因的改变NIH Public Access," 23(1), pp. 1–7. doi: 10.1016/j.brainres.2012.01.016.The.

Konarev, P. V. *et al.* (2003) "PRIMUS: A Windows PC-based system for small-angle scattering data analysis," *Journal of Applied Crystallography*, 36(5), pp. 1277–1282. doi: 10.1107/S0021889803012779.

Kosol, S. *et al.* (2013) "Structural characterization of intrinsically disordered proteins by NMR spectroscopy," *Molecules*, 18(9), pp. 10802–10828. doi: 10.3390/molecules180910802.

Kovalskaya, N. *et al.* (2014) "Application of a modified EDTA-mediated exudation technique and guttation fluid analysis for Potato spindle tuber viroid RNA detection in tomato plants (Solanum lycopersicum)," *Journal of Virological Methods*, 198, pp. 75–81. doi: 10.1016/j.jviromet.2013.12.015.

Kyung, J. K., Yeon, O. K. and Kang, H. (2005) "Characterization of transgenic Arabidopsis plants overexpressing GR-RBP4 under high salinity, dehydration, or cold stress," *Journal of Experimental Botany*, 56(421), pp. 3007–3016. doi: 10.1093/jxb/eri298.

Laemmli, U. K. (1970) "227680a0," *Nature*, 227, pp. 680–685.

Lancaster, A. K. *et al.* (2014) "PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition," *Bioinformatics*, 30(17), pp. 2501–2502. doi: 10.1093/bioinformatics/btu310.

Lee, N. and Wiegand, S. (2020) "Thermophoretic micron-scale devices: Practical approach and review," *Entropy*, 22(9), pp. 1–24. doi: 10.3390/e22090950.

Lima, L. M. T. R. *et al.* (2006) "Structural insights into the interaction between prion protein and nucleic acid," *Biochemistry*, 45(30), pp. 9180–9187. doi: 10.1021/bi060532d.

Lin, M. K. *et al.* (2009) "Analysis of the pumpkin phloem proteome provides insights into angiosperm sieve tube function," *Molecular and Cellular Proteomics*, 8(2), pp. 343–356. doi: 10.1074/mcp.M800420-MCP200.

Lindwall, G. *et al.* (2000) "A sparse matrix approach to the solubilization of overexpressed proteins," *Protein Engineering*, 13(1), pp. 67–71. doi: 10.1093/protein/13.1.67.

Lunde, B. M., Moore, C. and Varani, G. (2007) “RNA-binding proteins: Modular design for efficient function,” *Nature Reviews Molecular Cell Biology*, 8(6), pp. 479–490. doi: 10.1038/nrm2178.

Luo, M. and Reed, R. (2003) “ Identification of RNA Binding Proteins by UV Cross-Linking ,” *Current Protocols in Molecular Biology*, 63(1), pp. 1–10. doi: 10.1002/0471142727.mb2702s63.

Maizel, A. *et al.* (2020) “To move or not to move: roles and specificity of plant RNA mobility,” *Current Opinion in Plant Biology*, 57, pp. 52–60. doi: 10.1016/j.pbi.2020.05.005.

Manalastas-Cantos, K. *et al.* (2021) “ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis,” *Journal of Applied Crystallography*, 54, pp. 343–355. doi: 10.1107/S1600576720013412.

Mangeon, A., Junqueira, R. M. and Sachetto-Martins, G. (no date) “Functional diversity of the plant glycine-rich proteins superfamily,” *Plant Signaling and Behavior*, 5(2), pp. 99–104. doi: 10.4161/psb.5.2.10336.

Marques, A. F. *et al.* (2009) “Enhanced prion protein stability coupled to DNA recognition and milieu acidification,” *Biophysical Chemistry*, 141(2–3), pp. 135–139. doi: 10.1016/j.bpc.2008.12.011.

Maruri-López, I. *et al.* (2021) “Plant Stress Granules: Trends and Beyond,” *Frontiers in Plant Science*, 12(August), pp. 1–16. doi: 10.3389/fpls.2021.722643.

Meyerowitz, E. M. (1989) “Arabidopsis, a useful weed,” *Cell*, 56(2), pp. 263–269. doi: 10.1016/0092-8674(89)90900-8.

Meyerowitz, E. M. and Pruitt, R. E. (1985) “Meyerowitz1985,” 427(1983).

Morris, R. J. (2018) “On the selectivity, specificity and signalling potential of the long-distance movement of messenger RNA,” *Current Opinion in Plant Biology*, 43, pp. 1–7. doi: 10.1016/j.pbi.2017.11.001.

Mu, J., Pan, Z. and Chen, H. F. (2021) “Balanced Solvent Model for Intrinsically Disordered and Ordered Proteins,” *Journal of Chemical Information and Modeling*, 61(10), pp. 5141–5151. doi: 10.1021/acs.jcim.1c00407.

Mylonas, E. and Petoukhov, M. V (2007) “Structural characterization of flexible proteins using SAXS,”

*Journal of the American Chemical Society*, (2), pp. 5656–5664.

Notaguchi, M. (2015) "Identification of phloem-mobile mRNA," *Journal of Plant Research*, 128(1), pp. 27–35. doi: 10.1007/s10265-014-0675-6.

Nott, T. J. *et al.* (2015) "Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles," *Molecular Cell*, 57(5), pp. 936–947. doi: 10.1016/j.molcel.2015.01.013.

Omid, A. *et al.* (2007) "Characterization of phloem-sap transcription profile in melon plants," *Journal of Experimental Botany*, 58(13), pp. 3645–3656. doi: 10.1093/jxb/erm214.

Ostendorp, A. *et al.* (2017) "Functional analysis of Brassica napus phloem protein and ribonucleoprotein complexes," *New Phytologist*, 214(3), pp. 1188–1197. doi: 10.1111/nph.14405.

Ota, M. *et al.* (2013) "An assignment of intrinsically disordered regions of proteins based on NMR structures," *Journal of Structural Biology*, 181(1), pp. 29–36. doi: 10.1016/j.jsb.2012.10.017.

Pahlow, S. *et al.* (2018) "Phloem sap sampling from Brassica napus for 3D-PAGE of protein and ribonucleoprotein complexes," *Journal of Visualized Experiments*, 2018(131), pp. 1–11. doi: 10.3791/57097.

Panjkovich, A. and Svergun, D. I. (2016a) "Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis," *Physical Chemistry Chemical Physics*, 18(8), pp. 5707–5719. doi: 10.1039/c5cp04540a.

Panjkovich, A. and Svergun, D. I. (2016b) "SASpy: A PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data," *Bioinformatics*, 32(13), pp. 2062–2064. doi: 10.1093/bioinformatics/btw071.

Panjkovich, A. and Svergun, D. I. (2018) "CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data," *Bioinformatics*, 34(11), pp. 1944–1946. doi: 10.1093/bioinformatics/btx846.

Pantoja-Uceda, D. *et al.* (2021) "Phe-Gly motifs drive fibrillization of TDP-43's prion-like domain condensates," *PLoS Biology*, 19(4), pp. 1–17. doi: 10.1371/journal.pbio.3001198.

Parkin, I. A. P. *et al.* (2005) "Segmental structure of the Brassica napus genome based on comparative analysis with Arabidopsis thaliana," *Genetics*, 171(2), pp. 765–781. doi: 10.1534/genetics.105.042093.

Paterson, A. H. *et al.* (2001) "Brassica genomics: A complement to, and early beneficiary of, the Arabidopsis sequence," *Genome Biology*, 2(3), pp. 1–4.

Pazos, F. *et al.* (2013) "Protein intrinsic disorder in plants," *Frontiers in Plant Science*, 4(SEP), pp. 1–5. doi: 10.3389/fpls.2013.00363.

Prilusky, J. *et al.* (2005) "FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded," *Bioinformatics*, 21(16), pp. 3435–3438. doi: 10.1093/bioinformatics/bti537.

Psimadas, D. *et al.* (2012) "Molecular Nanomedicine Towards Cancer :," *Journal of pharmaceutical sciences*, 101(7), pp. 2271–2280. doi: 10.1002/jps.

Putnam, C. D. *et al.* (2007) "X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution," *Quarterly Reviews of Biophysics*, 40(3), pp. 191–285. doi: 10.1017/S0033583507004635.

Quérouil, S. *et al.* (2015) "Development and characterization of polymorphic microsatellite markers in neotropical fish of the genus Apistogramma (Perciformes: Labroidei: Cichlidae)," *Journal of Applied Ichthyology*, 31, pp. 52–56. doi: 10.1111/jai.12975.

Ray Mc Dermott, † *et al.* (2002) "Evidence That Ternary Complex (eIF2-GTP-tRNAi Met)– Deficient Preinitiation Complexes Are Core Constituents of Mammalian Stress Granules," *Molecular biology of the cell*, 13(6), pp. 2170–2179. doi: 10.1091/mbc.01.

Receveur-Brechot, V. and Durand, D. (2012) "How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective," *Current Protein & Peptide Science*, 13(1), pp. 55–75. doi: 10.2174/138920312799277901.

Reumann, S. *et al.* (2007) "Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms," *Plant Cell*, 19(10), pp. 3170–3193. doi: 10.1105/tpc.107.050989.

Ruiz-Medrano, R. *et al.* (2007) "Influence of cucumber mosaic virus infection on the mRNA population present in the phloem translocation stream of pumpkin plants," *Functional Plant Biology*, 34(4), pp.

292–301. doi: 10.1071/FP06300.

Ruiz-Medrano, R., Xoconostle-Cázares, B. and Lucas, W. J. (1999) “Phloem long-distance transport of CmNACP mRNA: Implications for supracellular regulation in plants,” *Development*, 126(20), pp. 4405–4419. doi: 10.1242/dev.126.20.4405.

Sachetto-Martins, G., Franco, L. O. and De Oliveira, D. E. (2000) “Plant glycine-rich proteins: A family or just proteins with a common motif?,” *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1492(1), pp. 1–14. doi: 10.1016/S0167-4781(00)00064-6.

Sanulli, S. and Narlikar, G. J. (2021) “Generation and Biochemical Characterization of Phase-Separated Droplets Formed by Nucleic Acid Binding Proteins: Using HP1 as a Model System,” *Current Protocols*, 1(5). doi: 10.1002/cpz1.109.

Sasaki, T. *et al.* (1998) “Detection of several mRNA species in rice phloem sap,” *Plant and Cell Physiology*, 39(8), pp. 895–897. doi: 10.1093/oxfordjournals.pcp.a029451.

Schmal, C., Reimann, P. and Staiger, D. (2013) “A Circadian Clock-Regulated Toggle Switch Explains AtGRP7 and AtGRP8 Oscillations in Arabidopsis thaliana,” *PLoS Computational Biology*, 9(3). doi: 10.1371/journal.pcbi.1002986.

Schmidt, R. (2010) “Dynamic Light Scattering for Protein Characterization,” *Encyclopedia of Analytical Chemistry*. doi: 10.1002/9780470027318.a9092.

Schöning, J. C. *et al.* (2008) “Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis,” *Nucleic Acids Research*, 36(22), pp. 6977–6987. doi: 10.1093/nar/gkn847.

Seidel, S. A. I. *et al.* (2013) “Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions,” *Methods*, 59(3), pp. 301–315. doi: 10.1016/j.ymeth.2012.12.005.

Silva, J. L. *et al.* (2008) “Intriguing nucleic-acid-binding features of mammalian prion protein,” *Trends in Biochemical Sciences*, 33(3), pp. 132–140. doi: 10.1016/j.tibs.2007.11.003.

Singh, D. *et al.* (2012) *Exome sequencing and advances in crop improvement*, *Advances in Genetics*. Elsevier. doi: 10.1016/B978-0-12-394395-8.00003-7.

Staiger, D. *et al.* (2003) “The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA,” *Plant Journal*, 33(2), pp. 361–371. doi: 10.1046/j.1365-313X.2003.01629.x.

Steffen, A., Elgner, M. and Staiger, D. (2019) “Regulation of Flowering Time by the RNA-Binding Proteins AtGRP7 and AtGRP8,” *Plant and Cell Physiology*, 60(9), pp. 2040–2050. doi: 10.1093/pcp/pcz124.

Strodel, B. (2021) “Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins: Energy landscapes of IDPs and protein aggregation,” *Journal of Molecular Biology*, 433(20), p. 167182. doi: 10.1016/j.jmb.2021.167182.

Svergun, D., Barberato, C. and Koch, M. H. J. (2015) “Announcement of winner of 2015 Steenkamp long-term impact award,” *International Journal of Research in Marketing*, 32(2), p. vi. doi: 10.1016/s0167-8116(15)00050-6.

Svergun, D. I. (1987) “Feigin_Svergun_1987.Pdf.”

Szostak, E. and Gebauer, F. (2013) “Translational control by 3’-UTR-binding proteins,” *Briefings in Functional Genomics*, 12(1), pp. 58–65. doi: 10.1093/bfgp/els056.

Takanashi, K. and Yamaguchi, A. (2014) “Aggregation of ALS-linked FUS mutant sequesters RNA binding proteins and impairs RNA granules formation,” *Biochemical and Biophysical Research Communications*, 452(3), pp. 600–607. doi: 10.1016/j.bbrc.2014.08.115.

Taratuta, V. G. *et al.* (1990) “J . Phys. Chem. 1990, 94, 2140-2144,” (1), pp. 2140–2144.

Tetyuk, O., Benning, U. F. and Hoffmann-Benning, S. (2013) “Collection and analysis of Arabidopsis phloem exudates using the EDTA-facilitated method,” *Journal of Visualized Experiments*, (80), pp. 1–11. doi: 10.3791/51111.

Thieme, C. J. *et al.* (2015) “Endogenous Arabidopsis messenger RNAs transported to distant tissues,” *Nature Plants*, 1(April), pp. 1–8. doi: 10.1038/nplants.2015.25.

Tian, S., Curnutte, H. A. and Trcek, T. (2020) “RNA granules: A view from the RNA perspective,” *Molecules*, 25(14). doi: 10.3390/molecules25143130.

Tong, X. *et al.* (2022) “Liquid–liquid phase separation in tumor biology,” *Signal Transduction and*

*Targeted Therapy*, 7(1). doi: 10.1038/s41392-022-01076-x.

Torchia, D. A. (2007) "Protein Dynamics from NMR Relaxation," *eMagRes*, 2007(9), pp. 7–10. doi: 10.1002/9780470034590.emrstm0419.

Trapnell, C. *et al.* (2010) "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nature Biotechnology*, 28(5), pp. 511–515. doi: 10.1038/nbt.1621.

Tria, G. *et al.* (2015) "Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering," *IUCrJ*, 2, pp. 207–217. doi: 10.1107/S205225251500202X.

Turgeon, R. and Wolf, S. (2009) "PHloem transport: Cellular pathways and molecular trafficking," *Annual Review of Plant Biology*, 60(December 2008), pp. 207–221. doi: 10.1146/annurev.arplant.043008.092045.

Urry, D. W. *et al.* (1992) "Hydrophobicity scale for proteins based on inverse temperature transitions," *Biopolymers*, 32(9), pp. 1243–1250. doi: 10.1002/bip.360320913.

Uversky, V. N. (2013) "Intrinsic Disorder-based Protein Interactions and their Modulators," *Current Pharmaceutical Design*, 19(23), pp. 4191–4213. doi: 10.2174/1381612811319230005.

Valcárcel, J. *et al.* (1996) "Interaction of U2AF65 RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA," *Science*, 273(5282), pp. 1706–1709. doi: 10.1126/science.273.5282.1706.

Walz, C. *et al.* (2002) "Evidence for the presence and activity of a complete antioxidant defence system in mature sieve tubes," *Plant Journal*, 31(2), pp. 189–197. doi: 10.1046/j.1365-313X.2002.01348.x.

Wang, T. *et al.* (2021) "RNA Motifs and Modification Involve in RNA Long-Distance Transport in Plants," *Frontiers in Cell and Developmental Biology*, 9(April), pp. 1–10. doi: 10.3389/fcell.2021.651278.

Wang, Y. *et al.* (2014) "Quantitative evaluation of colloidal stability of antibody solutions using PEG-induced liquid-liquid phase separation," *Molecular Pharmaceutics*, 11(5), pp. 1391–1402. doi: 10.1021/mp400521b.

Weber, S. C. and Brangwynne, C. P. (2015) “Inverse size scaling of the nucleolus by a concentration-dependent phase transition,” *Current Biology*, 25(5), pp. 641–646. doi: 10.1016/j.cub.2015.01.012.

Wheeler, J. R. *et al.* (2016) “Distinct stages in stress granule assembly and disassembly,” *eLife*, 5(Se), pp. 1–25. doi: 10.7554/eLife.18413.

Wienkoop, S., Baginsky, S. and Weckwerth, W. (2010) “Arabidopsis thaliana as a model organism for plant proteome research,” *Journal of Proteomics*, 73(11), pp. 2239–2248. doi: 10.1016/j.jprot.2010.07.012.

Wolozin, B. (2012) “Regulated protein aggregation: Stress granules and neurodegeneration,” *Molecular Neurodegeneration*, 7(1), pp. 1–12. doi: 10.1186/1750-1326-7-56.

Xoconostle-Cázares, B. *et al.* (1999) “Plant paralog to viral movement protein that potentiates transport of mRNA into the phloem,” *Science*, 283(5398), pp. 94–98. doi: 10.1126/science.283.5398.94.

Xu, F. *et al.* (2022) “The Receptor Kinase FER Mediates Phase Separation of Glycine-Rich RNA-Binding Protein 7 to Confer Temperature Resilience in Arabidopsis,” *bioRxiv*, p. 2022.03.06.483201. Available at: https://www.biorxiv.org/content/10.1101/2022.03.06.483201v1%0Ahttps://www.biorxiv.org/content/10.1101/2022.03.06.483201v1.abstract.

Xue, C. *et al.* (2017) “Thioflavin T as an amyloid dye: Fibril quantification, optimal concentration and effect on aggregation,” *Royal Society Open Science*, 4(1). doi: 10.1098/rsos.160696.

Yan, Y. *et al.* (2020) “A Plant SMALL RNA-BINDING PROTEIN 1 Family Mediates Cell-to-Cell Trafficking of RNAi Signals,” *Molecular Plant*, 13(2), pp. 321–335. doi: 10.1016/j.molp.2019.12.001.

Yang, J. *et al.* (2014) “The I-TASSER suite: Protein structure and function prediction,” *Nature Methods*, 12(1), pp. 7–8. doi: 10.1038/nmeth.3213.

Yang, L. *et al.* (2009) “The 3′-untranslated region of rice glutelin GluB-1 affects accumulation of heterologous protein in transgenic rice,” *Biotechnology Letters*, 31(10), pp. 1625–1631. doi: 10.1007/s10529-009-0056-8.

Yates, A. D. *et al.* (2022) “Ensembl Genomes 2022: An expanding genome resource for non-vertebrates,” *Nucleic Acids Research*, 50(D1), pp. D996–D1003. doi: 10.1093/nar/gkab1007.

Yoo, B. C. *et al.* (2004) “A systematic small RNA signaling system in plants,” *Plant Cell*, 16(8), pp. 1979–2000. doi: 10.1105/tpc.104.023614.

Zhang, S., Sun, L. and Kragler, F. (2009) “The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation1[W][OA],” *Plant Physiology*, 150(1), pp. 378–387. doi: 10.1104/pp.108.134767.

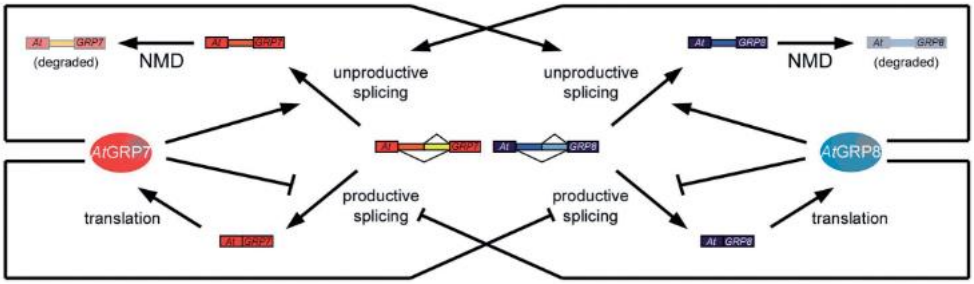
Zhang, W. *et al.* (2016) “TRNA-related sequences trigger systemic mRNA transport in plants,” *Plant Cell*, 28(6), pp. 1237–1249. doi: 10.1105/tpc.15.01056.

Zhao, Y. *et al.* (2021) “TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository,” *Journal of Translational Medicine*, 19(1), pp. 1–15. doi: 10.1186/s12967-021-02936-w.

Zheng, W., & Best, R. B. (2018). An extended Guinier analysis for intrinsically disordered proteins. *Journal of molecular biology*, *430*(16), 2540-2553.

# 7. Supplementary figures and tables

## 7.1 Introduction




**Supplementary figure 1 : model of interlocked *At*GRP8 and *At*GRP7 feedback loops.** (Schöning *et al.*, 2008)
*At*GRP8 and *At*GRP7 transcripts undergo productive splicing and translate. Increasing *At*GRP8 and *At*GRP7 protein levels promotes unproductive splicing of as_*At*GRP8 and as_*At*GRP7 (alternative spliced variants) and then these transcripts are degraded via non-mediated decay pathway (NMD). *At*GRP7 increasing levels can also promote the unproductive splicing of as_*At*GRP8 and later decay, as well as the other way around.

## 7.2 Materials and methods

**Supplementary table 1: primer list for genes of interest from RNA-sequencing**

| Gene of interest | size | Primers |
|---|---|---|
| At5g25610 (*RD22)* | 2187 bp | Fw: TAATACGACTCACTATAGGatggcgattcgtcttcct<br>Rev: ctagtagctgaaccacacaac |
| At1g64370 (PARCL) | 537 bp | Fw: TAATACGACTCACTATAGGatgcagtactacgaaaaccg<br>Rev: tcagtcgctgtcgctacc |
| At502120 (OHP) | 490 bp | Fw: TAATACGACTCACTATAGGatgagctcgtcgccgttatc<br>Rev: ttatagaggaagatcgagtcctttcc |
| At5g04020 (FIB) | 1386 bp | Fw: TAATACGACTCACTATAGGatggcgacggtaccattg<br>Rev: ttaagggtttaagagagagcttcc |
| At1g68520 (BBX14) | 1419 bp | Fw: TAATACGACTCACTATAGGatgatgaaaagtttggctagtgcg<br>Rev: ttagtgagcaacaccaattgaagatctc |
| At3g01700 | 411 bp | Fw: TAATACGACTCACTATAggatggcacgtctatttgtcgtag |

| | | |
|---|---|---|
| (AGP11) | | Rev: ttagagagagaagatgaagaatccgg |
| At1g02790 | 1545 bp | Fw: TAATACGACTCACTATAGGatgcagggtggtaggcttttg |
| (PG45) | | Rev: ttaaatagtgtagggtgtaggtaatgg |

**Supplementary table 2: RNA sequences for transcripts of interest measured through MST**

| RNA | Length (nt) | Sequence |
|---|---|---|
| *miRNA164* | 24 | GGgtggagaagcagggcacgtgca |
| *miNovel2* | 26 | Ggucuuguucugguuugguuuugaac |
| *miNovel106* | 26 | GGagatacgatctcttagcttttaac |
| *miNovel149* | 26 | GGattaattgtgctggtgtagacatc |
| *3'UTR GPR7* | 34 | GGauuuuguucugguucugcuuuagauuugaugu |

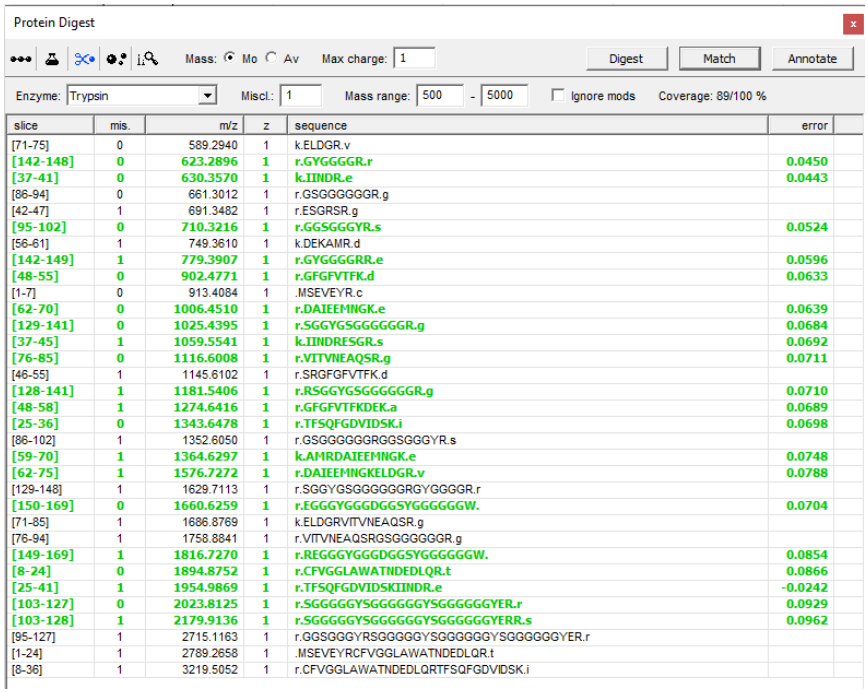
**Supplementary table 3: sample details SAXS measurements**

| **Sample details** | |
|---|---|
| Organism | *A. thaliana* |
| UniProt sequence ID | Q03251 |
| Extinction coefficient ε ($A_{280}$, 0.1%(w/v)) | 1.472 |
| MW (Da) | 16578 |
| Energy (eV) | 12400.4 |
| Solvent | 25 mM HEPES, 100 mM KCl, 1 mM DTT<br>pH 7.0 |
| Sample concentration | GRP8 monomer batch 2.2 mg/ml<br>GRP8 dimer batch 7.6 mg/ml<br>BSA batch 1.6 mg/ml<br>GRP8 monomer SEC-SAXS 3.6 mg/ml<br>GRP8 dimer SEC-SAXS 7.8 mg/ml |

**Supplementary table 4: sample data-collection information**

| **SAXS data-collection** | |
|---|---|
| Instrument/data processing | PETRA III Beamline P12 BioSAXS at DESY, Hamburg. PILATUS 6M detector |
| Wavelength (nm) | 0.099984 |
| Sample-to-detector distance (m) | 3 |
| Absolute scaling method | Relative scattering of pure water |
| Monitoring for radiation damage | Frame comparison |
| Exposure time (s/frame) | Batch – 0.1<br>SEC-SAXS – 0.0245 |
| Sample configuration | Monomer – P1<br>Dimer – P2 |
| Sample T (°C) | Batch – 20, 10, 5<br>SEC-SAXS - 20 |

## 7.3 Results




| slice | mis. | m/z | z | sequence | error |
|---|---|---|---|---|---|
| [71-75] | 0 | 589.2940 | 1 | k.ELDGR.v | |
| [142-148] | 0 | 623.2896 | 1 | r.GYGGGGR.r | 0.0450 |
| [37-41] | 0 | 630.3570 | 1 | k.IINDR.e | 0.0443 |
| [86-94] | 0 | 661.3012 | 1 | r.GSGGGGGGR.g | |
| [42-47] | 1 | 691.3482 | 1 | r.ESGRSR.g | |
| [95-102] | 0 | 710.3216 | 1 | r.GGSGGGYR.s | 0.0524 |
| [56-61] | 1 | 749.3610 | 1 | k.DEKAMR.d | |
| [142-149] | 1 | 779.3907 | 1 | r.GYGGGGRR.e | 0.0596 |
| [48-55] | 0 | 902.4771 | 1 | r.GFGFVTFK.d | 0.0633 |
| [1-7] | 0 | 913.4084 | 1 | .MSEVEYR.c | |
| [62-70] | 0 | 1006.4510 | 1 | r.DAIEEMNGK.e | 0.0639 |
| [129-141] | 0 | 1025.4395 | 1 | r.SGGYGSGGGGGGR.g | 0.0684 |
| [37-45] | 1 | 1059.5541 | 1 | k.IINDRESGR.s | 0.0692 |
| [76-85] | 0 | 1116.6008 | 1 | r.VITVNEAQSR.g | 0.0711 |
| [46-55] | 1 | 1145.6102 | 1 | r.SRGFGFVTFK.d | |
| [128-141] | 1 | 1181.5406 | 1 | r.RSGGYGSGGGGGGR.g | 0.0710 |
| [48-58] | 1 | 1274.6416 | 1 | r.GFGFVTFKDEK.a | 0.0689 |
| [25-36] | 0 | 1343.6478 | 1 | r.TFSQFGDVIDSK.i | 0.0698 |
| [86-102] | 1 | 1352.6050 | 1 | r.GSGGGGGGRGGSGGGYR.s | |
| [59-70] | 1 | 1364.6297 | 1 | k.AMRDAIEEMNGK.e | 0.0748 |
| [62-75] | 1 | 1576.7272 | 1 | r.DAIEEMNGKELDGR.v | 0.0788 |
| [129-148] | 1 | 1629.7113 | 1 | r.SGGYGSGGGGGGRGYGGGGR.r | |
| [150-169] | 0 | 1660.6259 | 1 | r.EGGGYGGGDGGSYGGGGGGW. | 0.0704 |
| [71-85] | 1 | 1686.8769 | 1 | k.ELDGRVITVNEAQSR.g | |
| [76-94] | 1 | 1758.8841 | 1 | r.VITVNEAQSRGSGGGGGGR.g | |
| [149-169] | 1 | 1816.7270 | 1 | r.REGGGYGGGDGGSYGGGGGGW. | 0.0854 |
| [8-24] | 0 | 1894.8752 | 1 | r.CFVGGLAWATNDEDLQR.t | 0.0866 |
| [25-41] | 1 | 1954.9869 | 1 | r.TFSQFGDVIDSKIINDR.e | -0.0242 |
| [103-127] | 0 | 2023.8125 | 1 | r.SGGGGGYSGGGGGGYSGGGGGGYER.r | 0.0929 |
| [103-128] | 1 | 2179.9136 | 1 | r.SGGGGGYSGGGGGGYSGGGGGGYERR.s | 0.0962 |
| [95-127] | 1 | 2715.1163 | 1 | r.GGSGGGYRSGGGGGYSGGGGGGYSGGGGGGYER.r | |
| [1-24] | 1 | 2789.2658 | 1 | .MSEVEYRCFVGGLAWATNDEDLQR.t | |
| [8-36] | 1 | 3219.5052 | 1 | r.CFVGGLAWATNDEDLQRTFSQFGDVIDSK.i | |

**Supplementary figure 2: mMass screenshot of GRP8 as a control for crosslinking.** Sequences matches found in through mass spectrometry to GRP8 represented in Green. Sequences missing represented in black. Final sequence coverage 89%.

Protein Digest

Mass: Mo Av Max charge: 1 Digest Match Annotate

Enzyme: Trypsin Miscl.: 1 Mass range: 500 - 5000 Ignore mods Coverage: 66/100 %

| slice | mis. | m/z | z | sequence | error |
|---|---|---|---|---|---|
| [71-75] | 0 | 589.2940 | 1 | k.ELDGR.v | |
| [142-148] | 0 | 623.2896 | 1 | r.GYGGGGR.r | 0.0366 |
| [37-41] | 0 | 630.3570 | 1 | k.IINDR.e | 0.0390 |
| [86-94] | 0 | 661.3012 | 1 | r.GSGGGGGGR.g | |
| [42-47] | 1 | 691.3482 | 1 | r.ESGRSR.g | |
| [95-102] | 0 | 710.3216 | 1 | r.GGSGGGYR.s | 0.0403 |
| [56-61] | 1 | 749.3610 | 1 | k.DEKAMR.d | |
| [142-149] | 1 | 779.3907 | 1 | r.GYGGGGRR.e | 0.0515 |
| [48-55] | 0 | 902.4771 | 1 | r.GFGFVTFK.d | |
| [1-7] | 0 | 913.4084 | 1 | .MSEVEYR.c | |
| [62-70] | 0 | 1006.4510 | 1 | r.DAIEEMNGK.e | |
| [129-141] | 0 | 1025.4395 | 1 | r.SGGYGSGGGGGGR.g | 0.0597 |
| [37-45] | 1 | 1059.5541 | 1 | k.IINDRESGR.s | 0.0567 |
| [76-85] | 0 | 1116.6008 | 1 | r.VITVNEAQSR.g | 0.0591 |
| [46-55] | 1 | 1145.6102 | 1 | r.SRGFGFVTFK.d | |
| [128-141] | 1 | 1181.5406 | 1 | r.RSGGYGSGGGGGGR.g | 0.0623 |
| [48-58] | 1 | 1274.6416 | 1 | r.GFGFVTFKDEK.a | 0.0561 |
| [25-36] | 0 | 1343.6478 | 1 | r.TFSQFGDVIDSK.i | 0.0551 |
| [86-102] | 1 | 1352.6050 | 1 | r.GSGGGGGGRGGSGGGYR.s | |
| [59-70] | 1 | 1364.6297 | 1 | k.AMRDAIEEMNGK.e | |
| [62-75] | 1 | 1576.7272 | 1 | r.DAIEEMNGKELDGR.v | 0.0745 |
| [129-148] | 1 | 1629.7113 | 1 | r.SGGYGSGGGGGGRGYGGGGR.r | |
| [150-169] | 0 | 1660.6259 | 1 | r.EGGGYGGGDGGSYGGGGGGW. | |
| [71-85] | 1 | 1686.8769 | 1 | k.ELDGRVITVNEAQSR.g | |
| [76-94] | 1 | 1758.8841 | 1 | r.VITVNEAQSRGSGGGGGGR.g | |
| [149-169] | 1 | 1816.7270 | 1 | r.REGGGYGGGDGGSYGGGGGGW. | |
| [8-24] | 0 | 1894.8752 | 1 | r.CFVGGLAWATNDEDLQR.t | |
| [25-41] | 1 | 1954.9869 | 1 | r.TFSQFGDVIDSKIINDR.e | |
| [103-127] | 0 | 2023.8125 | 1 | r.SGGGGGYSGGGGGGYSGGGGGGYER.r | 0.0746 |
| [103-128] | 1 | 2179.9136 | 1 | r.SGGGGGYSGGGGGGYSGGGGGGYERR.s | 0.0811 |
| [95-127] | 1 | 2715.1163 | 1 | r.GGSGGGYRSGGGGGYSGGGGGGYSGGGGGGYER.r | |
| [1-24] | 1 | 2789.2658 | 1 | .MSEVEYRCFVGGLAWATNDEDLQR.t | |
| [8-36] | 1 | 3219.5052 | 1 | r.CFVGGLAWATNDEDLQRTFSQFGDVIDSK.i | |

**Supplementary figure 4: mMass screenshot of GRP8 monomer crosslinked to miRNA164.** Sequences buried versus GRP8 control are represented by red squares. Sequences matches found in through mass spectrometry to GRP8 represented in Green. Sequences missing represented in black. Final sequence coverage 66%.

Protein Digest

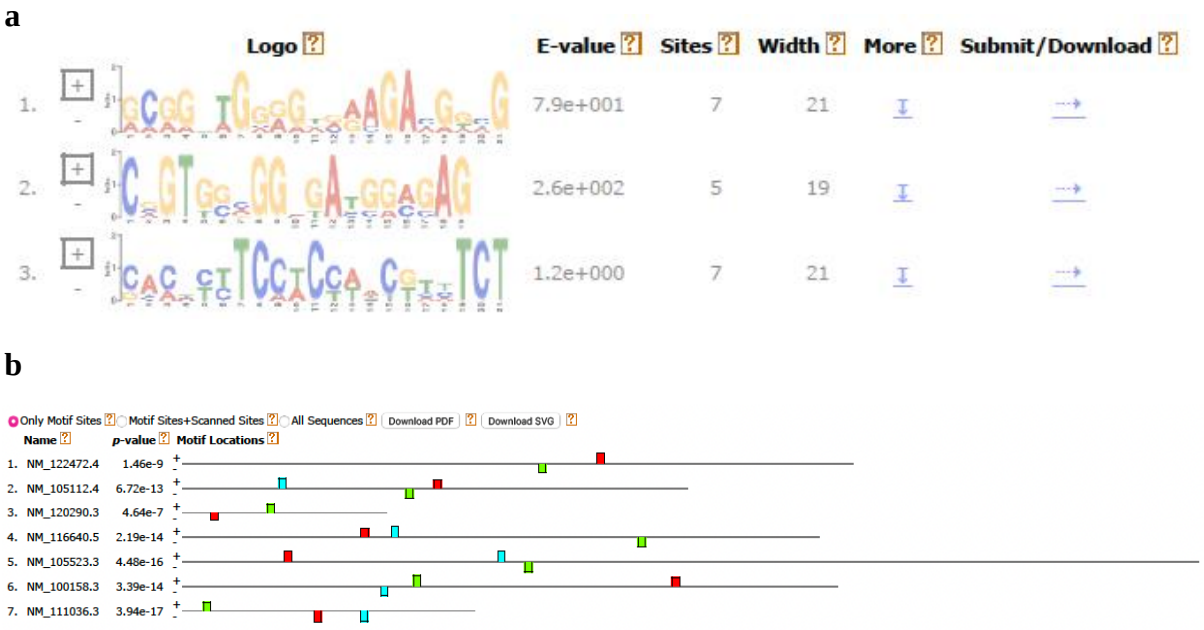
Mass: Mo Av Max charge: 1 Digest Match Annotate

Enzyme: Trypsin Miscl.: 1 Mass range: 500 - 5000 Ignore mods Coverage: 76/100 %

| slice | mis. | m/z | z | sequence | error |
|---|---|---|---|---|---|
| [71-75] | 0 | 589.2940 | 1 | k.ELDGR.v | |
| [142-148] | 0 | 623.2896 | 1 | r.GYGGGGR.r | 0.0524 |
| [37-41] | 0 | 630.3570 | 1 | k.IINDR.e | 0.0515 |
| [86-94] | 0 | 661.3012 | 1 | r.GSGGGGGGR.g | |
| [42-47] | 1 | 691.3482 | 1 | r.ESGRSR.g | |
| [95-102] | 0 | 710.3216 | 1 | r.GGSGGGYR.s | 0.0611 |
| [56-61] | 1 | 749.3610 | 1 | k.DEKAMR.d | |
| [142-149] | 1 | 779.3907 | 1 | r.GYGGGGRR.e | 0.0695 |
| [48-55] | 0 | 902.4771 | 1 | r.GFGFVTFK.d | 0.0701 |
| [1-7] | 0 | 913.4084 | 1 | .MSEVEYR.c | |
| [62-70] | 0 | 1006.4510 | 1 | r.DAIEEMNGK.e | 0.0710 |
| [129-141] | 0 | 1025.4395 | 1 | r.SGGYGSGGGGGGR.g | 0.0817 |
| [37-45] | 1 | 1059.5541 | 1 | k.IINDRESGR.s | 0.0820 |
| [76-85] | 0 | 1116.6008 | 1 | r.VITVNEAQSR.g | 0.0845 |
| [46-55] | 1 | 1145.6102 | 1 | r.SRGFGFVTFK.d | |
| [128-141] | 1 | 1181.5406 | 1 | r.RSGGYGSGGGGGGR.g | 0.0862 |
| [48-58] | 1 | 1274.6416 | 1 | r.GFGFVTFKDEK.a | 0.0866 |
| [25-36] | 0 | 1343.6478 | 1 | r.TFSQFGDVIDSK.i | 0.0875 |
| [86-102] | 1 | 1352.6050 | 1 | r.GSGGGGGGRGGSGGGYR.s | |
| [59-70] | 1 | 1364.6297 | 1 | k.AMRDAIEEMNGK.e | |
| [62-75] | 1 | 1576.7272 | 1 | r.DAIEEMNGKELDGR.v | 0.1013 |
| [129-148] | 1 | 1629.7113 | 1 | r.SGGYGSGGGGGGRGYGGGGR.r | |
| [150-169] | 0 | 1660.6259 | 1 | r.EGGGYGGGDGGSYGGGGGGW. | |
| [71-85] | 1 | 1686.8769 | 1 | k.ELDGRVITVNEAQSR.g | |
| [76-94] | 1 | 1758.8841 | 1 | r.VITVNEAQSRGSGGGGGGR.g | |
| [149-169] | 1 | 1816.7270 | 1 | r.REGGGYGGGDGGSYGGGGGGW. | |
| [8-24] | 0 | 1894.8752 | 1 | r.CFVGGLAWATNDEDLQR.t | 0.1346 |
| [25-41] | 1 | 1954.9869 | 1 | r.TFSQFGDVIDSKIINDR.e | |
| [103-127] | 0 | 2023.8125 | 1 | r.SGGGGGYSGGGGGGYSGGGGGGYER.r | 0.1172 |
| [103-128] | 1 | 2179.9136 | 1 | r.SGGGGGYSGGGGGGYSGGGGGGYERR.s | |
| [95-127] | 1 | 2715.1163 | 1 | r.GGSGGGYRSGGGGGYSGGGGGGYSGGGGGGYER.r | |
| [1-24] | 1 | 2789.2658 | 1 | .MSEVEYRCFVGGLAWATNDEDLQR.t | |
| [8-36] | 1 | 3219.5052 | 1 | r.CFVGGLAWATNDEDLQRTFSQFGDVIDSK.i | |

**Supplementary figure 5: mMass screenshot of GRP8 dimer crosslinked to miRNA164.** Sequences buried versus GRP8 control are represented by red squares. Sequences matches found in through mass spectrometry to GRP8 represented in Green. Sequences missing represented in black. Final sequence coverage 78%.

## 7.4 Discussion




**Supplementary figure 6: MEME-suite motif finder in the requested mRNA sequences.** a) motifs found with no significant values in the sequences of interest. b) positioning of the motifs in the sequences of interest. 1.NM_122472.4, RD22. 2.NM_105122.4, PARCL. 3.NM_120290.3, OHP. 4.NM_116640.5, FIB. 5.NM_105523.3, BBX14. 6.NM_100158.3, PG45. 7.NM_111036.3, AGP11. Motif 1 represented in red. Motif 2 represented in turquoise. Motif 3 represented in green.


**Supplementary figure 7: MEME-suite motif finder in the requested miRNA sequences.** a) motifs found with no significant values in the sequences of interest. b) positioning of the motifs in the sequences of interest. 1.miRNA164a. 2.miNovel2. 3.miNovel106. 4.miNovel149. 5.3'UTR-GRP7. Motif 1 represented in red. Motif 2 represented in turquoise. Motif 3 represented in green.

**Supplementary table 5: U/G content percentage for transcripts/genes of interest measured through MST**

| Transcript | U | G |
|---|---|---|
| *RD22* | 27% | 25% |
| *PARCL* | 27% | 22% |
| *OHP* | 30% | 19% |
| *FIB* | 25% | 22% |
| *BBX14* | 31% | 17% |
| *PG45* | 29% | 22% |
| *AGP11* | 28% | 19% |
| *GRP7* | 19% | 41% |
| *GRP8* | 19% | 41% |
| *miRNA164* | 8% | 50% |
| *miNovel2* | 50% | 31% |
| *miNovel106* | 35% | 19% |
| *miNovel149* | 31% | 35% |
| *3'UTR-GRP7* | 53% | 26% |

# 8. Declaration of oath

**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.

**Hamburg, den 18.11.22** **Unterschrift**

## 9. Confirmation of English language correction

I, Maiyah Rivers, confirm that I edited the PhD thesis, *Functional-structural characterization of phloem-mobile protein GRP8 in Arabidopsis thaliana* written and submitted by PhD candidate Francisca Méndez-Pinochet. As a U.S. born citizen and native English speaker, I attest that I have read the document for clarity of English grammar and punctuation and did not review this thesis for scientific accuracy.

I currently reside in Cranston, Rhode Island, United States and I am an employee of Tufts University in Boston, Massachusetts, United States, where I serve as the Assistant Director of Student Affairs at the School Museum of Fine Arts at Tufts.

Maiyah Rivers
11/19/2022

## 10. Acknowledgements

I would like to thank Julia for giving me this life-changing opportunity to move to Hamburg and taking me under her supervision to do my PhD within her laboratory.

I would like to express my gratitude to Richard for taking the time to talk to me when needed, checking on me and for making the extra effort to come over to Hamburg to be part of my PhD committee.

I would like to thank each person from AG Kehr that has been part of my PhD journey, especially the PhD students that have been an incredible family. Special thanks to Dr. Steffen Ostendorp for guiding me through the experimental part of my thesis and taking time for helping me with all the questions I came up with.

I'm extremely grateful to be part of the PLAMORF team. They are an amazing group of scientists that provided me with a lot of insightful discussions. Special thanks to Franziska Hoerbst who helped me spontaneously with last minute statistics.

Many thanks to Dr. Sven Falke at DESY for his time and help with the crystallographic trials and to Cy Jeffries at the EMBL Hamburg for the SAXS analysis collaboration.

I could have not undertaken this journey without my PhD colleagues, Kim Lühmann and Rita Fernandes. I was extremely lucky for having the opportunity to start our journeys together. Both became such a big support academically and personally in these last four years. Words cannot describe how grateful I am to share so many memories together.

I had the pleasure of working with wonderful people at the IPM, some for them who became friends. Special thanks to Mariana Motta, Sebastian Fricke, Linn Von Pein and Florian Pomrehn who made me feel like a bit more like home inside and outside the lab.

I would like to extend my sincere thanks to Maiyah Rivers-Gamble, who took the time to read through the whole thesis and correct my written English.

This endeavor wouldn’t been possible without the support and comfort of my friends at home and in Hamburg. They were fundamental for my wellbeing.

I would like to express my deepest appreciation for the unconditional support of my family that has been present no matter the distance. Specially because they have supported my academic path, even when that meant to move away from them.

Words cannot express my gratitude to Vincent and his family for taking me in as one more of the family making me feel at home and supported, especially towards the end of my PhD journey.